UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

a

LAPPEENRANNAN TEKNILLINEN YLIOPISTO
TEKNILLINEN TIEDEKUNTA
MATEMATIIKAN JA FYSIIKAN LAITOS

DISERTAČNÍ PRÁCE

Jazykově orientované modely pro podporu rozhodování

v oboru Aplikovaná matematika P1104
pro získání vědecké hodnosti Ph.D.

Univerzita Palackého
v Olomouci

Lappeenranta
University of Technology

Školitelé:
doc. RNDr. Jana Talašová, CSc.
doc. Pasi Luukka, Ph.D.
prof. Mikael Collan, Ph.D.

Vypracoval:
Mgr. et Mgr. Jan Stoklasa

Rok odevzdání: 2014

PALACKÝ UNIVERSITY, OLOMOUC
FACULTY OF SCIENCE
DEPT. OF MATH. ANALYSIS AND APPLICATIONS OF MATHEMATICS

and

LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
LUT SCHOOL OF TECHNOLOGY
DEPARTMENT OF MATHEMATICS AND PHYSICS

DISSERTATION THESIS

Linguistic models for decision support

A thesis submitted for the degree of Doctor of Philosophy

Univerzita Palackého
v Olomouci

Lappeenranta
University of Technology

Supervisors:
doc. RNDr. Jana Talašová, CSc.
doc. Pasi Luukka, Ph.D.
prof. Mikael Collan, Ph.D.

Author:
Mgr. et Mgr. Jan Stoklasa

Year of submission: 2014

Prohlašuji, že jsem tuto disertační práci zpracoval samostatně a že výsledky spoluautorů článků přiložených k této práci jsou jasně vymezeny a odlišeny od mých. Prohlašuji, že práce byla zpracována pod společným vedením mých školitelů doc. RNDr. Jany Talašové, CSc., prof. Mikaela Collana, Ph.D. and doc. Pasiho Luukky, Ph.D. v rámci double degree smlouvy o doktorském studiu mezi Univerzitou Palackého v Olomouci, Česká republika a Lappeenranta University of Technology, Finsko. Prohlašuji, že všechny použité zdroje jsem řádně citoval a uvedl v seznamu literatury.

V Olomouci, 4. listopadu 2014

Podpis: .....

I hereby declare that this thesis is my original work, that the credit of co-authors of the papers included in the thesis has been acknowledged and clarified and that I have written it under the joint supervision of doc. RNDr. Jana Talašová, CSc., prof. Mikael Collan, Ph.D. and doc. Pasi Luukka, Ph.D. under a doctoral double degree agreement between Palacký University, Olomouc, Czech Republic and Lappeenranta University of Technology, Finland. The literature used is listed in the list of references and duly cited in the text.

Olomouc, November 4, 2014

Signature: .....

Abstrakt

Jan Stoklasa
JAZYKOVĚ ORIENTOVANÉ MODELY PRO PODPORU ROZHODOVÁNÍ
Lappeenranta a Olomouc, 2014
352 stran (6 titulních stran, 1-136, 12 článků přiložených k práci 137-352)

Jazykově orientované modelování je relativně novou oblastí matematiky, která stále ještě prochází rychlým vývojem. Je ze své podstaty úzce spjato s teorií fuzzy množin a fuzzy logikou, nicméně pro úspěšnou tvorbu jazykově orientovaných modelů je potřeba znalostí také z ostatních oblastí matematiky, stejně jako z ostatních vědních disciplín - např. lingvistiky, behaviorálních věd apod. Tento přístup k matematickému modelování na sebe poslední dobou upoutal mnoho pozornosti, neboť nabízí nástroje pro matematickou reprezentaci nejběžnějšího komunikačního prostředku lidí - přirozeného jazyka. Přidáním jazykové úrovně do matematických modelů může vzniknout rozhraní pro snadnou komunikaci mezi matematickou reprezentací modelovaného systému a uživatelem daného modelu. Díky tomu, že je toto rozhraní vytvořeno na úrovni přirozeného jazyka, může být pro uživatele modelu dostatnečně srozumitelné a snadno použitelné, ale přitom si stále zachovávat schopnost předat uživateli modelu všechny potřebné (relevantní) informace a předejít tak nedorozuměním. Vytvoření dobře fungujícího jazykového rozhraní však není jednoduchým úkolem - je zapotřebí, aby propojení jazykové a matematické úrovně jazykově orientovaného modelu bylo správně vytvořeno a udržováno po celou dobu modelování.
Tato disertační práce se zaměřuje na vztah jazykové a výpočetní (matematické) úrovně matematických modelů pro podporu rozhodování. Pozornost je věnována několika podstatným otázkám matematické reprezentace významu jazykových výrazů, jejich správné transformaci do jazyka matematiky a v neposlední řadě "zpětnému překladu" matematických výstupů do běžného jazyka. V první části práce je shrnut pohled autora na jazykově orientované modelování pro podporu rozhodování a jsou navrženy doporučení pro tvorbu jazykově orientovaných modelů pro praktické použití v oblasti podpory rozhodování. Tato doporučení jsou základem metodologie tvorby jazykově orientovaných modelů použité při návrhu matematických modelů, které jsou jako další podstatný výstup práce prezentovány v druhé části práce (a to formou několika případových studií reálných problémů a představení odpovídajících matematických modelů, na jejichž vytváření se autor podílel).
Z teoretického pohledu jsou v první části práce studovány otázky matematické reprezentace významu jazykových termů, výpočtů s těmito reprezentacemi a jejich zpětného překladu do přirozeného jazyka (jazyková aproximace). Autor se zaměřuje na vhodnost matematických operací prováděných s matematickými významy jazykových termů, korespondenci matematické a jazykové úrovně modelů a správnou prezentaci vhodných výstupů uživatelům modelů. Diskutovány jsou zde také etické aspekty podpory rozhodování - zejména důsledky možné ztráty významu způsobené překladem matematických výstupů do běžného jazyka a otázky odpovědnosti za konečná rozhodnutí učiněná na základě výstupů modelů pro podporu rozhodování.)

V druhé části práce je prezentováno několik případových studií reálných problémů. Na jejich pozadí jsou popsány nové matematické výsledky a modely. Případové studie poskytují kontext a motivaci pro prezentované výsledky. Jako první je představen model pro podporu rozhodování v krizovém řízení, formulovaný jako problém fuzzy lineárního programování a je navrženo jeho možné heuristické řešení. V návrhu modelu je reflektována neurčitost vstupů, expertní znalost postupů při katastrofách a nezbytnost dosažení srozumitelných výstupů které jsou snadno interpretovatelné laiky (operátory zdravotnické záchranné služby) ve velice krátném čase.
Po analýze jazykové úrovně Saatyho analytického hierarchického procesu (AHP) jsou prezentovány další dvě případové studie založené na AHP - nejdříve je v kontextu hodnocení výstupů tvůrčí umělecké činnosti diskutována nutnost zavedení podmínky slabé konzistence matic intenzit preferencí. Na základě této podmínky je pak navržena adaptace AHP pro velké matice intenzit preferencí. Druhá případová studie pak využívá fuzzifikované AHP pro účely hodnocení - zde v kontextu začlenění peer-review komponenty do hodnocení výstupů výzkumu a vývoje.
V kontextu hodnocení lidských zdrojů je pak prezentován jazykově orientovaný model hodnocení akademických pracovníků založený na bázích fuzzy pravidel navržený tak, aby nebyla nutná jazyková aproximace jeho výstupů a aby výstupy byly snadno převoditelné na grafickou informaci. Tohoto bylo dosaženo využitím speciálního přístupu k přibližné dedukci.
Poslední případová studie je pak z oblasti humanitních věd - v rámci psychologické diagnostiky je navržen model pro interpretaci výstupů mnohorozměrných dotazníků. V tomto kontextu je zkoumána otázka kvality dat v klasifikačních úlohách. Je zde navržena modifikace receiver operating characteristic (ROC) metody pro posouzení fungování klasifikátorů, která umožňuje zohlednit rozdílnou kvalitu jednotlivých instancí dat při posuzování fungování klasifikátorů.
K práci je v rámci její třetí části přiloženo 12 publikací, jichž byl Jan Stoklasa autorem nebo spoluautorem. Tyto publikace shrnují dosažené matematické výsledky autora a umožňují detailnější náhled na modely a výstupy prezentované v druhé části práce.

Klíčová slova: jazykově orientované modelování, podpora rozhodování, fuzzy, hodnocení, MCDM, vícekriteriální rozhodování, klasifikace, slabá konzistence, umění, diagnostika, krizové řízení, zdravotnická záchranná služba, hodnocení pracovníků.

Abstract

Jan Stoklasa
LINGUISTIC MODELS FOR DECISION SUPPORT
Lappeenranta and Olomouc, 2014
352 pages (6 title pages, 1-136, and 12 papers appended to the thesis 137-352)

Linguistic modelling is a rather new branch of mathematics that is still undergoing rapid development. It is closely related to fuzzy set theory and fuzzy logic, but knowledge and experience from other fields of mathematics, as well as other fields of science including linguistics and behavioral sciences, is also necessary to build appropriate mathematical models. This topic has received considerable attention as it provides tools for mathematical representation of the most common means of human communication - natural language. Adding a natural language level to mathematical models can provide an interface between the mathematical representation of the modelled system and the user of the model - one that is sufficiently easy to use and understand, but yet conveys all the information necessary to avoid misinterpretations. It is, however, not a trivial task and the link between the linguistic and computational level of such models has to be established and maintained properly during the whole modelling process.

In this thesis, we focus on the relationship between the linguistic and the mathematical level of decision support models. We discuss several important issues concerning the mathematical representation of meaning of linguistic expressions, their transformation into the language of mathematics and the retranslation of mathematical outputs back into natural language. In the first part of the thesis, our view of the linguistic modelling for decision support is presented and the main guidelines for building linguistic models for real-life decision support that are the basis of our modeling methodology are outlined.

From the theoretical point of view, the issues of representation of meaning of linguistic terms, computations with these representations and the retranslation process back into the linguistic level (linguistic approximation) are studied in this part of the thesis. We focus on the reasonability of operations with the meanings of linguistic terms, the correspondence of the linguistic and mathematical level of the models and on proper presentation of appropriate outputs. We also discuss several issues concerning the ethical aspects of decision support - particularly the loss of meaning due to the transformation of mathematical outputs into natural language and the issue or responsibility for the final decisions.

In the second part several case studies of real-life problems are presented. These provide background and necessary context and motivation for the mathematical results and models presented in this part. A linguistic decision support model for disaster management is presented here - formulated as a fuzzy linear programming problem and a heuristic solution to it is proposed. Uncertainty of outputs, expert knowledge concerning disaster response practice and the necessity of obtaining outputs that are easy to interpret (and available in very short time) are reflected in the design of the model. Saaty's analytic hierarchy process (AHP) is considered in two case studies - first in the context of the evaluation of works of art, where a weak consistency condition is introduced and an adaptation of AHP for large matrices of preference intensities is presented. The second AHP

case-study deals with the fuzzified version of AHP and its use for evaluation purposes - particularly the integration of peer-review into the evaluation of R&D outputs is considered. In the context of HR management, we present a fuzzy rule based evaluation model (academic faculty evaluation is considered) constructed to provide outputs that do not require linguistic approximation and are easily transformed into graphical information. This is achieved by designing a specific form of fuzzy inference. Finally the last case study is from the area of humanities - psychological diagnostics is considered and a linguistic fuzzy model for the interpretation of outputs of multidimensional questionnaires is suggested. The issue of the quality of data in mathematical classification models is also studied here. A modification of the receiver operating characteristics (ROC) method is presented to reflect variable quality of data instances in the validation set during classifier performance assessment.
Twelve publications on which the author participated are appended as a third part of this thesis. These summarize the mathematical results and provide a closer insight into the issues of the practical applications that are considered in the second part of the thesis.

Keywords: linguistic modelling, decision support, fuzzy, evaluation, MCDM, classification, weak consistency, art, diagnostics, disaster management, medical rescue services, staff evaluation.

Preface

This text summarizes in many ways the last five years of my life. It partially contains the results of my research on linguistic fuzzy modelling, but partially also my view of mathematics, its possibilities and limitations. I am very glad that I had the opportunity to combine (as well as it was possible) my two areas of interest - psychology and applied mathematics. This led to my conviction that mathematicians may be needed in human sciences, and that a human sciences perspective can provide useful insights into mathematics. This was of much comfort to me as I could conclude that my choice of studies was not completely insane.

I am very glad I had the chance to meet all those great people who influenced the course of my research, who guided me and provided with resources and encouragement. Among all of them a special thanks is in order to Jana Talašová, Mikael Collan, Pasi Luukka, Mario Fedrizzi and Michele Fedrizzi - my dear colleagues and friends. Also big thanks to Iveta, Pavel, Tomáš, Věra and Jana - my fellow students who made much of the work easier by providing help, feedback and helping to create an atmosphere for sharing thoughts and ideas.

I would also like to thank to all those people closest to me - to my family and Janča - who had to wait until some work was done, who had to sleep in a room with lights on and computer humming during paper-writing nights, who had to reschedule their plans to let me meet a deadline, who never knew if plans will change. I know it was not easy. And I am afraid that it will not be much better now...

But I still think it was worth it!

Lappeenranta, June 2014

Jan Stoklasa

LIST OF THE ORIGINAL ARTICLES AND THE AUTHOR'S CONTRIBUTION

This thesis consists of an introductory part and several papers in refereed journals and conference proceedings. The papers and the author's contribution in them are summarized below.

I **Stoklasa, J.**, *A Fuzzy Approach to Disaster Modeling: Decision Making Support and Disaster Management Tool for Emergency Medical Rescue Services*. IN Mago, V. K. and Bhatia, N. (eds.) Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies, IGI Global, 2012. DOI: 10.4018/978-1-61350-429-1.ch028

II **Stoklasa, J., Talašová, J. and Holeček, P.**, Academic Staff Performance Evaluation - Variants of Models. *Acta Polytechnica Hungarica*, 8(3), 91-111, 2011.

III **Stoklasa, J., Jandová, V. and Talašová, J.**, Weak consistency in Saaty's AHP - evaluating creative work outcomes of Czech Art Colleges. *Neural Network World*, 23(1), 61-77, 2013.

IV **Stoklasa, J. and Luukka, P.**, Receiver operating characteristics and the quality of data. Submitted to *Psychometrika*, Springer. (May 2014)

V **Krejčí, J. and Stoklasa, J.**, Fuzzified AHP in the evaluation of R&D results. Submitted to *Central European Journal of Operations Research*, Springer. (April 2014)

VI **Stoklasa, J., Talášek, T. and Musilová, J.**,Fuzzy approach - a new chapter in the methodology of psychology? *Human Affairs*, 24(2), 189–203, 2014.

VII **Stoklasa, J., Talášek, T. and Talašová, J.**, AHP and weak consistency in the evaluation of works of art - a case study of a large problem. Accepted for publication in *International Journal of Business Innovation and Research*, Inderscience. (2014)

VIII **Collan, M., Stoklasa, J. and Talašová, J.**, On academic faculty evaluation systems - more than just simple benchmarking. *International Journal of Process Management and Benchmarking*, 4(4), 437–455, 2014.

IX **Talašová, J. and Stoklasa, J.**, Fuzzy approach to academic staff performance evaluation. *Proceedings of the* 28^{th} *International Conference on Mathematical Methods in Economics 2010*, 621–626, 2010.

X **Stoklasa, J. and Talašová, J.**, Using linguistic fuzzy modeling for MMPI-2 data interpretation. *Proceedings of the* 29^{th} *International Conference on Mathematical Methods in Economics 2011 - part II*, 653–658, 2011.

XI **Talašová, J. and Stoklasa, J.**, A model for evaluating creative work outcomes at Czech Art Colleges. *Proceedings of the* 29^{th} *International Conference on Mathematical Methods in Economics 2011 - part II*, 698-703, 2011.

XII **Stoklasa, J., Krejčí, J. and Talašová, J.**, Fuzzified AHP in evaluation of R&D outputs - a case from Palacky University in Olomouc, *Proceedings of the* 31^{st} *International Conference Mathematical Methods in Economics 2013*, 856–861, 2013.

The publications are ordered by type (from book chapters through papers in journals with IF, papers in refereed journals to refereed conference proceedings papers) and chronologically in each publication type category. Author's publications will be denoted by Roman numbers in the text.

J. Stoklasa is the sole author of Publication **I**, where a multiphase linguistic fuzzy mathematical model for decision support of the emergency medical rescue services is proposed and artificial results provided. The concept of an α-degree upper bound of a fuzzy number is proposed here to deal with fuzzy constraints. A heuristic solution to the fuzzy linear programming representation of the problem of finding the minimal number of ambulances needed to deal with the situation is proposed using the α-degree upper bound of a fuzzy number.

Publication **II** summarizes a linguistic fuzzy rule based methodology for academic faculty evaluation and compares the fuzzy rule base approach with other widely used aggregation approaches. J. Stoklasa is the main author of this publication, wrote the paper, proposed the mathematical evaluation model and significantly participated on the development of the whole evaluation methodology presented in this paper. First stages of this evaluation methodology - first attempts to define appropriate linguistic scales underlying the mathematical model were presented already in publication **IX** - J. Stoklasa participated in the development of the linguistic scales proposed in the paper and writing the paper. The evaluation methodology is still being developed and it is currently being implemented on several universities in the Czech Republic. A critical comparison of this approach to staff evaluation and its mathematical basis with several other models and approaches to this topic is provided in publication **VIII**. J. Stoklasa co-authored the paper and provided the HR-perspective, two of the case studies (the model used at Palacký University in Olomouc and the A&M University Kingsville) and participated in comparing and discussing the models.

Publication **XI** introduces an evaluation methodology for creative work outcomes of Czech art Colleges. J. Stoklasa is the co-author of the paper, participated on writing it and on the design and further development of the evaluation methodology. The solution required revisiting the standard consistency condition in Saaty's AHP [79, 83, 80] and an adaptation of the AHP method for large pairwise comparison matrices. Weak consistency as a minimum requirement on the consistency of the matrix of preference intensities is introduced here. This consistency condition is proposed so that it remains in accordance with the intuitive meanings of the linguistic terms of Saaty's scale. The concept of weak consistency is further studied and its properties investigated in **III**. J. Stoklasa is the main author of **III** and wrote most of the paper. The propositions presented in the paper were proved and the respective subsections of the paper were written by V. Jandová. In **VII** modifications of the model after the analysis of its pilot run, some implications of the use of weak consistency on the easy adjustability of the mathematical model and further development of the evaluation methodology are discussed. J. Stoklasa is the main author of the paper, wrote it, performed the analysis and proposed the modifications to the evaluation methodology. The evaluation methodology presented in these three papers is still being developed and fine-tuned, but a part of the subsidy from the state budget of the Czech Republic has been distributed among Czech art colleges based on the outputs of the evaluation methodology (implemented in the Registry of Artistic Results) since 2012.

Another evaluation model also from the field of tertiary education institutions is presented in papers **V** and **XII**. In **XII** a fuzzified AHP method as proposed by Krejčí et al. (see [53]) is applied to the evaluation of R&D results - particularly scientific monographs. An evaluation methodology combining a quality assessment of the publisher of the monograph with the peer-review evaluation of the monograph itself by a panel of experts is proposed. The fuzzified AHP and the respective linguistic scale are used not only to derive evaluation intervals for each book from a given cate-

gory of publishers, it is also used to visualize the preferences of the evaluators. J. Stoklasa is the main author of **XII**, wrote most of the text and participated significantly on the development of the evaluation methodology. In **V** J. Stoklasa as a co-author supplied the application part of the paper and participated at the conclusions. This paper presents the overview of the fuzzification of AHP, summarizes the development of the evaluation methodology for scientific monographs, discusses the role of the linguistic labels of the elements of Saaty's scale and the usefulness of the fuzzification of AHP presented in [53]. The evaluation methodology has been used to distribute funding for scientific monographs at the Faculty of Science of Palacký University in Olomouc (Czech Republic) in 2012.

In **VI** J. Stoklasa as the main author maps possible application areas for linguistic fuzzy modeling in humanities, with special focus on psychology and psychological diagnostics. J. Stoklasa wrote most of the text, provided the application examples and participated in discussion and in forming the conclusion part. The paper proposes possible focus areas for future research of the use of linguistic fuzzy modeling in psychology and humanities in general. A first step in this direction was made in **X**, where a decision support model based on linguistic fuzzy modeling is presented for psychological diagnostics. A fuzzy rule based classifier is proposed here to determine the presence or absence of a particular diagnosis. The topic of data quality is identified here as a necessary focus for future research. J. Stoklasa is the main author of the paper, wrote most of the text and proposed the mathematical model presented in the paper. In **IV** the issue of data quality and classifier performance is discussed in more details. A modification of the receiver operating characteristics (ROC - see [25, 30, 33]) is proposed here. This modification is capable of reflecting different quality of data in the validation set during the performance assessment of a classifier. The modification is illustrated both on artificial data and on real life data from psychological diagnostics setting (outputs of the classifier proposed in **X**). A "don't know principle" approach is briefly stated and discussed. J. Stoklasa is the main author of **IV**, wrote most of the text, proposed the modification of the ROC and performed the simulations.

$\emptyset$	empty set
$\mathbb{R}$	set of real numbers
$[a,\, b]$	closed interval; $a, b \in \mathbb{R}$
$(a,\, b)$	open interval; $a, b \in \mathbb{R}$
A, B	fuzzy sets
$A \cup B$	union of fuzzy sets
$A \cap B$	intersection of fuzzy sets
$A \subseteq B$	A is a subset of B
Deg$(A \subseteq B)$	degree to which A is a subset of B
Dist(A, B)	distance of A and B
$A \times B$	Cartesian product of fuzzy sets A and B
$R \circ S$	composition of fuzzy relations R and S
μ_A or $A(.)$	membership function of a fuzzy set A
$\mu_A(x)$, $A(x)$	degree of membership of x to A
$f:\ U \longrightarrow V$	mapping from a set U to a set V
f^{-1}	inverse function to a function f
$\mathcal{F}(U)$	set of all fuzzy sets on U
$A \in \mathcal{F}(U)$	A is a fuzzy set on U
$\mathcal{F}_N\,([a,\, b])$	set of all fuzzy numbers on $[a, b]$
(a_1, a_2, a_3)	fuzzy number A represented by a triplet of its significant values, Supp$(A) = (a_1, a_3)$, Ker$(A) = \{a_2\}$ (usually triangular shaped)
(a_1, a_2, a_3, a_4)	fuzzy number A represented by a quadruplet of its significant values, Supp$(A) = (a_1, a_4)$, Ker$(A) = [a_2, a_3]$ (usually rectangular shaped)
$\{{}^{A(x_1)}/_{x_1}, \ldots, {}^{A(x_n)}/_{x_n}\}$	fuzzy set A on a discrete universe $\{x_1, \ldots, x_n\}$
$\widetilde{a}$	fuzzy number representing the meaning of "about a"
Ker(A)	kernel of a fuzzy set A
Supp(A)	support of a fuzzy set A
A_α	α-cut of a fuzzy set A; $\alpha \in [0, 1]$
hgt(A)	height of a fuzzy set A
$Card(A)$	cardinality of a fuzzy set A
f_F	fuzzified mapping f
$\mathcal{P}(U)$	power set of U; set of all subsets of U
COG_A	center of gravity of A
$\mathcal{A}, \mathcal{B}, \mathcal{C}$	linguistic terms
$(\mathcal{V}, \mathcal{T}(\mathcal{V}), U, G, M)$	linguistic variable $\mathcal{V}$
$\mathcal{T}(\mathcal{V})$	set of all linguistic values (terms) of a linguistic variable $\mathcal{V}$

$M(\mathcal{C})$	meaning of a linguistic term $\mathcal{C}$
$\mathcal{R}$	rule base (linguistically defined function)
CI	consistency index of a matrix of preference intensities (Saaty)
CR	consistency ratio
RI_n	random consistency index of a matrix of order n

AFF	Number of people affected by a disaster (fuzzy number)
AHP	Analytic Hierarchy Process
AUC	Area Under Curve
CI	Inconsistency Index (Saaty)
CMRC	Current Medical Rescue Capacity
COA	Center Of Area
COG	Center Of Gravity
COM	Center Of Maxima
CR	Inconsistency Ratio (Saaty)
CWW	Computing With Words
CW1	Computing With Words - level 1
CW2	Computing With Words - level 2
ED	Explanatory Database
EMRS	Emergency Medical Rescue Services
FRB	Fuzzy Rule Base
GCL	Generalized Constraint Language
HR	Human Resource
HTC	Hospital Treatment Capacity
IS HAP	Information System for academic faculty performance evaluation
LHGA	Linguistic Hybrid Geometric Average
LOM	Left Of Maxima
LOMWGA	Linguistic Oredered Weighted Geometric Average
MCDM	Multiple Criteria Decision Making
MF	Membership Function
MMPI-2	Minnesota Multiphasic Personality Inventory - revised version
MOMI	Middle point Of the Mean Interval
MRC	Medical Rescue Capacity
MRS	Medical Rescue Services
NT1	Number of seriously injured people (fuzzy number)

OWA	Ordered Weighted Average
PA	Pedagogical Activities
R&D	Research and Development (RD in the evaluation of academic faculty performance)
ROC	Receiver Operating Characteristics
ROM	Right Of Maxima
RUV	Registry of artistic results (Registr Uměleckých Výstupů in Czech)
SH	Specialized Hospitals
WOWA	Weighted Ordered Weighted Average

PART I: OVERVIEW OF THE THESIS

CHAPTER I

Introduction

Decision making is a common activity in our everyday life. We are well suited for this purpose and as such we have developed a certain level of automatization when encountered with decision tasks. We may not be aware how many decisions need to be made at any moment, as many decisions are performed without our conscious activity - we simply reach a decision and act upon it (and usually in these cases we can not describe the process, how the decision was reached). Almost every activity involves decision making - what leg will make the first step today? Where do we step? How firmly do we have to grip our backpack to be able to lift it? Many similar decisions are made almost all the time. If our conscious attention was required for all these decisions, we would not be able to exist. And it is not only the "trivial" decisions that are performed automatically. Consider we are to be hit by a car - what do we do? We try to get out of its way. We do not spend time to think of what would be the best direction to jump, what consequences this might have in comparison with other possible courses of action, we simply do something to survive! Our brain decides for us (that is it is still a part of us that does the decision, but it is usually not the conscious one). What do these situations have in common? These are either tasks that need not be solved optimally (take the first step as an example) or tasks where any action (any decision to act) is better than inaction. In general these are situations where it makes no sense to devote time and our conscious activity to, or where there is no time to find the best solution and any solution that can be reached first is good. What follows from this is, that these solutions are made at risk of nonoptimality - either optimality does not matter or there is no time to achieve it. It is well possible that the ability of making quick but not necessarily optimal decisions enabled us to survive to this day. However, such decision making is possible only if at least some experience (or at least instinct) is available and some structure is recognized in the decision making situation. Such unconscious, quick and frequently imprecise decision making can not take place in a completely unfamiliar environment, in situations that are too complex (but yet not directly life threatening). It is also not appropriate when there is at least some (not necessarily much) time to make a "better" decision.

This is when the conscious decision making takes over. This is when it makes sense to start finding good or best decisions. This is also where mathematics can start being of some use to decision makers! Not that it would not be possible to construct a mathematical decision support system to tell us what to do to avoid being hit by a car. If we, however, consider the amount of information necessary to make a qualified decision, the time needed to process all the required inputs and to provide results and the format of the results that would be instantly comprehensible to the person

currently in danger, perhaps the only reasonable output would be "jump!" (that is if there would be enough time to get to this conclusion and to provide it to the person). Here we are getting to the first important issue we would like to stress in this thesis - the necessity to provide clear (which does not necessarily mean precise) and easy to understand results to decision makers. Something that linguistic modelling might be able to help with. The possibilities of achieving this goal will be discussed in the thesis in more details.

1.1 Objectives and research questions

When modelling systems in which human factor is involved, we need to be able to account for human experience and human ("expert") knowledge. A great deal of methods has so far been developed to build mathematical models from data, more and more data can be acquired and stored (and analysed and used to build models). In some situations, however, enough data might not be available. We might have just the expert knowledge to build on. And such knowledge has to be extracted to provide us with the information we need to build a mathematical model of the system - either through some guided interactive method, or through a dialogue with the person.

We need to realize that we understand our world, represent it and deal with it through language. Most of the knowledge transfer is made through words (and experience of course). It is words, perceptions and emotions we think in. And among these instances words allow us to share our view of the world, our emotions and experience with others (not in an exact manner, but in a close enough manner to work). And if this is how our knowledge is represented in our minds, we need tools to be able to deal with such representation mathematically. In fact we need to go even further than that - we need a whole methodology to be able to formally (mathematically) reflect such representation of the world around us.

There is also another part of mathematical modelling we need to consider when building models or decision support systems for practical use - the way we present the outputs to the users of the models, the interpretations we suggest, the information we provide concerning the models (in fact understandability, common sense compatibility, consistency and other factors are inherently included in the broad topic of "presentation of outputs"). And this issue becomes crucial when human factor is involved in the modelled system or in the decision making process. We need to realize that we do not build mathematical problems in a vacuum - in fact the models and their outputs might have huge impact on the systems at hand. And when impact is in question, the issue of responsibility rises inevitably. Although ethical issues in mathematical modeling (that is not in the sense of plagiarism which has been paid enough attention in the scientific circles so far, but in the sense of responsibility for the consequences of decisions that are made based on the outputs provided by our models) have been rather neglected recently, the onset of behavioral operations research (see e.g. Hämäläinen et al. [32]) confirms in our opinion the necessity of revisiting even these issues in mathematical modelling. This thesis will therefore strive to contribute to the discussion on ethical issues and responsibility in mathematical modeling and for the results an consequences of the use of mathematical models. The main research questions summarizing the ideas in the thesis can be formulated as follows:

What needs to be done and ensured to provide decision makers with mathematical models they can ***safely*** *use to facilitate their decisions, without compromising the authority of the decision maker in the process of reaching decisions? And how to ensure that the responsibility for the consequences*

of the decisions rests mainly with the decision maker and the decision making process is entirely under his/her control?

These questions are strongly connected with the design of mathematical models for decision support, expert knowledge representation, outputs representation, understandability of the outputs and the whole process of reaching outputs by the model, consistency of the model with "common sense", interpretability of the outputs and other topics that will be discussed in the thesis.

To contribute to the discussion on this topic we set the following two objectives:

- *to propose conditions or guidelines for the design of linguistic decision support models that allow the decision maker to remain responsible and in control in the decision making process (that is providing support for qualified decision making)* and
- *to discuss possible ways of outputs representation that provide the decision maker with as much information (not necessarily precise) on the outputs of the models as possible (and to develop new approaches to outputs representation in practical applications*),

that is apart from acknowledging the existence of these issues and discussing their significance. These objectives are set to enable more decision makers to make qualified and well informed decisions based on appropriately designed (and computed) outputs of mathematical models in the near future. Qualified decisions and accepting the responsibility for their consequences are possible only if the outputs of mathematical models are not misleading, the process of their computation can be verified (at least to some extent) by the decision maker and all the information potentially necessary to interpret the results is available to the decision maker. This includes all the assumptions of the model, limitations of its use, appropriate precision or uncertainty of the outputs and so on. The previously stated objectives of the thesis can be under much simplification summarized by Figure 1.1.

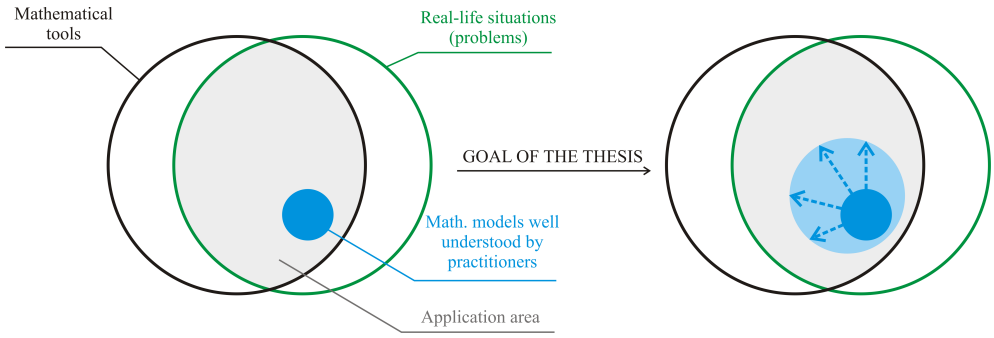


Figure 1.1: A graphical representation of two of the main objectives of the thesis - an expansion of the set of mathematical models for decision support that are well understood by laymen and provide results that are not against the "common sense" but still remain mathematically sound. This will require design of mathematical models with custom-made outputs, understandable and well representative interfaces between the model and its users and much understanding of the needs and requirements (and mathematical skills) of the model users.

There are many mathematical tools, theories and methods available and more are being developed. Many of these are applied to real life problems (although some of the real life problems either do

not need mathematical models, have not been modelled yet or can not be appropriately modelled by mathematical tools). I am well aware that to say that there are systems that mathematics can not model is perhaps too daring in a mathematical dissertation. On the other hand we need to consider that there is a great difference between a theoretical possibility of building a well fitting mathematical model and its actual usefulness (if we had all the resources to build it and to compute outputs). And we need to realize that in many cases there are opportunities for improvement in mathematical models - to achieve a user friendly and well working model with reasonable outputs is still a challenge.

On a general level the message of this thesis is the following: *for practical applications - that is for cases when mathematical models and their outputs will be used by non-mathematicians - we need to develop appropriate tools for expert knowledge extraction and representation, to provide interface with decision makers ("letting them see what is going on in the model") and most importantly to build models that provide outputs that can be actually used by the decision makers without the risk of misinterpretations. Such outputs need to provide information in a format that the decision makers can safely use, need to be understandable and in many cases intuitive. Linguistic and graphical outputs seem to be a promising direction on this way.*

In the spirit of the previously mentioned, we will also strive to fulfill the following objectives in this thesis:

- *to briefly summarize the state of the linguistic modelling for decision support in mathematics, with particular focus on the area of systems where human factor is involved*;
- *to contribute to the mathematical theory of linguistic (fuzzy) modelling for decision support and to suggest a unifying general view on the linguistic models for decision support, their design and connected issues including the ethical ones*;
- *to demonstrate the usability of linguistic modelling in real-life applications and decision making situations by presenting several working applications of linguistic models in various areas - ranging from the evaluation of works of art through disaster management to psychological diagnostics.* To do so, mathematical models and methods suitable for these situations had to be developed and are presented either directly in the text, or in Publications **I - XII**.

1.2 Scope

It is our aim to stress the importance of the proposed requirements on (linguistic) mathematical models - the importance of the decision maker in the whole process of designing decision support systems and models. The topics of expert knowledge representation, consistency of the mathematical and linguistic level of the models and appropriateness and clarity of outputs are therefore discussed on several places in the text and stressed in practical real life applications.

There are many mathematical tools that can be used for designing models of systems where human factor or language plays an important role. For example social science (where human factor and language as a means of communication are typical) adopted the statistical perspective long ago and many contributions to statistics have been motivated or directly originated in the field of humanities. Differential and difference equations are used to represent economical systems, quantitative linguistics have even adopted advanced mathematical tools to characterize language and text (e.g. fractal text analysis).

As the area of mathematical tools currently used to analyze and represent language and meaning and the area of mathematical models for multiple criteria and group decision making and evaluation is vast, we will focus only on a subgroup of the available methods, that is on methods

- capable of *providing decision support* and
- *able to deal with linguistic descriptions* of modelled systems and
- *applicable in real-life decision making situations* (this places some limits on the computational costs).

The thesis will therefore concentrate on methods and tools from fuzzy logic, fuzzy set theory and ordinal decision making, as these provide tools for expert knowledge representation, modeling of the meaning of linguistic terms and for providing easily interpretable (also linguistic) outputs. Multi-expert or group decision making and evaluation methods will not be considered explicitly - the focus of this thesis will be mainly on multiple criteria evaluation and decision making methods. This can be followed by an investigation of analogical issues in group decision making and evaluations and issues specific to this domain in future research.

We do not claim that the list of methods discussed here is complete, nor do we claim that the methods presented in this thesis are the only appropriate tools for the given area of interest. On the contrary - we acknowledge the need for other sophisticated mathematical tools (statistical, optimisation, etc.) in practical applications and in decision making. We want to contribute to the discussion on the appropriateness of mathematical methods used in particular settings, on the necessary development of new tools, on ethics in mathematical modelling, on responsibility - on the principles of mathematical modeling for practice in general. However, to make our point clear we need to narrow the scope of our investigation. We have selected some of the most widely used approaches to linguistic decision support, summarized them briefly and we use them in this thesis as a source of examples of the issues that linguistic modelling for decision supports must face. This choice allows the reader to find large amounts of practical examples of the use of these methods in various areas of human activity in the literature, to find examples of the issues discussed in this thesis from a familiar area and to consider the reasonability of the requirements set on linguistic mathematical models for decision support in this thesis.

We also hope that after reading this thesis, the reader will understand the reasons why at present point, linguistic modelling can not remain (or become, depending on the point of view) solely a mathematical discipline. Background in theoretical or applied mathematics, mathematical logic, linguistics are not enough to build models of sufficient quality for human users. Linguistic modelling will in our opinion require the development of mathematical "people skills" ranging from the ability to communicate well with the experts to describing the mathematical models comprehensibly, yet in sufficient details, to them. Experimental methodology will have to find its place within mathematics to confirm many of the assumptions our models have in specific situations. Providing understandable results of appropriate quality and (un)certainty from models that are not a "black box" is a logical prerequisite to transferring the responsibility for the decisions based on our mathematical models to the decision makers. We hope to contribute to these goals at least a bit as well in this thesis.

1.3 Structure of the thesis

The thesis is divided into three parts. *Part I* provides in the beginning of Chapter 2 a general overview on linguistic modelling for decision support. The modelling process is discussed in general, specific issues concerning linguistic modelling are discussed and a general approach to linguistic modelling for decision support is proposed. These guidelines for designing linguistic models for practice are formulated at the beginning of the thesis to explicitly state our view on linguistic (fuzzy) modeling for decision support. These guidelines will be further applied and discussed in the text of the thesis and in practical applications summarized in *Part II* and are also apparent throughout the Publications **I** to **XII** presented in *Part III* of the thesis.

Following the guidelines Chapter 2 provides an overview on the basic concepts of linguistic decision support ranging from the basics of fuzzy set theory (see e.g. [2, 11, 20, 22, 14, 50, 66, 89, 114, 131]) ordinal decision making (see e.g. [37, 107, 108, 109, 110, 113]) to linguistic fuzzy modelling ([2, 11, 14, 20, 22, 50, 66, 89, 114, 131]) and computing with words ([36, 45, 63, 112, 124, 125, 127, 128, 129]). Section 2.4 provides an overview of the classic approach to linguistic modeling using linguistic variables, linguistic scales, linguistic fuzzy rules and fuzzy inference. Methods for the construction of membership functions of fuzzy sets as well as linguistic approximation as crucial parts of the modelling process (at least from the linguistic modelling point of view) are discussed here.

Part II of the thesis presents examples of practical applications of the principles and guidelines for linguistic modelling outlined at the beginning Chapter 2. Chapter 3 summarizes a linguistic decision support model for the emergency medical rescue services in the Czech Republic. The multi-phase fuzzy decision support model was presented in Publication **I**. A heuristic approach to solve a fuzzy linear programming problem is described here and the concept of an $\alpha-$degree upper bound of a fuzzy number (introduced in Publication **I**) is applied to provide the decision maker with a means of expressing his/her attitude to violations of some constraints.

Chapter 4 discusses the linguistic level of Saaty's AHP method (see e.g. [80, 82, 83]) and the appropriateness of the linguistic labels of the elements of the fundamental scale (and its fuzzification proposed in [53] and further discussed in Publication **II** or **VIII**) in the context of Saaty's consistency condition. Adjustments to the linguistic labels and their meanings are proposed here. A weak consistency condition proposed and discussed in Publications **III** and **XI** is compared to the classic consistency condition proposed by Saaty and its connection with the linguistic level of the fundamental scale is discussed. Section 4.1 provides an example of the use of the weak consistency condition with large matrices of preference intensities - a methodology for the evaluation of works of art is proposed here (more can be found also in Publications **III**, **VII** and **XI** that describe the development of the evaluation model and the role of the weak consistency condition). Section 4.2 deals with the use of a fuzzified AHP [53] in the evaluation of scientific monographs (Publication **II** or **VIII**).

Chapter 5 provides insights to the use of linguistic modelling in HR management, a fuzzy rule-based model for academic faculty performance evaluation co-developed by the author that is currently being used on several universities in the Czech Republic is summarized here (its development is also described in Publications **II**, **VIII** and **IX**). A specific approach to fuzzy inference to obtain outputs that can be easily interpreted and graphically represented is summarized here - see also Publication **II**.

Finally Chapter 6 provides an example of the use of linguistic modelling in humanities (this topic is

discussed in details in Publication **VI**). A linguistic model for the interpretation of multidimensional questionnaire data is introduced in section 6.1 (context of psychological diagnostics). This inspired a more thorough investigation of the issue of data quality in classifier performance assessment, which is discussed in section 6.2 and in Publication **IV**.

A discussion of the results obtained in the thesis and of the fulfillment of the given goals follows in Chapter 7. *Part III* consists of 12 publications authored or co-authored by the author of the thesis.

CHAPTER II

Linguistic (fuzzy) modelling

It has been almost 50 years since Zadeh's introduction of fuzzy sets in [114] and the subsequent introduction of linguistic variables and linguistic modelling in [119, 120, 121] that links the mathematical and the linguistic description of the modelled system into one model. This field continues to develop quite rapidly (see e.g. [1, 10, 11, 13, 17, 44, 47, 45, 89] and many more examples both of theoretical research and practical applications of linguistic modeling) and its focus is broadening to operations with more complex language units - computing with words and perceptions (see e.g. [36, 124, 127, 128, 129]) as advertised by Zadeh and many other authors is a good example of this. Apart from perfecting the computational part with mathematical representations of meanings of words, more complex representations of meaning are being developed and used - e.g. interval type-2 fuzzy sets advocated by Mendel and other authors - see e.g. [1, 63, 73, 86, 101, 105]). Yager and others (see e.g. [29, 110, 109]) have been considering ordinal approach to linguistic modeling, where knowing the meaning of the linguistic terms (that is usually the membership functions of the respective fuzzy sets) is not necessary and the computations are carried out based on the ordinal information that is available.

Although much progress has been done in the development of complex and innovative methods for computing with words and perceptions (or simply linguistic modelling), there still seems to be one step missing somewhere at the very beginning of this journey. And it is this step we would like to focus on in this thesis. Representing formally the meaning of words and language phrases is not a trivial task. It is true that language is the main means of communication for people. Its inherent uncertainty and overlapping boundaries of meaning enable easy and not too complicated communication and information exchange (although the price for this is the risk of misunderstanding and imprecision of information transfer). Humans got used to dealing with the world in imprecise terms and information that is "precise enough" is sufficient for us to understand, decide, act - simply to survive. That is true in many (not all) cases and for many (not all) people in many (not all) contexts.

On the other hand there are well known inter-individual differences in meanings of words (see any study on connotative meaning) and the meaning of words varies even for one person depending on the context. This makes the formal (linguistic mathematical) modelling even more demanding. In fact the possibility of generalising our models, of using them in similar setting and situations, is compromised by the fact that meaning of linguistic terms is dependent on the context, on the problem within this context, on the person dealing with the problem and on all other persons who are participating in finding the solution. What is also interesting to note is, that the generalizabil-

ity is also complicated by the fact that *a simple translation of a linguistic fuzzy model to another language is not enough to ensure the model will be working in the new environment*. Meanings of words need to be revised and some linguistic terms might not have their equivalents in the target language (hence the number of elements of a linguistic scale might change). That is linguistic fuzzy modeling provides a different level of abstraction than classic mathematical modeling. Linguistic fuzzy models are not only context-dependent, but also in a sense they are culture-dependent. This is an interesting feature for a formal mathematical model.

There are many linguistic and logical aspects of the modeling of meaning of words, that can not be discussed here. Nor it is our intention to do so. Here, we aim to point out several problematic issues encountered in our experience with building decision support models for practice and propose a suitable solution to these issues. In a very simplified manner one of the main ideas of this thesis can be summarized in the following way: "Modelling systems with human component where the knowledge of the system is available primarily in linguistic form (this way many systems involving decision makers unfamiliar with mathematics are included) can be successful only if the linguistic description is respected or at least considered in all steps of the modelling process, where it is possible. Not doing so may result in a model that does not represent the reality well, or that provides good, but incomprehensible results to the decision maker." Comprehensibility of outputs (their interpretability and hence usefulness) for the decision maker is crucial (see e.g. [46, 47] for a discussion of these issues in the area of linguistic data summaries). And here we pose the first more ethical or methodological, than actually technical question - *should a well suitable method be used to solve a problem and propose a solution that only someone well familiar with the model and its underlying theory can interpret correctly? Or should a different (possibly less fitting) method be used to provide comprehensible results?* If we, for this moment, suppose that we are sure, that we can interpret all the result of our mathematical models correctly, without any misinterpretations, without disregarding pieces of information, without any risk of confusion - can we explain and interpret the results to any potential decision maker? Can we make sure that he/she understood everything? Do we do this? And will the decision maker be able to interpret well a similar, but slightly changed (or substantially changed) output? If an answer to at least one of the questions is NO, than the responsibility for the decision that is finally taken is ours and not the decision maker's. This is, however, a bit of an ethical problem, as in many cases we are to provide support, not final decisions. In this case we need to either sacrifice the "best" method and "precise" (meaning here absolutely adequate) results, or we ask the decision maker to make a decision which can not be considered qualified (since he/she does not have all the information or the insight required). Fortunately, a first step to revisit also ethical issues of mathematical modelling has already been made by Hämäläinen et. al. in [32] by identifying the need of behavioral operations research and it is our hope that this thesis might also help to proceed in this direction a bit further. In our opinion the decision maker, his/her needs, capabilities and limitations should always be in the center of modelling for decision support.

To sum it up - in linguistic modelling not only mathematical skills and rigor are required, but also the ability to communicate, explain, confirm our assumptions and adapt our methods to suit the reality as well as possible while still maintaining a high level of rigor are necessary. In practical situations, linguistic modelling may lead to the necessity of finding a proper tradeoff between mathematical elegance and a reasonable level of understandability and usefulness of the results for the decision makers. Linguistic modeling has been about compromises since the very beginning - to formally represent the meanings of words, some information has to be sacrificed (either while defining the membership functions of fuzzy sets, interval type-2 fuzzy sets or other models of meaning), or by

expressing preferences. It is up to us to ensure that the compromises are not too large and that an optimal balance between what we loose (in terms of information or precision) and what we get (in terms of information value and usefulness of the outputs of mathematical models) is achieved.

Let us now consider several suggestions or requirements on linguistic (fuzzy) modelling that can be found in the literature before we state our set of suggestions. Wenstøp [104, p.102] set the following set of conditions on his descriptive decision making model (on the auxiliary language used in his model):

i. It should contain provisions for operating with linguistic values *more or less in the same way as in natural language*.

ii. It should be *easy to learn to use and to understand*.

iii. It should be deductive.

iv. It should be implemented on a computer so that deductions can be performed automatically.

v. It should be versatile enough to give a fair description of a reasonably large class of systems.

These goals were set in 1980 - an implementation of say a fuzzy logic or a linguistic fuzzy model on a computer is not a problem nowadays, many software products are available, including specialised ones developed for complex decision making tasks under uncertainty (see e.g. [97]). What is of interest for us here is the emphasis on understandability and on the similarity of the language we use to model reality with the linguistic description. This is one of the issues that will be stressed in our proposal of the modeling framework for linguistic fuzzy models. Fuzzy set theory and fuzzy logic also provide quite versatile tools for linguistic modeling. There are many papers on the issue of deduction and approximate reasoning with fuzzy rules (see e.g. [21] for a discussion of various forms of meaning and the respective mathematical representation of fuzzy rules). That is in theory we should be able to deal with a large class of problems using these tools. What is important to see is, however, that when we are modelling meanings of words, we can expect the need of adapting our model to particular situation/problem much more frequently than in a non-linguistic modelling setting. We can even expect, that a linguistic model might have to be adapted for its use in a different language environment - as the meanings of "equivalent" linguistic terms in two different languages can not be expected to correspond fully. We can, however, say, that the requirements iii. to v. can be met by the use of fuzzy set theory. Fuzzy set theory also gives us useful tools to meet the requirements i. and ii. We also require sufficient understanding of meaning to achieve at least a "more or less" correspondence of mathematical representation with the linguistic description. It is also interesting to note that the requirements themselves are in fact formulated as fuzzy statements (*more or less in the same way, easy to learn, versatile enough*,...). This also seems to imply that the task of linguistic modelling is something that stands at the border between mathematics, computer science, linguistics, psychology and many other fields of science. There is a great potential for synergical effects but also for misunderstandings here. An excessive focus on a subset of these involved sciences (and hence angles of view) seems dangerous - a balanced utilisation of knowledge and experience from all these can, on the other hand, bring us very close to the desired synergy. Recent findings seem to support this claim - for example Trillas [100] stresses the importance of experimental research to find out which of the fuzzy set theories and concepts are appropriate in which types of situations and stresses the need of "testing them against some linguistic reality". We

might also be forced to experimentally verify our current tools and conceptions and perhaps relax some of our assumptions to get closer to the meaning of language and its modeling.

As far as we are concerned, multidisciplinary experimental research involving both the "hard" and the "soft" points of view should be (re)introduced into mathematics at least in the field of linguistic modelling. We need to find out which concepts work in which situations, which are more in accordance with a particular type of decision makers and so on. In fact we need a different approach - a different methodology to deal with human subjects and their linguistic representation of the world via mathematical terms than we need to work with data sets. Selection of proper tools can not be done solely on mathematical or theoretical assumptions - e.g. selecting a proper meaning of an intersection or union of fuzzy sets is not a matter of mathematical feasibility - what should matter more is its closeness to the description of the system provided by the decision maker - the fit to reality. Also Delgado et. al. [19] refer to the necessity of the coherence of the mathematical model with common-sense knowledge. And the common-sense knowledge will be in most cases represented linguistically.

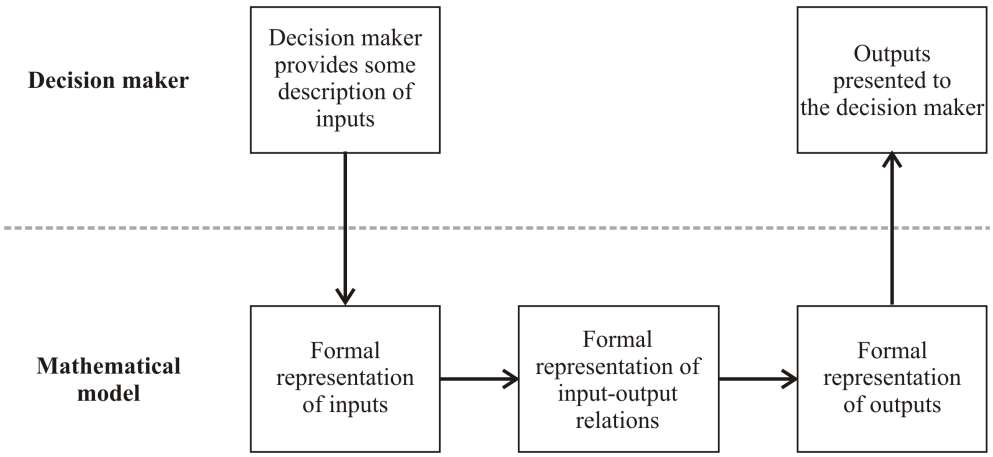


Figure 2.1: A diagram of the modelling process - general approach. Reproduced and modified from Publication **VI**.

Let us now consider an abstract representation of a modeling situation as shown in Figure 2.1. Let us consider, that we are able to identify the set of possible inputs and the decision maker has also specified the desired output. The relation between the inputs and the output(s) may not be known. Even more generally the set of possible output variables might not be known as well, only an outline of the decision we need to make based on the inputs may be present. When there is none or insufficient knowledge of the relation between inputs and outputs, it is up to us to find a way of finding at least a mathematical approximation of a possible relation. Many algorithms for rule extraction, machine learning and so on are available for such tasks. We just need the decision maker to be able to provide at least some input-output pairs to have a training set for our algorithm, or to specify based on what he is able to reach a decision. Under such circumstances, it is up to the mathematician to choose (or design) a proper modeling tool to find a relation between the inputs and outputs. Tasks that could be described by Figure 2.1 may include predictions of the behavior of the modelled system, finding a description of the mechanisms that guide the system's behavior and

similar. The whole modeling process can be divided into three parts, that will be further discussed in separate sections of this chapter:

- *Specification of inputs*, their availability, granularity, meaning and their *formal representation*. This also involves considering the uncertainty of the inputs and the role of the uncertainty in the process that is to be modelled.
- *Specification of outputs*, their form (numerical, linguistic, graphical), number, granularity, meaning, influence on the final decision and their *formal representation*. Interpretation of the results is either up to the decision maker, if these are self-explanatory (which is often not the case), or a suitable interpretation needs to be suggested by the mathematician.
- Choice of the modeling framework, methods and tools that allow us to obtain the predefined outputs based on the available inputs and the actual *design of the mathematical model*. As was already mentioned, this phase is usually the domain of mathematicians and when no information on the relation between the inputs and outputs is available, it remains with the mathematician as long as the model fits the reality well.

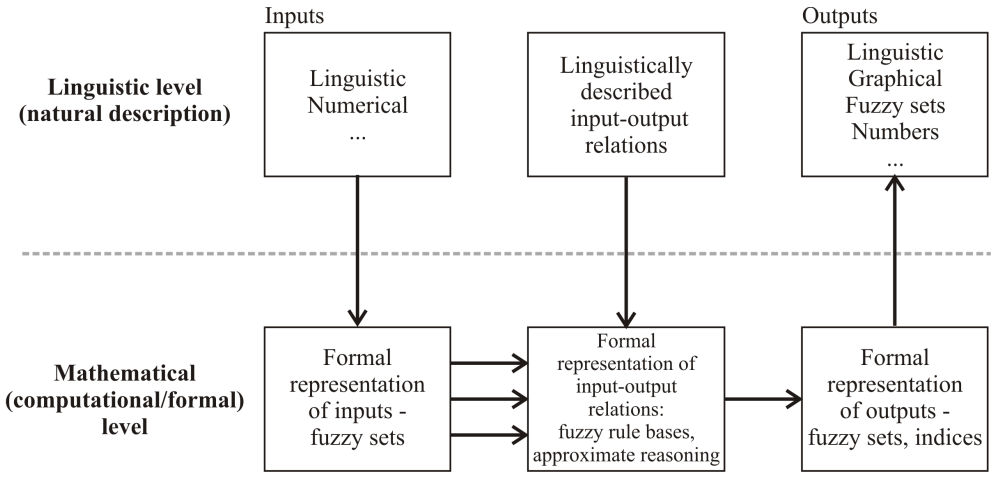


Figure 2.2: A diagram of the modelling process in multiple criteria decision making or evaluation, where description of inputs, outputs or their relationship is provided in a linguistic form. Reproduced and modified from Publication **VI**.

When, however, the relation between the inputs and outputs is known, we are closer to the field of multiple criteria evaluation and multiple criteria decision making. Let us from now on suppose, that the system we are modelling has a human component - that is at least a part of the description of the relationship between inputs and outputs, or the inputs or outputs are in linguistic form. This way we move to problem whose representation is summarized in Figure 2.2. This calls for the tools that will be briefly summarized in Sections 2.1 and 2.2 and further discussed in the following sections. The meanings of the linguistic terms used to describe the situation need to be clarified or at least specified to a level that allows some formal representation in the given context (and the context needs to be well specified to avoid misinterpretations). It is not always necessary to represent the meanings of words by fuzzy sets (or other tools of fuzzy modeling). In these cases (see e.g. [109])

at least the information on the ordering of the linguistic terms representing evaluations of attributes needs to be provided. The same is true for the outputs.

Although the diagram in Figure 2.2 now contains all the necessary elements, it is not yet complete. The linguistic modelling provides two levels of description - a linguistic level, that remains a good and comprehensible representation of the modelled system for the decision maker and the computational level, where formal representations are used and computations are performed. These two levels cannot exist separately in any step of the modelling process. What needs to be understood is

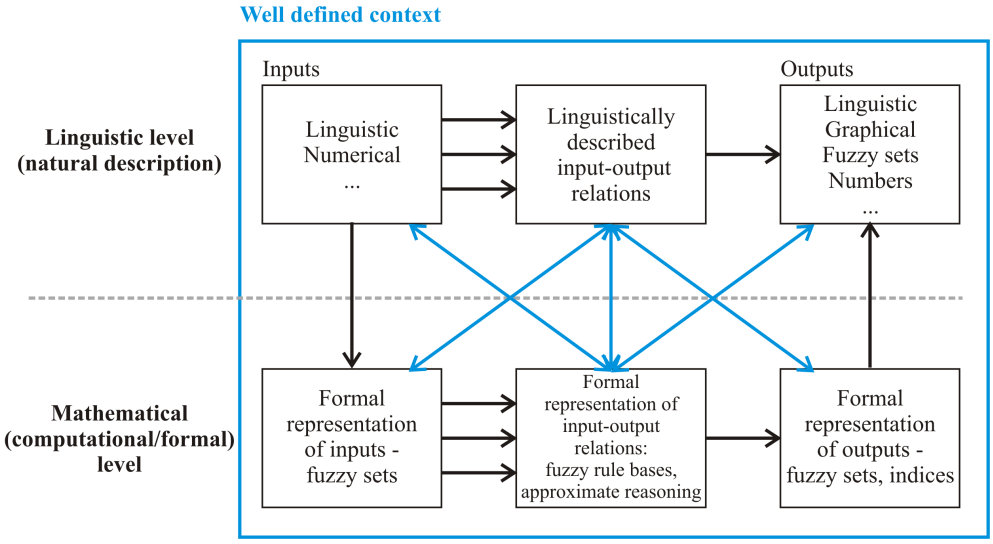


Figure 2.3: A diagram of our approach to the modelling process in multiple criteria decision making or evaluation, where description of inputs, outputs or their relationship is provided in a linguistic form. Apart from the well defined context, coherence of the mathematical representation on each step (when possible) with the linguistic description of the system is required. This requires to design the mathematical model in such a way, that the outputs of the formal level can be interpreted and easily understood in the "natural description" level. Reproduced and modified from Publication **VI**.

the fact, that the linguistic modelling is much more demanding as far as information from and interactions with decision maker are concerned. Trillas [100, p. 1484] supports this claim by stating that we should "...go deeper in the relations of fuzzy logic and language by *always identifying meaning with use*...". If we choose to leave out the decision maker or the linguistic level of the model from our considerations even for one step in the model, we might not be able to find a suitable and appropriate interpretation of the results. Any operation performed with the mathematical representations of the objects (e.g meanings of the linguistic terms) should not be in conflict with the linguistic description. In laymen terms, *what we do in the mathematical level must make sense in the linguistic level. If some operation or output seem counterintuitive to the decision maker when translated into the linguistic level, we are not representing the reality well enough.* That is we do not require absolute precision, we require something like the coherence with common-sense as discussed in [19]. The modeling situation becomes much more complicated from the modeling point of view, as is illustrated in Figure 2.3. On the other hand if contact with the linguistic level is maintained throughout the modeling process, the translation of the outputs of the model back into the natural

description level is straightforward and the outputs are provided in an easily interpretable form. The decision maker can understand how the results were obtained. The model is no longer a black box, modifications of the model are possible and the model is much more suitable as a decision support tool, as its functioning can be understood by the people using the model - at least at the linguistic level. In the following chapters, we will provide several examples of real life applications of linguistic modelling, where the importance of the contact with the linguistic level and the coherence of the model with common-sense will become more apparent. We suggest the following guidelines for linguistic modelling, particularly when the users of the decision support models are not well familiar with (or fond of) mathematics:

- The *context of the modeling needs to be specified and understood* by the decision maker and by the mathematician designing the model. A change of context may imply changes of meanings of the linguistic terms used to describe the system. Such change requires revision of the model before its use can continue.

- All the *meanings of the input terms* (or at least the ordering of the linguistic terms, if membership functions are not used) need to be *understood and accepted by the decision maker.* In fact it seems reasonable to require the meanings to be as intuitive for the decision maker as possible (particular examples can be found in **I**, where decision support during the first phase of disasters is considered - the importance of avoiding misinterpretations is obvious, when human lives are at stake). Particularly in case of those terms that are used to describe the functioning of the system. If the meanings are for some reason modelled differently than the decision maker expects (or is used to), we are risking misinterpretations and measures need to be taken to prevent them.

- *The decision maker has to understand the outputs of the model*. Particularly when imprecision or uncertainty is involved, providing numbers as output is not advised (at least our experience suggests against this practice). That is although defuzzification might be necessary in fuzzy controllers, in systems with a human component it is better to provide results that are self-explanatory and easy to interpret for the decision maker. Optimally, the outputs should be customized for a given decision maker to best serve the purpose. It is surprising that many decision makers require numerical outputs, although their interpretation is not easy and decisions based on these outputs can be biased (take a center of gravity representation of a fuzzy number for example). We propose to use graphical, color or linguistic outputs when uncertainty is present. The possibilities of graphical outputs from fuzzy decision support models are far from being exhausted at present. We will provide an example of this from the context of human resource management as proposed in **II** and discussed further in **VIII**.

- *No operations should be performed with the representations of the linguistic terms that would contradict common sense when transformed into the linguistic level*. This does not mean that we can model only predictable situations. This means that when any discrepancy between the reasonability of the linguistic and the corresponding computational operation occurs, it has to be clarified and resolved before proceeding to the next step. A good example of this is the consistency condition in Saaty's AHP and the introduction of the weak consistency condition discussed in **III**, **VII** and **XI**.

- *Black-box decision support systems*, that is systems that provide outputs without the user knowing how they were obtained (or at least based on what reasoning/rules) *are dangerous*

in the hand of laymen. Even more so when uncertain results are presented as seemingly certain. That is unless the users completely understand the results and there is no risk of wrong interpretation. If there is a way of preserving the uncertainty of uncertain results without compromising the decision support function, this should be at least attempted.

Just a remark to the last item - in many cases, we are used to decide based on uncertain or incomplete information. If information is insufficient to make a qualified decision, we should seek more information. If we provide decision makers in these situations with outputs that seem precise, although they are in fact only carriers of condensed meaning (as for example the mean value and dispersions are for random variables), sooner or later we can expect these values to be treated as precise. If we are not willing to share the responsibility for the decisions based on the outputs of our models, we should not provide results that might imply otherwise. If we for example rank the alternatives based on their fuzzy-number-evaluations using the center of gravity method and do not provide the decision maker with the fuzzy numbers, we have filtered the information for him. In situations where not much is at stake, this might be acceptable. In situations, where higher values are dealt with, we should, at least in my personal opinion, provide results that really leave the burden and responsibility of decisions with the decision maker. That is provide him/her with maximum information he/she can get concerning the situation, but not to make the decision for him. I for one would not like to loose a job in the future because the COG of my fuzzy evaluation was a bit lower than the one of my colleague's and an HR manager interpreted this fact in a way that I was worse. He might be right, but will he be able to justify his decision? I will conclude this remark with a simple statement, that might not be widely accepted, but which I deem very important - *decision support systems should provide support for decisions, not make them for decision makers*. Linguistic fuzzy modelling is, in my opinion, more than capable of doing so.

2.1 Basic concepts underlying (linguistic) fuzzy modelling

Before we begin our analysis of the various issues concerning linguistic modeling, let us first summarize the key concepts and unify the notation that will be used through the thesis. Let us begin with fuzzy sets as introduced by L. A. Zadeh in [114] in 1965.

Let U be a nonempty set (a universe of discourse). A *fuzzy set A* on U is defined by a mapping $\mu_A : U \to [0,1]$, where μ_A is called a *membership function* of A. For simplicity, we will denote a fuzzy set and its membership function by the same symbol in the text (that way the membership function of a fuzzy set A will be denoted $A(.)$). For a fuzzy set A and for any $x \in U$ we call the value $\mu_A(x) = A(x)$ a *degree of membership* of x to A. The set of all fuzzy sets on Uwill be denoted $\mathcal{F}(U)$. Clearly a membership function of a fuzzy set can be seen as a generalization of characteristic function of a set on the given universe. Crisp sets can therefore be represented by fuzzy sets in the following way. Let B be a crisp set on U and $\chi_B : U \to \{0,1\}$ its characteristic function, then B can be represented by a fuzzy set $\tilde{B}$ on U with a membership function $\mu_{\tilde{B}} = \chi_B$ for all $x \in U$.
Remark: It is well known that a fuzzy set A on U can be defined in a more general way, that is by a mapping $\mu_A : U \to L$, where L is a residuated lattice (see e.g. [18, 66]). That is the degrees of membership need not be real numbers from $[0,1]$. As this thesis deals with linguistic modeling, the interval $[0,1]$ however plays an important role, as it allows an easy interpretation of the degree of membership - a degree of compatibility of a given element of the universe with a fuzzy sets, which will be used to represent meanings of linguistic terms (as will be discussed later).

Several authors also in connection with the modelling of meanings of linguistic terms (expressions from natural language) recommend the use of a more complex construction - type-2 fuzzy sets [1, 63, 64, 86, 119]. A *type-2 fuzzy* set $\ddot{A}$ is a fuzzy set, whose membership degrees are fuzzy sets on [0,1], formally $\mu_{\ddot{A}} : U \rightarrow \mathcal{F}([0,1])$.

Let $A \in \mathcal{F}(U)$, the *kernel* of A is a crisp set $\mathrm{Ker}(A) = \{x \in U \mid A(x) = 1\}$. The kernel consists of such elements of the universe that are absolutely compatible with the fuzzy set (that is with the feature or linguistic label whose meaning it represents). The *support* of A is a crisp set $\mathrm{Supp}(A) = \{x \in U \mid A(x) > 0\}$. The support is a collection of such elements of the universe that are at least to some (understand nonzero) degree compatible with the fuzzy set. If the support of a fuzzy set A is finite, that is if $\mathrm{Supp}(A) = \{x_1, \ldots, x_n\}$, we can write $A = \left\{ {}^{A(x_1)}/_{x_1}, \ldots, {}^{A(x_n)}/_{x_n} \right\}$, this notation means, that the degree of membership of x_1 to A is $A(x_1)$, ... , and the degree of membership of x_n to A is $A(x_n)$.

An $\alpha-$*cut* of A is a crisp set $A_\alpha = \{x \in U \mid A(x) \geq \alpha\}$ for any $\alpha \in [0,1]$. An $\alpha-$cut of a fuzzy set A is such a subset of U that its elements are compatible with A at least to a degree α. A *height* of a fuzzy set A is defined as $\mathrm{hgt}(A) = \sup\{A(x) \mid x \in U\}$, it therefore corresponds with the supremum of α for which A_α is a nonempty set. There is an apparent connection between the membership function of a fuzzy set A and the system of its $\alpha-$cuts. This can be summarized by the following theorem proven in [65].

Theorem 2.1.1 (Representation theorem) Let $A \in \mathcal{F}(U)$. Then for any $x \in U$

$$A(x) = \sup\{\alpha \in [0,1] \mid x \in A_\alpha\}. \tag{2.1}$$

It follows from (2.1) that a fuzzy set A on U can be characterised by a collection of crisp sets - its $\alpha-$cuts. This is summarized by the following theorem, that introduces an alternative definition of fuzzy set that is equivalent to the Zadeh's definition stated at the beginning of this section (see [76] for the proof).

Theorem 2.1.2 Let us consider a nonempty set U. Then a fuzzy set A on U is understood as a collection of crisp sets $A_\alpha \subseteq U$, $\alpha \in [0,1]$ such that
1. $A_0 = U$

2. $A_\beta \subseteq A_\alpha$ for all $0 \leq \alpha < \beta \leq 1$

3. $A_\beta = \bigcap_{0 \leq \alpha < \beta} A_\alpha$,
where $\bigcap$ the standard intersection operator.

Let $A = \{A_\alpha\}_{\alpha \in [0,1]}$ be a fuzzy set on U. Then a mapping $\mu_A : U \longrightarrow [0,1]$ defined as $\mu_A(x) = \sup\{\alpha \in [0,1] | \, x \in A_\alpha\}$ is called a membership function of the fuzzy set A.

The fact that a fuzzy set can be represented by a system of crisp sets means that mathematical operations with crisp sets can be also used to define operations with fuzzy sets (that is some properties of crisp sets set can be generalized to fuzzy sets by requiring them to hold for all $\alpha-$cuts of the respective fuzzy set).

Let us consider $A, B \in \mathcal{F}(U)$. We say that A *is equal to* B if $A(x) = B(x)$ for all $x \in U$. A fuzzy set A is a *(fuzzy) subset of a fuzzy set* B, formally $A \subseteq B$, if $A(x) \leq B(x)$ for all $x \in U$. The fuzzy subsethood might be an important notion in linguistic modelling, when meanings of generic (general) and subordinate (specific) terms are considered. In some cases, we might need the

meaning of "very young" to be a fuzzy subset of "young" - we will discuss these issues more when linguistic hedges are introduced. As will become apparent further in the text, e.g. for some problems (classification tasks) and also in rule base design it might be useful to model the meanings of pairs of linguistic terms such as "young" and "very young" as separate fuzzy sets that partially overlap but neither one is a fuzzy subset of the other. Again we need to deal with the fact that each problem requires many decisions concerning the correct way of modelling the linguistic terms used in its description. There are no "absolute truths" and in many cases theory has to retrieve a bit in order to make place for the demands of practice. We do not claim that mathematical modeling should loose its rigor. On the contrary! Our aim is to point out that each new problem might need a development of a new way of thinking, dealing with meanings and new tools to reflect reality by mathematical means well enough while still formally correctly. The assumptions stemming from linguistics, mathematical logic and many other disciplines need to be questioned and their acceptance justified. The level of meaning is a level of human systems and it needs to be approached as such - context and many specific issues need to be accounted for before the meaning is properly captured.

As fuzzy sets will be used to represent meanings of linguistic terms or properties of objects, basic set operations with fuzzy sets need to be defined. Even here we encounter the specific feature of fuzzy set theory and particularly of linguistic fuzzy modeling that uses fuzzy sets to model meanings of linguistic terms. It is the fact that both intersection and union of fuzzy sets can be modelled not by one, but by a family of binary operations. Out of these families of binary operations (T-norms and T-conorms are used) we need to choose those that fit most the particular situation. Usually a requirement of using mutually dual norms and co-norms is present. The reason for this is e.g. that we require the DeMorgan's laws to hold ($A \cup_o B = \overline{(\bar{A} \cap_o \bar{B})}$, where A, B are fuzzy sets, $\cup_o$ is a generalized union, $\cap_o$ is a generalized intersection and the bar over symbols denotes a negation). There are however authors (see e.g. [100]) that are starting to question this practice in particular applications and starting to explore non-dual theories.

Definition 2.1.3 (Generalized intersection of fuzzy sets) Let A and B be fuzzy sets on U. Their *generalized intersection* $T(A, B)$ is also a fuzzy set on U, such that $T(A, B)(x) = i(A(x), B(x))$, for all $x \in U$, where i is a mapping $i : [0, 1] \times [0, 1] \to [0, 1]$ with the following properties:
1) $i(\alpha,\ 1) = \alpha$, for all $\alpha \in [0, 1]$ (*boundary condition*)
2) $i(\alpha,\ \beta) = i(\beta,\ \alpha)$, for all $\alpha, \beta \in [0, 1]$ (*commutativity*)
3) $i(i(\alpha,\ \beta),\ \gamma) = i(\alpha,\ i(\beta,\ \gamma))$, for all $\alpha, \beta, \gamma \in [0, 1]$ (*associativity*)
4) if $\alpha \leq \alpha'$ and $\beta \leq \beta'$, then $i(\alpha,\ \beta) \leq i(\alpha',\ \beta')$, for all $\alpha, \alpha', \beta, \beta' \in [0, 1]$ (*monotonicity*)
Each binary operation that fulfills the properties 1) - 4) is called a *triangular norm* or a *T-norm* for short. These four conditions will be called the *axiomatic skeleton of a T-norm*.

To restrict the class of T-norms, additional requirements can be formulated - among the most common ones (see e.g. [50]) are:

5) i is to be continuous (*continuity*),
6) $i(\alpha,\ \alpha) < \alpha$, for all $\alpha \in [0, 1], \alpha \neq 1$ (*subidempotency*),
7) if $\alpha < \alpha'$ and $\beta < \beta'$, then $i(\alpha,\ \beta) < i(\alpha',\ \beta')$, for all $\alpha, \alpha', \beta, \beta' \in [0, 1]$ (*strict monotonicity*).

A continuous T-norm satisfying the requirement of subidempotency is called an *Archimedean T-norm*.

A generalized union of fuzzy sets can be defined analogically.

Definition 2.1.4 (Generalized union of fuzzy sets) Let A and B be fuzzy sets on U. Their *generalized union* $S(A, B)$ is also a fuzzy set on U, such that $S(A, B)(x) = u(A(x), B(x))$, for all $x \in U$,

where u is a mapping $u : [0,1] \times [0,1] \rightarrow [0,1]$ with the following properties:
1) $u(\alpha,\, 0) = \alpha$, for all $\alpha \in [0,1]$ (*boundary condition*)
2) $u(\alpha,\, \beta) = u(\beta,\, \alpha)$, for all $\alpha, \beta \in [0,1]$ (*commutativity*)
3) $u(u(\alpha,\, \beta),\, \gamma) = u(\alpha,\, u(\beta,\, \gamma))$, for all $\alpha, \beta, \gamma \in [0,1]$ (*associativity*)
4) if $\alpha \leq \alpha^{'}$ and $\beta \leq \beta^{'}$, then $u(\alpha,\, \beta) \leq u(\alpha^{'},\, \beta^{'})$, for all $\alpha, \alpha^{'}, \beta, \beta^{'} \in [0,1]$ (*monotonicity*)
Each binary operation that fulfills the properties 1) - 4) is called a *triangular conorm* or a *T-conorm* for short. T-conorms are also denoted as S-norms in the literature. Conditions 1) - 4) are called the *axiomatic skeleton of an S-norm*.

The difference between a T-norm and an S-norm is in the boundary condition. The conditions of continuity, subidempotency and strict monotonicity can be defined analogically to the T-norm case. Usually, a T-conorm (S-norm) that is dual to a given T-norm has to fulfill the following condition:

$$S(\alpha, \beta) = 1 - T(1 - \alpha, 1 - \beta) \text{ for any } \alpha, \beta \in [0,1]. \tag{2.2}$$

The concept of duality can however be introduced on a more general level by defining a strong negation. Any strong negation can then be used in the duality formula (2.2) to substitute the standard fuzzy negation that can be expressed as $\bar{A}(x) = 1 - A(x)$ for all $x \in U$, where $A, \bar{A} \in \mathcal{F}(U)$.

Definition 2.1.5 (Strong negation) Let A be fuzzy a set on U. A *strong negation* of A, $N(A)$ is also a fuzzy set on U, such that $N(A)(x) = n(x)$, for all $x \in U$, where n is a mapping $n : [0,1] \rightarrow [0,1]$ which satisfies the involutive property $n(n(x)) = x$ for all $x \in U$, is continuous and strictly decreasing. It is also assumed that $n(0) = 1$ and $n(1) = 0$.

Obviously for all $x, y \in U$ it holds that if $x < y$ then $N(x) > N(y)$. For a more elaborate discussion on T-norms, S-norms and negations, we refer to [22, 28, 48, 50, 65, 70]. Among the most widely used pairs of dual T-norms and S-norms, we can recall e.g.:

- *Gödel T-norm* (also called the minimum T-norm): $T_{\min}(\alpha, \beta) = \min\{\alpha, \beta\}$
  and
  Gödel S-norm (also called the maximum S-norm): $S_{\max}(\alpha, \beta) = \max\{\alpha, \beta\}$
- *product T-norm*: $T_{\text{prod}}(\alpha, \beta) = \alpha \cdot \beta$
  and
  probabilistic sum S-norm: $S_{\text{sum}}(\alpha, \beta) = \alpha + \beta - \alpha \cdot \beta$
- *Łukasiewicz T-norm*: $T_{\text{Łuk}}(\alpha, \beta) = \max\{\alpha + \beta - 1, 0\}$
  and
  Łukasiewicz S-norm: $S_{\text{Łuk}}(\alpha, \beta) = \min\{\alpha + \beta, 1\}$.

The choice of a proper pair of dual T-norm and S-norm for a particular problem has received some attention in the literature (see e.g. [21, 100]). In linguistic modelling the appropriateness of choice has to be assessed against the reasonability of conclusions (and their interpretation) that are suggested by the chosen T-norm and S-norm pair.

We shall now remind the extension principle (see [119] or other publications on fuzzy set theory such as [22, 50, 66, 76, 131]) to have means of extending classical functions to be used to describe relationships between fuzzy objects.

Definition 2.1.6 (Extension principle) Let f be a mapping $f : U \to V$. A *fuzzification of this mapping* is a the mapping $f_F : \mathcal{F}(U) \to \mathcal{F}(V)$ such that assigns to any $A \in \mathcal{F}(U)$ a fuzzy set $f_F(A) \in \mathcal{F}(V)$ with a membership function defined as follows:

$$f_F(A)(y) = \begin{cases} \sup\{A(x) \mid y = f(x), x \in U\}, & \text{if } f^{-1} \neq \varnothing \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Alternatively (see [76]), the extension principle can be formulated using the $\alpha-$cuts in the following way. Let us consider a mapping $f : \mathcal{P}(U) \to \mathcal{P}(V)$, where $\mathcal{P}(U)$ and $\mathcal{P}(V)$ are power sets of U and V respectively. Its fuzzification is a mapping $f^F : \mathcal{F}(U) \to \mathcal{F}(V)$ defined for each $A \in \mathcal{F}(U)$ as $f^F(A)(y) = \sup\{\alpha \in [0,1] \mid y \in f(A_\alpha)\}$ if such A_α exists and 0 otherwise. As proven in [76] both these representations of the extension principle are equivalent.

Let $A \in \mathcal{F}(U)$ and $B \in \mathcal{F}(V)$. We can generalize the classical concept of Cartesian product of sets to a *Cartesian product of fuzzy sets* in the following way. The Cartesian product of A and B is a fuzzy set on $U \times V$ with a membership function defined for any $(x, y) \in U \times V$ as $(A \times B)(x, y) = \min\{A(x), B(y)\}$. It is straightforward to generalize this notion for a Cartesian product of n fuzzy sets. An n-ary fuzzy relation on a universe $U_1 \times \cdots \times U_n$ is any fuzzy set $R \in \mathcal{F}(U_1 \times \cdots \times U_n)$. Let $Q \in \mathcal{F}(U \times V)$ and $R \in \mathcal{F}(V \times W)$ then the composition of these two fuzzy relations is a fuzzy set $(Q \circ R) \in \mathcal{F}(U \times W)$ with a membership function defined for any $(x, z) \in U \times W$ as $(Q \circ R)(x, z) = \sup_{y \in V} \min\{Q(x, y), R(y, z)\}$.

In linguistic fuzzy modeling as well as in multiple criteria decision making the membership degree of an n-tuple $(x_1, \ldots, x_n)$ to R, $x_1 \in U_1, \ldots, x_n \in U_n$, can be interpreted as the strength of the relationship modelled by$R$ between $x_1, \ldots, x_n$. A binary fuzzy relation R defined on $U \times U$, that is reflexive ($R(x, x) = 1$, for all $x \in U$), symmetric ($R(x, y) = R(y, x)$, for all $x, y \in U$) and transitive ($R(x, z) \geq \sup_{y \in U}\{\min\{R(x,\ y),\ R(y,\ z)\}\}$, for all $x, y, z \in U$) is called a *fuzzy equivalence relation*. A reflexive and transitive relation, that is antisymmetric ($(R(x,\ y) > 0) \wedge (R(y,\ x) > 0) \Rightarrow x = y$, for all $x, y \in U$) is called a *fuzzy partial ordering relation*.

It might be interesting to note here, that e.g. for the classic Saaty's AHP method (see [80, 77]) based originally on a multiplicative matrix of preference intensities it is possible to transform the multiplicative matrix into an additive one that can be viewed as a fuzzy relation representing the preferences of the experts.

We now need to define a special type of fuzzy sets that are usually used to represent uncertain quantities and amounts - fuzzy numbers.

Definition 2.1.7 (Fuzzy number) Let B be a fuzzy set on $\mathbb{R}$, such that all the following conditions are met
1) B is normal, that is $\text{hgt}(B) = 1$
2) B_α is a closed interval for all $\alpha \in (0, 1]$
3) $\text{Supp}(B)$ bounded
then B is called a *fuzzy number* on $\mathbb{R}$, denoted as $B \in \mathcal{F}_N(\mathbb{R})$.

Condition 3) is in some contexts not present in the definition of the fuzzy number, this way allowing all the real numbers to have a nonzero membership degree to a given fuzzy number. This can be advocated by the uncertainty or inability of determining the lower and upper bound for the support of the respective fuzzy number. Uncertain quantities will be used in our models as well (see e.g. the EMRS decision support system application presented in Publication **I**). From our experience

experts are able to provide at least some bounds for the support of the respective fuzzy numbers by specifying which values are in their opinion not compatible at all with the concept/quality that the given fuzzy number should represent. We agree that an artificial crisp boundary is therefore introduced in the model by bounding the support of the fuzzy number. On the other hand it is well known that for example the human ability to discriminate among very low values is low - hence omitting those elements of the universe with very low membership degrees from the support of the fuzzy number will not introduce much error. What we gain by bounding the support of the fuzzy number is an easy representation of its important characteristics. We can now represent each fuzzy number $B \in \mathcal{F}_N(\mathbb{R})$ by a quadruple of characteristic values $B \sim (b_1, b_2, b_3, b_4)$, where $b_1 \leq b_2 \leq b_3 \leq b_4$ and $(b_1, b_4) = \text{Supp}(B)$, $[b_2, b_3] = \{x \in \mathbb{R} \mid B(x) = 1\} = \text{Ker}(B)$ and $B(x) = 0$ for all $x \in (-\infty, b_1] \cup [b_4, \infty)$. If $[b_1, b_4] \subseteq [a, b]$ we call B a *fuzzy number on an interval* $[a, b]$. The set of all fuzzy numbers on an interval $[a, b]$ will be denoted $\mathcal{F}_N([a, b])$.

To fully characterize a fuzzy number $B \in \mathcal{F}_N([a, b])$, $B \sim (b_1, b_2, b_3, b_4)$, we need to specify the shape of the left part of the membership function ($B_L(x)$) between b_1 and b_2 and the right part of the membership function ($B_R(x)$) between b_3 and b_4. In accordance with theorem 2.1.2 we can now use the pseudoinverse functions to $B_L(x) : [b_1, b_2] \to [0, 1]$ and $B_R(x) : [b_3, b_4] \to [0, 1]$ (that is the functions $\underline{b}(\alpha) : [0, 1] \to [b_1, b_2]$ and $\overline{b}(\alpha) : [0, 1] \to [b_3, b_4]$, $\alpha \in [0, 1]$ respectively) to represent the fuzzy number B in the following way: $B = \{[\underline{b}(\alpha), \overline{b}(\alpha)]\}_{\alpha \in [0,1]}$. In this representation $[\underline{b}(\alpha), \overline{b}(\alpha)] = B_\alpha$ for all $\alpha \in (0, 1]$ and $[\underline{b}(0), \overline{b}(0)] = B_0 = [b_1, b_4]$. If $B_L(x)$ and $B_R(x)$ are linear functions, we call B a *linear fuzzy* number on $[a, b]$, more precisely if $b_2 \neq b_3$ we call B a *rectangular fuzzy number*, if $b_2 = b_3$ we call B a *triangular fuzzy number*. If $b_1 = b_2$ and $b_3 = b_4$ then B is a fuzzy representation of a crisp interval and if $b_1 = b_2 = b_3 = b_4$ then B is a fuzzy representation of a single real number. Thanks to the possibility of representing fuzzy numbers by their $\alpha-$cuts we can introduce arithmetics with fuzzy numbers by interval arithmetics (see e.g. [20, 50] for alternative ways based on the extension principle).

Fuzzy numbers can be utilized to capture the meaning of some linguistic terms (particularly those for which the universe of discourse can be represented by an interval). It means that at least one element of the universe must be fully compatible with the linguistic label whose meaning is modelled by the fuzzy set (there might be exceptions from this - for terms like "do not know", "information is missing" and so on). We require this so that at least one good representant of the meaning of the linguistic term exists. Linguistic terms for which no representant that is fully compatible with them exists (or is at least conceivable) are difficult, if not impossible, to interpret correctly. More discussions on these issues in those frameworks that use membership functions to capture meaning - e.g. Zadeh's framework of linguistic modeling can be found e.g. in [119, 120, 121, 122, 127, 128, 129].

Remark: In cases where discrete universes are present (also when the universe of discourse is not a subset of $\mathbb{R}$) we require the fuzzy set representing the meaning of a linguistic term to be at least normal (the same is usually required also in the Computing With Words (CWW) paradigm [112]). Formally we can define a discrete fuzzy number in the following way.

Definition 2.1.8 (Discrete fuzzy number [103]) Let A be a fuzzy set on $\mathbb{R}$. A is called a discrete fuzzy number if its support is finite, i.e. there exist $x_1, x_2, \ldots, x_n \in \mathbb{R}$, where $x_1 < x_2 < \cdots < x_n$ such that $\text{supp}(A) = \{x_1, x_2, \ldots, x_n\}$ and there exist natural numbers k, l with $1 \leq k \leq l \leq n$ such that:
1) $A(x_i) = 1$ for any natural number i, $k \leq i \leq l$; ($\{x_k, x_{k+1}, \ldots, x_i, \ldots, x_l\} = \text{Ker}(A)$),
2) $A(x_i) \leq A(x_j)$ for any natural number i, j such that $1 \leq i \leq j \leq k$,

3) $A(x_i) \geq A(x_j)$ for any natural number i, j such that $k \leq i \leq j \leq n$.

The ordering of fuzzy numbers is not a trivial task - many approaches can be found in the literature ranging from the ordering based on the centers of gravity through multi-stage methods based on various characteristics of fuzzy numbers to fuzzy rule based systems. Each proposal so far has its strengths and limitations and new methods are still being developed to suit best the demands of particular problems (more can be found e.g. in [3, 4, 7, 8, 13, 22, 50, 69, 76, 89, 116, 119, 131]).

Using fuzzy numbers we can introduce fuzzy partitions on intervals, thus enabling granularity. The main idea of this concept, that is very beneficial in linguistic modelling, is the representation of uncountable sets (intervals) by a finite set of object (fuzzy numbers), that each partially describe a part of the original set, even overlap partially. When linguistic labels are assigned to these fuzzy numbers, we obtain means for representing complex relationships by a finite number of rules in linguistic form (a reasonably simple but yet accurate enough approximation of the relationship can be obtained). The idea of granularization can be illustrated on the concept of a fuzzy scale (which can be found in the literature also as *Ruspini fuzzy partition* (see e.g. [78, 96]).

Definition 2.1.9 (Fuzzy scale) Let $A_1, \ldots, A_n \in \mathcal{F}_N([a, b])$. We say that these fuzzy numbers form a *fuzzy scale* of $[a, b]$, if the fuzzy numbers are numbered in accordance with their ordering and for all $x \in [a, b]$ the following holds

$$\sum_{i=1}^{n} A_i(x) = 1. \tag{2.4}$$

The property (2.4) ensures, that the membership of each element of the universe will be fully divided among two neighboring fuzzy numbers, or a full membership to one of them will occur. If the fuzzy numbers represent meanings of linguistic terms, then (2.4) translates in the fact, that any element of the universe can be fully described by the linguistic terms available (its full compatibility is full to one of the linguistic terms or divided between two neighboring linguistic terms, each of which partially fits the element as its description). As such Ruspini fuzzy partitions will be used as a basis for the structure of a linguistic scale.

Some issues concerning the scale proposed by Saaty [81] for inputting preferences in AHP and the linguistic descriptors used for this purpose are discussed Chapter 4 and in Publications **III**, **VII** and **XI**.

2.2 Several frameworks for linguistic modeling

Let us now consider a multiple criteria decision making problem where we are to choose a best alternative from a set of n alternatives $\{A_1, \ldots, A_n\}$. Let us suppose, that each of these alternatives can be evaluated according to each of m criteria $\{C_1, \ldots, C_m\}$. Let us also assume, that the decision maker provides the evaluations of each alternative against each criterion in linguistic terms from a predefined set of s linguistic terms $\{\mathcal{E}_1, \ldots, \mathcal{E}_s\}$ - this set could include terms like "very good", "average", "unsatisfactory" etc. Based on these evaluations, we need to determine an overall evaluation of each alternative and then choose the alternative with the best evaluation.

There are now several issues that deserve our attention. As evaluations are provided using linguistic terms, we need to know how strong the information that is provided actually is. We might be

able to assign to each linguistic term a crisp value or an interval of values (this is the case that is considered in standard multiple criteria decision making and hence will not be discussed here). We can also be aware of the thing that linguistically expressed evaluations are uncertain (in the sense of vagueness of their meaning) and hence their meaning might be represented by normal fuzzy sets or fuzzy numbers - in this case we are in the field of linguistic fuzzy modeling and decision making, linguistic variables and scales can be used (fuzzy sets, type-2 fuzzy sets are used to represent meaning of linguistic terms). This classic approach proposed by Zadeh [114, 119, 120, 121, 122] will be discussed later in section 2.4.

We can also suppose that the evaluation process (or a system we want to model) is described by more complex propositions - e.g. evaluations might be in the form "usually this is very good" or even more complex ones. In this case propositions can be seen as introducing constraints on implicit variables and inference is based on a constraint propagation algorithm. These issues will be briefly discussed in section 2.4.3 in the framework of computing with words and perceptions (see e.g. [36, 45, 61, 63, 124, 125, 127, 128, 129]). In both these frameworks (that would fall in the category of "extension principle based" approaches according to [19]) the meaning of the linguistic terms has to be specified - knowledge of the membership functions of fuzzy sets that represent the meaning of linguistic expressions is required.

On the other hand we can easily find examples of situations from our own experience, when the meaning of the words we use to evaluate alternatives might not be easy to specify (even the universe of discourse on which to define the meaning of the linguistic terms might not be apparent). It is however still possible to determine a desired output, even when the decision maker is unable to specify the meanings of linguistic terms he/she uses to evaluate each alternative against each criterion. In this case (that belongs to the category "symbolic" approaches according to [19]) the set of linguistic terms is required to be linearly ordered (for each pair of linguistic evaluations we need to know which one is better) and no more information is required. This is the case of ordinal decision making proposed by Yager (see [29, 107, 108, 109, 110]). As will be discussed in section 2.3, computing the overall evaluation of each alternative is limited to the use of such aggregation operators, that can function with ordinal scales, that is $\min$, $\max$, ordinal OWA [110, 113].

In general the more information is provided, the more complex mathematical methods can be used to aggregate partial evaluations - when the evaluation scale moves from ordinal to ratio or interval scales (see [77]), more information is required from the decision maker concerning the meaning of linguistic terms he uses, but more complex computations can be performed with the meanings. Linguistic decision support models therefore involve finding the best tradeoff between the precision of information required from the expert concerning vague meanings of linguistic terms (and the related complexity of the mathematical methods that can be used) and the ability of the resulting model to capture the reality well.

2.3 Ordinal linguistic modeling

Let us now consider the situation, where a decision maker provides evaluations of alternatives $\{A_1, \ldots, A_n\}$ with respect to the criteria $\{C_1, \ldots, C_m\}$ in a linguistic form. That is each alternative is evaluated according to each criterion using one linguistic term from the set of available evaluation terms $\{\mathcal{E}_1, \ldots, \mathcal{E}_s\}$. Let us also consider that the set of evaluation terms is linearly ordered (a natural ordering of the terms based on their meaning in the given language exists) and indexed according to this ordering, that is $\mathcal{E}_1 < \mathcal{E}_2 < \cdots < \mathcal{E}_s$, where $a < b$ means "a *is a worse*

evaluation than b" and $\{\mathcal{E}_1, \ldots, \mathcal{E}_s\}$ is considered to be an *ordinal linguistic evaluation scale*. The ordering of the terms of the scale provides sufficient information to determine overall evaluation of a given alternative [109, 110]. We need to stress here that in this approach to linguistic multiple criteria decision making, there is no need to specify the meaning of the linguistic terms or to model them formally. No excessive precision of the decision maker is required. As Yager states in [109] it allows one to escape the "*tyranny of numbers*". Such an approach does not require much information or precision from the decision maker. On the other hand only operations available for ordinal scales are admissible.

Let us now summarize Yager's methodology introduced in [107] and further developed and commented e.g. in [9, 109, 110]. For better clarity and without any loss of generality, let us assume that for $s = 7$ we have the following ordinal linguistic evaluation scale $\{\mathcal{E}_1, \ldots, \mathcal{E}_7\} = \{unsatisfactory,$
$very\_low, low, medium, high, very\_high, perfect\}$. It is obvious, that from the natural ordering of the linguistic terms it follows, that $\mathcal{E}_i > \mathcal{E}_j$ if $i > j$. There are two operators that are implicit in this scale, which are

$$\min(\mathcal{E}_i, \mathcal{E}_j) = \mathcal{E}_i \quad \text{if } \mathcal{E}_i \leq \mathcal{E}_j \tag{2.5}$$

$$\max(\mathcal{E}_i, \mathcal{E}_j) = \mathcal{E}_i \quad \text{if } \mathcal{E}_i \geq \mathcal{E}_j. \tag{2.6}$$

For any alternative A_k, $k = 1, \ldots, n$ we can obtain from the decision maker an evaluation vector $P_k = (\mathcal{P}_{k1}, \mathcal{P}_{k2}, \ldots, \mathcal{P}_{km})$, where $\mathcal{P}_{kl}$ is the linguistic evaluation of an alternative k according to criterion l, $\mathcal{P}_{kl} \in \{\mathcal{E}_1, \ldots, \mathcal{E}_7\}$ for any $k = 1, \ldots, n$ and $l = 1, \ldots, m$.

According to [109] the decision maker now defines a vector of importances of criteria using the ordinal linguistic evaluation scale $\{\mathcal{E}_1, \ldots, \mathcal{E}_7\}$. That is $I = (\mathcal{I}_1, \ldots, \mathcal{I}_m)$, where $\mathcal{I}_r \in \{\mathcal{E}_1, \ldots, \mathcal{E}_7\}$ for all $r = 1, \ldots, m$. It is, however, questionable, whether the same linguistic term set will be suitable to describe both partial evaluations of an alternative with respect to a criterion and importances of the criteria. If the use of the linguistic terms for both purposes is not counterintuitive for the decision maker, then there is no need to object to this approach. To aggregate the partial evaluations into an overall evaluation of an alternative with respect to the importances of the criteria, we now implement the following requirement *for all* criteria: *"if a criterion is important, then it should have a high score"*. We can see that the requirement is in a form of an implication and as such it can be rewritten into the following form *for all* criteria: "*either a criterion is not important, or it has a high score*". We see that we will need to define a negation for the elements of the ordinal linguistic evaluation scale. This can be done as

$$\text{Neg}(\mathcal{E}_i) = \mathcal{E}_{s-i+1} \quad \text{for all } i = 1, \ldots, s \tag{2.7}$$

We can see that such a negation behaves in a predictable way, as e.g. $\text{Neg}(medium) = medium$, $\text{Neg}(low) = high$ or $\text{Neg}(perfect) = unsatisfactory$. Using this negation, the computation of the overall evaluation $\mathcal{P}_k$ of an alternative k was described in [107, 109] for all $k = 1, \ldots, n$ as:

$$\mathcal{P}_k = \min_r \left\{ \max \left\{ \text{Neg}(\mathcal{I}_r), \mathcal{P}_{kr} \right\} \right\}. \tag{2.8}$$

This way it is possible to obtain a linguistic overall evaluation for each alternative $\mathcal{P}_k \in \{\mathcal{E}_1, \ldots, \mathcal{E}_7\}$ for all $k = 1, \ldots, n$. Each alternative will at the end be characterised by a linguistic evaluation, that

is easy to understand for the decision maker, no linguistic approximation is needed to obtain easily interpretable results. There are obvious limitations given by the finite set of the linguistic evaluations, we may therefor obtain several alternatives that are evaluated as best and further analysis might be needed to solve the decision making problem. We should also note that when the aggregation is defined by (2.8) we assume that all the criteria need to have good evaluation for an alternative to be evaluated well. In [9] a generalisation of the presented aggregation is suggested using S- and R-implication operators. [38] also discusses several possible approaches to the aggregation of linguistic (ordinal) information.

For decision making under ignorance (with unknown probabilities of the states of nature) with linguistic evaluations, Yager in [110] proposes an ordinal version of the OWA operator originally introduced in [108] to aggregate the linguistic evaluations of alternatives under different states of nature. This allows us to implement the well known decision making strategies under ignorance (the optimistic maximum of maxima criterion, pessimistic maximum of minima criterion, Hurwitz criterion etc.) into linguistic modeling with ordinal linguistic evaluation scales.

Definition 2.3.1 (Ordinal OWA operator [110]) Let $\mathcal{E} = \{\mathcal{E}_1, \ldots, \mathcal{E}_s\}$ be a set of linguistic terms, such that $\mathcal{E}_i > \mathcal{E}_j$ if $i > j$. A mapping $F : \mathcal{E}^n \to \mathcal{E}$ is called an *ordinal OWA operator* of dimension n if it has an associated weighting vector $\mathbf{w} = (w_1, w_2, \ldots, w_n)$, such that
1) $w_j \in \mathcal{E}$ for all $j = 1, \ldots, n$,
2) $w_j \geq w_i$ if $j > i$,
3) $\max_j\{w_j\} = \mathcal{E}_s$
and where

$$F(a_1, \ldots, a_n) = \max_j\{\min\{w_j, b_j\}\}, \tag{2.9}$$

where b_j is the jth largest of the a_j.

For example the optimistic strategy characterizing each alternative by its maximum achieved evaluation through all the states of nature would be represented by a weighting vector $\mathbf{w}_{opt} = (\mathcal{E}_s, \mathcal{E}_s, \ldots, \mathcal{E}_s)$, the pessimistic strategy characterising each alternative by its worst evaluation is represented by a weighting vector $\mathbf{w}_{pes} = (\mathcal{E}_1, \ldots, \mathcal{E}_1, \mathcal{E}_s)$. The definitions of several other aggregation operators for linguistic (ordinal) information such as LOWGA (linguistic ordered weighted geometric average), LHGA (linguistic hybrid geometric average) and can be found in [106], where an aggregation method for linguistic preference relations in a multiple-criteria multi-expert decision making setting is proposed.

To summarize, the ordinal (symbolic) approach to linguistic modelling allows us to perform computations directly with linguistic labels, provided that these are elements of linearly ordered scales. There is no need to linguistic approximations, as the output of the model is always an element of the linguistic scale. This approach is less demanding on the input information from the decision maker, our computation options are however limited to using operations based on negation, maximum and minimum. Let us therefore now consider the situations, where more information concerning the meaning of linguistic terms is available and normal fuzzy sets can be constructed to represent the meaning.

2.4 Linguistic modeling with linguistic variables

In [119, 120, 121] L. A. Zadeh introduced the concept of a linguistic variable as a means for modelling systems that are described linguistically. In general a linguistic variable is a variable, whose

values are words or expressions from a natural language (artificial languages can also be considered). Its formal definition follows.

Definition 2.4.1 (Linguistic variable [120]) A *linguistic variable* is characterized by a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is the name of the linguistic variable, $\mathcal{T}(\mathcal{V})$ is the set of its *linguistic values*, X is a universe on which the meanings of the linguistic values of $\mathcal{V}$ are defined as fuzzy subsets of X, G is a syntactic rule (usually in the form of a grammar) for generating the names of the values of $\mathcal{V}$ and M is a semantic rule which associates each term $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ with its meaning $A = M(\mathcal{A}) \in \mathcal{F}(X)$.

In the text of the thesis linguistic values of linguistic variables (linguistic terms) will be denoted, whenever possible, by capital calligraphic letters and the fuzzy sets representing their meaning by a plain capital letters. As was mentioned before, we will also for simplicity use the same symbol for the fuzzy set and its membership functions. We might add that when we use linguistic variables, the membership function and membership degree of an element of a universe to a fuzzy set on this universe can be interpreted as a *degree of compatibility* of the linguistic label the meaning of which the fuzzy set represents with the element of the universe. If for example a fuzzy set were to represent "*heavy objects*" on the universe describing weight (in grams), then the degree of membership of *700g* to the fuzzy set representing the meaning of *heavy objects* can be interpreted either as a measure to which a *700g* object possesses the attribute of "*heaviness*", or more intuitively as a *compatibility of the label "heavy object" with an object that has 700g*. That is the membership functions describe how well the linguistic term they represent describes a given element from the universe of discourse. It therefore makes sense to require that the fuzzy sets used to represent meanings of linguistic terms be normal. This way the existence of at least one "typical example" of an element of X fully compatible with the given linguistic term is ensured.

The concept of a linguistic variable links together the (intuitive) description of the variable by expressions from a natural language with the mathematical representation of their meaning. Compared with the previous approach to linguistic modeling, much more information is required from the decision makers to be able to construct the mapping $M : \mathcal{T}(\mathcal{V}) \rightarrow \mathcal{F}(X)$ that captures well the meaning of the linguistic terms. Several methods for the construction of fuzzy sets representing the meaning of linguistic terms will be summarized in section 2.4.1. We also need to realize that there are interindividual differences in meaning, the meaning of linguistic terms depends on context, situation, people involved as well as on the terms that are available in $\mathcal{T}(\mathcal{V})$.

In the ideal case, if the mathematical level (fuzzy numbers) represented completely appropriately the meanings of the linguistic terms and expressions, the decision maker would not have to deal with the mathematical level of the model and it would suffice for him to remain on the linguistic level. If however, we are not sure, whether the meanings of the linguistic terms are in line with what the decision maker thinks the meanings should be, interaction with the model solely through the linguistic level is not enough. In every application of linguistic modeling we should make sure, that the meanings are not "counterintuitive" for the decision maker and that they reflect well the intentions of a decision maker. When we design linguistic models that incorporate the meaning of linguistic terms, each new user has to be well acquainted with the model and the meanings that are being used, unless he/she participated on their definitions (or unless the definitions are fairly similar throughout the population - then again how do we know this?). When meanings are imposed on the decision maker by the mathematical model, we can never be sure whether the user really uses the linguistic terms in the meaning modelled within the system or whether his/her own interpretation of the meaning prevails. Linguistic models designed to be accessed through the linguistic level should

therefore be custom made for a given decision maker or at least well explained and adapted, when necessary to avoid confusions of meaning.

Regardless of these issues, linguistic modeling is, at least in our opinion, a very powerful tool. It provides means for information granulation - a linguistic variable with a finite set of linguistic terms defined over an uncountable universe of discourse can be used to introduce a fuzzy partition of the universe (see the following definition).

Definition 2.4.2 (Linguistic scale) Let $(\mathcal{V},\, \mathcal{T}(\mathcal{V}),\, [a,b],\, G,\, M)$ be a linguistic variable, let $\mathcal{T}(\mathcal{V}) = \{\mathcal{A}_1, \ldots, \mathcal{A}_n\}$ be the set of its linguistic terms and let $M(\mathcal{A}_i) = A_i,\, i = 1, \ldots, n$ form a fuzzy scale on $[a,b]$. Then $\mathcal{V}$ is called a *linguistic scale*.

In other words $A_1, \ldots, A_n$ are fuzzy numbers, and it holds that $\sum_{i=1}^{n} A_i(x) = 1$ for all $x \in [a,b]$, that is $A_1, \ldots, A_n$ form a Ruspini fuzzy partition of $[a,b]$. Linguistic scales can prove very useful, as each element from the universe X can be fully characterized by one or two linguistic labels - either one label is fully compatible, or the compatibility is fully divided between two linguistic labels. This way an uncountable universe of discourse can be characterized by a finite set of linguistic terms (or the respective fuzzy sets).

Depending on the problem we are dealing with, we might require the fuzzy numbers modelling the meanings of the linguistic values to be uniformly distributed along the universe or to have a more detailed partition of certain parts of the universe (e.g. in social science answers that are extreme are expected to be less uncertain, whether neutral answers are more uncertain, thus requiring the partition to have more values at the edges of the universe and less in the middle [1, 89]), to be non symmetrical and so on (see e.g. [38]). The required number of the elements of the scale may differ among problems, but also different decision makers may be used to a different level of detail. Herrera and Martínez [39, 60] suggest a 2-tuple linguistic model that enables different decision makers to use scales with different numbers of terms and still is able to aggregate the information provided by them and to interpret in back in terms that are understandable to all the decision makers involved. Massanet et al. [62] discuss a similar problem for discrete universes (meanings modelled by discrete fuzzy numbers).

Several structures have been proposed in literature to refine the initial granulation of the universe of discourse. In [96] an enriched and expanded fuzzy scale were proposed to enable more detailed description of the universe and particularly to provide refining linguistic terms for the last phase of linguistic modeling, that is the retranslation of our results back into the linguistic level - linguistic approximation.

Definition 2.4.3 (Enriched linguistic scale [96]) Let $(\mathcal{V},\, \mathcal{T}(\mathcal{V}),\, [a,b],\, G,\, M)$ be a linguistic variable, let the set of its linguistic values $\mathcal{T}(\mathcal{V})$ by composed of *a set of elementary terms:*

$$\mathcal{T}_{elem.}(\mathcal{V}) = \{\mathcal{T}_1,\, \mathcal{T}_2,\, \ldots\, ,\, \mathcal{T}_s\}\,,\; M(\mathcal{T}_i) = T_i \in \mathcal{F}_N([a,b]),\; i = 1,\, 2,\, \ldots\, ,\, s, \tag{2.10}$$

that form a linguistic scale on $[a,b]$ and of *a set of derived terms:*

$$\mathcal{T}_{der.}(\mathcal{V}) = \{\text{definitely}\, \mathcal{T}_1,\, \text{more or less}\, \mathcal{T}_1,\, \ldots\, ,\, \text{definitely}\, \mathcal{T}_s,\, \text{more or less}\, \mathcal{T}_s\}\,, \tag{2.11}$$

that is $\mathcal{T}(\mathcal{V}) = \mathcal{T}_{elem.}(\mathcal{V}) \cup \mathcal{T}_{der.}(\mathcal{V})$, where for the meanings of the derived linguistic terms

$$M\,(\text{definitely}\, \mathcal{T}_i) = T_i^- \in \mathcal{F}_N([a,b]),\, M\,(\text{more or less}\, \mathcal{T}_i) = T_i^+ \in \mathcal{F}_N([a,b]),\, i = 1,\, 2,\, \ldots\, ,\, s,$$

(2.12)

the following three conditions hold:

1) $T_i^-(x) < T_i(x) < T_i^+(x)$ for all $i = 2, 3, \ldots, s$, and $x \in (x_1^i,\, x_2^i)$,
2) $T_i^-(x) < T_i(x) < T_i^+(x)$ for all $i = 1, 2, \ldots, s-1$, and $x \in (x_3^i,\, x_4^i)$,
where $T_i \sim (x_1^i,\, x_2^i,\, x_3^i,\, x_4^i)$, $i = 1, 2, \ldots, s$. And

3) for an odd s the sequences of fuzzy numbers:

$$T_1^-,\, T_2^+,\, T_3^-,\, T_4^+,\, \ldots,\, T_s^-$$

and

$$T_1^+,\, T_2^-,\, T_3^+,\, T_4^-,\, \ldots,\, T_s^+$$

for an even s the sequences of fuzzy numbers:

$$T_1^-,\, T_2^+,\, T_3^-,\, T_4^+,\, \ldots,\, T_s^+$$

and

$$T_1^+,\, T_2^-,\, T_3^+,\, T_4^-,\, \ldots,\, T_s^-$$

form a fuzzy scale on $[a, b]$. Then we say that the values of $\mathcal{V}$ form an enriched linguistic scale on $[a, b]$.

We can see that the definition of the enriched linguistic scale does not assume any particular representation of the meaning of the operators "definitely" and "more or less". Obviously $T_i^- \subset T_i \subset T_i^+$ for all $i = 1, \ldots, s$. We require that from the meanings of the derived terms a linguistic scale can be constructed. This allows us to have a Ruspini fuzzy partition of the universe defined also by the sequences of the derived terms presented in the definition. As such each element of the universe can be described by a single label or two labels amongst which the compatibility of the element of the universe is divided.

Two *linguistic hedges* are used to derive new terms from $\mathcal{T}_{elem}(\mathcal{V})$ - "definitely" and "more or less". We expect these hedges here to be modelled by modifying the meanings of the original linguistic terms. To use the Zadeh's terminology introduced in [115, 117] and further discussed and elaborated in e.g. [13, 16, 55, 65, 66, 89, 119] we would expect the effect of "definitely" to be modelled by a *concentration* operator and the effect of "more or less" by a *dilation* operator. Zadeh proposed the following definition of concentration of a fuzzy set $A \in \mathcal{F}(U)$: $\mu_{CON(A)}(y) = \mu_A^2(y)$ for all $y \in U$ (which can also be denoted as $CON(A)(y) = A^2(y)$ for all $y \in U$). The result of a concentration operation is a fuzzy set that has the same support and kernel, but is less uncertain (uncertainty of $A \in \mathcal{F}([a, b])$ can be defined e.g. by its cardinality Card$(A) = \int_a^b A(x)dx$). The "more or less" could be represented by a dilation operator defined by Zadeh as $DIL(A)(y) = A^{0.5}(y)$ for all $y \in U$. Dilation makes the original fuzzy set more fuzzy, as would be expected of the meaning of the linguistic modifier "more or less".

We need to realize, that the definitions of dilation and concentration provided by Zadeh are somehow intuitive, but if these are applied with no empirical evidence demonstrating their appropriateness, we are arbitrarily affecting the uncertainty of the meanings of the predefined terms. This could make the retranslation back into the linguistic level (or linguistic approximation) difficult. As was stated in

the first section of the thesis, each step of building a fuzzy linguistic model should remain in contact with the linguistic level and the "intuitiveness" of the representation of meanings of linguistic terms should be maintained as well as possible. That is there is probably a good reason why *definitely* $\mathcal{A}$ should be less uncertain than $\mathcal{A}$ and why *more or less* $\mathcal{A}$ should on the other hand be more uncertain that $\mathcal{A}$. This follows from the natural understanding of the linguistic terms and as a general idea could be accepted. The mathematical representation as $A^2(y)$ however does not follow directly from the natural understanding (we simply do not think in exponential representations) and has to be therefore justified.

Simple translation into another language might change the meaning of the hedge. Let us for example assume, that a language A has a single term to describe "very" and we would like to use our model developed in a language A environment in another environment, where language B is used. Let us assume that there are 4 different expressions of "very" in the language B none of which completely fits the previous language A. In this case we either chose one of the available expressions from that language B which is close enough (in terms of meaning) to the original "very" and keep the fuzzy set modeling the meaning unchanged and thus risk that in some instances the meaning will not fit appropriately, or we need to construct new (well fitting) fuzzy set representing the meaning of the chosen linguistic term.

There is one more reason why, in our opinion, the modifying operators should not be applied automatically without any thought. Let us consider that we have a linguistic scale that describes the age of a person. This linguistic scale has two values - *young* and *old* (their meanings modelled on a universe [0,130] by fuzzy numbers, that form a Ruspini fuzzy partition of this universe). What we need to realize first is, that the meaning of the linguistic term *young* is not only dependent on the individual who is using the term and hence the model (a child might have a very different opinion of who is young than a person of 60), there is also serious context dependence (*old* might mean something different in a country with low life expectancy than in a developed country; this becomes even more apparent when defining the concept of *old* e.g. for moths and people). The meaning might also be dependent on the purpose of the model. These are well known (but sometimes not well reflected) issues that are inevitable when dealing with the meaning of linguistic terms.

We should also realize, that the meaning of the two elements of our linguistic scale is also influenced by the number of elements of the scale. It is important to notice that the meaning is not absolute - it can change depending on many variables. Let us for now assume, that we add a third value to the linguistic scale - *middle-aged*. The meanings of the terms *old* and *young* could remain unchanged, but this would probably not seem very intuitive to us now. The fact that another linguistic value was added means that we can now discriminate more at least at a part of the universe and hence the uncertainty of some of the values can be reduced. This changes nothing about our concept of who is *old* and who is *young*, but in this particular situation given the terms we can use we adjust their meaning so that their usefulness and discrimination power is maximized. We need to realize that adding or removing terms of the scale changes the meanings of the other terms. Now what if we added to our initial two terms *old* and *young* a third term - *definitely old*? Using automatically the the operator of concentration, we would get $M(definitely\ old) \subset M(old)$. This is in some cases an appropriate model. But if by the introduction of the term *definitely old* into the term set the decision maker starts classifying people into three distinct classes - young, old and very old, then a model for which $\mathrm{Ker}(M(definitely\ old)) \cap \mathrm{Ker}(M(old)) \neq \varnothing$ is not an appropriate model. What is even more important, when the decision maker provides descriptions concerning *old people*, we should know, whether these apply to *definitely old* as well, or whether *definitely old* are treated as a separate category. That is the choice of the representation of the meaning of the linguistic hedge

definitely here has to be done with respect to the way how the decision maker uses the linguistic terms. If a concentration operator was necessary e.g. to keep the structure of the linguistic scale, then a fitting linguistic label should be found for the result this operation - that is one that does not contradict the intuitive use of this term by the decision maker.

More operators representing linguistic hedges (e.g *very*, *slightly*, *highly; plus*, *minus* as called by Zadeh [117] accentuators and deaccentuators; *sort of* and so on, which can be modelled by modifications and combinations of the concentration, dilation, negation, intensification and normalization operators) can be applied directly to the fuzzy set to modify its membership function. We have already defined the operator appropriate for modelling the meaning of the linguistic hedge *not* as it can be represented by a strong negation. Some binary operators to model *and* and *or* have also been presented as T-norms and S-norms. More on the modeling and use of linguistic hedges can be found e.g. in [1, 38, 45, 55, 65, 89, 117, 119, 121].

For the purpose of representation of very uncertain information and for linguistic approximation purposes, the structure of an extended linguistic scale may also prove useful.

Definition 2.4.4 (Extended linguistic scale [96]) Let $(\mathcal{V},\, \mathcal{T}(\mathcal{V}),\, [a,b],\, G,\, M)$ be a linguistic variable with the set of linguistic values $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1,\, \mathcal{T}_2,\, \ldots\, ,\, \mathcal{T}_s\}$, $T_i = M(\mathcal{T}_i),\ i = 1,\, 2,\, \ldots\, ,\, s$, which defines a linguistic scale on $[a,b]$. A linguistic variable $(\mathcal{V}',\, \mathcal{T}(\mathcal{V}'),\, [a,b],\, G,\, M)$ defines an extension of $\mathcal{V}$ if its term of linguistic values $\mathcal{T}(\mathcal{V}')$ complies with the following:

1) the set of elementary terms of $\mathcal{V}'$ is given as $\mathcal{T}_{elem.}(\mathcal{V}') = \mathcal{T}(\mathcal{V})$

2) the rest of its linguistic values are defined as linguistic labels of the following fuzzy numbers:

$$T_1 \cup_L T_2,\, T_2 \cup_L T_3,\, \ldots\, ,\, T_{s-1} \cup_L T_s$$

$$T_1 \cup_L T_2 \cup_L T_3,\, T_2 \cup_L T_3 \cup_L T_4,\, \ldots\, ,\, T_{s-2} \cup_L T_{s-1} \cup_L T_s$$

$$\ldots$$

$$T_1 \cup_L T_2 \cup_L \ \cdots\ \cup_L T_s,$$

where "$\cup_L$" is the Łukasiewicz union.

3) assigning of linguistic values to the fuzzy numbers defined by 2), that is the mapping M^{-1}, follows these rules:

$$M^{-1}(T_1 \cup_L T_2 \cup_L \ \cdots\ \cup_L T_s) = M^{-1}([a,b]) = \text{"}\textit{indeterminate}\text{"} \tag{2.13}$$

and for all $i,\, j \in \{1,\, 2,\, \ldots\, ,\, s-1\}\,,\ i < j$, except for the case when $i = 1$ and $j = s$ we get:

$$M^{-1}(T_i \cup_L T_{i+1} \cup_L \ \cdots\ \cup_L T_j) = \text{"}\textit{from } \mathcal{T}_i \textit{ to } \mathcal{T}_j\text{"}. \tag{2.14}$$

2.4.1 Construction of membership functions

Let us now focus on the methods of obtaining the meaning of linguistic terms - that is on the construction of the membership functions of fuzzy sets. In linguistic modelling, fuzzy sets (frequently fuzzy numbers) can represent both uncertain quantities as well as the meanings of linguistic terms and expressions. One of the very first steps therefore has to be the definition of the mapping M, that assigns meaning to linguistic terms. This can be done either directly from data, if appropriate data is available in sufficient quantity and quality, or by extracting the meaning somehow directly

from the people that use the respective linguistic terms. Proper methods for the construction of membership functions (MF) of fuzzy sets are still being studied and developed in various fields of science - see e.g. [130] for a construction of MF representing spatial soil information using typical values/instances, descriptive knowledge and purposive sampling; [88] for a construction of MF of fuzzy sets representing time periods of given historical events based on data from the internet; [34] integrating human estimation, probability density functions and Shannon entropy to construct membership functions using mathematical programming; [42] designing MF of fuzzy concepts based on the context model and modal logic; [102] combining objective information and subjective opinions of experts using PERT and special relativity theory and many others.

We will summarize here briefly several more general approaches to the construction of MF. A naive approach would, for example, be to proceed from the end backwards - to define e.g a fuzzy partition on a given universe (with a desired granularity, uniformly distributed, symmetrical or not etc.) based on the requirements of the methods we plan to use. Then we would have to ask the decision maker "how would you call (something like) this?" while pointing at a particular fuzzy number and describing what information is conceived in its membership function (that is if we wanted to obtain a qualified and well informed answer from the decision maker). This way we might obtain mathematically well treatable objects with some linguistic label attached. The problem with such approach is, that the linguistic labels might not be frequent ones in the given language, they might also be too complicated (as a result of the process of trying to capture as much of the shape of the fuzzy set as possible). Although the model might get a linguistic level this way, the linguistic level might not work well.

If at least a *partial information on the membership degrees of some elements from the universe is available* (a data set containing samples of elements of the universe with assigned membership degrees), we can use this information to construct the MF (see e.g. [50, 87]). We can then use some interpolation technique to find a suitable polynomial function that fits the available data. We can also make an assumption on the required shape of the membership function (e.g. rectangular, triangular, bell-shaped etc.) expressed in a parameterized form and use some curve fitting algorithm to set the parameters of the curve to fit the data as well as possible (least-square methods can be used). Alternatively we can also use some learning algorithm to find the most suitable parameters or to construct the membership function (e.g. neural networks have been used for this purpose). In case of such algorithms as neural networks, the membership function is then represented by the neural network itself - after the learning phase, the neural network provides for each element of the universe a membership degree.

If no such data set is available, we need to construct the MF representing the meaning of a linguistic term $\mathcal{A}$ based on the *contact with the decision maker or a group of decision makers*. [50] distinguishes *direct* and *indirect* methods of the construction of membership functions. Direct methods involve asking questions on the compatibility of certain elements x of the universe U with a given linguistic term, the meaning of which we are trying to construct (see [87] for an overview).

Perhaps the most simple *direct method* would be to randomly select elements from the universe and ask the decision maker "how much $\mathcal{A}$ is x" (in fact how much is the linguistic term $\mathcal{A}$ - for example "*old*" - compatible with a particular element x of the universe - for example *11 years*). Membership degrees from [0,1] can be used directly, or some scoring scale (10 point, 5 point) can be used. In agreement with [87] we can call this method *direct rating*. If more decision makers are involved, their opinions can be aggregated (e.g. by a weighted average respecting the importance of each decision maker or by more complex methods). We can also reverse this method and ask the

decision maker for which elements from U is the compatibility with $\mathcal{A}$ equal to a given value.

Another method that might be considered direct is the use of the *semantic differential* proposed by Osgood [68]. This method used for the measurement of the connotative meaning (although measurement might be a strong term) of linguistic terms is well suited for abstract concepts and terms. The semantic differential method determines the position of a given term or object in the semantic space. Based on the assessment of the object using various pairs of bipolar adjectives, each object is represented by a vector in n-dimensional semantic space (in the original method, three dimensions were considered - evaluation, potency and activity). The degree of compatibility of a given element x from the universe (e.g. $x = Jan$) with the linguistic term $\mathcal{A}$ the meaning of which we are trying to find (e.g. $\mathcal{A} = brave$) can be determined by calculating a normalized distance of the vectors representing *Jan* and *brave* in the semantic space and subtracting the result from 1.

We can also ask the decision maker two questions - one concerning the kernel of the fuzzy set $A \sim (a_1, a_2, a_3, a_4)$ and one concerning its support. To determine the kernel of $M(\mathcal{A})$ that is an interval $[a_2, a_3]$ we need the decision maker to answer a question "which elements from U are completely compatible with $\mathcal{A}$?". To determine the support of $M(\mathcal{A})$ that is an interval (a_1, a_4) a question "which elements from U are not compatible at all with $\mathcal{A}$?" - these will not be in the support of $M(\mathcal{A})$. What remains is to define the shape of the membership function. Assuming that the universe U is an interval scale (see e.g. [77]) we can for example add linear functions connecting the points $(a_1, 0)$ and $(a_2, 1)$ forming the left part of the fuzzy number (A_L) and the points $(a_3, 1)$ and $(a_4, 0)$ forming the left part of the fuzzy number (A_R). If there is some information available on the expected shape of A_L or A_R, it can be utilized.

If more decision makers are available, each can be asked whether in his/her opinion "x is $\mathcal{A}$". If a sufficient number of decision makers is available, than the compatibility degree of x with $\mathcal{A}$ can be computed as the ratio of the number of positive answers to the number of decision makers that have been asked. Obviously, this "voting" strategy is of no use if individual meaning is to be determined. If however something like collective meaning is of interest, this method might provide useful results.

Direct methods are however very demanding on the precision of the decision maker. We need to obtain precise membership degrees of many elements of the universe. It is questionable, whether a decision maker is capable of such precision. If not some of his/her answers may be misleading, inconsistent, arbitrary etc. This might be an issue connected particularly with abstract concepts such as beauty, friendship - that can not be easily quantified. *Indirect methods* therefore derive the membership degree indirectly by asking questions that can be better assessed and answered by the decision maker. Instead of asking a decision maker about an exact degree to which x is $\mathcal{A}$, we can ask how much more $\mathcal{A}$ is x than y. Pairwise comparison methods (e.g. Saaty's AHP [80, 82]) can be applied on the elements of universe and their compatibility with $\mathcal{A}$ assessed pairwise. Elements of the priority vector computed by the AHP methodology (after normalization) then represent the membership degrees of each of the elements that were compared to the fuzzy set representing the meaning of $\mathcal{A}$.

Once we are able to assign fuzzy numbers as meanings of the linguistic terms, we can perform mathematical operations with the fuzzy numbers. We have already discussed how more complex expressions in natural language can be constructed using appropriate T-norms and S-norms and negations to model connectives and, or, not and various linguistic hedges. We have stressed the need for careful selection of proper mathematical tools to model the meaning of linguistic terms

and the connectives that are used to represent complex statements. We can also use interval arithmetics to perform computations with the meanings of the linguistic terms (add, subtract, multiply, divide the fuzzy numbers) and using the extension principle we can even use much more complex computations.

2.4.2 Fuzzy rules and rule bases

We can now proceed to the representation of expert knowledge that is obtained in a linguistic form. When we ask an expert (a decision maker) to explain how a given system works, we usually obtain a system of IF-THEN rules describing the relation between the inputs and outputs. A set of such rules describing a given system is called a *linguistic (fuzzy) rule base (FRB)*. In multiple criteria decision making and evaluation, the rules describe what values of criteria should lead to a good overall evaluation of a given object (and based on these overall evaluations the best object can be chosen), in classification tasks, linguistic fuzzy rule bases can be used to describe prototypes of each class (rules define a typical element of the class) and the fulfillment of the rules can be interpreted as membership degree to a given class. A more detailed discussion on the use of linguistic rule bases for classification purposes is provided e.g. in [40, 92] and several issues concerning its use in human sciences and psychology are discussed in Publications **VI** and **X**. Publication **IV** also discusses issues of (fuzzy) classifier performance assessment under variable quality of data. The concept of a linguistic fuzzy rule base can be summarized by the following definition.

Definition 2.4.5 (Linguistic fuzzy rule base) Let $(\mathcal{X}_j, \mathcal{T}(\mathcal{X}_j), U_j, G, M)$ for $j = 1, \ldots, m$ and $(\mathcal{Y}, \mathcal{T}(\mathcal{Y}), V, G, M)$ be linguistic variables. Let $\mathcal{C}_{i,j} \in \mathcal{T}(\mathcal{X}_j)$ be the linguistic values of the linguistic variable j and let their meanings $C_{i,j} = M(\mathcal{C}_{i,j})$ be fuzzy numbers on U_j for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Let $\mathcal{D}_i \in \mathcal{T}(\mathcal{Y})$ a $D_i = M(\mathcal{D}_i)$ be fuzzy numbers on V for all $i = 1, 2, \ldots, n$. Then $\mathcal{R}$:

Rule 1: If $\mathcal{X}_1$ is $\mathcal{C}_{1,1}$ and ... and $\mathcal{X}_m$ is $\mathcal{C}_{1,m}$, then $\mathcal{Y}$ is $\mathcal{D}_1$

Rule 2: If $\mathcal{X}_1$ is $\mathcal{C}_{2,1}$ and ... and $\mathcal{X}_m$ is $\mathcal{C}_{2,m}$, then $\mathcal{Y}$ is $\mathcal{D}_2$

...

Rule n: If $\mathcal{X}_1$ is $\mathcal{C}_{n,1}$ and ... and $\mathcal{X}_m$ is $\mathcal{C}_{n,m}$, then $\mathcal{Y}$ is $\mathcal{D}_n$

is called a *linguistically defined function* (*linguistic fuzzy rule base*) describing the relationship between linguistic variables $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ and $\mathcal{Y}$.

Linguistic fuzzy rule bases have proven to be a powerful tool in fuzzy control due to the granulation they provide (a (Ruspini) fuzzy partition is defined on each universe that characterizes a potentially uncountable universe be a finite set of linguistic terms, whose meanings are modelled by fuzzy numbers). The use of linguistic granules provides not a precise description of the system, but an approximate one. It represents a reaction to the incompatibility principle formulated by Zadeh e.g. in [118, p. 28]: "*As the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (relevance) become almost mutually exclusive characteristics.*" In this sense fuzzy rule bases capture the "essence" of the system by a small enough set of rules. Not all the details are reflected, but the description can be made *precise enough* to suffice for the purposes of controlling the system or of understanding the system. In many ways this representation is very close to how the world is represented in our minds. We use words with imprecise meanings (and undefined boundaries) whose interpretation is dependent on many variables and yet we are able to

exist well in this world. Such reasoning leads to the realisation, that approximate (or precise just enough) representation of systems might be enough to understand it.

Fuzzy control also benefits from the granularity issue - that is from the representation by a finite number of relatively simple rules. This way computations can be done quickly and results obtained in short (close to real) time in many cases. From the linguistic modeling point of view linguistic rule bases allow us to describe the world around us - various systems and processes - in words, that is in the most natural way we know. As our knowledge of the world is imprecise, the tools modeling it should be able to reflect this fact. But how exactly is the knowledge of the world represented in our mind? We have introduced the fuzzy rule base which describes a modelled system by IF-THEN rules. We need to understand that the IF-THEN relation does not necessarily mean that the rules expressed in natural language in the form "IF $\mathcal{X}$ is $\mathcal{C}$ THEN $\mathcal{Y}$ is $\mathcal{D}$" are implications.

In [21] various types of representations of fuzzy IF-THEN rules are discussed in more detail. Generally two main categories of fuzzy IF-THEN rules can be distinguished based on their interpretation [21, 121]. If we understand an IF-THEN rule as a special kind of an IF-THEN-ELSE rule, this distinction becomes apparent. Let us now assume, that $C \in \mathcal{F}_N(U)$ and $D, E \in \mathcal{F}_N(V)$, where $C = M(\mathcal{C})$, $D = M(\mathcal{D})$ and $E = M(\mathcal{E})$. A fuzzy rule "IF $\mathcal{X}$ is $\mathcal{C}$ THEN $\mathcal{Y}$ is $\mathcal{D}$" can now be rewritten into a more general form

$$\text{IF } \mathcal{X} \text{ is } \mathcal{C} \text{ THEN } \mathcal{Y} \text{ is } \mathcal{D} \text{ ELSE } \mathcal{Y} \text{ is } \mathcal{E}. \tag{2.15}$$

Obviously as the rule is provided in natural language, we know what to infer (what the result will be) if the antecedent part holds (that is if $\mathcal{X}$ is $\mathcal{C}$). But what should we infer if $\mathcal{X}$ is not $\mathcal{C}$? There are two possibilities.

Either *we do not know* (that is the information concerning the output is not available, hence $M(\mathcal{E}) = E = \varnothing$). In this case the rule, however, does not behave as an implication in the logical sense. In accordance with [21, 121] we can represent the rule (2.15) using the meanings of the respective linguistic terms as

$$(C \times D) \cup (\neg C \times \varnothing) = (C \times D), \tag{2.16}$$

where $\times$ is the Cartesian product of fuzzy sets, $\cup$ is a union operator and $\neg$ is a negation. Such rules can be interpreted (see [21]) as "the more $\mathcal{X}$ is $\mathcal{C}$, the more *possible* $\mathcal{Y}$ is $\mathcal{D}$". It is easy to see that such rules are represented as a fuzzy number on $U \times V$. As the linguistic level suggests, the rule does not contain any information on what happens, if $\mathcal{X}$ is not $\mathcal{C}$. Rather than an implication, it represents a *piece of data* - it can therefore in the context of expert knowledge extraction be understood as a *"prototypical experience"* (the expert knows of a situation when $\mathcal{C}$ and $\mathcal{D}$ occurred). Adding more rules to the rule base is then equivalent to adding more data into a data set or adding more experience into an experience pool. The mathematical representation of such *conjunction-based* rules should therefore behave as a data accumulation procedure. All the pieces of data put together then form the rule base. It is obvious that such knowledge is in a form that either one prototypical example will be used to derive results, *or* another one, *or* another one... The fuzzy rule base from definition (2.4.5) - a result of such data accumulation procedure is therefore best modelled by a union of fuzzy sets:

$$\mathcal{R} = \bigcup_{i=1}^{n} (\mathrm{C}_{i,1} \times \ldots \times \mathrm{C}_{i,m} \times D_i)\,. \tag{2.17}$$

On the other hand, the knowledge provided by the expert (decision maker) might be in the form of a proper implication, that is if $\mathcal{X}$ is not $\mathcal{C}$ we might conclude that *anything is possible*. This means

that $M(\mathcal{E}) = E = V$. The fuzzy rule (2.15) can now be rewritten in the form

$$(C \times D) \cup (\neg C \times V) = (C \rightarrow D). \tag{2.18}$$

where $\rightarrow$ is a fuzzy implication (e.g. Lukasiewicz implication $I_L(\alpha, \beta) = \min\{1, 1 - \alpha + \beta\}$, Kleene-Dienes implication $I_{KD}(\alpha, \beta) = \max\{1 - \alpha, \beta\}$, Yager implication $I_Y(\alpha, \beta) = \beta^\alpha$ etc.). For a discussion on the problem of selecting proper fuzzy implication to model the rules see e.g [50]. Such rules can be interpreted (see [21]) as "the more $\mathcal{X}$ is $\mathcal{C}$, the more *certain* $\mathcal{Y}$ is $\mathcal{D}$". Any rule here is in fact a constraint restricting the set of possible outcomes. Adding rules to a rule base therefore means refining of knowledge and as such elimination of possible solutions. A proper mathematical model of the rule base from definition (2.4.5) with these implication-based rules therefore is

$$\mathcal{R} = \bigcap_{i=1}^{n} [(\mathrm{C}_{i,1} \times \ldots \times \mathrm{C}_{i,m}) \rightarrow D_i]. \tag{2.19}$$

The choice of a proper mathematical model of a fuzzy rule base depends on the information that is contained in the linguistic description of the rules and in the expert knowledge of the decision maker. Once we have established the mathematical model of a rule base, we need an inference algorithm that would be able to assign appropriate output to an input in the form

$$\mathcal{X}_1 \text{ is } \mathcal{C}_1^{'} \text{ and } \ldots \text{ and } \mathcal{X}_m \text{ is } \mathcal{C}_m^{'}, \tag{2.20}$$

that can be represented as

$$(\mathrm{C}_1^{'} \times \cdots \times \mathrm{C}_m^{'}). \tag{2.21}$$

The process of approximate reasoning - obtaining an output $\mathcal{D}^{'} = M^{-1}(D^{'})$ from a rule base $\mathcal{R}$ for an input $(\mathcal{C}_1^{'} \times \cdots \times \mathcal{C}_m^{'})$ can be schematically represented in the following way:

$$D^{'} = (\mathrm{C}_1^{'} \times \cdots \times \mathrm{C}_m^{'}) \circ \mathcal{R}, \tag{2.22}$$

where $\circ$ is a composition of fuzzy relations. The linguistic description $\mathcal{D}^{'}$ of the result of approximate reasoning $D^{'}$ is obtained by linguistic approximation (or retranslation). This is an important part of the linguistic modeling process using fuzzy rule bases - it is important to convey all the information represented by $D^{'}$ to the decision maker. In Publication **II** we present a fuzzy rule based decision support system for HR management (academic faculty performance evaluation), where the fuzzy inference is designed to enable graphical outputs as well as easily interpretable linguistic description of the outputs without the need for complicated linguistic approximation procedures. In **VII** this decision support system and its outputs are compared with several systems for faculty performance evaluation used at universities and the benefits of our approach are discussed.

Classic examples of the use of conjunction-based fuzzy rule bases representable by (2.17) can be found e.g. in [57, 58, 93, 94], context dependent fuzzy inference is discussed e.g. in [12], implication-based fuzzy rule bases representable by (2.18) are well discussed e.g in [65, 66].

2.4.3 Computing with words and perceptions

Before we concentrate on the last step of linguistic modelling in general - that is the translation of results (outputs of our mathematical models) back into the linguistic level of description, one

more approach to linguistic modelling, that emerged from Zadeh's concepts of linguistic variables and linguistic modeling [115, 118, 119, 120, 121, 123] should be mentioned. It is the methodology of *computing with words and perceptions* [124, 125, 127]. This can be viewed as a next step in Zadeh's interpretation of linguistic modeling - a step from the modeling using linguistic variables and fuzzy rule bases to mathematical models able to represent complex expressions from natural language and perform computations and inference with them. In this light the linguistic modelling described so far can be viewed as computing with words - level 1 (CW1). What we are about to explore here is the second level of computing with words (CW2) also known as computing with words and perceptions (CWP) which has received much attention in the recent years (see e.g. [29, 36, 39, 45, 59, 60, 61, 62, 63, 73, 86, 101, 105, 126, 128, 129] including also several applications and theoretical frameworks for CW2 based on type-2 fuzzy sets). CW1 and fuzzy logic (in broad sense) provide necessary tools and foundation for CW2 [124].

In computing with words and perceptions (more particularly in what Zadeh calls the computational theory of perceptions) words are used as labels of perceptions, which are inherently fuzzy in nature. In other words perceptions are represented as propositions in natural language. The knowledge contained in a proposition in a natural language is seen as a (fuzzy) constraint on some (one or more) implicit variables. Propositions formulated in a natural language are first transformed into a Generalized Constraint Language (GCL), subsequently inference is performed based on the constraint propagation mechanism and the obtained outputs are at the end retranslated back into the natural language (linguistic approximation).

Let us consider a proposition p in a natural language - for example $p =$ *the temperature today is pleasant*. According to [125] p (the information it contains) can be viewed as a network of fuzzy constraints. After aggregating these constraints an overall fuzzy constraint can be obtained represented in general by an expression

$$X \text{ is} r \ R, \tag{2.23}$$

where R is some fuzzy relation constraining the variable X and r is a variable the values of which can be:

- e - in this case (2.23) in an equality constraint
- d - (2.23) is a disjunctive (possibilistic) constraint, R is a possibility distribution of X, in other words R is a fuzzy set on X. This is one of the most common cases and therefore the "d" is usually omitted from the notation and we write X is R.
- p - (2.23) is a probabilistic constraint - R is a probability distribution
- u - "usually" which can be interpreted as "usually $(X$ is $R)$" etc.

The expression (2.23) is called *the canonical form of* p. The main purpose of such representation is to make explicit the fuzzy constraint that is implicit in p. The meaning of p is thus defined in two steps in a process called *precisiation*. First a procedure is needed that works on an explanatory database (ED) and provides the constrained variable X. A second procedure is needed to provide the restricting relation R based on ED. ED is a collection of relations in terms of which the meaning of p is defined (simply speaking the explanatory database is the information that is needed to define the meaning of p) - see [123] for more details.

The following constraint propagation rules can be used [124, 125, 127]. In the following A and B are fuzzy relations, $A \subset U$, $B \subset V$, conjunction and disjunction are defined using a T-norm and an S-norm respectively (e.g. via $\min$ and $\max$). For simplicity, the rules are represented as inference rules, antecedent constraints are above the horizontal line and consequent constraints below the line.

Conjunctive rule 1:

$$\frac{\begin{array}{c} X \text{ is } A \\ X \text{ is } B \end{array}}{X \text{ is } A \cap B}$$

Conjunctive rule 2:

$$\frac{\begin{array}{c} X \text{ is } A \\ Y \text{ is } B \end{array}}{(X, Y) \text{ is } A \times B}$$

Disjunctive rule 1:

$$\frac{\begin{array}{c} X \text{ is } A \\ \text{or} \\ X \text{ is } B \end{array}}{X \text{ is } A \cup B}$$

Disjunctive rule 2 ($A \subset U, B \subset V$):

$$\frac{\begin{array}{c} X \text{ is } A \\ Y \text{ is } B \end{array}}{(X, Y) \text{ is } (A \times V) \cup (B \times U),}$$

where $A \times V$ and $B \times U$ are cylindrical extensions [22, 50, 66, 76, 114, 131] of A and B respectively.

Conjunctive rule for isc:

$$\frac{\begin{array}{c} X \text{ is}c\ A \\ X \text{ is}c\ B \end{array}}{X \text{ is}c\ A \cup B}$$

Disjunctive rule for isc:

$$\frac{\begin{array}{c} X \text{ is} c\ A \\ \text{or} \\ X \text{ is} c\ B \end{array}}{X \text{ is} c\ A \cap B}$$

Compositional rule of inference:

$$\frac{\begin{array}{c} X \text{ is } A \\ (X, Y) \text{ is } B \end{array}}{Y \text{ is } A \circ B,}$$

where $\circ$ is the composition of A and B. There are many other rules, among which of particular interest to us is the generalized extension principle as the principle rule for constraint propagation. Let us start with the notation of the extension principle (equivalent to the definition 2.1.6).
Extension principle:

$$\frac{X \text{ is } A}{f_F(X) \text{ is } f_F(A)}$$

where $f : U \to V$, and $f_F(A)(y) = \sup\{A(x)|y = f(x), x \in U\}$ for all $y \in V$.

Generalized extension principle:

$$\frac{f(X) \text{ is } B}{q(X) \text{ is } q(f^{-1}(B)),}$$

where $f : U \to V$, $q(X)$ is a query concerning X, $q(X)(y) = \sup\{B(f(x))|y = q(x), x \in U\}$. Using the generalized extension principle, we are able to compute the membership function of a query "*(what is) the average height of Swedes*" (to use Zadeh's example [127]) based on the perception that "*most Swedes are tall*". To do so, we obviously need to precisiate this perception using some explanatory database that would contain the heights of all Swedes and the information on how much each of these heights is compatible with the term "*tall*" as well as a fuzzy constraint representing the meaning of "*most*". The generalized extension principle describes how to compute a fuzzy set that would represent the meaning of average height of Swedes. This fuzzy set can then be interpreted linguistically (see the following section).

In practical applications *a basic interpolative rule* is often used as a special case of the compositional rule of inference. This way e.g conjunction-based fuzzy rule bases can be included into this framework of computing with words and perceptions in the following way. Let us consider simple fuzzy rules (which can each be interpreted as a piece of data - hence conjunction-based representation is appropriate for them) in the form IF X is A_i THEN Y is B_i, $i = 1, \ldots, n$. These rules form a fuzzy rule base $\mathcal{R}$. Then the basic interpolative rule (fuzzy inference rule) can be e.g. formulated as

$$\begin{array}{c} \mathcal{R} \text{ is } \bigcup_i (A_i \times B_i) \\ X \text{ is } A \\ \hline Y \text{ is } \bigcup_i \min\{m_i, B_i\}, \end{array}$$

where $m_i = \sup(A_i \cap A) = \text{hgt}(A_i \cap A)$ is a measure of the degree to which an input A matches the antecedent part of rule i, and $\min\{m_i, B_i\}$ is a fuzzy set on V with a membership function $\min\{m_i, B_i\}(y) = \min_{y \in V}\{m_i, B_i(y)\}$. This way we obtain the Mamdani style fuzzy inference mechanism [57]. CW2 is a more general framework for linguistic modelling, than CW1 (linguistic modelling described in the previous sections). It generally provides mathematical tools for computing answers to questions posed in natural language. The final step, however, is to translate these computed results back into linguistic terms. This will be the focus of the following section.

2.4.4 Linguistic approximation, defuzzification or other courses of action?

We have already stressed on several places in the text, that one of the main advantages of linguistic fuzzy modelling is the ability of the model to provide outputs in terms from natural language. This is very important for decision support systems, evaluation systems and generally everywhere where the outputs will be handled and acted upon by humans. Of particular importance are easy to understand (we can even say "intuitively clear") outputs in situations, where high values are endangered and when there is not much time to react based on the outputs of the mathematical models. Outputs provided in natural language are an important asset, as human beings are used to conveying information through words. As was already discussed, it is the uncertainty of the meaning, the level of inexactness that makes natural language a powerful tool. In many cases an imprecise (but still precise enough) piece information is much more useful than an exact number, as it can be understood. And it is understanding of the results that we need to achieve in multiple criteria decision making, evaluation, classification - generally all instances of decision support for practice.

So far we have discussed all the important steps of process of linguistic modelling but one. Several ways of representing meaning of linguistic terms, approaches to defining membership functions of fuzzy sets, operations with these mathematical representations of meaning of various complexity ranging from simple addition to the construction of linguistic fuzzy rule bases have been briefly summarized and discussed in the light of the approach to linguistic fuzzy modelling suggested at the beginning of the thesis. And more of these issues will be discussed on practical problems and their solutions in the following chapters. What was not yet addressed is the last step - the conversion of the outputs of the linguistic models (usually in the form of fuzzy sets, crisp numbers or intervals or their combinations) back into natural language. We have discussed how to construct a mathematical model for a given linguistic term (for a given decision maker under given circumstances for a particular purpose etc. - many of the issues and possible shortcomings of these procedures were at least mentioned here). Now we need to construct an inverse operation to the one of assigning meaning. We need to find a way how to assign a given fuzzy set (output of our mathematical model) an interpretation. There are several general ways of doing so that come to mind, some of which are listed in the following text.

Presenting crisp numbers (or generally crisp sets) - At the end of the modelling process, where the linguistic (fuzzy model) was built on linguistic description of the system, we transform its

outputs into numbers or intervals. That is we replace fuzzy outputs by crisp ones and remain at the mathematical level. No transformation of the outputs back to the linguistic level is done. Some decision makers even require numbers as inputs.

For example in evaluation models, it is convenient for some decision makers to obtain a single number evaluation. It is too easy to rank real numbers, to make averages, to use such outputs to support decisions. We need to understand that due to our education that does not involve fuzzy methods or any representation of uncertainty (except for some basic ideas on statistical view of uncertainty due to randomness), we are used to working with crisp numbers. And sometimes even a belief that mathematical models are able to provide definite or true or objective results is present among decision makers. But frankly a decision support model will never be infallible - it will always have limitations, it will be only as good as the knowledge that was available for its construction (and its quality might be a bit lower due to the limitations and assumptions of the mathematical methods used). Let us consider the following situation - we are to choose the best candidate for a given managerial position. There are two candidates A and B. We have constructed a sophisticated mathematical model reflecting many important criteria, some of which are qualitative and hence assessed linguistically, that at the end produces an evaluation of each of the candidates. We can consider this evaluation to be a real number from $[0, 1]$ representing the degree to which the candidate meets our requirements - 1 representing "*all requirements fully met*" and 0 "*no requirements met*" (that is the more the better). What if the evaluations are as follows: $e_A = 0.8$ and $e_B = 0.801$. Which one do we choose? The fallacy of using crisp numbers in human decision making is that based on these evaluations we might be inclined to select B, because $0.801 > 0.8$. There is no flaw in the conclusion itself, but in the reason why we have reached it. We should understand, that the numbers represent some characteristics of some uncertain evaluations that were computed by the evaluation model. As such the actual value of each evaluation might be dependent on the method that has been chosen to transform the fuzzy evaluation into a crisp one. This would however mean, that we can see both evaluations as indistinguishable. But there is still the issue that $0.801 > 0.8$, so "clearly" (based on what we have learned at school) one is better than the other. How do we explain this to the candidates? Should the decision maker use the reasoning that the larger value is better? Should we even provide this value to him to put this problem before him? The main idea of this example is, that if we present the outputs from linguistic (fuzzy) models as crisp, they will most probably be treated as such. So far the question of responsibility has not been even posed, but in our opinion if we provide results that can be misused, we are at least partially responsible for the consequences. Complex defuzzification methods can also be constructed that are able to reflect some kind of meta-knowledge we might have on the correct way of defuzzifying the outputs of the model in the given situation [111]. Still our current view is, in accordance with what has been declared in Chapter 2, that crisp outputs should not be provided as the only information from linguistic fuzzy models. In fact if we reduce all the uncertainty to zero in the last step, it is questionable whether it was necessary to reflect it from the beginning - but this is rather a rhetorical question. Crisp numbers can, however, be presented in connection with other types of outputs - linguistic, graphical, etc. - to provide additional information.

There is one area of linguistic fuzzy modelling where crisp outputs are necessary and appropriate - this is the area of fuzzy control [14, 43, 51, 52, 67]. If we see fuzzy controllers as tools for controlling real life systems, there is no objective need of retranslation of the outputs back into the linguistic level. In fact fuzzy controllers usually aim on automatization of the control

of the system [58, 93]. That is based on some (usually measured) inputs, that are fuzzified and processed by some (linguistic) fuzzy mechanism are transformed into fuzzy outputs, that need to be defuzzified in order to provide crisp values to mace a control intervention. The intervention changes the parameters of the system, which are then again measured, fuzzified, processed and new intervention (if necessary) is performed. It might be useful to be able to incorporate expert knowledge into the description of the system (see [58]). Since in this branch of linguistic modeling crisp results are required we will briefly summarize several possible ways of obtaining them. The process of obtaining crisp representatives of fuzzy sets is called *defuzzification*. Let us consider that a linguistic mathematical model provided an output A which is a general fuzzy set on an interval $[a, b]$, that is $A \in \mathcal{F}([a, b])$. There are several ways of *defuzzifying* this output, usually based on heuristic ideas (see e.g. [56] for defuzzification on various types of scales). First let us define a core of a fuzzy set (see [56]) . Let $C \in \mathcal{F}([a, b])$, then the set $\text{core}(C) = \{x \mid x \in [a, b] \text{ and } \nexists y \in [a, b] \text{ such that } C(y) > C(x)\}$ is called the *core of a fuzzy set* C. In case of $C \in \mathcal{F}([a, b])$ we can simplify the notion to $\text{core}(C) = \{x \mid x \in [a, b] \text{ and } C(x) = \text{hgt}(C)\}$. We can now define the defuzzification of C that is $d \in [a, b]$ using one of the following methods

$$d_{LOM} = \min\{x \mid x \in \text{core}(C)\} \text{ or} \tag{2.24}$$

$$d_{ROM} = \max\{x \mid x \in \text{core}(C)\} \text{ or} \tag{2.25}$$

$$d_{COM} = \frac{d_{ROM} + d_{LOM}}{2}, \tag{2.26}$$

where LOM, ROM and COM stand for *left of maxima*, *right of maxima* and *center of maxima* respectively. We can also use the whole information contained in the shape of the fuzzy set (that is in its membership function). This leads to the "area family" of methods, among which e.g.

$$d_{COA}, \text{ such that } \int_a^{d_{COA}} C(x)dx = \int_{d_{COA}}^b C(x)dx \text{ or} \tag{2.27}$$

$$d_{COG} = \frac{\int_a^b x \cdot C(x)dx}{\int_a^b C(x)dx} \text{ or} \tag{2.28}$$

$$d_{MOMI} = \frac{\int_0^1 \underline{c}(\alpha) + \overline{c}(\alpha)d\alpha}{2} \text{ (in case } C \in \mathcal{F}_N([a, b])) \tag{2.29}$$

can be used. Here COA, COG and $MOMI$ represent the *center of area*, *center of gravity* and *middle point of the mean interval method* respectively (d_{MOMI} can also be interpreted as an expected value of a fuzzy variable in some contexts [41]). In Publication **I** we have proposed the concept of a *α-degree upper bound of the fuzzy number C* summarized by the following definition.

Definition 2.4.6 Let $h \in [a, b]$ be a real number and C be a fuzzy number with a non zero membership function on $[a, b]$ and zero membership function outside of this interval. Then we say, that h is an α*-degree upper bound of the fuzzy number C*, $\alpha \in [0, 1]$ if and only if

$$\alpha = \frac{\int_a^h C(t)dt}{\int_a^b C(t)dt}. \tag{2.30}$$

We can see, that specifying α we can get a defuzzification of C as well based on (2.30). It is also obvious, that the center of area (COA) method represented by (2.27) is a special case of (2.30) for $\alpha = 0.5$. The use of definition (2.4.6) to interpret fuzzy constraints is discussed in Chapter 3 and also in Publication **I**, where a complex decision support tool for decision support of medical rescue services during disasters is proposed. Many other approaches to defuzzification also on discrete universes can be found e.g. in [56, 131]. It is obvious that the methods presented here in general do not provide the same results. The choice of defuzzification method has to be done with respect to the purpose of the model.

Presenting fuzzy sets as outputs themselves - This is another "extreme" approach. This way no information will be lost and the whole process of careful propagation of uncertainty carried by the membership functions of fuzzy sets will make sense, as the information would be fully exploited at the output. The tricky part here is, that the decision maker might not be able to interpret a fuzzy set well enough. Although much can be achieved by training and experience with fuzzy tools, fuzzy sets (unlike linguistic labels whose meaning they represent) are still quite an unnatural way of representation of meaning. That is we risk that the information conveyed by the fuzzy set output might get distorted by its interpretation by the decision maker. Still, as an addition to a linguistic or numerical information, presenting a fuzzy set output can significantly increase the understanding of the outputs by the decision maker. As fuzzy sets are used to reflect and model uncertainty (inherent in linguistic descriptions and labels in linguistic fuzzy modelling) it seems appropriate to add fuzzy set outputs into the set of outputs provided to the decision maker, particularly when the uncertainty may have an influence at the decision. But it can not be the only output provided. Graphical outputs in the form of fuzzy sets are deemed as "*... undoubtedly the most compact and exhaustive of the decision making problem. However it is not well accepted by the clinical user because it must be analysed to be properly understood.*" in the context of a task of finding abnormalities in electrocardiographic signal (clinical diagnostics) in [16, p. 155]. This only stresses our point that outputs must be customized for each decision maker.

It is natural language that we understand best. One of the main reasons of the usefulness of language is that a finite set of terms is enough to describe any situation or system. Although the description might be uncertain, it is precise enough for us to comprehend and possibly also act upon. Here we encounter an important discrepancy - a linguistic fuzzy model (consider e.g. a FRB) can in theory provide infinitely many different fuzzy sets as outputs. The set of linguistic terms for comprehensible description of these outputs is however finite (it usually does not contain many terms). As there can usually be no one to one mapping that would assign each output a unique linguistic label, assigning linguistic labels to the outputs results in a partial loss of information. A set of possible outputs of the model is possibly assigned a single linguistic label. This process is called *linguistic approximation* (or *retranslation* in the framework of CWW). Let us now consider several examples of possible approaches to linguistic approximation.

Building models that result directly in outputs with known linguistic labels - Obviously this is not an actual way of linguistic approximation, as there is nothing to approximate when we obtain linguistic outputs of the model. Models of this type are built in a way does not use any mathematical representation of the actual meaning of linguistic terms. Ordinal scales of linguistic terms are used (ordered sets of linguistic terms). Ordinal models presented in section 2.3 could serve as an example of such models. The approximation is built directly into the aggregation or inference mechanism.

The price for easy interpretability of outputs of such mathematical models is a significantly restricted set of possible outcomes. Generally just one of the terms from a predefined set of linguistic output terms $\mathcal{T} = \{\mathcal{T}_1, \ldots, \mathcal{T}_m\}$ can be obtained. As we have already discussed, the set of mathematical tools that are available for such modelling is also limited to operations that can be performed on ordered sets. Is some cases, particularly when precision is not necessary and crude results are enough, this approach is a tool of choice, as the results are selfexplanatory for the decision maker. The cooperation with the decision maker on defining the rules, term sets and the relations between object is essential to obtain results that are intuitive and well understood (both in the sense of ease of understanding and in the sense of avoiding misinterpretations). As there is no reduction or alteration of the meaning of the outputs in the last step of the modelling (in the interpretation of retranslation phase), there are also not many ways how to distinguish two cases with the same linguistic output (evaluation) and slightly different results. The difference of these two instances gets lost in the process of deriving outputs. This is, however, not a problem in situations where the decision maker plays an active role in the process of formulation of conclusions. From the management point of view, models that provide only crude linguistic results might be an impulse for the decision maker to participate in the process more actively, to assume responsibility for the final decisions.

Finding the most proper linguistic term from a predefined term set - This is a classical case of linguistic approximation. Let us consider a linguistic variable $(\mathcal{Y}, \mathcal{T}(\mathcal{Y}), [a,b], G, M)$ and let $\mathcal{T}(\mathcal{Y}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_m\}$ be the set of its linguistic terms that are *understood by the decision maker*. Let $T_i = M(\mathcal{T}_i) \in \mathcal{F}_N([a,b])$ be the meaning of each of the linguistic terms modelled by a fuzzy number on $[a,b]$. Let us consider an output of our mathematical model $O \in \mathcal{F}([a,b])$ that is a fuzzy set on $[a,b]$. The problem of *linguistic approximation* is one of finding a mapping $M^{-1} : \mathcal{F}([a,b]) \to \mathcal{T}(\mathcal{Y})$ that would meet certain conditions. These conditions should assure the meaningfulness of the linguistic approximation, that is reasonability of the process of assigning linguistic labels to fuzzy sets. The main goal is to prevent counterintuitive linguistic approximations - that is to find $\mathcal{T}' \in \{\mathcal{T}_1, \ldots, \mathcal{T}_m\}$ such that $\mathcal{T}'$ summarizes the information contained in O in the most reasonable way. There are several interpretations of the "reasonability" of linguistic approximation (see e.g. [36, 99, 112]).

- The first one might be that an it is reasonable to approximate the meaning of the output by a linguistic term that is more general, that is to assign such $\mathcal{T}'$ for which $O \subseteq M(\mathcal{T}')$. This follows from the entailment principle that can be summarized in the following way. From "V is A" we can infer "V is B" such that $A \subseteq B$ (see e.g. [55, 112, 123]). Let us recall, that for any $A, B \in \mathcal{F}([a,b])$ the *inclusion* is defined as $A \subseteq B$, if $A(x) \leq B(x)$ for all $x \in [a,b]$. Such crisp concept of inclusion might be too restrictive - it is easy to imagine a situation, when $O \nsubseteq M(\mathcal{T}_i)$ for any $i = 1, \ldots, m$. Is such cases we would

not be able to assign a linguistic label, which is not acceptable. We can, however, define a softer concept of inclusion in the following way.

Definition 2.4.7 (Degree of inclusion [112]) Let $A, B \in \mathcal{F}([a,b])$. The degree to which A is a subset of B, denoted as $\text{Deg}(A \subseteq B)$ can be computed as

$$\text{Deg}(A \subseteq B) = \min_{x \in [a,b]} \{I(A(x), B(x))\}, \tag{2.31}$$

where I is a fuzzy implication operator.
Remark: Examples of fuzzy implications can be found in the discussion of (2.18). Having defined the degree of inclusion, the linguistic approximation can now be formulated as a problem of finding a $\mathcal{T}' \in \{\mathcal{T}_1, \ldots, \mathcal{T}_m\}$, such that

$$M(\mathcal{T}') = T' = \arg \max_{i=1,\ldots,m} \text{Deg}(O \subseteq T_i). \tag{2.32}$$

By requiring the meaning of the target linguistic term to be a superset of the output fuzzy set, we in fact assess something that might be called the *validity* of our linguistic approximation or retranslation. But there is still a question how close is our approximation.

- We can also require the the meaning of $\mathcal{T}'$ to be as close to O as possible. For this we need some measure of *distance (or closeness)* of fuzzy sets. For example if the outputs of the models are fuzzy numbers (i.e. if $O \in \mathcal{F}_N([a,b])$) we can define

$$\text{Dist}_N(O, T') = \int_0^1 \frac{\alpha^w \left(|\underline{o}(\alpha) - \underline{t'}(\alpha)| + \left|\overline{o}(\alpha) - \overline{t'}(\alpha)\right| \right) d\alpha}{2(b-a)}, \tag{2.33}$$

where $w > 0$ is a parameter and the term α^w reflects that differences in high α levels are more significant than differences in low α levels. Degani and Bortolan [16] discuss various possible distance measures for fuzzy numbers.
If general fuzzy set outputs are provided by the model (that is if $O \in \mathcal{F}([a,b])$), we need to define the distance in a different manner, for example as

$$\text{Dist}_1(O, T') = \int_a^b |O(x) - T'(x)|\, dx \text{ or} \tag{2.34}$$

$$\text{Dist}_2(O, T') = \int_a^b (O(x) - T'(x))^2\, dx \text{ or} \tag{2.35}$$

$$\text{Dist}_3(O, T') = \left(\int_a^b (O(x) - T'(x))^r\, dx \right)^{\frac{1}{r}},\ r > 0. \tag{2.36}$$

There are many other approaches of defining a distance between two fuzzy sets on the same universe. The problem of linguistic approximation now translates into finding $\mathcal{T}' \in \{\mathcal{T}_1, \ldots, \mathcal{T}_m\}$, such that

$$M(\mathcal{T}') = T' = \arg \min_{i=1,\ldots,m} \text{Dist}_k(O, T_i), \tag{2.37}$$

where k is an index of the chosen distance measure.

- We could also need the linguistic approximation to *suggest particular perceptions*. Although this might seem a bit on the edge of manipulation, in some cases we might want to stress some aspects of the fuzzy set more than the others (e.g. risk). See [112] for an account on this issue.
- There is also the possibility of *combining all the previously mentioned criteria* of reasonability of linguistic approximation. This way we can approach the task of finding a proper linguistic label for the output of the mathematical model as a multiple criteria decision making problem. Importance of each criterion can be specified, even complex relations of the criteria can be reflected - various aggregation operators and their fuzzifications can be used including linguistic fuzzy rule bases.

Using a predefined set a linguistic terms and a syntactic rule - This approach is very similar to the previous one. We again consider a linguistic variable $(\mathcal{Y}, \mathcal{T}(\mathcal{Y}), [a, b], G, M)$ and let $\mathcal{T}_{elem.}(\mathcal{Y}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_m\}$ be the set of its elementary linguistic terms to be used to describe the output of the model, that are *understood by the decision maker*. Let us also consider the syntactic rule G that allows us to construct the term set $\mathcal{T}_{der.}(\mathcal{Y})$ by constructing derived terms based on $\mathcal{T}_{elem}(\mathcal{Y})$ using e.g. the *negation*, *and*, *or*, and/or *linguistic hedges*. The enriched linguistic scale and the extended linguistic scale presented in definitions 2.4.3 and 2.4.4 respectively are a good example of such structure. This way, we can now search for a proper linguistic label for the output of our model in the term set $\mathcal{T}(\mathcal{Y}) = \mathcal{T}_{elem.}(\mathcal{Y}) \cup \mathcal{T}_{der.}(\mathcal{Y})$.

Although more linguistic terms are available using the derived term set, we need to keep in mind that too complex constructions of derived terms might bring us closer to minimizing the distance or maximizing the validity, but at the same time might gradually loose intuitive meaning and interpretability for the decision maker. An optimum tradeoff between complexity of the newly constructed linguistic terms and their understandability for the decision maker has to be achieved.

Constructing mathematical models in a way that ensures easy interpretation of outputs - This is an approach we have proposed in Publications **II** and **IX** in the context of HR management. Also Publication **VIII** provides insights into this topic by comparing the results achieved in **II** and **IX** with other approaches to faculty evaluation. In Chapter 5 we discuss a linguistic fuzzy rule-based multiple criteria evaluation model of academic faculty performance proposed in Publications **II** and **IX**. By reflecting the needs of the decision maker during the design process of the mathematical model, the linguistic approximation phase can simplify significantly. This system was designed to provide understandable outputs on various level of aggregation. On the computational level, the evaluations are obtained as real numbers on intervals with known interpretation. Linguistic scales are defined on these intervals for the purposes of linguistic approximation. Each output is interpreted using its membership degree to one or two adjacent meanings of linguistic terms from the term set. The *linguistic approximation here is in fact direct linguistic interpretation* of the outputs.

Linguistic models can be specifically constructed to avoid at least some information loss due to linguistic approximation by planning the outputs, consulting their appearance and future use with the decision maker and also by providing also intermediate results that are well interpretable. See Chapter 5 for more discussion on this topic. A necessary requirement on such approach to modelling is to maintain the link between the linguistic and the computational level of the model as well as possible.

Providing graphical instead of numerical results - It is surprising how much attention has been paid to the linguistic level of fuzzy models, when as well as by words the uncertainty can be e.g. in evaluation problems expressed graphically - using color scales. In Publication **II** and chapter 5 a linguistic fuzzy model that provides outputs on a color scale (where each color is also assigned a linguistic label) is presented. From our experience the alternative representation that also enables an intuitive/emotional grasp of the results is very desirable. Graphical results are able to communicate the uncertainty while remaining intuitively comprehensible. This ability should, in our opinion, be better harnessed in linguistic fuzzy modeling - specifically in the phase of linguistic approximation (retranslation or interpretation) of results. Alternative methods of presenting the outputs of linguistic models seem to be an interesting and promising line of future research in this area.

Combination of the previously mentioned methods - Custom made approaches to presenting results suiting well the decision makers and reflecting their needs might ensure proper use of the models and allow the decision makers to make their decisions supported by the outputs of the model. From our perspective it is important to realize, that *a mathematical model that makes a decision for the decision maker is not well designed*. The decision maker plays an important rule in the decision making process (consider e.g. medical decision making and disaster management in chapter 3, staff performance evaluation in chapter 5). When the context of the problem has to be taken into account (even if only marginally), then the decision support model should not provide outputs that might substitute the final decision made by the person responsible for it. Here we have the ethical problem of decision support and responsibility again. *If we (that is mathematical models designed by us) provide decisions, we are responsible for the consequences*. On the other hand if we (or the models and tools designed by us) provide information, suggestions, non dominated alternatives, promising courses of action, warnings etc., the decision remains with the decision maker and we are just facilitating the process - this is what in our opinion decision support means. Achieving this is, however, subject to our ability of providing well interpreted, unbiased results that can be understood and used, that are not more precise than the outputs of the model suggest and that carry all the information necessary for an informed choice of alternative or course of action.

There are many issues concerning multiple criteria decision making and evaluation with linguistic fuzzy models that could not be discussed here. For example the methods chosen to obtain the final ordering of fuzzy numbers (or fuzzy evaluations of alternatives) have large impact on the outputs and their interpretability (see e.g. [4, 5]) as well. Any choice of a method for ordering of fuzzy evaluations of alternatives usually leads to a partial reduction of information that is present in the fuzzy set representation of the evaluations. It is however not the purpose of this thesis to discuss every issue connected with linguistic fuzzy modeling. We have concentrated so far on the concepts that will be further discussed in the next chapters where linguistic mathematical models and methods developed by or with the participation of the author for practical applications.

PART II: APPLICATIONS OF LINGUISTIC FUZZY MODELLING

CHAPTER III

Linguistic modelling in disaster management

In Publication **I** a decision support tool for the emergency medical rescue services is proposed. Its main purpose is to provide a "second opinion" to the operator of the emergency center while evaluating the emergency call. The main reason why this decision support system was designed was the need of providing supporting information to the decision makers in situations, when classical procedures seize to work and a shift of priorities is required. These situations are called disasters and can be characterised by the following properties. In general disasters are events that occur suddenly, unexpectedly, infrequently and have a huge impact on people, their lives, health and/or property. Usually some important values such as lives are threatened during disasters. In the medical rescue service, a shift from "saving all" strategy to "try to give a chance to survive to as many as possible" strategy takes place. As lives are threatened and the condition of injured people can deteriorate quickly without medical care, it is crucial to get well equipped medical personnel to them as quickly as possible. There is little time to make decisions concerning the numbers of people and resources required to successfully deal with the situation (in the initial phase). On the other hand underestimating the severity of the situation can lead to insufficient personnel on the disaster site. We need to realize that although it might be possible to call for reinforcements at any time, it takes time to get them to the disaster site. Overestimating the severity of the disaster can also have consequences, as the resources of medical rescue services are limited in a given region and drawing all the ambulances and rescue teams to the disaster site results in inability of performing everyday duty. This means that the decisions concerning the sufficiency of the number of your own resources and people has to take place at the very beginning of the disaster response.

Another typical feature of decision making during disasters is, that the decision makers usually do not have much experience with such situations - disasters do not happen very often and each disaster is unique. There is an apparent lack of experience (although the operators might have participated in some exercises of disasters) and the operators making the decisions are far out of their comfort zone. Great responsibility makes it harder to decide for the operators. To complicate matters more, the quality of information concerning the disaster is usually very low. An emergency call is usually the first and only source of information - the information is provided in linguistic form, it consists of unqualified estimates of numbers of people injured and it can not be easily verified at the firs step of disaster response. Any decision support therefore has to be able to accept any type of inputs - precise, fuzzy, linguistic and provide such interface, that is as intuitive as possible for the decision maker.

Before we proceed, let us formulate the problem at hand and the necessary output, that has to be reached. The model has been designed for the emergency medical rescue services (EMRS) in the Czech Republic, which are responsible not only for providing care to injured people at the disaster site, but also for the transportation of these people to appropriate medical facilities. This obviously drains the resources of the EMRS. Resources are limited hence based on the emergency call, which provides initial information on the emergency situation, a decision has to be made on how many ambulances and emergency teams (teams with a doctor that are equipped to treat seriously injured and teams without a doctor that are equipped to treat moderate injuries are usually distinguished) are needed to deal with the situation (to provide appropriate care to all and to transport them into suitable hospitals). If the resources are deemed insufficient, reinforcements need to be summoned. The first decision of the operator could therefor be summarized in the following way: *"Do I send some of the teams we have here or all? If all, is it enough? If not, how many more teams and ambulances are needed? Whom do I need to contact?"*. This decision should be made within one minute after the emergency call, as this is the time when first ambulances should be leaving their garages. The desired output can in fact be one of the values of the linguistic variable summarized in Figure 3.1 representing from what area do the ambulances be summoned - ranging from the local ambulances through the whole region to the surrounding states. In this case a linguistic output (that is one of the linguistic values of the scale) is a desirable output. Although more information can be provided by the model, this initial piece of information is the most important one during the first minute and has to be obtained quickly and understood by the operator without problems. If the assistance of EMRS from the whole state is needed, we do not need to know how many ambulances are required exactly - the operator will simply notify all the regions to send what they can. That is the main purpose of the decision support is to provide a check whether the situation was understood (its severity assessed) well according to the emergency call. The aim is to prevent underestimation and overestimation of the situation. But the final decision is operator's responsibility.

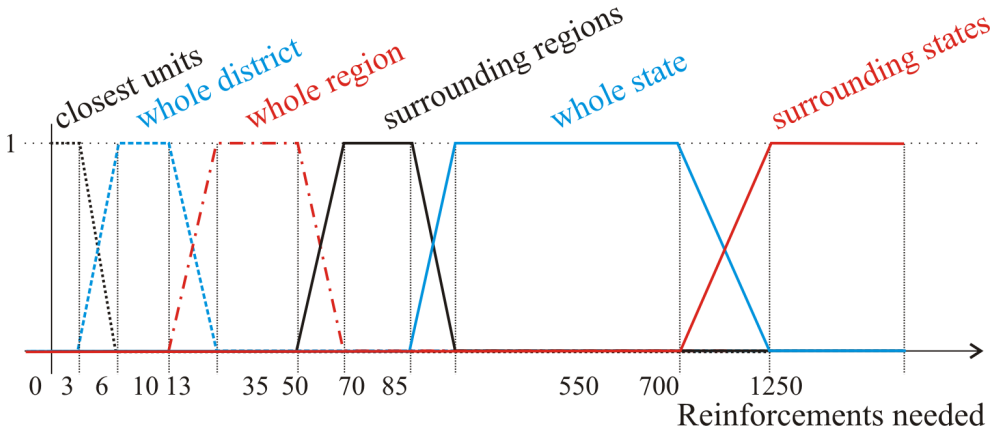


Figure 3.1: Terms of the linguistic scale representing the desired output information and their meanings. Reproduced and modified from Publication **I**.

Let us now consider the types of inputs that can be expected. We can start with exact numbers - these, however, are usually not available (at least not in a reliable quality). Still the model should be able to accept crisp inputs should a precise information be available. Crisp numbers and intervals can be easily represented by fuzzy numbers.

Linguistic description provided by the witness of the disaster during the emergency call (such as "a bus crashed into a large group of people, there was an explosion and everybody is hurt" or "a train collided with a fallen bridge construction") is more frequent. There is no time for conversions of words into fuzzy sets in decision makers head. It is however possible to enable the operator to input linguistic terms directly by pressing the appropriate linguistic labels on a touch screen. A set of linguistic terms with pre-assigned meanings modelled by fuzzy numbers (as they represent uncertain quantities) has therefore been proposed - see e.g. Figure 3.2 or Publication **I**. The meanings can be

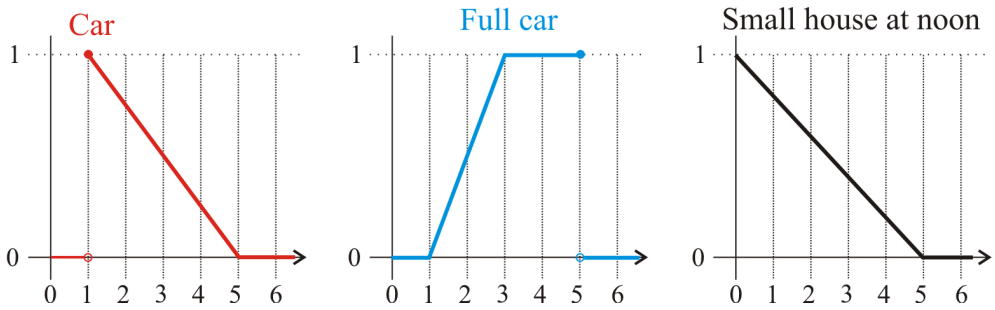


Figure 3.2: Several linguistic units available as inputs to the model. Meanings associated with the linguistic labels are modelled by fuzzy numbers and represent an uncertain amount of people that might be involved in the disaster. Meanings are time and region specific. By simple addition of fuzzy numbers an estimation of the number of people involved in the disaster can be obtained - again as a fuzzy number. Reproduced and modified from Publication **I**.

easily adjusted to reflect specific features of a given region, more linguistic units can be added. We also need to realize, that in disaster management, *meanings of such terms can be even time dependent*. If we want to model the meaning of a linguistic term "full train", the respective fuzzy number has to represent a higher amount of people during rush hours than for example during the night. This way we need to use such meaning, that is appropriate for a given time of day. To eliminate the time dependency of the meanings of linguistic terms, time specifications would have to be added to each label in the term set. This way the number of labels in the term set would increase. Finding proper linguistic labels on a touch screen could thus be complicated and the usefulness of such inputs may be lost. Fortunately it is no problem for the computer to use the meanings of the terms that are adequate for the given time of the day. Modification can also be done for example by predefined fuzzy rule bases as in the adjustment of MRC (see Figure 3.9). It is also crucial to ensure that the linguistic labels are accurate (capture the reality in the given context well, they can be fuzzy of course) and that the operator is familiar with their meaning. If the meanings are not intuitive, using them in a disaster decision support may be risky as the outputs provided by the model might be confusing rather than helping for the operator.

We can not expect to have all the possible labels predefined. If based on the description present in the emergency call it is not possible to use the predefined terms, but the operator is still able to assess the severity of the situation and describe it using one of the values of the linguistic scale (as an equivalent for the possible extent of the given situation) presented in Figure 3.4, we can use the meaning of such a term as an input into the model. In cases when the predefined terms do not fit the situation, the operator has to be able to input an uncertain amount of people involved in the disaster himself - using a fuzzy number as in Figure 3.3 It is not difficult to train operators to be able to

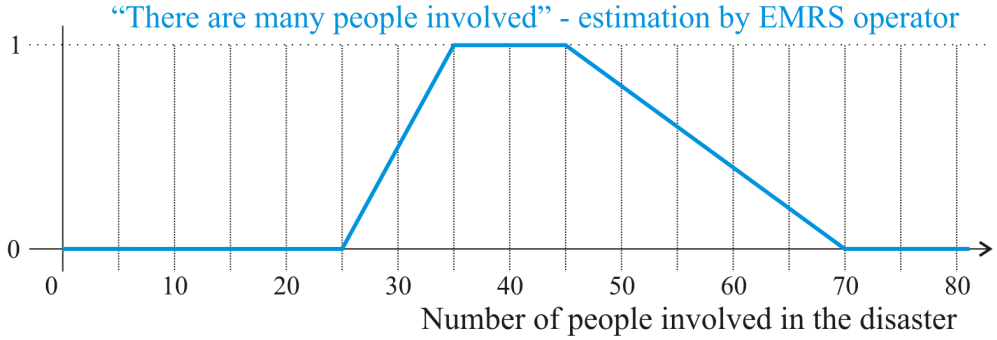


Figure 3.3: An estimate meaning of the information from the emergency call represented by a fuzzy number. The possibility of inputting fuzzy numbers by the operator requires a bit more knowledge and skill from the decision maker, but it is necessary for the model to remain robust.

input fuzzy numbers, however, the validity of such inputs may be low and the need of specifying four significant values might however be contra-productive. Nevertheless there are situations when this is necessary.

Now that we know the desired output and the types of inputs that can be used, let us formulate the actual decision making problem that has to be solved. Let us suppose that we know how many teams with a doctor and without a doctor we have available and let us also assume that each of these teams has an ambulance at its disposal. Our goal is to assess how many ambulances are needed to deal with the disaster (that means to provide care to all the people that are injured with respect to the severity of their injuries and to transport them into proper medical facilities). For more details of the EMRS procedures and responsibilities of the Czech Republic see e.g. [90, 91]. Let us consider n hospitals closest to the disaster site and let $I = \{1, \ldots, n\}$ be the set of indices of these hospitals, indices reflect the ordering of the hospitals based on their distance from the disaster site (1 is assigned to the closest hospital). We can now distinguish between specialized hospitals (SH) that is hospitals with specialised units, such as e.g. burn unit, and standard hospitals (H). Let us define the set of indices of specialised hospitals as $I_{SH} = \{i_1, \ldots, i_r\}$, $I_{SH} \subseteq I$ and the index set of standard hospitals $I_H = I - I_{SH} = \{j_1, \ldots, j_s\}$. Now the problem needs to be split into two branches. One will be for the seriously injured who need to be transported within the first hour to specialised hospitals and need to be treated by teams with a doctor. The second one will be for moderately injured, who can be treated by teams with or without a doctor and can be transported to any hospital except for the closest one to the disaster site, which is reserved to walking patients. The time to deal with patients with moderate injuries is six hours. These are conditions and time limitations set by the emergency procedures currently at place. We will also assume a limitation of one patient per ambulance during transport. To determine the minimum number of ambulances x_{T1} needed to deal with the seriously injured during the first hour of the disaster response, the following optimisation problem formulated in Publication **I** has to be solved:

$$x_{T1} = \sum_{k=1}^{r_1} x_{i_k} \to \min, \tag{3.1}$$

where x_{i_k} is the number of ambulances transporting patients from the disaster site to the specialised

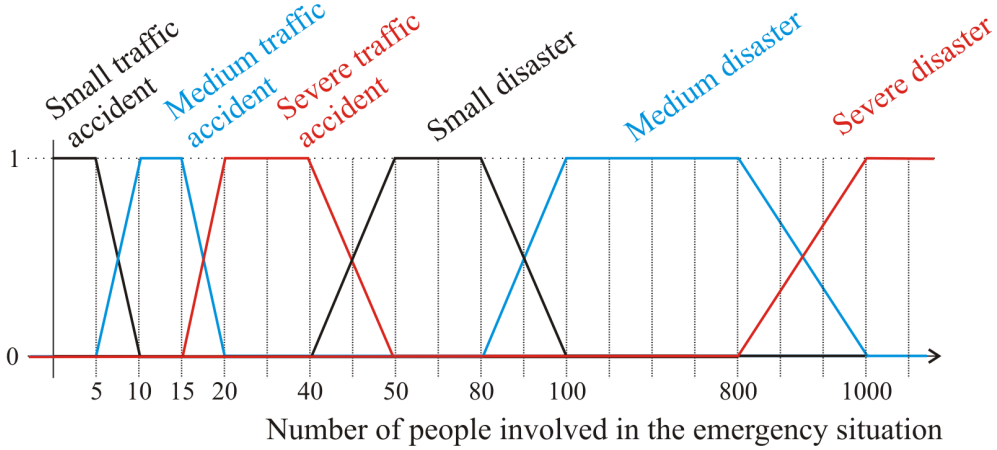


Figure 3.4: A linguistic scale for the description of the severity of situation and the meanings of its terms represented by a Ruspini fuzzy partition [78] of the universe representing the number of people affected by the emergency situation.

hospital i_k under the conditions

$$\sum_{k=1}^{r_1} trav_{i_k} \cdot x_{i_k} \geq NT1, \tag{3.2}$$

and

$$trav_{i_k} \cdot x_{i_k} \leq HTC_{SH_{i_k}}, \quad k = 1, \ldots, r_1, \tag{3.3}$$

where $trav_{i_k}$ represents the number of journeys to the specialised hospital i_k and back to the disaster site that can be made in one hour (an average speed of the ambulances has to be assumed here and the distance to the hospitals known; obviously only specialised hospitals for which $trav_{i_k} < 0.5$ are considered - these are the only ones reachable within 1 hour), $NT1$ is a fuzzy number representing the expected number of seriously injured people at the disaster site and HTC_{i_k} is the hospital treatment capacity of the specialised hospital i_k (expressed as a real number) representing the number of seriously injured patients that can be treated by the hospital per hour. Condition (3.2) ensures that enough ambulances are available to transport all the seriously injured people into specialised hospitals. Condition (3.3) then ensures that the hospital treatment capacities of the hospital are not exceeded. The problem described by (3.1) to (3.3) is in fact a fuzzy linear programming problem (see e.g. [76] for more details).

The index $r_1 \leq r, r_1 \in I_{SH}$ is set to meet the condition

$$\sum_{k=1}^{r_1} HTC_{SH_{i_k}} \geq NT1, \tag{3.4}$$

that is r_1 is the lowest possible index of a specialised hospital for which this condition is fulfilled. The seriously injured will be transported to the closest specialised hospitals (this way the number of ambulances can be minimized), but the HTC limitations need to be respected. The relation $\geq$ in

(3.2) and (3.4) has to be defined, as on the left side of the inequality, we have a crisp number and on the right side a fuzzy number. In Publication **I** we have suggested an α-degree upper bound for a fuzzy number, summarized here by definition (2.4.6). The definition allows us to introduce one more degree of freedom to the decision making bounded by fuzzy constraints. We can allow the constraints to be met fully by setting $\alpha = 1$, or to be partially violated. This gives us an opportunity of reflecting the decision makers attitude to risk at least in some way.

We can now use the α-degree upper bound of a fuzzy number to interpret (3.2) and (3.4) in a way that the left side of this equation has to be a α-degree upper bound of the fuzzy number $NT1$. Introducing (2.30) we now have a means of allowing partial violation of the condition. We can still require the number of ambulances to be higher than or equal than any value $d \in \mathbb{R}$ such that $NT1(d) > 0$. This is achieved by setting $\alpha = 1$, which is consistent with a very cautious approach that represents a decision maker not willing to tolerate any violations of (3.2). Publication **I** presents a heuristic procedure for solving the fuzzy linear programming problem represented by (3.1) to (3.3) using (2.30).

We now need to specify how to obtain the estimation of the number of seriously injured people at the disaster site $NT1$. We have also not addressed the issue of taking care of the injured at the disaster site yet. So far we have formulated an optimisation problem to determine how many ambulances will be needed to transport stabilized patients to specialised hospitals. Now we also need to assess how many teams with a doctor are required to stabilize the seriously injured people at the disaster site before their transport to hospitals.

Let us start with the former issue, that is determining the fuzzy number $NT1$ - an estimation of the number of people that have been seriously injured. If the inputs are provided by any of the four means suggested at the beginning of this chapter (either by selecting predefined linguistic terms from term set, selecting a prototypical emergency situation from the linguistic scale in Figure 3.4, providing a fuzzy number estimate of the number of people involved in the emergency situation or providing an exact number of people), we have either one or several fuzzy numbers that represent how many people are involved. In case we have several fuzzy numbers, we can easily add them together to obtain a fuzzy number representing the number of people involved in the situation (let us denote the resulting fuzzy number AFF). This way we obtain an estimation of the number of people that have been affected by the disaster represented by AFF, but not all of these people will be injured. Let us now for simplicity assume that we are dealing with disasters that result mainly in mechanical injuries (situations with prevailing chemical, radiation and explosive injuries would be treated differently in some aspects). Research suggests (see e.g. [15]) that about 10% of the people affected by the disaster will be seriously wounded and about 20% will be moderately wounded. As the percentages are rough estimates, it might be useful to represent them by fuzzy numbers as well. Let us therefore define $\widetilde{10\%}, \widetilde{20\%} \in \mathcal{F}_N([0,1])$ as triangular fuzzy numbers $\widetilde{10\%} = (0, 0.1, 0.2)$ and $\widetilde{20\%} = (0.1, 0.2, 0.3)$.

A remark: there is no direct reason for this particular fuzzification - different supports of the fuzzy numbers and different shapes of their membership functions may be chosen. That is we acknowledge, that a different fuzzification can be performed. It may also be argued, that we are introducing unnecessary uncertainty into the model by fuzzifying the percentages and as the AFF already is a fuzzy quantity, we might use the percentages as crisp. We can not agree with such line of reasoning. The percentages are presented as estimates, and both in literature and in practice are treated as uncertain quantities. It seems illogical to treat them as precise numbers. Linguistic fuzzy modelling

should reflect the expert knowledge and the nature of the data as well as possible - if a quantity a is treated as uncertain in the decision making process, its equivalent "about a" (that is $\tilde{a}$) should be used in the model. By fuzzifying the percentages, we aim to preserve the uncertainty inherent in the decision making situation. This might complicate the last step of the modelling - that is the retranslation of results back into linguistic terms, but at least we will have mathematical outputs that are not more precise than they should be under the circumstances. There is another reason why the introduction of $\widetilde{10\%}$ and $\widetilde{20\%}$ can prove to be beneficial. This follows from the two-layer design of linguistic fuzzy models. The use of linguistic labels like "about 10%" and "about 20%" allows us to build a single model on a linguistic level for all disasters resulting mainly in mechanical injuries. If the experience or historical data suggest, that these number vary less e.g. for large traffic accidents and more e.g. for earthquakes, we can represent the meanings of "about 10%" and "about 20%" in accordance with the given situation (more uncertain for earthquakes, less uncertain for traffic accidents) while maintaining a simple linguistic level description of the whole family of disasters resulting mainly in mechanical injuries. In many cases (see e.g. [15]) in disaster management, the KISS (Keep It Simple Stupid) principle is stressed. In disaster management having a general linguistic level of description of a family of problems while at the mathematical levels the meanings of the linguistic terms are adjusted to fit best the situation at hand is more than desirable. Linguistic modelling seems to provide tools for at least some level of generalization (= simplicity as only a few different linguistic descriptions are required), while on the computational level specificity can be maintained by switching to appropriate meanings - that is reflecting the context dependency of meaning. Let us get back to the determination of the number of people that have been injured. We can now compute the number of seriously injured people as $NT1 = AFF \cdot \widetilde{10\%}$ using fuzzy number arithmetics (see [49, 22, 50, 131] for more). Similarly the number of people with moderate injuries can be computed as $NT2 = AFF \cdot \widetilde{20\%}$. What remains now is to determine how many medical teams will be needed to provide care for the injured people at the disaster site.

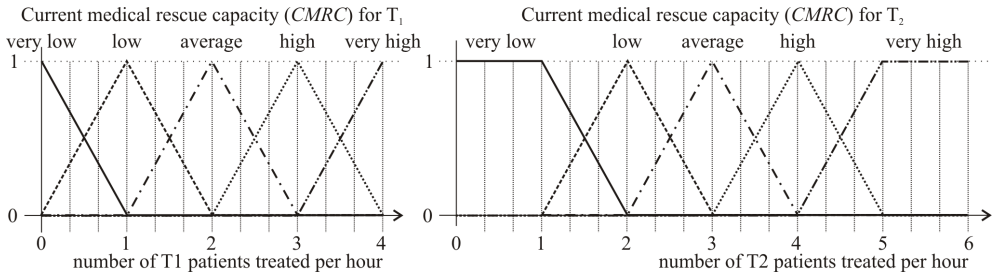


Figure 3.5: Linguistic variables and the meanings of their values for *CMRC* description for T1 patients (left) and T2 patients (right).

We will use the Medical Rescue Capacity (*MRC*) which describes how many people can be treated by a given EMRS team. We need to distinguish between care provided to seriously injured people (let us denote this group T1) and care provided to the moderately injured (T2). In [15] lists of procedures that need to be performed with T1 and T2 patients are provided. It is clear that T1 patients need more time to be stabilized than T2 patients. As the procedures are standardized, a standard time to perform them can be determined under ideal conditions. Based on this information (see [15, 90]) we can assume that the maximum number of T1 patients that can be treated per

hour is 4 and for T2 patients this number is 6. Hence we have upper bounds for the *MRC* for each category of patients. We can assume, that these values are achievable under ideal conditions - that is if the EMRS team is well-trained and coordinated, if the weather conditions (including e.g. temperature, wind strength, visibility) are acceptable and if the fatigue does not play any role. It seems reasonable to reflect all these variables in the process of determining each EMRS team's current *MRC* (*CMRC*). We can introduce several linguistic variables to deal with this issue and then describe their influence on *CMRC*. Linguistic variables and fuzzy rules are used as only linguistic description of the influence of these variables on the *CMRC* of each team is available at present time - that is we need to rely on the experience of practitioners. The linguistic variables (see [91] for details on their development) will be:

- Quality of the team (TQ) with linguistic values *average*, *good*, *great*. The meanings of these terms are summarized in Figure 3.6.

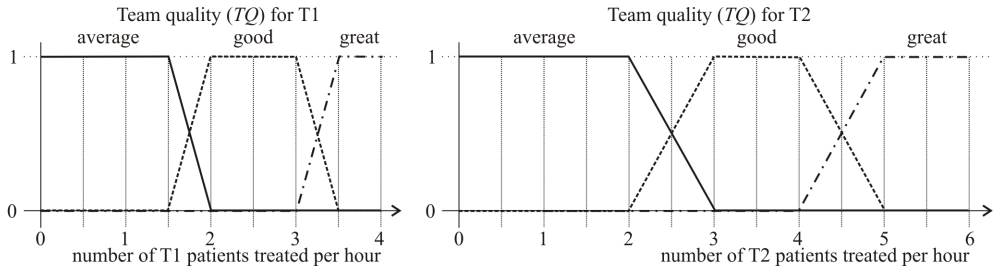


Figure 3.6: Linguistic variables and the meanings of their values for *TQ* description for T1 patients (left) and T2 patients (right). Reproduced and modified from Publication **I**.

Remark: It is important to stress, that the linguistic labels must provide intuitive interface between the expert providing his knowledge and the mathematical model. We can note here, that the lowest team quality, that in fact comprises also such performance that would be deemed intolerable in practice is "average". The reason for this (and also an explanation why a team whose performance is 0 patients per hour is deemed 100% compatible with the label "average") is, that the linguistic rules are provided by practitioners, who are reluctant to use worse linguistic labels for the performance of their colleagues. It is therefore obvious that the linguistic labels and their meanings must be tailor-made for each situation (and well understood, if not intuitive, for the person that uses them to express his/her knowledge, experience, evaluations). On the other hand, the *CMRC*, which is also measured in terms of patients that can be treated per hour, but is influenced also by weather and other external influences (see below) can be described also as *very low* or *low*. We point this out here to show, that the understandability of the model to the practitioners on the linguistic level has to be superior to formal/mathematical neatness that might demand unified term sets and so on. We simply need to make every effort to carry the decision makers meaning as well as possible into the model without creating risk of misunderstandings and misinterpretations.
We can also see, that since *TQ* has the same linguistic values for T1 and T2 patients, it will be possible to define a single rule base for both T1 and T2 patients. Only during the inference concerning *CMRC* for T1 patients the meanings of *TQ* defined for T1 will be used, and to compute the *CMRC* for T2 patients, the meanings of *TQ* concerning T2 will be used. Single

rule base is present and several instances of computations for various meanings (contexts T1 and T2) can be carried out.

- Weather (*WE*) also play its role. In the rule base used for CMRC determination, we will use an aggregated evaluation of weather described by a linguistic variable with terms: *dangerous*, *problematic*, *unpleasant*, *ideal*. Meanings of these terms are summarized in Figure 3.8 on the left. The evaluation of the weather presented on the x scale (that forms the universe of discourse for this linguistic variable) is computed based on three weather characteristics - rainfall, temperature and wind-strength (see [91] or Publication **I** for more). For each of these characteristics an evaluation function is defined (Figure 3.7) and the final weather evaluation is computed as a geometrical mean of these three partial evaluations. As is discussed in Publication **I** it is also possible to assume that the weather characteristics will not be measured, but provided as "qualified estimations" using linguistic terms modelled by fuzzy numbers on the respective universes (amount of rain, temperature, speed of wind). In this case the final weather evaluation that will be used as an input for the rule base defined below will be a fuzzy number computed using the extension principle as

$$(T{\cdot}W{\cdot}R)_F^{\frac{1}{3}}(y) = \max\{T(w_1), W(w_2), R(w_3) \mid y = (w_1{\cdot}w_2{\cdot}w_3)^{\frac{1}{3}}, w_1, w_2, w_3 \in [0,1]\}. \quad (3.5)$$

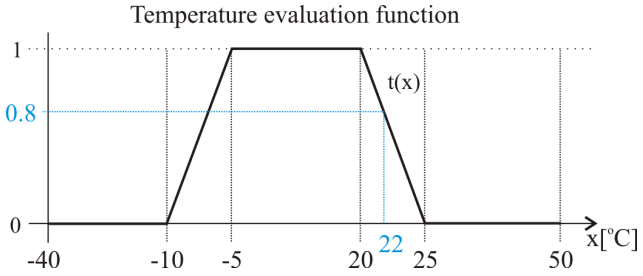


Figure 3.7: The partial evaluation function for temperature.

- Time on Duty (*ToD*) - this variable reflects the fatigue that can affect the performance of the team. Four linguistic values are available: *begun*, *in the middle*, *ending*, *overtime* with their meanings summarized in Figure 3.8 on the right.

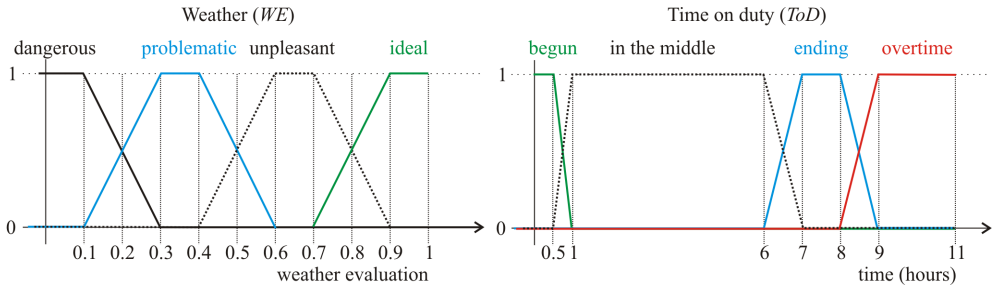


Figure 3.8: Linguistic variables and the meanings of their values for the evaluation of weather (*WE*; on the left) and for the description of Time on duty (*ToD*; on the right).

Using these linguistic variables a fuzzy rule base describing the relationship between *TQ*, *WE*, *ToD* and *CMRC* has been designed. All the combinations of values of *TQ*, *WE* and *ToD* were assigned a linguistic value of *CMRC*. 48 fuzzy rules were formulated such as:

IF *TQ* is *average* and *WE* is *ideal* and *ToD* is *begun* THEN *CMRC* is *average*.
...
IF *TQ* is *average* and *WE* is *problematic* and *ToD* is *ending* THEN *CMRC* is *low*.
...
IF *TQ* is *good* and *WE* is *unpleasant* and *ToD* is *ending* THEN *CMRC* is *high*.
...
IF *TQ* is *great* and *WE* is *ideal* and *ToD* is *begun* THEN *CMRC* is *very high*.

The fact that the meanings of the linguistic values of *TQ*, *WE* and *ToD* form a Ruspini fuzzy partition of the respective universe ensures, that there are no unexpected or unnecessary overlaps of the rules. As the meanings of the values of *CMRC* represent uncertain numbers of patients, the generalized Sugeno fuzzy inference mechanism as proposed in [96] was used. A discussion can be made at this point on the appropriateness of the conjunction-based model of a fuzzy rule base used for fuzzy inference here. Based on the discussion with the expert providing the rules, it became obvious that the information obtained through these rules was in the form of a lower bound for the possibility of the output. In fact disaster managers tend to the more pessimistic view of the world which includes supposing the worst to be prepared for it, should it come true. Their knowledge also seems to be based more on data accumulation - the only way to learn about disasters is to study historical data - that is to gain prototypical representations (examples) of disasters from the past. It therefore seems reasonable to use a conjunction-based model for the rule base in accordance with [21] instead of an implication based one. It is however much advised to pay more attention to the exact nature of the expert knowledge to be sure the models we use to represent it are absolutely appropriate. Experimental work on this topic is one of the future research aims of the author.

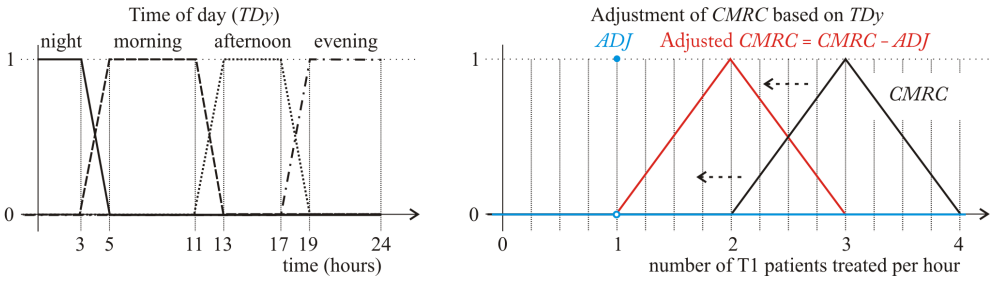


Figure 3.9: Linguistic values of the variable Time of day and their meanings (on the left). The effect of the fuzzy rule base computing the necessary adjustment of *CMRC* based on *TDy* (on the right) - the output of the rule base is denoted *ADJ* and represented as a fuzzy singleton.

It is also obvious that the rule base is already quite large to represent expert knowledge well enough - to define the consequent parts of 48 rules is not an easy task for practitioners. There is however one more variable we have not yet reflected in the model - the time of day (*TDy*). It seems reasonable to think that the *MRC* of teams will be higher when they can see what they are doing or have not just finished eating lunch. If we wanted to include another variable into the rule base (its linguistic terms

and their meanings are presented in Figure 3.9 on the left), the number of rules would increase from 48 to 192. We can not expect the decision maker to provide consequent parts of 192 rules (and to do it consistently). We can, however, ask him to explain how the time of day may affect the *CMRC*. After doing so we were provided by the following 4 fuzzy rules:

IF *TDy* is *night* THEN modify *CMRC* by 1. (*expect 1 less patient to be treated per hour*)
IF *TDy* is *morning* THEN modify *CMRC* by 0.
IF *TDy* is *afternoon* THEN modify *CMRC* by 0.3.
IF *TDy* is *evening* THEN modify *CMRC* by 0.6.

The rule base reflects the fact, that during morning the conditions can be assumed to be ideal. In the afternoon, fatigue begins to appear, and during the evening and night we can expect the tams to be significantly slower. We have used classic Takagi-Sugeno fuzzy reasoning approach [95] to compute a real number *adj* that is transformed into a fuzzy singleton *ADJ*. For a given time of the occurrence of the disaster *t*, we get

$$adj = 1 \cdot night(t) + 0 \cdot morning(t) + 0.3 \cdot afternoon(t) + 0.6 \cdot evening(t), \tag{3.6}$$

where *night*, *morning*, *afternoon* and *evening* are considered to be the fuzzy numbers representing the meanings of the respective linguistic terms. The *adjusted CMRC* is then computed by subtracting *ADJ* from *CMRC* - see Figure 3.9. Let us consider than *n* EMRS teams are available at the current EMRS centre. An average MRC (*AMRC*) can be computed simply by $AMRC = \sum_{i=1}^{n} adjusted\ CMRC_i$, where *adjusted CMRC*$_i$ is the adjusted current medical rescue capacity of team i. If we consider only the teams with doctors (these are the teams that will be required to provide care for the seriously injured) in the computation of *AMRC*, we can easily compute the number of emergency teams with doctors needed at the disaster site as $TEAMS = NT1/AMRC$ using fuzzy number arithmetic.

We now have all we need to assess the sufficiency of our teams and resources and the need for reinforcements. The fuzzy number representing the number of emergency teams with doctors required during the first hour of the disaster response is ($x_1 + TEAMS$). If we compare this fuzzy number with the number of teams we have currently available (the α-degree upper bound of a fuzzy number can be used), we can determine whether our resources are enough to deal with the disaster. If not then the difference between our current available number of teams and this fuzzy number is the number of reinforcements that is needed. This difference can be approximated by one of the terms of the output linguistic scale represented in Figure 3.1.

For the moderately injured the computation is analogical and the formulation of the optimisation problem represented by (3.1) to (3.3) for seriously injured is transforms into the form of (3.7) to (3.9). For T2 patients both teams with and without a doctor can be used - that is after the first hour when doctors tend to seriously injured the start helping the moderately injured as well and the time limit is five hours (during each hour a 1/5 of the moderately injured has to be treated and transported to a hospital). Also all kinds of hospitals can be used to transport patients in (MF stands now for any medical facility - specialised od standard hospital). No patients will be transported to the hospital closest to the disaster site (MF_1) as it will remain reserved for patients that might be able to get there on their own.

$$x_{T2} = \sum_{k=2}^{q} x_k \rightarrow \min, \tag{3.7}$$

where x_k is the number of ambulances transporting patients from the disaster site to any hospital $k \in I$ under the conditions

$$\sum_{k=2}^{q} trav_k \cdot x_k \geq \frac{NT2}{5}, \tag{3.8}$$

and

$$trav_k \cdot x_k \leq HTC_{MF_k}, \quad k = 2, \ldots, q, \tag{3.9}$$

where $trav_k$ represents the number of journeys to the medical facility k and back to the disaster site that can be made in one hour, $NT2$ is a fuzzy number representing the expected number of moderately injured people at the disaster site and HTC_{MF_k} is the hospital treatment capacity of the medical facility k. The index $q \leq n, q \in I$ is set to meet the condition

$$\sum_{k=2}^{q} HTC_{MF_k} \geq \frac{NT2}{5}. \tag{3.10}$$

The conditions (3.8) and (3.10) are presented here in a simplified form (an assumption that $trav_k \geq 1$ has been made). When medical facilities for which $trav_k < 1$ are involved, the problem can not be approached by splitting the 5 hours interval into 5 equal parts and the whole time has to be considered and the formulation of the model for moderately injured has to be adjusted. Publication **I** provides a numerical example of the proposed model.

The mathematical part of this decision support system (including the heuristic approach to solving the fuzzy linear programming problem) is designed to be able to provide a linguistic output concerning the need for reinforcements. As the decision has to be reached and understood very quickly, we tried to minimize the need of transformation any information into unnecessary complicated mathematical representation. Linguistic inputs are therefore accepted. The decision support system can also identify the hospitals that need to be informed about the occurrence of the disaster to prepare for receiving patients (this information is obtained during the heuristic solution) and estimates the number of people affected by the disaster and the number of seriously and moderately injured people that can be expected. The tools of linguistic modeling provide not only means for dealing with linguistic inputs and outputs, but also for a simple and easy to understand description of the relationships between the variables relevant in the decision making process. Even without the mathematical level, the model can still be used to describe the operation of the emergency medical rescue services - at least the a part of it that is concerned with the sufficiency of resources. The model has been designed in cooperation with the practitioners working at the EMRS in the hope of easing the decision making process of EMRS operators during disasters by providing them with a second opinion. The next step of the development of this system will necessarily involve its testing and fine-tuning on real life data.

CHAPTER IV

Linguistic modelling and AHP

In this chapter, we will try to view the Analytic Hierarchy Process (AHP) method proposed by T. L. Saaty [79, 80, 82, 83] from the linguistic modelling perspective. AHP is a widely used multiple criteria decision making tool and it has received much attention of practitioners and researchers since its introduction. A more complex version of this method able to reflect dependencies of criteria - the Analytic Networ Process (ANP) [85] - has also been developed and widely used. The AHP method is based on splitting the multiple criteria decision making problem we want to model into a hierarchical structure of subproblems and then solving each of these using pairwise comparisons. At this place we will focus on these subproblems on the lowest level of the hierarchy which are the keystone of the method. These subproblems need to be solved in AHP in order to construct evaluations on higher levels of the hierarchy. The main tool to capture the preference structure of a decision maker on a set of alternatives is pairwise comparison.

Let us consider we have a single decision maker who needs to evaluate each alternative from a given set of n alternatives $\{A_1, \ldots, A_n\}$ and let us assume that only one evaluation criterion is considered. This evaluation criterion does not have to be quantitative (measurable) - AHP can deal with qualitative criteria as well. It is convenient to express the preference structure on this set by comparing pairs of alternatives and assessing which one is more preferred to the other and also assessing the strength of this preference. Our aim in the AHP is to obtain evaluations $h_1, \ldots, h_n$ of these alternatives. Based on these evaluations a reciprocal square matrix H of the dimension $n \times n$ can be constructed, $H = \{h_{ij}\}_{i,j=1}^{n}$, such that $h_{ij} = h_i/h_j$ and obviously $h_{ij} = 1/h_{ji}$. The value h_{ij} then represents the relative preference of alternative i over alternative j and can be linguistically interpreted that A_i is h_{ij} times more important than A_j. Usually the multiple criteria decision problem is one of finding such evaluations $h_1, \ldots, h_n$, as they are not known in advance. To do so an estimation of the matrix H, a reciprocal square matrix of preference intensities $S = \{s_{ij}\}_{i,j=1}^{n}$, $s_{ij} = 1/s_{ji}$, can be constructed by a decision maker. In case of more decision makers the matrices of preference intensities can be aggregated into a single overall matrix of intensities of preferences. This can be done for each element of this overall matrix by computing a geometrical mean of all the values provided by various decision makers for the respective pairwise comparison.

Based on the matrix S the evaluations $h_1, \ldots, h_n$ can be computed as the arguments of minimum of

the following expression

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(s_{ij}-\frac{h_i}{h_j}\right)^2. \tag{4.1}$$

The solution to this problem (the evaluations $h_1,\ldots,h_n$) can be found as the components of the eigenvector corresponding to the largest eigenvalue of S. Alternatively the logarithmic least squares method can be applied and the solutions found in the form

$$h_i=\sqrt[n]{\prod_{j=1}^{n}s_{ij}},\qquad i=1,\ldots,n. \tag{4.2}$$

The elements s_{ij} are estimations of the actual values of h_{ij} and are provided by the decision maker as answers to questions "what is the intensity of preference of A_i over A_j for you according to the given criterion?" or alternatively "how much more preferred is A_i than A_j?".

We can observe that the questions are not easy to answer - at least not in a precise manner. We can not expect a decision maker to provide an answer such as "A_3 *is 3.56 more preferred than* A_5". Even if the decision maker might be able to provide such a precise answer, can we be sure that such a value is reliable? Saaty suggests the scale presented in table 4.1 to be used to express preference intensities. Linguistic labels are suggested to represent five elements of the scale. The numerical

Table 4.1: Saaty's scale - 9 numerical values of its elements for expressing preference (or indifference in case of 1) of alternative i over alternative j and their suggested linguistic labels.

s_{ij}	linguistic labels of the numerical intensities of preferences
1	alternative i is **equally preferred as** alternative j
3	alternative i is **slightly/moderately more preferred than** alternative j
5	alternative i is **strongly more preferred than** alternative j
7	alternative i is **very strongly more preferred than** alternative j
9	alternative i is **extremely/absolutely more preferred than** alternative j
2,4,6,8	correspond with the respective intermediate linguistic meanings obtained by joining the respective two linguistic labels $\mathcal{T}_k$ and $\mathcal{T}_l$ by "between" into the label "between $\mathcal{T}_k$ and $\mathcal{T}_l$"

values of the scale can be derived from the Weber-Fechner logarithmic law of response to stimuli [81] (the stimulus-response theory states that the higher the initial intensity of a stimulus is, the larger has the increase of the intensity be to be registered) and the maximal number of the scale (that is 9) is the result of the requirement of homogeneity (the requirement is for the number of the elements to be small and of the same order of magnitude - see [81] for more details). We will not discuss here the reasonability of the assumption that any evaluation follows the Weber-Fechner logarithmic law of response to stimuli, which was in fact originally intended to describe responses to stimuli on physiological level as this is more a fundamental question of decision theory. We aim to draw attention to the linguistic modeling aspects of the AHP method and Saaty's scale.

Before we get to the requirements on the matrix S, let us first remark that there is some arbitrariness involved in the construction of the linguistic labels of Saaty's scale. First the numerical values representing preference or indifference were determined and their number was set as 9, which resulted

in the following set of numerical values: $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. If we interpret the numbers in the suggested way, that is as describing how many times is one alternative preferred to another one, we need to explain why there is no greater value than 9. There is in fact no natural maximum of the number of times something can be preferred to something else. When we now assign a linguistic label "*extremely (or absolutely) preferred to*" to "*9 times more preferred than*". This can, obviously, introduce some confusion to decision makers. What is more, we define "*3 times more preferred*" to be just "*slightly preferred to*" which also does not seem to correspond with our intuition. It seems as if an ordered set of 5 linguistic labels expressing various levels of preference was constructed and then simply assigned to the numerical values. There is no indication of any process of finding proper meaning for the selected terms. We can agree with the proposed ordering of the linguistic labels representing the intensity of preference. The meanings of the linguistic terms are not equidistant, as the numerical scale represents ratios (multiples), not absolute differences in intensity of preference. Again, we should ask whether this is intuitive or understandable for the decision maker that uses the linguistic labels. The fact that Saaty's scale is being applied in various countries suggests that the linguistic labels had to be translated into other languages. It is questionable, whether the meanings of the translations maintain the same position (and relative position) on the universe $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

It seems to us that there is no strong correspondence between the linguistic labels and the numbers that are supposed to represent their meaning (other than that a stronger linguistically described preference is assigned a larger number as a meaning than a weak linguistically described preference). The linguistic terms are used as a tool to acquire the intensities of preferences via natural language. Then they are transformed into numerical representation by a predefined, but not very well grounded mechanism (from the linguistic modeling point of view the assignment of numbers seems rather rough). That is we do not argue against the Saaty's fundamental scale represented by its numerical values, we are trying to show that the linguistic level of the AHP works differently than the numerical one. This might not be a big problem, if just one of the levels is used to obtain S - that is if the decision maker provides either just numbers, or just linguistic values. If, however, these two levels are combined and the decision maker has to deal with the fact that e.g. "9 times more = absolutely more", or "3 times more = slightly more" problems can occur due to the possible ambivalence of the assigned meaning.

It is also interesting that the connection of the numerical and computational level is lost in the moment when the matrix S is obtained. This issue can be further illustrated on the concept of consistency. The *consistency condition* for matrices of preference intensities can be expressed as

$$s_{ik} = s_{ij} \cdot s_{jk,} \text{ for all } i, j, k = 1, 2, \ldots, n. \tag{4.3}$$

It is known that using the values $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and their reciprocals in the matrix of preference intensities, the consistency condition might not be fulfilled for expertly defined matrices of preference intensities, particularly if these are of larger order. There are many approaches to the assessment of consistency of matrices of preference intensities [6] and more are being developed [26, 74, 75]. We remind here the approach proposed by Saaty, that defines the inconsistency index CI based on the spectral radius ($\lambda_{\max}$) of S by (4.4).

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{4.4}$$

For a perfectly consistent matrix $\lambda_{\max} = n$ and hence $CI = 0$. For other matrices we can define the inconsistency ratio $CR = CI/RI_n$, where RI_n is a random inconsistency index computed as an

average of inconsistency indices of randomly generated reciprocal matrices of preference intensities of the order n. As long as $CR < 0.1$ the matrix S is considered to be consistent enough.

Now let us examine the consistency condition (4.3) from the perspective of linguistic modeling. To be compatible with the linguistic labels used in Saaty's scale, the condition should "make sense" also when we substitute the linguistic labels into it. Let us consider $s_{ik} = 3$ and $s_{kj} = 3$ then based on (4.3) we need $s_{ij} = 3 \cdot 3 = 9$. If we transform this into the linguistic level, we get if "A_i *is slightly more preferred than* A_k" and "A_k *is slightly more preferred than* A_j" then "A_i *is extremely/absolutely more preferred than* A_j". This is rather counterintuitive - we would expect a much smaller preference between A_i and A_j induced by two slight preferences. The consistency condition (4.3) is not well defined for the linguistic labels (or the linguistic labels are not well defined). In any cases if the decision maker provides information in linguistic form only and (4.3) is required, we declare as consistent something that is counterintuitive. To use the linguistic level for inputs more safely, a weaker consistency condition (see definition 4.0.8), that reflects the linguistic labels well has been proposed in Publication **XI** and further elaborated in Publication **III**.

Definition 4.0.8 (Weak consistency condition [Publication **III**]) Let $S = \{s_{ij}\}_{i,j=1}^{n}$ be a matrix of preference intensities. We say, that S is *weakly consistent*, if for all $i, j, k \in \{1, 2, \ldots, n\}$ the following holds:

$$s_{ij} > 1 \wedge s_{jk} > 1 \quad \Longrightarrow \quad s_{ik} \geq \max\{s_{ij}, s_{jk}\}; \tag{4.5}$$

$$(s_{ij} = 1 \wedge s_{jk} \geq 1) \vee (s_{ij} \geq 1 \wedge s_{jk} = 1) \quad \Longrightarrow \quad s_{ik} = \max\{s_{ij}, s_{jk}\}. \tag{4.6}$$

The properties of this condition are discussed in more details in Publication **III**. It is important to note here, that this condition is reasonable on both the numerical and the linguistic level of description. In situation when "A_i *is slightly more preferred than* A_k" and "A_k *is slightly more preferred than* A_j" we just require A_i to be "*at least slightly more preferred than* A_j". Such a condition is obviously much weaker than (4.3). It is therefore seen as a minimum requirement on the consistency of expertly defined matrices of preference intensities.

Using the notion of weak consistency a mathematical model for the evaluation of works of art created by Czech art colleges and faculties has been developed. The underlying mathematical model is described in details in Publications **III** and **XI** and its further development and some of the advantages of the concept of weak consistency are presented in Publication **VII**. The Registry of Artistic Performances (RUV in Czech), that uses the evaluation methodology and the mathematical model proposed in these publications, is currently being used in the Czech Republic for distributing a part of the subsidy from the state budget to Czech public universities. The model will be briefly summarized at the end of this chapter.

In our opinion AHP is a good example of a method that seems to provide a linguistic level of description (see the linguistic labels of the Saaty's scale in table 4.1) but does not comply with the requirements set by us on linguistic methods in the section on linguistic fuzzy modelling. Let us now analyze the scale a bit more. It is possible to transform a *multiplicative* (elements interpreted in terms of ratios/multiples) pairwise comparison matrix $S = \{s_{ij}\}_{i,j=1}^{n}$ into an *additive* (elements interpreted in terms of differences) pairwise comparison matrix $Z = \{z_{ij}\}_{i,j=1}^{n}$ using the following transformation (see e.g. [27, 77]):

$$z_{ij} = \frac{1}{2}(1 + \log_9 s_{ij}). \tag{4.7}$$

The resulting matrix Z then carries the same information concerning the preferences of the decision maker. Z is additively reciprocal, that is $z_{ij} = 1 - z_{ji}$, $z_{ij} \in [0, 1]$ and $z_{ii} = 0.5$ for all $i, j = 1, \ldots, n$. As such the matrix Z can be interpreted as a fuzzy relation, its elements z_{ij} representing the degree of preference of A_i over A_j. Obviously, $z_{ij} = 0.5$ is interpreted as indifference between A_i and A_j, $z_{ij} = 1$ is interpreted as A_i is absolutely preferred to A_j, and $z_{ij} = 0$ is interpreted as A_j is absolutely preferred to A_i.

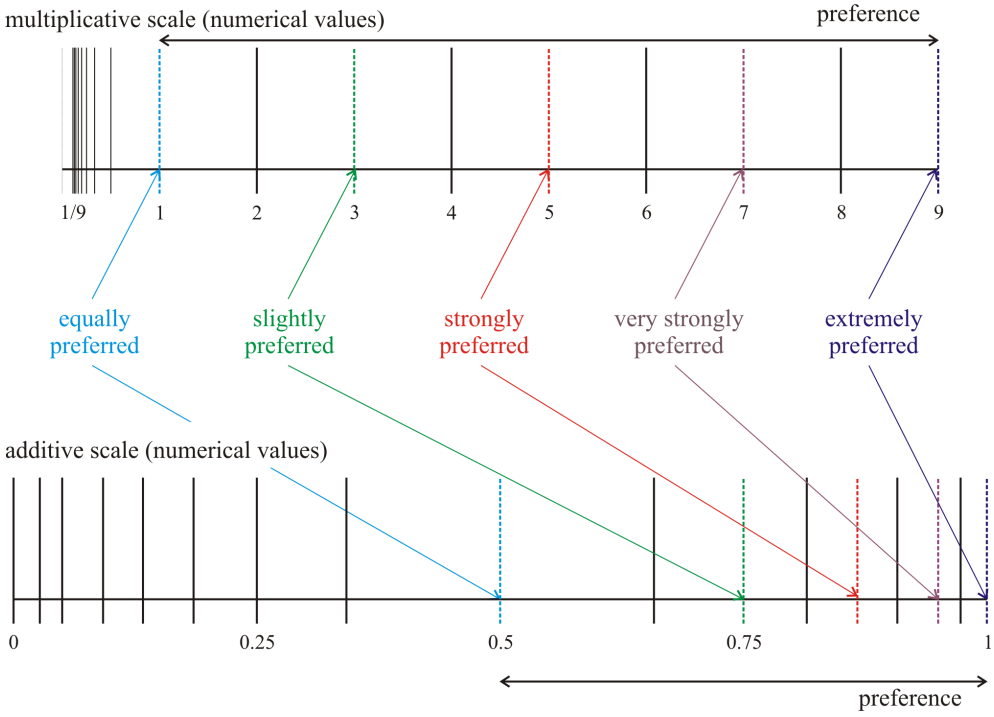


Figure 4.1: Transformation of the numerical values corresponding with the linguistic labels (in colour) and the intermediate numerical values and their reciprocals from Saaty's multiplicative scale into the values of the additive scale.

As we already know the meanings of the linguistic labels in the multiplicative case, we can now transform them into the additive representation using (4.7) and see, whether they "make sense" in the additive case, where a natural minimum and maximum of the degree of preference exists. Figure 4.1 summarizes the results of such transformation. We can see, that at least the meaning of the linguistic label "slightly preferred" seems to be misplaced. If we look at its numerical equivalent on the additive scale, we can see that the meaning of slightly preferred is half the way between indifference and extreme preference. At least in our opinion this does not correspond with the intuitive meaning of slight preference - the meaning of slight preference should in our opinion be much closer to indifference.

The reason for this might be the process of construction of the multiplicative scale and its linguistic labels as well as choosing the number 9 as its maximum value. It might be possible to construct the meanings of the available linguistic labels to be more intuitive (and thus making the linguistic labels more compatible with the multiplicative consistency condition) by constructing the meanings in the additive model instead. That is to define an appropriate meaning of each of the five linguistic labels

used in Saaty's scale as a number from $[0.5, 1]$. Then these values would have to be transformed back to the multiplicative universe (model) and a closest integer value from $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ would be assigned to them. The result of such an approach, if the meanings of the linguistic labels were uniformly distributed on $[0.5, 1]$ is summarized in Figure 4.2. We can see that after this modification, the multiplicative consistency condition requires much more reasonable relations in the linguistic level - if we again consider that "A_i *is slightly more preferred than* A_k" and "A_k *is slightly more preferred than* A_j" then has to be "A_i *is between strongly and very strongly more preferred than* A_j" (numerically $s_{ik} = 2$ and s_{kj}=2 and therefore $s_{ij} = 2 \cdot 2 = 4$) - this seems to us much closer to the intuitive expectation of the aggregated preference than in the original case. The uniform distribution of meanings of the five linguistic labels on $[0.5, 1]$ is just an example to illustrate the proposed approach. The meanings of these terms would have to be set in accordance with the understanding of these linguistic labels by the decision maker.

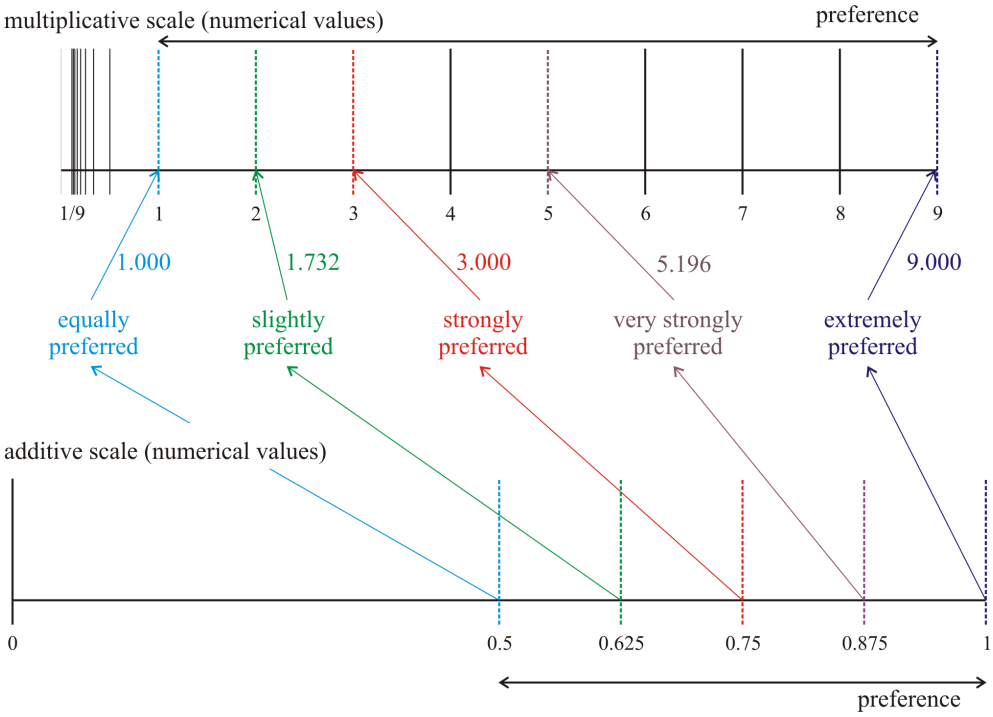


Figure 4.2: Transformation of the meanings of the linguistic labels of Saaty's scale that are considered to be uniformly distributed on $[0.5, 1]$ in the additive approach back to numerical values of the multiplicative scale. For each linguistic label an exact numerical value of its meaning after transformation is presented and the closest integer is assigned as its meaning in the multiplicative case.

Defining the meanings of the linguistic labels of Saaty's scale in the additive model on a universe with natural minimum and maximum seems reasonable - the decision maker is asked to simply find a fixed point on the interval $[0.5, 1]$ (that is between indifference and absolute preference) for each linguistic label. This way we are able to assign linguistic description to 5 real numbers representing the strength of a preference (we can also consider the intermediate linguistic terms thus obtaining 9 linguistically interpretable values). This, however, does not fully use the potential of granulation provided by fuzzy sets. That is e.g. in the additive case some values from the interval $[0.5, 1]$ have

no linguistic interpretation. We can consider the real number assigned to a linguistic term as its meaning to be a typical (best) representative of the given linguistic term. If we now move on the interval [0.5,1] away from this value, the further we get from it, the less compatible the original linguistic label is with the number we obtain. Up to a point where we reach a number representing the meaning of another linguistic term, that is we arrive at a number that is fully compatible with another linguistic label (and as such is no longer compatible with the previous linguistic label at all). The same holds for the multiplicative scale (here we would move in the interval $[1, 9]$). This would suggest that it might be reasonable to define a Ruspini fuzzy partition on the given universe (either in the additive or in the multiplicative case).

Remark: In the following text, we will consider possible fuzzification of the Saaty's scale. Before we do so, several quotations of Saaty might be in place. Saaty [84] claims that the scale with linguistic labels summarized in Table 4.1 is already fuzzy and no further fuzzification is necessary. In fact his statements against the fuzzification of AHP are very strong (e.g. [84, p. 973]): *"Fuzzy set practitioners have been leading a parasitic existence (a parasite according to Webster's dictionary is one depending on another and not making adequate return - "we might add most of the time") by looking at all numbers as if they are subject to uncertainty and fuzzifying them purportedly to improve consistency without either giving good reasons for doing it because we know that good consistency does not imply greater validity, or proving that the results thus obtained are more valid than is obtained directly from the judgments. In sum, we note that making poor judgments leads to poor outcomes and fuzzifying poor judgments still leads to poor outcomes."* We agree with Saaty that fuzzifying poor judgements might make no sense. However we have already provided several examples of discrepancies between the linguistic labels and their meanings in Saaty's scale (Table 4.1). Although Saaty claims that the linguistic level of the scale already reflects the fuzziness of the meaning of linguistic terms, the meanings do not appear to be intuitive and seem to be ill defined in connection with the consistency condition (4.3). We have also proposed how more reasonable meanings can be assigned to the linguistic labels. In context of what has already been presented in this chapter, Saaty's refusal of including (linguistic) fuzzy perspective to the AHP is strange at least - he even comes to the conclusion [84, p. 970] that "*... one should never use fuzzy arithmetic on AHP judgment matrices*". On one hand he seems to object against unreasonable introduction of fuzziness into the method (which can be understood), on the other hand he seems to neglect the appropriateness of the meaning of the linguistic labels that are used. Linguistic labels are modelled by integers, and this is translated as "being already fuzzy" (to be precise [84, p. 962]*"When judgments are allowed to vary in choice over the values of a fundamental scale, as in the Analytic Hierarchy Process, these judgments are themselves already fuzzy."*). We do not wish to insist on introducing fuzziness into the AHP, our aim is to show that the linguistic modelling perspective, as well as the fuzzy modelling perspective could be of benefit to AHP. What will be presented in the following text is therefore a suggestion of how the Saaty's scale could be fuzzified in a *reasonable way*. It is also important to see that any results that might have been obtained by the crisp version of AHP can be obtained using the fuzzification of the scale and of the computations presented here and in [53] and Publications **V** and **XII**. This is ensured by interpreting the numerical values of Saaty's scale as typical values that form the kernel of the respective fuzzy number.

Let us for now consider that the linguistic labels suggested by Saaty are appropriate (although we have suggested a procedure of obtaining more appropriate labels or at least of assigning the given set of linguistic labels a more intuitive meaning, this has to be done in cooperation with a decision maker. Without a particular decision maker, the Saaty's scale as summarized in table 4.1 will suffice as an example). The meaning of each linguistic label can be modelled by a fuzzy number, the kernel

of which would contain a single point from the universe - the typical value, and the support would be bounded by the typical values of the neighboring linguistic labels.

Let us remark that even in the multiplicative case considered by Saaty, this does not contradict the idea of having 9 elements in the linguistic scale. We would still have 8 linguistic labels for the intensities of preferences + 1 linguistic label for indifference. This approach just provides us with a means of linguistically interpreting any intensity of preference expressed as a number on $[1, 9]$ or its reciprocal value. This way we can obtain the meanings of the linguistic terms as fuzzy numbers as summarized in Figure 4.3 for the case of the multiplicative scale.

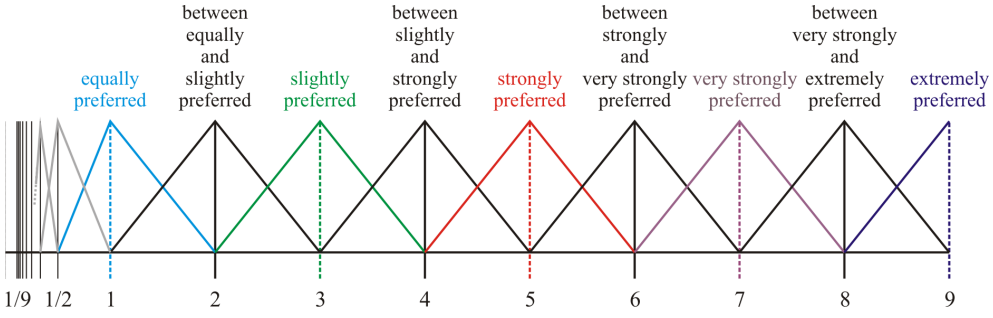


Figure 4.3: Meanings of the linguistic labels used in Saaty's scale as modelled by fuzzy numbers (triangular fuzzy numbers are used here as an example).

Here the fuzzy numbers representing the meanings of the linguistic labels (including the intermediate values) are considered to be triangular (see [53]). We need to realize that in such case the reciprocals of these values of the scale will not be triangular fuzzy numbers (also $M(\text{equally preferred}) \neq 1/M(\text{equally preferred})$), but can be approximated by triangular fuzzy numbers.

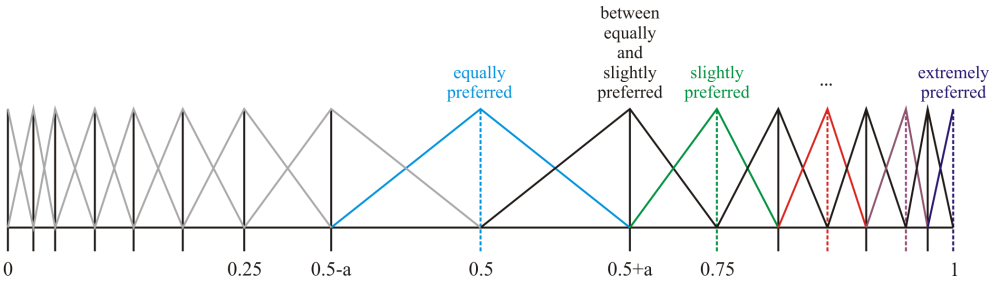


Figure 4.4: Meanings of the linguistic labels defined on the additive scale as triangular fuzzy numbers (the numerical values in the kernels of these fuzzy numbers are the transformed numerical values of the original Saaty's multiplicative scale; in this case $a = 0.158$).

The fuzzy numbers representing the meanings of the linguistic labels can be defined so that the reciprocity condition of Saaty's matrix is fulfilled when using these fuzzy numbers as elements of the matrix. To obtain meanings of the linguistic labels as fuzzy numbers that would all be of the same type and that would fulfil the multiplicative reciprocity condition, we can again start with the additive model. We can define here the meanings of the linguistic labels using triangular fuzzy

numbers (e.g. as presented in Figure 4.4). The additive reciprocity condition requires the meanings of reciprocal linguistic labels to be symmetric about 0.5. Triangular fuzzy numbers fit well the additive reciprocity condition. Let us assume that the meaning of the label "*equally preferred*" is modelled by a triangular fuzzy number $EP_Z \sim (0.5-a, 0.5, 0.5+a)$, where $a \in (0, 0.5]$, that is $EP_Z = \{[a(\alpha-1)+0.5, a(1-\alpha)+0.5]\}_{\alpha\in[0,1]}$. By its transformation back to the multiplicative model, we obtain the meaning of the label "*equally preferred*" as a fuzzy number $EP_S = \{[9^{2a(\alpha-1)}, 9^{2a(1-\alpha)}]\}_{\alpha\in[0,1]}$. It is easy to see that now $EP_S = 1/EP_S$ that is the multiplicative reciprocity condition if fulfilled for this term (see Figure 4.5).

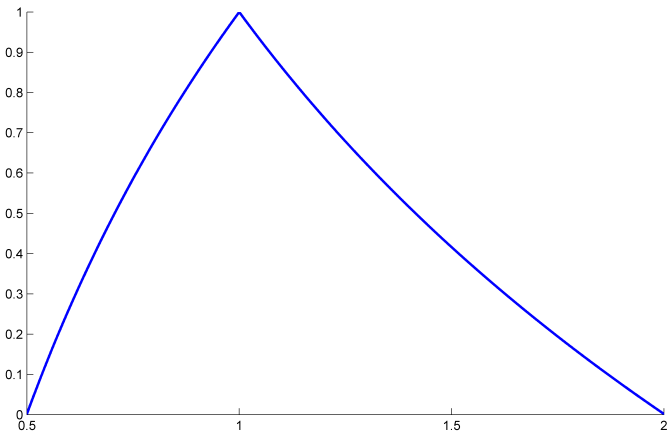


Figure 4.5: Membership function of the fuzzy number EP_S representing the meaning of "*equally preferred*" constructed by transforming the triangular fuzzy number EP_Z defined in the additive model into the multiplicative model.

Such construction results in shapes of all the fuzzy numbers representing the meanings of the linguistic labels in the multiplicative model being similar to the fuzzy number in Figure 4.5. Regardless of the shape of the membership functions of the fuzzy numbers representing the meaning of the linguistic terms, we can focus on the triplets of significant values that define the kernel and the support of each of these fuzzy numbers (see Publications **V** and **XII** for a discussion of the possible benefits of this representation). Let us now consider the matrix of preference intensities $\tilde{S} = \{\tilde{s}_{ij}\}_{i,j=1}^{n}$ the elements of which are fuzzy numbers $\tilde{s}_{ij} \sim (s_{ij1}, s_{ij2}, s_{ij3})$. The significant values of the fuzzy numbers representing the evaluations of alternatives $\tilde{h}_i = (h_{i1}, h_{i2}, h_{i3})$ for all $i = 1, \ldots, n$ can now be computed (see [53] or Publication **V**) using the logarithmic least squares method and respecting the multiplicative reciprocity condition:

$$h_{i1} = \min \left\{ \frac{\sqrt[n]{\prod_{j=1}^{n} s_{ij}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} s_{kj}}} \; ; \; \begin{array}{l} s_{kj} \in [s_{kj1}, s_{kj3}], \forall j > k, \\ s_{jk} = \dfrac{1}{s_{kj}}, \forall j < k \\ s_{jj} = 1, \forall j \end{array} \right\}, \tag{4.8}$$

$$h_{i2} = \frac{\sqrt[n]{\prod_{j=1}^{n} s_{ij2}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} s_{kj2}}}, \tag{4.9}$$

$$h_{i3} = \max \left\{ \frac{\sqrt[n]{\prod_{j=1}^{n} s_{ij}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} s_{kj}}} ; \begin{array}{ll} s_{kj} \in [s_{kj1}, s_{kj3}], \forall j > k, \\ s_{jk} = \dfrac{1}{s_{kj}}, \forall j < k \\ s_{jj} = 1, \forall j \end{array} \right\}. \tag{4.10}$$

Publications **V** and **XII** showcase the usefulness of the fuzzification of AHP also in the context of evaluation. A fuzzy AHP based methodology for the evaluation of R&D (research and development) results is proposed in these publications and the role of fuzzified AHP in guiding decision makers through the decision making process is discussed.

4.1 Registry of Artistic Performances

Let us now consider the usefulness of the linguistic modelling perspective in practical applications. Publications **VII** and **XI** summarize the mathematical model used in the Registry of Artistic Performances (RUV) for the evaluation of works of art. The aim of the evaluation methodology used in RUV was to propose such evaluation strategy, that would be able to include peer-review evaluation as well as several measurable criteria in the evaluation of works of Arts. Although evaluation of arts is a difficult task, there are many reasons why this issue needs to be addressed. Among the most pressing ones we can include the funding of arts and art colleges, assessment of the efficiency of programmes aimed on the development or support of specific fields of arts and perhaps most importantly the promotion of quality in artistic performance.

RUV therefore seeks to collect information on the outputs of Czech art colleges and faculties, to assess the quality (novelty) and impact of the works of art produced by their employees and students and to provide appropriate information for funds distribution among these institutions. A second (but not secondary) aim of RUV is to provide comparisons - to enable each institution to see the outputs of other institutions, their quality and reception and to strive to be even better. As such RUV is conceived as a tool for quality promotion in Czech art colleges and faculties. The outputs of the multiple criteria evaluation have been used in the official methodology for funds distribution from the state budget among Czech public universities since 2012. In 2014 over 70 million of CZK (2.5 million EUR) have been distributed among art colleges and faculties based on the outputs of our evaluation model and its importance is expected to grow as the outputs of even more subjects are expected to be evaluated through RUV (nonartistic faculties whose aim and teaching is close to the field of arts). The evaluation methodology and its underlying mathematical model we have developed has thus been accepted as an appropriate tool for this purpose on the national level. In a broader perspective the evaluation methodology developed for RUV could also be adapted for

the evaluation of R&D outputs, where the quality assessment (also through peer-review evaluation) plays an important role as well and has to be integrated with other criteria.

In our opinion the reason why such a level of acceptance of the evaluation model (and the whole methodology) was achieved was the ongoing cooperation between the stakeholders (artists from the art colleges and faculties) and us. Each step of the development of the evaluation model (and the whole evaluation methodology) has been carried out according to the principles of linguistic modelling described in the introductory section on linguistic fuzzy modelling and although fuzzy sets were not used in the development of the mathematical model, the linguistic level of AHP and natural understanding of the meaning of each step of the computations were crucial. The concept of *weak consistency* was proposed to facilitate the process and has proven to be a valuable tool in dealing with pairwise comparison matrices of large order. Publication **III** provides a detailed description of its properties, development and use. The need of constant interaction with the stakeholders also resulted in the development of a new approach for obtaining pairwise comparison matrices of preference intensities from the decision makers. Let us briefly summarize the key features of the proposed mathematical model. More details are presented in Publications **III**, **VII** and **XI**.

Although evaluation of works of art might be considered an intangible task, a consensus on how to at least approach this task has been achieved. The arts sector in the Czech Republic has been divided into seven segments - *architecture, design, film* (now called *audiovisual arts*), *fine arts, literature, music* and *theatre*. The works of art produced in any of these segments are assessed according to three criteria - three criteria were specified (and agreed upon) by the artists themselves. Each work of art is therefore to be evaluated based on its relevance, extent and reception. It is easy to see that such criteria are not measurable and in fact not trivial to define precisely. On the other hand we can see that intuitively these three aspects might be able characterize a piece of art. For each of these three criteria a set of three linguistic evaluations was defined, these linguistic labels are each assigned a capital letter (the closer the letter to the beginning of the alphabet, the better evaluation it represents with respect to a given criterion). The criteria and their linguistic values are summarized below (see also Publications **III**, **VII** and **XI**):

Relevance or significance of the piece - this criterion reflects the novelty or innovativeness of the piece of art. It is intended to capture the "quality" of the piece in terms of its contribution/significance to arts and society. Innovativeness or contribution is not measurable (in fact to define attributes that describe an innovative piece of art is very difficult and to find such that would fit any segment of arts even more so). In fact if the evaluation of arts is to remain reasonable to the artists (and acceptable to them), it can not be stripped of the subjective component. The evaluation according to this criterion therefore requires expert (peer review) assessment - by experts from the respective segment of artistic production. Although it is not quantifiable, it permits the human component to enter the evaluation process. This makes the evaluation more "trustworthy" to the artists and more reasonable, as a criterion that is abstract to a certain level is present and it is not subject to some computation, by to expert assessment. Three values of this criterion are distinguished:

- **A** - a new piece of art or a performance of crucial significance;
- **B** - a new piece of art or a performance containing numerous important innovations;
- **C** - a new piece of art or a performance pushing forward modern trends.

Extent of the piece - this criterion is a bit easier to quantify. It is included to reflect the amount of (creative) work needed to produce the piece, the costs associated with it, number of people

involved in the creation of the piece and so on and of course the actual extent of the result. Its values are specified on a general level linguistically in the following way:

K - a piece of art or a performance of large extent;

L - a piece of art or a performance of medium extent;

M - a piece of art or a performance of limited extent.

Institutional and media reception/impact of the piece - a criterion that reflects how the particular piece of arts was received, whether it is known only close to the area of its origin/presentation or whether a wider audience has knowledge about it. Out of these three criteria the values of this one are considered the easiest to specify - lists of institutions and/or events are provided, with each institution classified as having (being associated with) an influence on the region only, on the national or even international level. Its linguistic values are:

X - international reception/impact;

Y - national reception/impact;

Z - regional reception/impact.

The values of the criteria are defined primarily linguistically. The reason for this is the need to be able to assess any piece of art according to all the criteria. Art is about creating something new, about breaking form, about being innovative - if the values of the criteria were too specific, it would be easy to "fall out" of them and not to be able to use any of the values for a new piece of art. The linguistic level here provides the required flexibility of the values of the criteria (if defined generally enough the same linguistic values can be used in all the segments of art and applied to various types of artistic production).

We also need to realize that since there is a significant amount of subjectivity in the evaluation of works of art, the evaluation needs to be performed by several evaluators. Too abstract values of criteria permit too different interpretations and this might complicate the aggregation process of evaluations provided by different evaluators - the same value of criterion might be understood differently by each evaluator and thus represent different evaluation of the piece of art. In such a case there would be no possibility of realizing these discrepancies in meaning from the perspective of the mathematical model.

It is necessary to keep the general linguistic descriptions of the values of criteria to be able to evaluate any possible work of art and to have a universal description of the values of criteria, that would be acceptable and meaningful for all the segments of art. But at the same time we need to provide some means of "*calibrating each evaluators interpretation of these general values*" to be able to unify (at least roughly) the understanding of the values among various evaluators. This has to be done not only among evaluators from one segment of art, but also across all the segments. We can provide at least leads as to what meaning should be assigned to the linguistic values of criteria. For the criterion of extent it can be done for example by specifying the number of pages for literature or duration of a play for theatre at first (the incompatibility principle seems to manifest itself here as well - the more precise the specification, the more special cases that need to be treated differently appear; the process of specifying the values to guide the interpretation of values K, L and M took more than 2 years and the process will continue). For institutional and media reception, we have specified lists of institutions and events. For the relevance and significance criterion, at least typical

examples of real-life outputs are available in each segment for each type of outputs relevant for the given segment.

Thanks to the linguistic values of the criteria that provide a general enough description of the each criterion's value, it is possible to obtain an evaluation of any piece of art (from any segment) according to all the criteria. The evaluation can be summarized by a triplet of letters - e.g. AKY or CMZ. 27 categories of works of art can thus be defined using the three criteria and their values. Due to the general linguistic descriptions of the values of all criteria, we now consider 27 general types of outputs in the whole arts sector. Each such output needs to be assigned a score. The sum of scores of all the outputs produced by one institution can then be used as a measure of its artistic performance. This measure can be used to make strategic decisions and to distribute funding.

Considering the situation, pairwise comparisons were chosen to obtain the information on the preferences on the set of 27 abstract categories of works of art. Pairwise comparisons have the advantage that we require the decision maker to choose a category that is preferred to the other for each pair of categories, and we can even reflect the strength of the preference. Saaty's AHP provides a good base for this task, as the evaluations obtained by this method are exponential - in the sense that the difference in scores between two best categories is much larger than the difference between any other pair of neighboring categories (see Publication **VII**). This way top quality can be effectively appreciated. In the Czech republic this is also in accordance with the evaluation of scientific papers.

There are however several issues that prevent us from directly using the AHP method. If we take a look at the three criteria and their values, we can see that the criteria might be partially dependent. Classic AHP approach would require us first to find the weights of all the criteria and the relative weights of each value within each criterion and then aggregate these to obtain an evaluation of any of the 27 categories. When there are dependencies among the criteria, the ANP can be used [85], which would require working with the abstract linguistic definitions of criteria, which can be a significant problem for the experts. Comparing abstract categories is in general a difficult task. If we consider we are talking about abstract categories describing works of art originating from seven different segments of artistic production, we can see the complexity of such task. Alternatively, we can work directly with the 27 categories - these are no longer interdependent and each can be represented by a real life example. Using the examples transforms the task of comparing general categories into comparing specific "typical examples" of works of art belonging to these categories. This can be done by the evaluators.

The AHP is however not intended for direct comparisons of so many objects - first of all $n(n-1)/2$ comparisons need to be made when n categories are considered (351 in our case). This may be too much for the evaluators. Also consistency of the preferences of evaluators expressed by their matrix of preference intensities might be an issue. To asses, whether the matrix of preference intensities is consistent enough using the $CR < 0.1$ condition, we need a complete matrix of preference intensities - all the 351 comparisons need to be made before we can check whether the matrix is consistent enough. If the matrix is insufficiently consistent, either modification of its elements need to be made in cooperation with the evaluators, or the input process repeated. With so many pairwise comparisons this is not convenient at all. Ideally the consistency of the matrix should be checked after each input and the decision maker should not be allowed to input such intensities of preferences that are very inconsistent (counterintuitive). Also we should remind that the criteria and their levels are defined linguistically - linguistic descriptions of intensities of preferences should therefore be expected as inputs and a consistency condition compatible with them should be used.

Weak consistency (see definition 4.0.8) has therefore been introduced in Publication **XI**. In Pub-

lication **III** its properties are discussed in more details. Among the most important ones are the possibility of checking its fulfillment after each input of an element into the matrix of preference intensities and compatibility with the linguistic labels of preference intensities used in Saaty's scale.

If the categories are ordered in accordance with their significance from the most preferred one to the least preferred one (that is a_i is preferred to a_j whenever $i < j$), the weak consistency condition can be formulated in a very simple way - the values in the matrix of preference intensities S need to me nondecreasing from left to right in each row and from the bottom up in each column. This is an easy rule that can be followed by all evaluators. The initial ordering of the categories can be obtained using the pairwise comparison method (see Publication **III** or **XI**). Weak consistency is understood here as a minimum requirement on the consistency of expert preferences. It can be maintained during the input process so that after 351 comparisons a weakly consistent matrix of preference intensities is obtained. The scores of the categories are then computed from this matrix using the eigenvector method and linearly transformed so that the maximum value is 305 (in accordance with the maximum score of a scientific journal in the R&D evaluation methodology).

The evaluation of each work of art involves the assessment of each work of art by its creator (and his/her university or faculty representative). This way an initial category (triplet of letters) is assigned to the output. In the second step this evaluation is assessed by the board of the given segment of art, who also suggest an appropriate category (after this, the university/faculty can reconsider its initial assessment or keep it). These two suggestions of categories then go to two independent experts, who assess the work of art as well and assign a category that is appropriate in their opinion. In the end four categories are suggested. The final evaluation is determined by a majority opinion on the levels of the three criteria. In indecisive cases the evaluation by independent evaluators is favoured. We can see that the evaluation procedure is in fact a process of finding an appropriate category (defined by the values of the three evaluation criteria) for the given work of art, where independent external evaluators play an important role. When the list of institutions and events for the institutional reception is available and so are the leads for the assessment of the extent, there is no reason why the categorization according these two criteria should differ. In this case the independent evaluators can concentrate an the relevance and significance of the piece - providing a "peer-review quality assessment" of the piece of art.

Publication **VII** also discusses several implications of the use of weak consistency in the adjustments of the model, provides a detailed description of its development and discusses outputs of the analysis of the performance of the proposed mathematical model on real-life data (3902 works of art produced in 2012).

4.2 A case of R&D outcomes evaluation using fuzzified AHP

The integration of a more subjective component into the evaluation process of R&D outputs, represented usually by some sort of a peer-review assessment is also desirable. The reason for the development of R&D evaluation systems is the need of effective allocation of research funding. Governments and various institutions have a limited amount of money - hence high quality research has to be identified to allocate funding here. This can be done in many different ways - see various national methodologies for R&D outputs assessment. If e.g. scientific papers are considered, there are two different approaches to quality assessment that come to mind. The first considers the quality of the media the paper is published in (e.g. the impact factor of the journal). This information is only an indirect assessment of the quality of the paper itself, as it is based on the number of citations of

papers published in the given journal in the past. In this case the quality of the journal is extrapolated from the fact, that it got through the same review process with the same scientific quality requirements as the papers in the past. On the other hand there are measures of the impact of a paper on the scientific community - e.g. the number of its citations. But the number of citations does not indicate whether the citations were favourable or critical (in fact in both cases the purpose of science - to promote critical thinking and to extend our knowledge of the world might be achieved). There are also various conditions in various fields of science - in experimental fields, the citations of a paper can be expected to occur quickly after it has been published, in more theoretical research citations can start to appear after e.g. 20 years. As each measure of scientific quality has its strengths and weaknesses, the peer-review approach to quality assessment in R&D is being discussed. It is after all a common practice in international evaluations of research units, a part of the feedback process in large R&D projects and so on.

In the Czech Republic we have been faced with this problem as well. The national R&D evaluation methodology (e.g. for the evaluation of results achieved in 2012) used a different approach to quality assessment for different types of R&D outputs. For scientific journals, the impact factor was assumed to be a sufficient measure of quality (each paper was assigned a score based on the ranking of the journal it was published in in a sequence of journals in the given field ordered in a descending order according to the impact factor). The evaluation scale for scientific papers was chosen to be exponential in the Czech national R&D evaluation methodology (for funding purposes) - papers in top journals were assigned significantly more points than papers in journals with lower impact factor (see e.g. Publication **II** or **VIII** for the exact formula for computing the scores and a discussion how the exponential character of the scale can influence academic faculty evaluation models). In any case, the quality of the paper (and the research it presented) was reflected at least by the impact factor of the journal.

On the other hand the same methodology assigned a fixed amount of points to any scientific monograph. This was regarded as not suitable for the purpose of quality promotion, as regardless of the quality of the scientific monograph (only formal criteria had to be met - number of pages, etc.), it will be assigned the same score. The element of motivation is however missing in such evaluation - nothing stimulates the authors to publish quality monographs, as the quality is not assessed and hence not reflected. This was the reason why the Faculty of Science of Palacký University in Olomouc introduced a peer review assessment of scientific monographs. Publications **V** and **XII** present the mathematical model based on fuzzified AHP designed to assist in the peer review process by providing evaluation intervals and "default evaluations" as a starting point for the peer review process.

The main idea of the evaluation methodology presented in Publications **V** and **XII** is to combine the assessment of the quality of the media (represented here by the scientific reputation of the publishing house), and the quality of the monograph assessed in a peer-review process at the university. Four categories of publishers were defined based on their scientific reputation by the board of experts (this is obviously dependent also on the field of science) - from *Category 1* being of the highest reputation to *Category 4* being of the lowest scientific reputation, but still considered a scientific publisher. The categories were defined by an iterative heuristic clustering procedure, after which 4 groups of publishers were found and characterised by a general linguistic description. This was done to provide some limits to the subsequent peer-review assessment. Each category was then assigned an interval of possible scores that any book published by a given publishing house might be assigned. The choice of the particular score within this interval was to be done based on the peer-review assessment of its quality. This way much of the subjectivity of the peer review process

has been removed by providing limitations for possible scores.

Initially, the group of experts provided the following suggestion of the intervals of scores.

$$\begin{array}{ll} \text{Category 1:} & 50-75 \text{ points,} \\ \text{Category 2:} & 30-40 \text{ points,} \\ \text{Category 3:} & 15-20 \text{ points,} \\ \text{Category 4:} & 5-10 \text{ points.} \end{array} \tag{4.11}$$

These intervals were presented to the academic senate as being a result of a discussion of the board of experts (and consensus of the board of experts was reached). It was, however, not clear what the intervals represent and whether they are in accordance with the desired goal. The matrix of preference intensities was therefore suggested by us to be used to visualize the preference structure of the board of experts. The following matrix of preference intensities was constructed using the middle values of each interval to represent a typical evaluation (see Publication **V** for a detailed discussion of the whole process):

$$S = \begin{pmatrix} 1 & 1.79 & 3.57 & 8.34 \\ \frac{1}{1.79} & 1 & 2.00 & 4.67 \\ \frac{1}{3.57} & \frac{1}{2.00} & 1 & 2.33 \\ \frac{1}{8.34} & \frac{1}{4.67} & \frac{1}{2.33} & 1 \end{pmatrix} \quad \rightarrow \quad \text{rounded } S = \begin{pmatrix} 1 & 2 & 4 & 8 \\ \frac{1}{2} & 1 & 2 & 5 \\ \frac{1}{4} & \frac{1}{2} & 1 & 2 \\ \frac{1}{8} & \frac{1}{5} & \frac{1}{2} & 1 \end{pmatrix} \tag{4.12}$$

The indices of the elements in the matrix represent the numbers of categories. From the elements directly above the main diagonal we can see, that *Category 1* is a bit closer to *Category 2* (only 1.79 times more preferred) than *Category 2* to *Category 3* (2 times preferred) and so on. Now let us try to use the linguistic level of Saaty's scale to interpret the values. As only integers are assigned linguistic labels by Saaty, we can round the values in the matrix to obtain linguistically interpretable values. Doing so, we realize that "*each category is between equally preferred and slightly preferred than the next (worse) category*". That is if we consider the linguistic labels as proposed by Saaty to be appropriate for the numerical values. Even if not, we can interpret the preference structure to tell us that "*an average book published by a publisher of higher category can be compensated by two average books published by a publisher from the next lower category*". This interpretation is too crisp to fit the purpose of evaluation. We are talking about average books.

The notion can be generalized in the following way "*a book published by a publisher of higher category can be compensated by* ***about two*** *books published by a publisher from the next lower category*". This way we are consistent with the original requirement of obtaining evaluation intervals. We now, however, need an appropriate representation of "about two" which can be obtained by using a fuzzy number $(1, 2, 3)$ in accordance with the scale presented in Figure 4.3 (see also e.g. Publication **V** or [24, 53]).

In the context of evaluating the research of their colleagues, the board of experts preferred to limit the possible effects of the peer review to such a level that would not produce many conflicts, but that would still provide means for the encouragement of quality publications. The mathematical model for their decision support was therefore designed to provide a default evaluation of an "average" or typical book published by a publisher from a given category. This is achieved by utilizing the fuzzy

number representation of the intensities of preferences (the scale presented in Figure 4.3) and using (4.8), (4.9) and (4.10) to compute the evaluations from a fuzzified matrix $\tilde{S}$.

We can define several alternatives of the fuzzified matrix $\tilde{S}$ based on the information summarized in (4.12). We can e.g. use the fuzzified scale in Figure 4.3 to construct a matrix $\tilde{S}_1$:

$$\tilde{S}_1 = \begin{pmatrix} 1 & (1,2,3) & (3,4,5) & (7,8,9) \\ \left(\frac{1}{3},\frac{1}{2},1\right) & 1 & (1,2,3) & (4,5,6) \\ \left(\frac{1}{5},\frac{1}{4},\frac{1}{3}\right) & \left(\frac{1}{3},\frac{1}{2},1\right) & 1 & (1,2,3) \\ \left(\frac{1}{9},\frac{1}{8},\frac{1}{7}\right) & \left(\frac{1}{6},\frac{1}{5},\frac{1}{4}\right) & \left(\frac{1}{3},\frac{1}{2},1\right) & 1 \end{pmatrix}. \tag{4.13}$$

The evaluations of each category - fuzzy numbers represented by a triplet of their significant values - can be computed in the form of (4.14). The evaluation weights computed using (4.8), (4.9) and (4.10) have been multiplied by 100 to avoid decimals.

$$\begin{aligned} \tilde{h}_{\tilde{S}_1 1} &= (41,53,61) \\ \tilde{h}_{\tilde{S}_1 2} &= (19,28,40) \\ \tilde{h}_{\tilde{S}_1 3} &= (9,13,20) \\ \tilde{h}_{\tilde{S}_1 4} &= (5,6,9)\,. \end{aligned} \tag{4.14}$$

The supports of these fuzzy numbers can be interpreted as *intervals of possible scores* for any book published by a publisher from the given category. The values in the kernels can be interpreted as representing "*a default score*" - a score assigned to a typical book published by a publisher from the given category (neither really great nor really bad in comparison with other books published by the publisher from the category). The values from the kernels would also be obtained by classic AHP if not a fuzzy scale, but Saaty's original scale was used (they correspond to the evaluations that would be computed from the rounded S in (4.12)). The classic AHP is in fact included in the computations and its results are among the results obtained by the fuzzified AHP.

We should note that the scale used to express (4.14) is different from the scale used in the intuitive expression of the evaluation intervals in (4.11). The values in (4.11) and (4.14) are therefore not directly comparable. We can compare the relations and gaps between the intervals. We can see that there are now no large gaps between the evaluation intervals, there is even a small overlap between the interval of possible evaluation for *Category 2* and *Category 3*.

When the preferences of the board of experts were visualized to them using the matrix of preference intensities and the values in the matrix interpreted linguistically in the previously described way, the evaluators realized that their intentions are better captured by (4.14) than by the original evaluation intervals. Publications **V** and **XII** describe the development of the mathematical model for the evaluation of R&D outcomes more in detail and discuss the benefits of using fuzzified AHP.

It was interesting to see that as the discussion with the board of experts continued and their understanding of the abilities of fuzzy set representation of preference intensities increased, they even reconsidered their original view of 2 books from lower category compensating 1 book from a higher one into the following statement: "*a book published by a publisher from a higher category can be compensated by **about 3** books published by publishers from the neighboring lower category*", where about 3 is represented as $(1,3,5)$ - this statement expresses a stronger preference of publishing in higher category publishing houses (typically 3 typical books published by lower category publishers can compensate 1 published in a higher category publishing house, but even 5 might be needed to compensate it if they are of low quality) and also more uncertainty. On the other hand the previous approach is included here as well, as this representation of preference intensity still admits the

possibility that a single book from a lower category publisher (a good one) can compensate for one book published by a higher category publisher. In Publications **V** and **XII** the use of a fuzzified version of the Saaty's scale $\{1, 3, 5, 7, 9\}$ without intermediate values is explained. Based on this information expressed linguistically and modelled by $(1, 3, 5)$ as an intensity of preference between neighboring categories, we can construct a fuzzy matrix of preference intensities $\tilde{S}_2$, where the information provided in the linguistic form is reflected by $\tilde{s}_{12} = \tilde{s}_{23} = \tilde{s}_{34} \sim (1, 3, 5)$, the reciprocal values need to be $\tilde{s}_{21} = \tilde{s}_{32} = \tilde{s}_{43} \sim (\frac{1}{5}, \frac{1}{3}, 1)$, there are ones on the main diagonal and the rest of the matrix is completed so that the matrix maintains maximum multiplicative consistency for the values in the kernels of the fuzzy sets.

$$\tilde{S}_2 = \begin{pmatrix} 1 & (1,3,5) & (3,5,7) & (7,9,9) \\ \left(\frac{1}{5}, \frac{1}{3}, 1\right) & 1 & (1,3,5) & (3,5,7) \\ \left(\frac{1}{7}, \frac{1}{5}, \frac{1}{3}\right) & \left(\frac{1}{5}, \frac{1}{3}, 1\right) & 1 & (1,3,5) \\ \left(\frac{1}{9}, \frac{1}{9}, \frac{1}{7}\right) & \left(\frac{1}{7}, \frac{1}{5}, \frac{1}{3}\right) & \left(\frac{1}{5}, \frac{1}{3}, 1\right) & 1 \end{pmatrix} \tag{4.15}$$

From (4.15) we can compute the following fuzzy-number-evaluations of the four categories of publishers.

$$\begin{aligned} \tilde{h}_{\tilde{S}_2 1} &= (38, 58, 69) \\ \tilde{h}_{\tilde{S}_2 2} &= (14, 25, 45) \\ \tilde{h}_{\tilde{S}_2 3} &= (6, 11, 23) \\ \tilde{h}_{\tilde{S}_2 4} &= (4, 5, 10)\,. \end{aligned} \tag{4.16}$$

The intervals of possible scores for each category (supports of the fuzzy numbers) now overlap significantly. This is a major shift of attitude of the board of experts. Surprisingly, this result was at the end accepted as best capturing the intentions of the board of experts. The reason why the intervals overlap is understandable to the board of experts as the preference structure is represented by the matrix of preference intensities (and thus made explicit) and the main underlying idea behind the evaluation is summarized linguistically - "*a book published by a publisher from a higher category can be compensated by* ***about 3*** *books published by publishers from the neighboring lower category, also 1 good book might be enough, and possibly as many as 5 books of below average quality*". The mathematical model has been used for the evaluation of scientific monographs at the Faculty of Science, Palacký Univeristy in Olomouc in 2013.

The linguistic modeling as well as the possibility of adding uncertainty of linguistic description into the mathematical model can result in more intuitive and desirable outputs of the mathematical model. To reach this a continuous discussion with the decision makers is needed. We can see that the fuzzified AHP provides the same information as the classic AHP if only 1-cuts (α-cuts for $\alpha = 1$) of all the fuzzy numbers in $\tilde{S}$ are considered. It also provides additional information on other possible outcomes of the evaluation process if different values of α are considered. The resulting evaluations can then be interpreted as intervals of possible scores (supports of the evaluations) with a default evaluation available (value in the kernel of the resulting evaluation). This makes the peer-review process subsequent to the first phase of the evaluation based on initial category of publisher possible, but not necessary. If the peer-review is not needed, a default value is available as an evaluation. However if a shift of the evaluation from the default value is required it can be done within the predefined interval. This way the subjectivity of the peer-review is allowed but still restricted by providing reasonable boundaries.

CHAPTER V

Linguistic modelling in HR management

As was already mentioned in the section on linguistic approximation and retranslation, it is important to provide outputs from linguistic models in such a form, that is intuitively understandable to the decision maker. It has to contain all the necessary information and should not appear to be more exact or more uncertain than it really is. This could prove to be a challenge for classic linguistic approximation and retranslation methods. It is, however, possible to design the linguistic models directly in such a way that would not require complicated linguistic approximation.

In this chapter, we will summarize the main principles of the academic faculty performance evaluation system (Information System for Academic Faculty Evaluation - IS HAP in Czech) that has been originally developed for the Faculty of Science, Palacký university in Olomouc - more details can be found in Publications **II**, **VIII** and **IX**. Let us note, that the IS HAP has been under constant development since Publication **II**, hence the linguistic scales used in the model have undergone some changes since the first publication - the up-to-date version of the model will be summarized in this chapter. The main features of IS HAP from the linguistic modelling point of view are the following:

- The *evaluation process is described linguistically* using a linguistic fuzzy rule base for the aggregation of partial evaluations in two main areas of interest - PA (pedagogical activities) and RD (research and development). A specific approximate reasoning mechanism was introduced in Publication **II** that similarly to the Sugeno-Yasukawa approximate reasoning [94] uses crisp representation of the consequent parts of fuzzy rules. This allows for simple interpretation of the outputs without unnecessary loss of information. HR management is a typical area where humans are involved both as evaluators and as evaluatees - any mathematical model should therefore be understandable (in terms of required inputs, process and outputs) to all the stakeholders.

- The evaluation is carried out in *multiple steps* - evaluations on different levels of aggregation are available - from the source data through partial evaluations in PA and RD to the overall evaluation. The *outputs of each level are available in linguistic form*. Linguistic scales are used to describe partial and overall evaluations.

- Aggregated evaluations are provided to the decision maker as *graphical outputs* as well - this allows the evaluator to get a quick overall idea about a large group of evaluated staff members in a very short time.

- As the final evaluation is the responsibility of the evaluator, the IS HAP is *designed as a decision support system*. The outputs are provided in graphical and linguistic form, and although carry enough information to form an overall opinion, they need to be interpreted by the evaluator - this way engaging him/her actively into the evaluation process.
- A threshold is set for the performance in each of the areas of interest, above which every performance is considered excellent. This *excellence threshold* prevents the construction of rankings of people. HR management and particularly faculty evaluation is understood as a management tool. It should identify strengths of the faculty, possibilities for improvement and provide information necessary to motivate and manage people, plan their development, provide a feedback to them and ensure the organisation achieves its goals.
- the use of linguistic modelling tools makes the outputs of the model intuitive and easy to understand. On the other hand the tools used (linguistic scales, linguistic FRB) make the evaluation model *easy to adjust* to the needs of any institution.

The IS HAP and the evaluation model that will be briefly summarized in the following text is already being used on several universities in the Czech Republic and is being implemented on others. The main ideas it is based on - simplicity of description, intuitiveness of outputs and easy adjustability to the needs of the institution combined with a sound mathematical base provided by a linguistic fuzzy model - make it an attractive HR management tool. Let us now take a closer look at the mathematical model that we have developed for academic faculty evaluation.

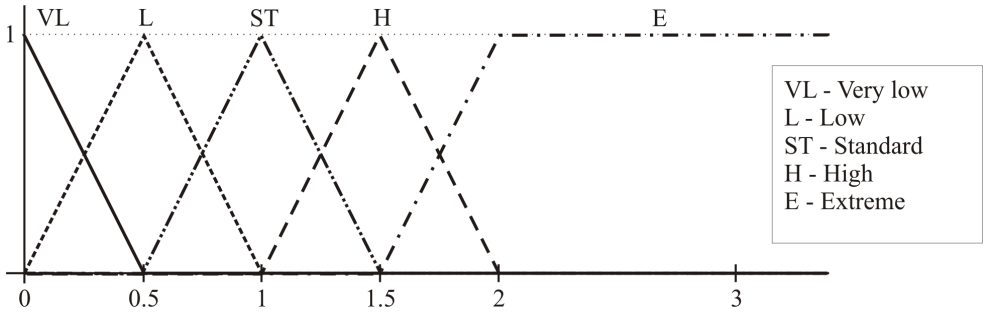


Figure 5.1: Meanings of the linguistic labels used to describe the performance of an academic faculty member in the area of pedagogical activities. The values of the universe are in terms of multiples of standard score for PA and given academic position. Reproduced from Publication **VI**.

The following requirements were set on the model by the evaluators:

i. It should be able to include, *all activities of faculty members* relevant to the well being of the university/faculty;

ii. Only easy to verify and *objective data* should directly influence the outputs. This requirement was set to provide a firm basis for the evaluation. The absence of "softer" information is reflected in the form in which the outputs of the evaluation are provided - graphical and linguistic inputs require elaboration and addition of other relevant pieces of information, the

evaluator is seen as an active part of the evaluation process and the IS HAP provides information for the evaluation (not the evaluation itself).

iii. It should be easy to work with - that is easy to understand, easy to input data, easy to interpret outputs. The new evaluation model should provide no reasons for its rejection by the academic faculty members or evaluators (dean, heads of departments, HR management unit).

iv. It should provide means for *reflecting the benefit of a particular academic faculty member* to the faculty/department. The main purpose of a staff evaluation system is not repression, but further development.

v. It should provide means for *flexible and complex evaluation*, that could be modified according to the needs of the Faculty of Science or its departments.

First attempts to develop an evaluation system in 2006 (see [98]) involved experimenting with OWA and WOWA aggregation operators ([113]). The use of such tools has proven to be too unpredictable (incomprehensible) for the evaluators and also for the people that were to be evaluated. An analysis and comparison of various aggregation operators and their appropriateness for HR management purposes is available in Publication **II**. In the end we have decided for "the obvious" - to try linguistic

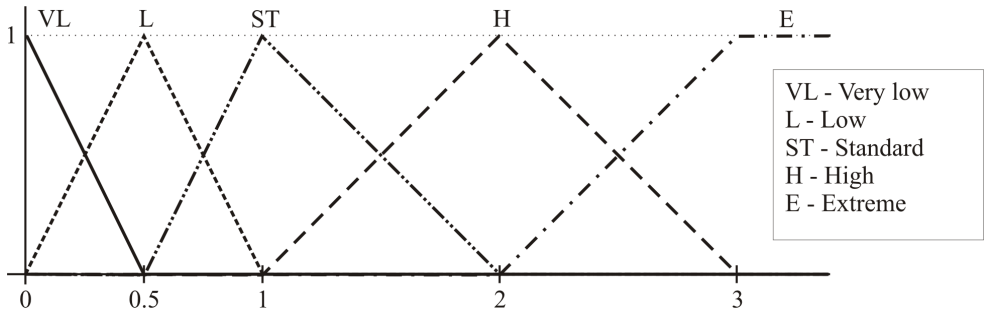


Figure 5.2: Meanings of the linguistic labels used to describe the performance of an academic faculty member in the area of research and development. The values of the universe are in terms of multiples of standard score for RD and given academic position. Reproduced from Publication **VI**.

modelling to design the evaluation model to fit these requirements. This decision has proven well, as the linguistic model was well received by the academic faculty members.

The performance of each academic faculty member is evaluated in both pedagogical (PA), and research and development (RD) areas of activities. Input data are acquired from a form, filled in by the faculty members, where particular activities are assigned a score according to their importance and time-consumption. Three areas are taken into consideration for pedagogical performance evaluation: a) lecturing, b) supervision of students, and c) work associated with the development of fields of study. The evaluation of research and development activities is based on the methodology for R&D evaluation valid in the Czech Republic; other important activities (grant project management, editorial board memberships etc.) are also included. Both PA and RD areas are assigned a standard score - different for senior assistant professors, associate professors, and professors.

The performance of each academic faculty member is then assessed in both areas by summing up the scores of activities performed by the faculty member and subsequently expressed in terms of multiples of standard score in the given category and position. Partial evaluations in PA and RD are obtained, which are linguistically interpreted using linguistic scales - Figures 5.1 and 5.2 summarize the linguistic terms of these scales and the fuzzy numbers representing their meaning. Each value of a standard score multiple can be linguistically interpreted by its membership degree to maximally two neighboring fuzzy numbers (retranslation/linguistic approximation is not necessary). For example a performance of 4.5 times the standard in RD will be interpreted as "*extreme*" (compatibility degree 1), whereas a performance of 1.1 times the standard score in PA would be interpreted as "*80% standard and 20% high*" or alternatively e.g. "*between standard and high, but much closer to standard*". We can see that although the same set of linguistic terms is used to describe the performance in PA and in RD, the meanings of the linguistic terms are different. This is a result of the difference in the character of the scales.

Overall performance of an academic faculty member in PA and RD		Research and Development performance				
		Very low	Low	Standard	High	Extreme
Pedagogical Activities performance	Very low	Unsatisfactory	Unsatisfactory	Substandard	Standard	Very good
	Low	Unsatisfactory	Unsatisfactory	Substandard	Very good	Excellent
	Standard	Substandard	Substandard	Standard	Very good	Excellent
	High	Standard	Very good	Very good	Excellent	Excellent
	Extreme	Very good	Excellent	Excellent	Excellent	Excellent

Unsatisfactory
Substandard
Standard
Very good
Excellent

Figure 5.3: Description of an aggregation of partial evaluations in PA and RD into overall performance. Certain level of compensation is allowed - "*very low*" performance in one area can be compensated by an "*extreme*" performance in the other to obtain "*very good*" overall performance assessment. Reproduced and modified from Publication **VI**.

Overall performance of an academic faculty member in PA and RD		Research and Development performance				
		Very low	Low	Standard	High	Extreme
Pedagogical Activities performance	Very low	Unsatisfactory	Unsatisfactory	Substandard	Substandard	Standard
	Low	Unsatisfactory	Unsatisfactory	Substandard	Standard	Very good
	Standard	Substandard	Substandard	Standard	Very good	Excellent
	High	Substandard	Standard	Very good	Excellent	Excellent
	Extreme	Standard	Very good	Excellent	Excellent	Excellent

Unsatisfactory
Substandard
Standard
Very good
Excellent

Figure 5.4: Description of another aggregation of partial evaluations in PA and RD into overall performance. Much less compensation than in Figure 5.3 is allowed - "*very low*" performance in one area results in "*standard*" overall performance assessment at best.

The activities in the PA area can be quantified (or at least approximately quantified) in terms of time consumption (how much time does it take to prepare for a lecture, how much time does it take to supervise a student, etc.). As such there is a natural limit to the maximum amount of time that a given faculty member can devote to pedagogical activities (there are only 24 hours a day). On the other hand in the RD area, the scores of activities were assigned in accordance with the R&D evaluation methodology valid in the Czech Republic. In essence quality affects the score of an output (via the impact factor of journals - see the second section in Chapter 4 or Publication **II** for more details) - e.g. the same type of publication can be assigned 10 or up to 305 points. It is therefore possible to reach higher multiples of standard score than in the PA area. It is clear that although both scales are expressed in standard score multiples, their values are in fact directly incomparable. Although transformations of the values of these scales can be made (this possibility is explored in Publication **II**), it is difficult to find appropriate interpretation for the resulting overall evaluation. Retranslation or linguistic approximation would in this case be problematic.

Before we proceed to the method of aggregation used in IS HAP, we should also notice, that an "excellence threshold" is set up by the definition of the meaning of linguistic term "*extreme*" in both areas (and the respective linguistic scales). We can see that any performance better than twice the standard score in PA or three times the standard score in RD will be considered *extreme* in the respective area (and this label will be absolutely compatible with such a performance). In linguistic level there is no better value than "*extreme*", but the numerical value of the multiple of standard score is still available in the model - the information is not lost, it is only deemed unnecessary for the purposes of the HR management. We can now define the aggregation of the partial evaluations in PA and RD using a base of fuzzy rules in the following way:

- In the first step we need to specify the ordered sets of linguistic terms describing inputs and output of the evaluation. We have already done so for the inputs specifying the linguistic scales depicted in Figures 5.1 and 5.2. At this point, we only need the linguistic term sets of these linguistic scales, the scales themselves will be defined at a later step. We also need to specify the linguistic term set for the output variable. Note, that the number of linguistic terms will influence the level of detail we will be able to achieve at the linguistic level. For now it would suffice to define $\mathcal{T}(\text{Overall performance}) = \{\text{Unsatisfactory, Substandard, Standard, Very Good, Excellent}\}$.

- We can now ask the evaluator to describe the desired output for each combination of inputs still on the linguistic level. We need to obtain an intuitively acceptable linguistic description of the evaluation. This description can also provide information to the academic faculty members that are to be evaluated concerning the possibility of compensating lower performance in one area by a better performance in the other area. The linguistic level will also serve as an interface for the evaluator to make adjustments to the evaluation methodology. This way it is easy to implement even very complex aggregation using natural language. Figures 5.3 and 5.4 provide two different examples of possible setup of the aggregation. In both these figures 25 linguistic rules are defined.

- We need to specify the universes for all input variables and "typical values" of their linguistic terms. That is each linguistic term of each input linguistic variable is assigned one element from the universe of discourse of this variable, that is most compatible with the linguistic term. In our case the universe is achievable multiples of a standard score, that is $[0, PA_{\max}]$ and $[0, RD_{\max}]$ where $PA_{\max}$ and $RD_{\max}$ are the largest possible values of standard score

multiples in PA and RD respectively. This way we obtain the set of most typical values for PA performance partial evaluation as $\{0, 0.5, 1, 1.5, 2\}$, where "2" is understood as "2 and more". For RD performance partial evaluation we get $\{0, 0.5, 1, 2, 3\}$, where "3" is understood as "3 and more". The same has to be done for the output variable overall performance. We define its universe as $[0, 2]$. Note, that the elements from this universe are not easily interpretable - it is difficult to assign a meaning to them. Let us therefore assume that it is just an arbitrary interval and let us distribute the typical values of the output linguistic terms evenly in this interval obtaining $\{0, 0.5, 1, 1.5, 2\}$. These values now identify *evaluation categories* and the evaluation using the already defined linguistic rule based can be considered to be a classification problem.

- To be able to compute the firing strength of each rule we now define the meanings for the linguistic labels (or values) of the linguistic variables *PA performance partial evaluation* and *RD performance partial evaluation*. The meanings will be modelled by linear fuzzy numbers on the respective universes, the kernels of these fuzzy numbers will consist of the "typical values" of the respective linguistic terms and the fuzzy numbers will be defined to form a Ruspini fuzzy partition of the respective universe. The resulting linguistic scales are summarized in Figures 5.1 and 5.2. We define a linguistic scale for the output variable analogically, the result of this process is summarized in Figure 5.5.

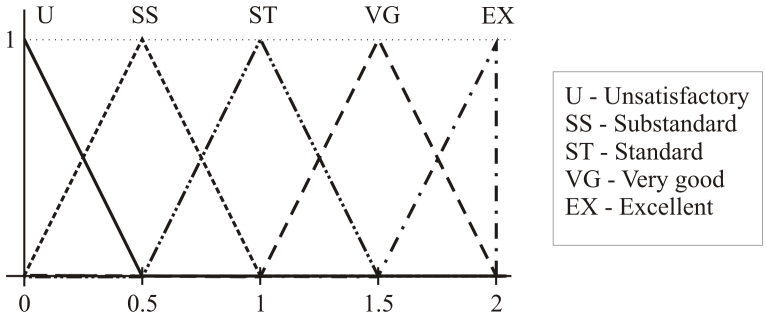


Figure 5.5: Meanings of the linguistic labels used to describe the output of the aggregation - the *overall performance* of an academic faculty member. Reproduced Publication **VI**.

- As was proposed in Publication **II**, we now define the output $eval$ for each pair of inputs $pa \in [0, PA_{\max}]$ and $rd \in [0, RD_{\max}]$ in the following way

$$eval(pa, rd) = \frac{\sum_{i=1}^{k} A_i(pa)B_i(rd)ev_i}{\sum_{i=1}^{k} A_i(pa)B_i(rd)} = \sum_{i=1}^{k} A_i(pa)B_i(rd)ev_i, \tag{5.1}$$

where A_i is a fuzzy number representing the meaning of the given value of *PA performance partial evaluation* in rule i, B_i is a fuzzy number representing the meaning of the given value of *RD performance partial evaluation* in rule i and ev_i is *the typical value for the linguistic output of rule* i, $i = 1, \ldots, k$ (in our case $k = 25$). The product T-norm is used to define the fuzzy rule base model. Our approach can be seen as similar to the fuzzy inference mechanism presented in [94], where ev_i would be defined as the COG of the respective fuzzy number from the output linguistic scale.

The mathematical model is constructed in this manner to allow us to find the interpretation of the output value $eval(pa, rd)$ easily. Note, that we require the meanings of the linguistic terms to form Ruspini fuzzy partitions of the respective universes and that we consider crisp inputs. We should also require the rule bases to be reasonable - in the case of evaluation we should require the evaluation function $eval$ represented by the rule base to be nondecreasing in both arguments (Figure 5.6 provides a plot of the evaluation functions represented by the rule bases depicted in Figures 5.3 and 5.4). We can also see that simple changes in the linguistic level of the model can define complex evaluation functions on the computational level of the linguistic model. An informed user is therefore able to make adjustments to the evaluation model by himself by changing the linguistic definition of the aggregation rules.

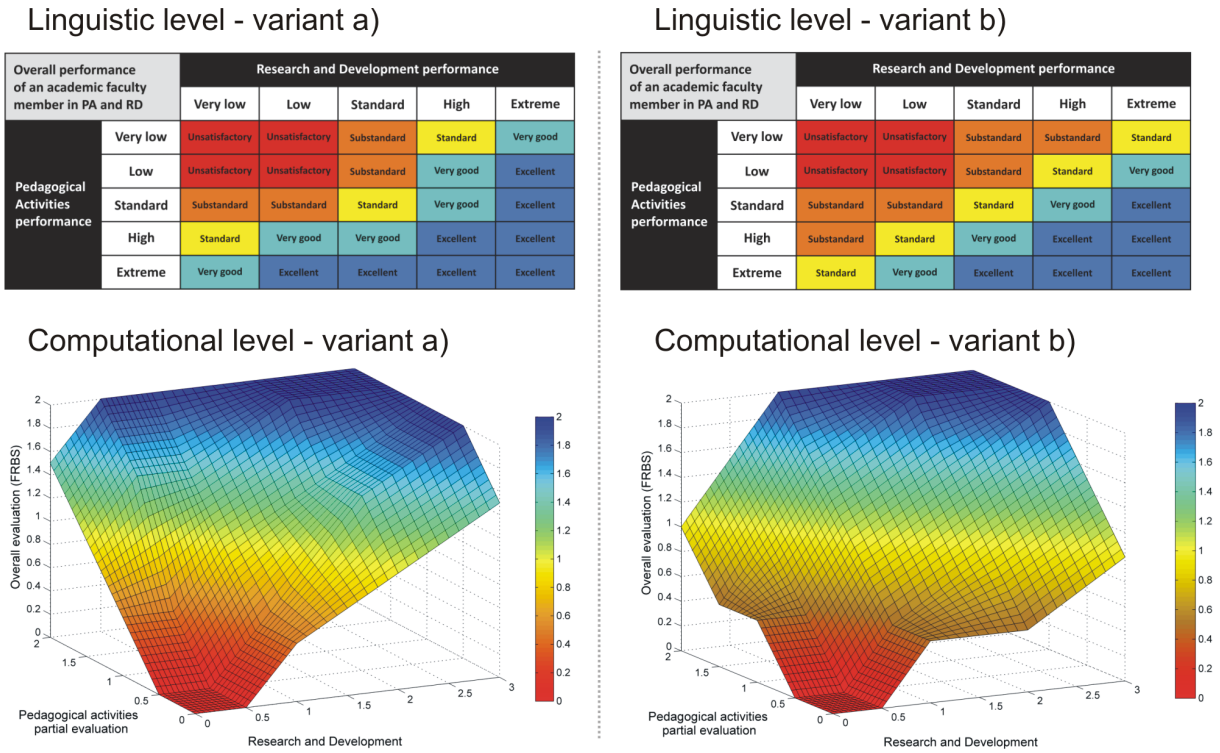


Figure 5.6: Comparison of the shape of the evaluation functions defined by the rule bases presented in Figures 5.3 and 5.4. Variant a - rule base is reproduced and modified from Publication **VI** and the evaluation function plot is reproduced and modified from Publication **II**.

The real number $eval(pa, rd)$ can be linguistically interpreted using the linguistic scale presented in Figure 5.5. An overall evaluation 1.7 is then interpreted as "*60% very good and 40% excellent*" (or alternatively *"between very good and excellent, but closer to very good"*). Another possibility of interpreting the output of the evaluation is using the colour of the respective value of the evaluation function (the basic color scale that is used in IS HAP is presented on the right side in Figures 5.3 and 5.4). This can either be done by providing a mixed colour (made of 60% of the colour associated with *very good* and 40% of the colour associated with *excellent*). Representing the outputs of the evaluation in colors provides the evaluator with visual and clear overall information on the status/evaluation of each academic faculty member. Partial evaluations (even multiples of standard scores in numerical form) and the raw data provided by the faculty members are also available. This

way a rough piece of information presented in colour or in linguistic form might be enough for quick orientation, but definitely not enough to make any decision.

More areas of interest can be reflected in the evaluation, as well as e.g. the load of managerial activities. The system provides information valuable for the evaluation process, while still leaving the responsibility of evaluation with the evaluator - consistently with our approach to linguistic decision support presented in the introductory chapter. Figure 5.7 provides a screenshot of an actual output of a software implementation of this evaluation model into IS HAP.

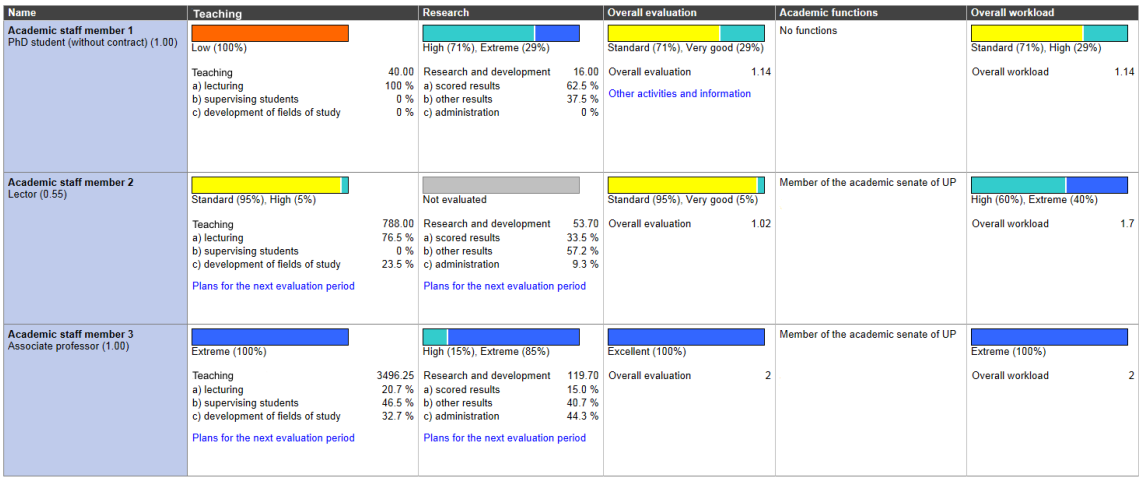


Figure 5.7: A screenshot of the overview of evaluations provided by the software implementation of the evaluation model in IS HAP - colour and linguistic outputs are available. Reproduced from Publication **VI**.

CHAPTER VI

Linguistic modelling in humanities

There are many challenges in designing mathematical models for sociology, management, psychology - humanities in general - that are given by the fact that the systems being modelled have a human component. Information needs to be extracted from humans or provided to them without unnecessary loss of meaning in the context of humanities. Linguistic modelling can prove to be very useful in this area. Publication **VI** maps several possible application areas of linguistic modelling in psychology and humanities in general and provides an overview of the available literature on linguistic (fuzzy) modelling in humanities. There seem to be many promising areas that could benefit significantly from the use of mathematical models with a linguistic level, able to handle uncertainty of natural language expressions. For the purpose of this thesis, our focus will be on the application of linguistic fuzzy modelling in diagnostics setting (an example of psychological diagnostics using the Minnesota Multiphasic Personality Inventory - 2 (MMPI-2) will be presented). A mathematical model for the interpretation of data from multidimensional psychological questionnaires that has been outlined in Publication **X** will be summarized here. In this publication we have introduced a linguistic fuzzy model for the interpretation of MMPI-2 protocols as a fuzzy classifier. The diagnostics criteria as well as the expert knowledge of a diagnostician are represented by a base of linguistic fuzzy rules.

Psychological diagnostics has some specific features in comparison with medical diagnostics. There are many concepts in psychology that are not directly measurable. The information, that is available in a diagnostics situation is therefore obtained either from client's documents, from the diagnostic review with the client or from various psychological diagnostics methods. Among these methods psychological inventories and questionnaires are frequent tools (the client provides agree/disagree answers or yes/no answers respectively). A great deal of information relevant for the diagnostics process is provided by the client and using natural language expressions. The client can, however, act as a filter and deliberately distort the information he/she provides (also unintentional distortion is conceivable - this also lowers the consistency of the information and makes its interpretation or use for diagnostics purposes difficult).

Various tools for the detection of deliberate distortion of data provided by the client (e.g. various lie scores) have been developed in psychology and are integrated into complex psychological instruments. In Publication **IV** we propose a method of reflecting the information concerning the level of data distortion (the quality or validity of the information provided by a client) in the process of performance assessment of diagnostics methods. As diagnostics situations can be seen as a classi-

fication problems (can be simplified to a binary classification task: client does or does not have a given diagnosis), diagnostics tools can be treated as classifiers. We propose a modification of the receiver operating characteristics (ROC) analysis and the area under the ROC curve (AUC). The modified ROC reflects the quality (or validity) of data provided by the client in the following way: misclassifications of data instances of low quality (validity) influence the performance rating of the classifier less than misclassifications of high-quality data (valid data).

6.1 Psychological diagnostics as a classification task - MMPI-2 interpretation

As was already mentioned in the introduction to this chapter, inventories and questionnaires are frequent sources of data in psychology. The outputs obtained by these methods are based on patient's answers to single test items, usually in the way of agree/disagree or yes/no. The items are usually designed so that it is not obvious what the psychologists are trying to find out. We can not be sure if the items were well understood by the patient (such words as "many", "usually", "small" etc. can be interpreted differently by different patients). There is also no absolute guarantee that a certain answer to a certain item in the test indicates that the patient/client is really ill and should be assigned a diagnose. The results of such "measurement", although represented frequently by real numbers (various indices), are uncertain, or at least their interpretation is not straightforward. Linguistic modelling can be of significant use when creating decision support models to assist practitioners in diagnostics situations.

We will consider here the possibilities of dealing with multidimensional data originating from questionnaires (particularly MMPI-2) for which the validity rate is known or can be computed. The term *validity* is used in connection with the MMPI-2. Alternatively the term *quality of data* can be used. Validity or quality of data refers here to our ability to come to reasonable conclusions based on the data - that is conclusions that describe well the person who provided the data. Distorted data (lies, deliberate faking of pathology, inconsistency) are considered to be of low quality (validity).

Multidimensional test methods (questionnaires or inventories) in psychology provide outputs usually in the form of scores of certain scales - e.g. T-scores. T-score is a score used for setting up norms for standardized psychological tests in the USA, it is a result of a linear transformation of normalized standard scores (T-scores have normal distribution, $\mu = 50$, $\sigma = 10$). It must be also stressed here, that the term "*scale*" has a different meaning in psychology than in mathematics. In the context of psychological diagnostics it describes a set of questions (items) of a particular questionnaire that are used to identify the presence or strength of a certain psychological phenomenon. A raw score of a scale is usually computed as the number of items answered in the pathological direction. Unless indicated as fuzzy or linguistic, the term scale will be used in this section of the chapter in the psychological sense.

The questionnaires or inventories usually measure more features (pathological) simultaneously. This makes it more difficult for the patient to recognize, which items identify which pathology. Apart from the items detecting the presence of pathology, items for monitoring the presence of misinterpretations and the distortion of answers are also present. Based on the previously mentioned characteristics we can compute one overall characteristic that determines the validity of the outputs of the questionnaire. In the proposed mathematical model, a fuzzy set representing an ideally valid output (MMPI-2 protocol) is defined. The measure of validity of each protocol is defined as a membership degree of the protocol in this fuzzy set.

Information concerning the validity of the data has not been directly included into the mathematical models for the evaluation of questionnaire data so far. We will present a mathematical model here that allows for inclusion of this information into the diagnostics process using linguistic rules.

Minnesota multiphasic personality inventory (MMPI)

MMPI (and its revised version MMPI-2) are test methods for psychological differential diagnostics developed by Hathaway and McKinley in 1930s (first introduced in [35]). According to Greene [31] it is the most widely used diagnostics tool for psychopathology assessment worldwide. The MMPI-2 version of the test consists of 567 statements (such as "There are many flowers in my house/flat."). The tested person provides "agree/disagree" answers to all these statements. Even here fuzziness can be traced - in the 567 statements, there are many uncertain expressions, such as "usually", "many", "rarely", etc.

The answers of a tested person saturate 10 basic clinical scales, 7 basic validity scales, many content scales and supplementary scales. A large amount of indices is also provided. Each clinical scale corresponds with one particular psychological diagnosis and was constructed so that its discriminating powers between healthy and ill people were maximized. In the process of choosing items with sufficient diagnostics value and identifying the diagnose these items suggest, more than one control group was compared to the criterion group. In addition to the group of healthy individuals, other control groups were used to minimize the influence of age and socioeconomic status; a group of non-psychological patients and a group of psychological patients with different diagnoses were also used as control groups. The resulting 10 clinical scales are reliable identifiers of 10 psychological diagnoses.

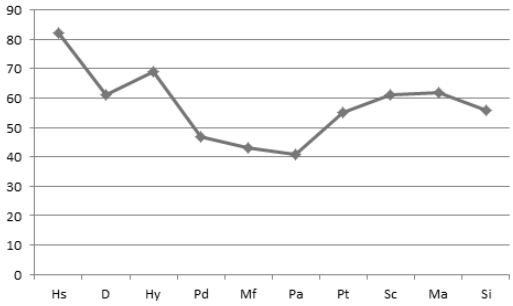


Figure 6.1: MMPI-2 profile illustrating the T-scores of 10 clinical scales.

In practical use the MMPI-2 protocol is computer evaluated and provides the diagnostician with more than 80 scales T-scores. A plot of the T-scores of the 10 clinical scales may be also provided (see Figure 6.1). Greene provides interpretation guidelines for these outputs in [31].

Validity in MMPI-2

The criteria for a valid protocol remain the same worldwide. Validity in MMPI-2 is not a single concept. Instead, several ways of possible data distortion (intentional or not) are monitored. A valid protocol is such that its validity is acceptable according to all the validity criteria. To better understand the need for a data quality measure in humanities, we will briefly summarize how validity is understood in MMPI-2. There are five basic measures of the concept of validity in MMPI-2 (the following descriptions are in accordance with the Greene's MMPI-2 interpretation manual [31]):

- *Unanswered items* (the "UI" scale) - the number of unanswered items within the first 370 items should not exceed 30.
- *Consistency* (the VRIN and TRIN scales) - two scales check if the person undergoing the test is consistent in his/her answers.
- *Desire to appear better* (the L scale) - this scale is saturated by such items, that describe qualities highly appreciated in the society but not frequently present (e.g. "I have never stolen anything.").
- *Bizarre answers* (F and Fb scales) - these scales monitor answers that were infrequent in the normalisation process of the MMPI-2. A valid protocol does not differ from the statistical norm for these items (norm is set by the standardisation process, this scale is saturated only by items, where 90% of the standardisation sample answered in the same way).
- *Underreporting or overreporting of pathology* (the UO scale) - this scale measures the effort of the patient to hide or exaggerate his/her symptoms.

If we wish to define a valid MMPI-2 protocol, it is such that has enough items answered ("UI" scale score is acceptable), the patient is consistent in his/her answers (VRIN and TRIN scores are acceptable), the patient is not lying (L score is acceptable), there are not many bizarre answers (F and Fb scores are acceptable) and the patient is not trying to hide or exaggerate his/her true state (UO scale score is acceptable). For each of the 7 mentioned validity scales an interval of acceptable scores is available in [31] (e.g. less than 30 unanswered items is acceptable). The statistical approach used to construct MMPI-2 introduces the well known counterintuitive "boundary problem" - the existence of a crisp threshold between norm and deviation (e.g. 30 unanswered items is acceptable, but when 31 items are unanswered, the protocol is considered invalid in the crisp sense). To overcome this problem and to make the proposed decision support tool able to reflect expert's experience in validity assessment, fuzzy numbers are used to represent the acceptable scores for each validity scale. A fuzzy set of valid protocols can thus be defined.

Fuzzy model for data validity assessment

The MMPI-2 interpretation guide [31] suggests crisp thresholds for all the validity scales scores for a valid protocol. It however acknowledges that there are some "grey areas" where the score is borderline. In these cases it is not possible to determine whether the protocol is valid or not and the interpretation of such protocol is therefore questionable. This is where the fuzzy approach provides a solution. Let $ui, vrin, \ldots, uo$ be the scores of the seven validity scales UI, VRIN,..., UO. The protocol validity can be defined as the truth value vr of the following compound statement: "*ui* is *acceptable* and *vrin* is *acceptable* and ... and *uo* is *acceptable*". This value can be understood as the membership degree of a protocol to the fuzzy set of valid protocols.

We can define the acceptable scores of all the validity scales by the following linguistic terms: *UI_acceptable, ...,UO_acceptable*. Meanings of these linguistic expressions can be modeled by linear fuzzy numbers $M(UI\_acceptable), \ldots, M(UO\_acceptable)$ that are defined on the respective scores domains - see Figure 6.2. The prototype of a valid protocol VL can then be defined as the following Cartesian product of fuzzy numbers:

$$VL = M(UI\_acceptable) \times M(VRIN\_acceptable) \times M(TRIN\_acceptable) \times M(L\_acceptable) \times M(F\_acceptable) \times M(Fb\_acceptable) \times M(UO\_acceptable). \quad (6.1)$$

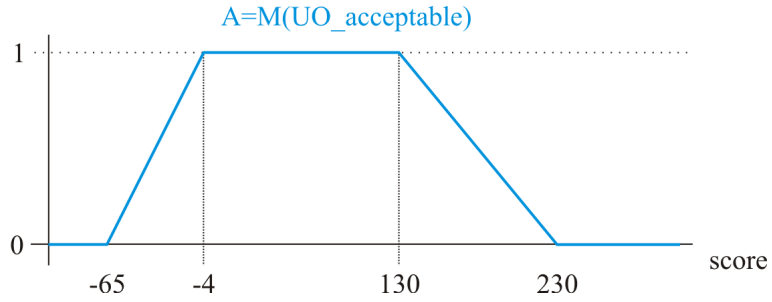


Figure 6.2: Fuzzy number representing the meaning of an "acceptable score" for the UO scale. Reproduced and modified from Publication **X**.

Using fuzzy numbers to describe the acceptable scores of all the validity scales, we can deal with the "grey areas", but we can also reflect the expert knowledge of the diagnostician in the process of defining the membership functions of fuzzy numbers representing the meaning of "*acceptable score*" for a given validity scale (let us just remark that the diagnostics criteria from the test manual should not be violated). If we consider a set of all MMPI-2 protocols $PR = \{PR_1, PR_2, \ldots, PR_k\}$, we can define the validity rate vr for each MMPI-2 protocol in PR in the following way:

$$vr_i = VL(ui_i, vrin_i, trin_i, l_i, f_i, fb_i, uo_i) = \min\{UI\_acceptable(ui), \ldots, UO\_acceptable(uo)\} \tag{6.2}$$

where $ui_i, vrin_i, \ldots, uo_i$ are the scores of the validity scales for PR_i, $i = 1, \ldots, k$. A fuzzy set of valid protocols VLP on PR can be defined in the following way:

$$VLP = \{^{vr_1}/_{PR_1}, ^{vr_2}/_{PR_2}, \ldots, ^{vr_k}/_{PR_k}\}. \tag{6.3}$$

The validity rate of each protocol is a valuable piece of information for diagnostic purposes. It determines whether the results of the diagnostics method can be interpreted at all, which other diagnostic methods should be used to confirm the results and provides a piece of information concerning the patient's attitude to the test, which can also be beneficial in understanding him/her better.

Mathematical model for fuzzy classification - diagnostics decision making

In accordance with [54] we can define a fuzzy classifier for the purposes of this paper in the following way. Let x be a vector in an n-dimensional real space $\mathbb{R}^n$ and let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ be a set of class labels (crisp, linguistic or fuzzy). A fuzzy classifier is an if-then inference system (a fuzzy rule based system) IS which either

a) yields a single class label (crisp, linguistic or fuzzy) for x:

$$IS : \mathbb{R}^n \rightarrow \Omega, \tag{6.4}$$

b) or for a discrete Ω maps $\mathbb{R}^n$ into a fuzzy set on Ω :

$$IS : \mathbb{R}^n \rightarrow F(\Omega) = \{\Omega_x \mid x \in \mathbb{R}^n\} \tag{6.5}$$

Table 6.1 Sample data - 20 MMPI-2 protocols, for each 10 clinical scales scores, 7 validity scales scores, diagnosis.

Protocol	UI	L	F	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	Si	O/U	Fb	VRIN	TRIN	Diagnosis
PR_1	0	38	57	65	59	57	45	43	50	49	54	63	40	15	47	75	40	H
PR_2	1	50	48	51	61	55	65	56	59	59	56	44	44	-6	57	75	58	H
PR_3	14	58	71	69	50	55	62	43	53	61	67	79	47	108	84	47	76	H
PR_4	0	50	60	42	81	51	63	58	56	94	84	41	89	212	94	58	53	H
PR_5	22	66	48	71	78	78	52	59	44	61	59	48	68	55	54	51	64	NH
PR_6	0	45	50	49	39	50	35	65	56	40	44	36	57	65	72	58	58	H
PR_7	0	50	54	51	50	55	49	59	44	55	54	56	31	-22	57	59	52	H
PR_8	6	54	71	82	61	69	47	43	41	55	61	62	56	132	57	43	58	NH
PR_9	0	54	80	52	44	40	40	46	56	55	50	70	62	158	57	51	70	H
PR_{10}	0	70	51	83	67	72	58	46	53	61	64	46	53	-49	43	43	46	NH
PR_{11}	0	50	72	68	59	60	45	56	50	66	66	90	54	139	61	65	58	H
PR_{12}	1	50	68	57	42	43	50	53	44	39	51	52	53	121	60	55	58	H
PR_{13}	8	37	81	75	97	73	78	53	82	102	83	51	85	196	97	58	58	NH
PR_{14}	1	66	65	64	63	68	59	65	56	56	53	57	57	61	58	65	58	NH
PR_{15}	5	54	57	70	59	59	34	61	44	49	50	56	62	62	54	59	40	H
PR_{16}	2	46	54	65	50	61	52	66	47	49	43	38	41	-32	43	43	52	H
PR_{17}	1	54	61	68	65	55	30	61	56	57	53	51	76	105	50	51	58	H
PR_{18}	0	37	60	75	85	68	58	49	50	64	62	38	68	158	61	47	63	NH
PR_{19}	0	58	69	78	89	75	60	56	62	88	81	53	76	205	76	47	48	NH
PR_{20}	1	50	74	61	69	80	68	65	56	56	60	58	63	115	54	69	48	NH

such that for all $x \in \mathbb{R}^n$ it holds that $\sum_{i=1}^{c} \Omega_x(\omega_i) = 1$. Thus IS distributes the full membership of x among the classes. The fuzzy set Ω_x can be interpreted as "appropriate class label for x".

Using linguistic fuzzy rules we can construct a classifier based on expert knowledge only. Although the knowledge of the expert is based on interpretation manuals as well as on working with particular protocols, the original expert's training set of protocols is not available any more. The classifier will be taught by the expert (expert will describe the diagnostics process linguistically) and then tested on real life data. Table 6.1 provides an overview of 20 protocols of real life patients (provided by the Faculty Hospital in Olomouc, all the ethical issues have been dealt with appropriately, NH represents "not healthy" - converse symptoms are present and confirmed by other diagnostics methods, H represents "healthy" - converse symptoms are not present). This data set will serve for testing purposes, as the expert that taught the classifier never came into contact with these patients. The information comprised in Table 6.1 was not used while constructing the classifier.

For the purposes of this paper we will consider only one diagnosis - the dissociative (conversion) disorders classified according to the international classification of diseases as F44.4-F44.7. This diagnostics category describes weakness or paralysis of parts of the body, losses of sensation, impaired vision or hearing with no known somatic cause. According to the MMPI-2 classification manual three clinical scales are important for this diagnosis - Hs (Hypochondrias), D (Depression) and Hy (converse hysteria). We need to check whether these 3 scales' T-scores are above all the other scales' T-scores (the *location condition*) and if their mutual relationship is such that their plot forms a "converse V" - (in other words that Hs and Hy are higher that D and neither Hs nor Hy is too dominant) - the *shape condition*. Only valid protocols should be assessed - the *validity rate* determination for each protocol has already been described in the previous section. In addition to the result of the classification, the validity rate of a particular protocol is obtained as a second result of our model. The validity rate serves as an additional piece of information for the practitioner - it tells the practitioner how much the results of the classification can be trusted. The validity rate can also be used (as a data quality measure) in the modified version of the ROC that will be presented in the next section to assess the performance of our classifier.

We will now define the rules for converse symptoms presence identification based on the MMPI-2 protocol in order to show how a fuzzy classifier can be constructed for such purpose. We also want to explore the role of the validity rate of data in the classifier performance assessment process. Let us notice that the validity rate assessment as well as the determination of the location and shape appropriateness rates are classical multiple criteria evaluation problems with partial evaluation functions modeled by membership functions of fuzzy sets (see [2, 96]). We begin with the *location condition*. It considers the relative elevation of the scales Hs, D and Hy and the remaining seven clinical scales (Pd, Mf, Pa, Pt, Sc, Ma, Si). The basic requirement that the scores of the three clinical scales Hs, D and Hy should be higher than the scores of other scales was further specified by an expert psychologist into the following condition: "If (*D-Sc* is *big enough*) and (*Pt-Mf* is *acceptable*) and (*Pt-Pa* is *acceptable*) and (*Pt-Ma* is *acceptable*) and (*D* is *elevated*) and ($Hy - \max(Pd, Pa, Pt, Ma, Si)$ is *acceptable*, then the *location* is *appropriate*". We need to specify the meaning of "*acceptable*" and "*big enough*" score differences for all the relevant pairs of scales (scores). This is similar to the validity rate determination. Let lar_i be the fulfillment rate of the previously mentioned linguistic fuzzy rule. We can define a fuzzy set of location-appropriate protocols LAP on PR:

$$LAP = \{{}^{lar_1}/_{PR_1}, {}^{lar_2}/_{PR_2}, \ldots, {}^{lar_k}/_{PR_k}\}, \tag{6.6}$$

where lar_i is called the *location appropriateness rate* of PR_i, $i = 1, \ldots, k$.

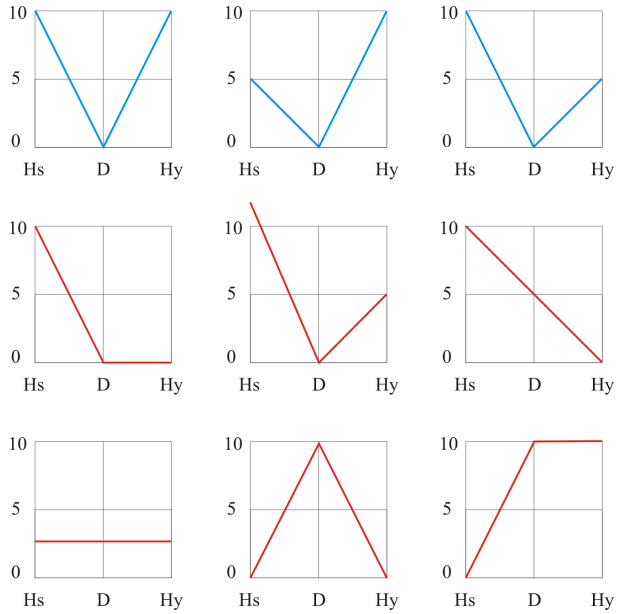


Figure 6.3: Examples of obvious "converse V" shapes (top row; shape appropriateness rate $sar = 1$) and indistinct "converse V" shapes (middle and bottom row; shape appropriateness rate $sar = 0$). Reproduced from Publication **X**.

The second step is to assess the shape of the plot of scores of the Hs, D and Hy clinical scales. According to [31], it should resemble the shape of the letter V. Figure 6.3 presents examples of obvious and indistinct "converse V" shapes. Mathematically, the required shape can be described by the conjunction of the following fuzzy conditions:

i. $(Hs - D)$ is significant and $(Hy - D)$ is significant.

ii. $(Hs - D)$ is very significant or $(Hy - D)$ is very significant.

iii. The (Hs_Hy_ratio) is acceptable.

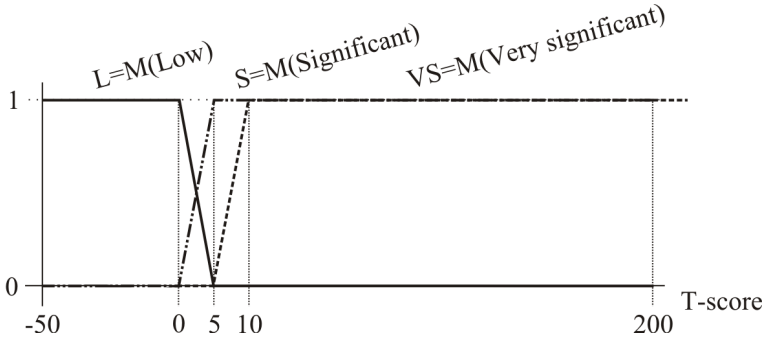


Figure 6.4: Fuzzy numbers representing the meanings of linguistic terms "significant" and "very significant".

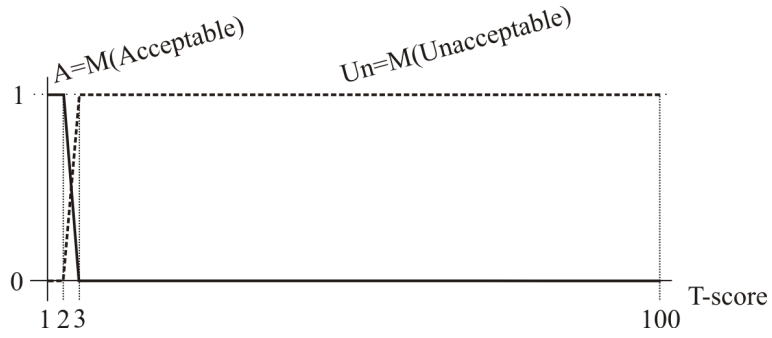


Figure 6.5: Fuzzy numbers representing the meanings of linguistic terms used to describe the Hs_Hy_ratio.

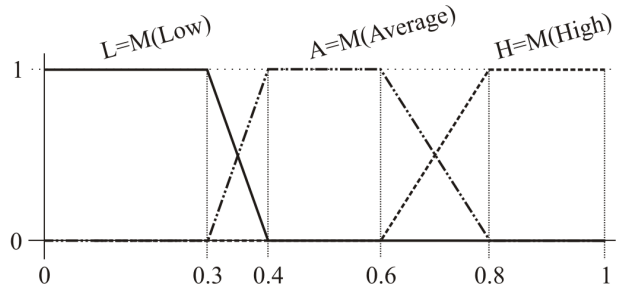


Figure 6.6: Meanings of the linguistic terms for lar and sar values description. Reproduced and modified from Publication **X**.

Here Hs, Hy and D are the respective clinical scales T-scores, (Hs_Hy_ratio) is described by the following formula:

$$(Hs\_Hy\_ratio) = \begin{cases} \frac{\max(|Hs-D|,|Hy-D|)}{\min(|Hs-D|,|Hy-D|)} & \text{if } \min(|Hs-D|,|Hy-D|) \neq 0, \\ 100 & \text{else.} \end{cases} \tag{6.7}$$

The conjunction is modeled by the $\min$ t-norm. Based on these conditions we get the *shape appropriateness rate* (sar) for each protocol as the minimum of the fulfillment rates of the three conditions. Figures 6.4 and 6.5 show the meanings of the linguistic terms "*significant*", "*very significant*" and "*acceptable*". We can define a fuzzy set of shape-appropriate protocols SAP on PR in the following way:

$$SAP = \{^{sar_1}/_{PR_1}, ^{sar_2}/_{PR_2}, \ldots, ^{sar_k}/_{PR_k}\}. \tag{6.8}$$

Once we have established the values of lar and sar (we know how to compute the rate of fulfillment of the location and shape conditions), we can start constructing the fuzzy classifier and its fuzzy rule base. We define linguistic scales for lar and sar input variables (see Figure 6.6). Their linguistic values will be used in the fuzzy rule base for final diagnosis determination at the end of this section. In our example we use the same linguistic scale for lar and sar. An approximate reasoning algorithm similar to the Sugeno-Yasukawa approach is used in this model (see Publication **X** for more details). In contrast with the Sugeno-Yasukawa approach (see [94]) we do not represent the fuzzy

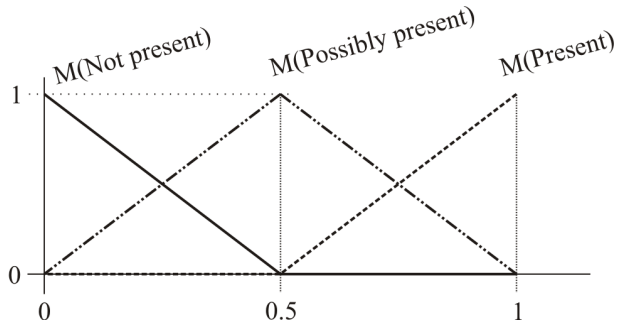


Figure 6.7: Meanings of the linguistic terms used for diagnosis determination.

Table 6.2 Fuzzy rule base applied in the classifier

Rule number	lar	sar	Output (conv. symp. are:)
1	High	High	Present
2	High	Average	Present
3	High	Low	Possibly present
4	Average	High	Present
5	Average	Average	Possibly present
6	Average	Low	Not present
7	Low	High	Possibly present
8	Low	Average	Not present
9	Low	Low	Not present

numbers on the consequent parts of the rules (the meanings of the linguistic terms $\mathcal{Y}_i$, $i = 1, \ldots, n$) by their centers of gravity but instead by other typical values - elements from the kernels of the respective linear fuzzy numbers (triangular fuzzy numbers are used in the presented application - see Figure 6.7).

The output of the modified Sugeno-Yasukawa approach y^{SY} can then be interpreted linguistically in the following way: if y^{SY} lies in the intersection of supports of two neighboring fuzzy numbers $Y_{i_k}, Y_{i_{k+1}}$ then the output y^{SY} can be characterized as being $Y_{i_k}(y^{SY}) \cdot 100$ percent of $\mathcal{Y}_{i_k}$ and $Y_{i_{k+1}}(y^{SY}) \cdot 100$ percent of $\mathcal{Y}_{i_{k+1}}$. If y^{SY} lies in the kernel of Y_{i_k}, it is interpreted as being $\mathcal{Y}_{i_k}$. Interpretability of the results is highly important, when the users of the method are laymen - psychologists. The modified Sugeno-Yasukawa approach used in Publication **X** provides easily interpretable results. Hence it is a suitable tool for decision support systems where the process as well as the results of the decision making need to be described linguistically. The rules presented in Table 6.2 were defined in cooperation with an expert diagnostician and reflect the diagnostics criteria suggested by Greene in [31] as well as the diagnostician's experience.

The described mathematical model works as a continuous fuzzy classifier. It assigns to every protocol PR_i a real value dg_i from $[0, 1]$, $i = 1, \ldots, k$. For the purposes of the rule base construction, we choose just 3 important values from $[0, 1]$ - 0 can be interpreted as "*conversion symptoms not present*", 1 as "*conversion symptoms are present*" and 0.5 represents "*conversion symptoms are possibly present*". Based on these three output values a fuzzy scale used in the rule base (see Table 6.2 reproduced from Publication **IV**) was constructed. For the meanings of the linguistic values of this scale see Figure 6.7. If we know the crisp inputs lar and sar for the current protocol, we can

Table 6.3 Fuzzy classifier outputs summary

Protocol	Diagnosis	Validity (vr)	Location (*lar*)	Shape (*sar*)	dg
PR_1	H	0.500	0.0	0.000	0.00
PR_2	H	0.500	0.0	0.000	0.00
PR_3	H	0.500	0.0	0.000	0.00
PR_4	H	0.000	0.0	0.000	0.00
PR_5	NH	0.960	1.0	0.000	0.50
PR_6	H	1.000	0.0	1.000	0.50
PR_7	H	0.705	0.0	0.000	0.00
PR_8	NH	0.980	1.0	0.375	0.88
PR_9	H	0.720	0.0	0.000	0.00
PR_{10}	NH	0.262	1.0	0.000	0.50
PR_{11}	H	0.910	0.0	0.000	0.00
PR_{12}	H	1.000	0.0	0.000	0.00
PR_{13}	NH	0.000	0.0	0.000	0.00
PR_{14}	NH	0.960	0.2	0.000	0.00
PR_{15}	H	1.000	0.0	0.000	0.00
PR_{16}	H	0.541	0.0	1.000	0.50
PR_{17}	H	1.000	0.0	0.000	0.00
PR_{18}	NH	0.720	0.0	0.000	0.00
PR_{19}	NH	0.250	0.0	0.000	0.00
PR_{20}	NH	1.000	0.2	0.000	0.00

obtain a crisp output dg. The calculation of dg is analogous to the application of Sugeno's inference mechanism to the elements from the kernels of fuzzy numbers forming the output fuzzy scale. We compute the weighted average of the important values (0, 0.5 and 1) corresponding with the consequent parts of the rules, where the weights are the degrees of fulfillment of the antecedent parts of the rules with the given values lar and sar. The real valued output dg_i is interpreted linguistically in terms of the linguistic scale presented in Figure 6.7 (e.g. for an output $dg_i = 0.8$ the converse symptoms will be interpreted as being 40% *possibly present* and 60% *present*). This in addition to the validity rate of the particular protocol is an appropriate output for the psychologists, as it can be easily understood and interpreted. Note that as there is much uncertainty in the diagnostics process, a "don't know" type category is present ("*possibly present*") is available to describe the outputs linguistically.

6.2 Classifier performance assessment - reflecting data quality in ROC

In the previous section, the information concerning validity rate of each protocol was provided as an additional piece of information to the conclusion suggested by a fuzzy-rule-based classifier and presented in terms of natural language. The question of quality of the mathematical model (or in fact the expert knowledge represented by the model) is now at hand. How well does the classifier described by the base of linguistic fuzzy rules (see Table 6.2) perform on a given set of data? As classifiers are widely used mathematical tools, there are many tools for classifier performance assessment (see e.g. [33, 72] for an overview).

The diagnostics situation described in the previous section has one important specific feature - we

know, that the data based on which we are trying to assign a diagnosis may be distorted and we have a measure of such distortion. The validity rate vr provides such measure - the higher the validity rate, the higher the quality of data. In general when classification problems are considered, we assume that a data set that contains information necessary to design a well working classifier (mathematical model of a classifier) is available. That is we assume that the data we have available are not distorted, or that there is no way of us knowing the extent of their distortion. A mathematical model is then designed to perform the classification task. It is reasonable to assume that the so called training set based on which the classifier is constructed or "taught" contains high-quality data (not distorted or of high validity). The training set can be, in the case of medical or psychological diagnostics, substituted by diagnostics criteria (derived for a given method during its standardization, by the international classification of diseases or diagnostics manuals) or by the knowledge of an expert diagnostician. This would be represented by the conversion symptoms diagnostics criteria for MMPI-2 and several years of experience with the method in the previous section. The constructed classifier (we do not need to consider a specific method or mathematical tool to build it now) represents the best way we are able to assign classes to objects given the mathematical methods we have chosen and given the data we had at our disposal while designing the model. It can be expected to work well on "usual" (not deliberately distorted) data.

As was already illustrated in the previous section of this chapter - in psychology, it is possible to obtain data about a client, that clearly indicate he/she should be assigned a given diagnosis, although he/she is in fact healthy. This means that given the data obtained from the diagnostics method, the diagnosis suggested by the classifier (= using the diagnostics criteria) is correct. The client is just not well described by the data (he might have lied, omitted ceratin important answers, might tried concealing some symptoms). In this case it is not the classifier (classification rules) that is the source of the possible misclassification. It is the fact that the data and the person described by the data do not correspond well - hence the diagnosis based on the data might not fit the person. If we are faced with data with zero validity - that is with data the quality of which is so low that we can not conclude anything concerning the person who provided them, then no diagnosis should be assigned. But what if the data are distorted only a little bit? In psychology, it is necessary to work even with partially distorted data, as better data may not be available.

In Publication **IV** we propose a classifier performance assessment approach that reflects the quality of data (represented by vr in the MMPI-2 case) in the following way. *Misclassifications of data instances of low quality is considered less serious than misclassifications of data instances of high quality.* In other words it is "OK" for the classifier to assign wrong diagnosis, as long as the data do not correspond with the person who provided them well (and we have a measure of this correspondence). If the data however reflect the important features of the person well, then the diagnosis must be correct to consider the performance of the classifier as good. In Publication **IV** we propose a modification of a classic tool frequently used for binary classifier performance assessment in medical and psychological setting - the receiver operating characteristics (ROC). We also discuss an interesting influence of the answer to the question "what class should be assigned if the quality of data is low" - this is discussed as the "*don't know principle*" in the paper. A numerical study on artificial data as well as an application on the outputs of the classifier summarized in the previous section are provided in the paper as well.

To briefly summarize the key ideas presented in Publication **IV** let us first remind the key concepts of ROC [23, 25, 71]. Let us consider a situation when we need to classify a set of objects $X = \{x_1, \ldots, x_n\}$ each of which is either "*healthy*" or "*not healthy*". This way we can define a set of healthy instances $H \subseteq \{x_1, \ldots, x_n\}$ and a set of not healthy instances $NH \subseteq \{x_1, \ldots, x_n\}$;

obviously H and NH are disjunctive and their union is the set of all instances.

Let us assume that we have a classifier that provides two outputs - *positive* (P - represented by the numerical value 1) and *negative* (N - represented by the value 0). We can in fact consider the classifier to provide as an output any real number from $[0, 1]$. In such a case the classifier is called a *continuous classifier* and a threshold $t \in [0, 1]$ has to be specified in order to achieve a conclusion. Let us now suppose that the continuous classifier we are considering assigns to each instance $x_i \in \{x_1, \ldots, x_n\}$ a value $dg_i \in [0, 1]$. For a given $t \in [0, 1]$ if $dg_i < t$ then N is assigned, otherwise P is assigned). For every possible value of t the following four mutually disjunctive subsets of X can be defined:

TP_t (true positive) - set of instances where *not healthy* were classified as *positive* (*P*),

FP_t (false positive) - set of instances where *healthy* were classified as *positive* (*P*),

FN_t (false negatives) - set of instances where *not healthy* were classified as *negative* (*N*),

TN_t (true negatives) - set of instances where *healthy* were classified as *negative* (*N*).

We can now compute several important characteristics, such as *sensitivity* (TP_rate_t) and *specificity* ($1 - FP\_rate_t$).

$$TP\_rate_t = \frac{Card(TP_t)}{Card(NH)} = \frac{Card(TP_t)}{Card(TP_t \cup FN_t)} \tag{6.9}$$

$$FP\_rate_t = \frac{Card(FP_t)}{Card(H)} = \frac{Card(FP_t)}{Card(TN_t \cup FP_t)} \tag{6.10}$$

$$FN\_rate_t = \frac{Card(FN_t)}{Card(NH)} = \frac{Card(FN_t)}{Card(TP_t \cup FN_t)}, \tag{6.11}$$

$$TN\_rate_t = \frac{Card(TN_t)}{Card(H)} = \frac{Card(TN_t)}{Card(TN_t \cup FP_t)}. \tag{6.12}$$

Here $Card(X)$ denotes the cardinality of a set X, i.e. the number of its elements. We can now plot the TP_rate_t and FP_rate_t in a graph (TP_rate_t is usually depicted on the vertical axis, FP_rate_t on the horizontal axis). Discrete classifiers produce a single point in the ROC space, for continuous classifiers an ROC curve can be plotted $ROC\_curve = \{(x, y) \mid x = FP\_rate_t, y = TP\_rate_t$, for all $t \in [0, 1]\}$. The area under this curve (*AUC*) is used as a classifier performance measure. The closer the *AUC* is to 1, the better the performance of the classifier. Values of *AUC* close to 0.5 represent random classification, classifiers with $AUC < 0.5$ are considered worse than random assignment of classes.

This classifier performance measure can be modified to reflect the quality of data in the following way - see publication Publication **IV** for a detailed discussion. Another modification for fuzzy signals was presented in [71] and is discussed in connection with the classic ROC and the modification we propose to integrate data quality into ROC in Publication **IV**.

If we get back to the psychological diagnostics decision support system (FRB classifier) presented in the previous section of this chapter, we can use the validity rate of a given MMPI-2 protocol vr_i as

a measure of data quality. Let us call $q_i \in [0,1]$ a *quality rate* of a data instance i (we set $vr_i = q_i$ for all $i = 1, \ldots, n$). We can define a fuzzy set of quality data instances as $\widetilde{QDI} = \{{}^{q_1}/_{x_1}, \ldots, {}^{q_n}/_{x_n}\}$. This fuzzy set plays an important role in the proposed modification of the classic ROC analysis. We now need to define the sets of not healthy instances, healthy instances, true positive, false positive, false negative and true negative instances respectively as fuzzy sets (see Publication **IV**). For any value $t \in [0,1]$:

- $\widetilde{NH} = \{{}^{s_1}/_{x_1}, \ldots, {}^{s_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $s_i = 1$ iff x_i is a not healthy instance and $s_i = 0$ otherwise,
- $\widetilde{H} = \{{}^{1-s_1}/_{x_1}, \ldots, {}^{1-s_n}/_{x_n}\}$,
- $\widetilde{TP_t} = \{{}^{\alpha_1}/_{x_1}, \ldots, {}^{\alpha_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $\alpha_i = 1$ iff $[(dg_i \geqq t) \wedge (s_i = 1)]$ and $\alpha_i = 0$ otherwise
- $\widetilde{FP_t} = \{{}^{\beta_1}/_{x_1}, \ldots, {}^{\beta_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $\beta_i = 1$ iff $[(dg_i \geqq t) \wedge (s_i = 0)]$ and $\beta_i = 0$ otherwise
- $\widetilde{FN_t} = \{{}^{\gamma_1}/_{x_1}, \ldots, {}^{\gamma_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $\gamma_i = 1$ iff $[(dg_i < t) \wedge (s_i = 1)]$ and $\gamma_i = 0$ otherwise,
- $\widetilde{TN_t} = \{{}^{\delta_1}/_{x_1}, \ldots, {}^{\delta_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $\delta_i = 1$ iff $[(dg_i < t) \wedge (s_i = 0)]$ and $\delta_i = 0$ otherwise.

It is easy to see that $\alpha_i + \beta_i + \gamma_i + \delta_i = 1$ for all $i = 1, \ldots, n$. Now we can define the modified characteristics of the classifier for any value of threshold $t \in [0,1]$:

$$TP\_rate_t = \frac{Card(\widetilde{TP_t} \cap \widetilde{QDI})}{Card(\widetilde{NH} \cap \widetilde{QDI})} = \frac{Card(\widetilde{TP_t} \cap \widetilde{QDI})}{Card[(\widetilde{TP_t} \cup \widetilde{FN_t}) \cap \widetilde{QDI}]}, \tag{6.13}$$

$$FP\_rate_t = \frac{Card(\widetilde{FP_t} \cap \widetilde{QDI})}{Card(\widetilde{H} \cap \widetilde{QDI})} = \frac{Card(\widetilde{FP_t} \cap \widetilde{QDI})}{Card[(\widetilde{TN_t} \cup \widetilde{FP_t}) \cap \widetilde{QDI}]}, \tag{6.14}$$

$$FN\_rate_t = \frac{Card(\widetilde{FN_t} \cap \widetilde{QDI})}{Card(\widetilde{NH} \cap \widetilde{QDI})} = \frac{Card(\widetilde{FN_t} \cap \widetilde{QDI})}{Card[(\widetilde{TP_t} \cup \widetilde{FN_t}) \cap \widetilde{QDI}]}, \tag{6.15}$$

$$TN\_rate_t = \frac{Card(\widetilde{TN_t} \cap \widetilde{QDI})}{Card(\widetilde{H} \cap \widetilde{QDI})} = \frac{Card(\widetilde{TN_t} \cap \widetilde{QDI})}{Card[(\widetilde{TN_t} \cup \widetilde{FP_t}) \cap \widetilde{QDI}]}. \tag{6.16}$$

The cardinality of the fuzzy sets is computed as the sum of the membership degrees of all the elements to the respective set. The intersection can be modelled using the $\min$ operator and the union of fuzzy sets using the $\max$ operator. The ROC curve can be constructed analogically to the crisp case. Again we can use the area under the ROC curve (AUC_M, the subscript M identifies that the AUC is computed for the ROC curve constructed using our modification of the method) as a performance measure of the classifier. The modified ROC introduced in Publication **IV** is a

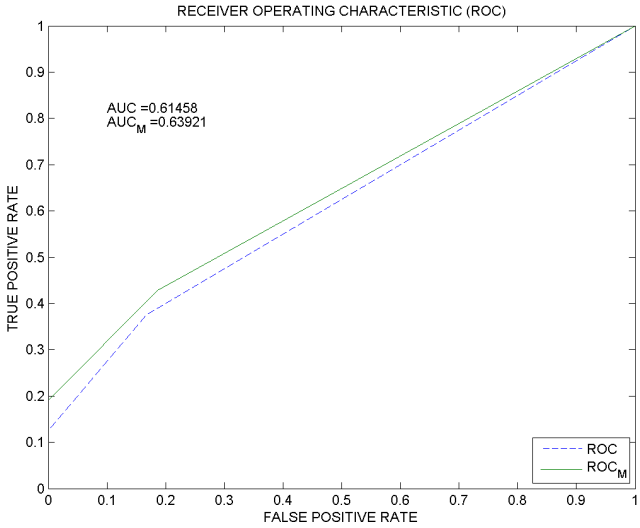


Figure 6.8: MMPI-2 diagnostics: ROC curve and *AUC* for the classic ROC analysis compared to the results (ROC curve and AUC_M) computed using the modified ROC analysis suggested in this paper on real-life data. Reproduced from Publication **IV**.

generalization of the classic ROC analysis. These two approaches coincide for $q_1 = q_2 = \cdots = q_n$.

This way we have introduced the quality of data into ROC analysis. High-quality data now influence the classifier performance measure *AUC* more than low-quality data. To see how the proposed approach performs on data a simulation study of its behavior has been conducted in Publication **IV**. Also the behavior of the classifier presented in previous section of this chapter has been explored in the publication. Let us therefore only briefly present the comparison of the modified ROC with the classic ROC on the assessment of the classifier conversion symptoms detection based on MMPI-2.

Table 6.3 (reproduced from Publication **IV**) summarizes the outputs of the classifier dg_i, the actual diagnosis (confirmed by additional diagnostics techniques) and the validity rate of each MMPI-2 protocol vr_i that is considered to represent the quality of the data q_i, $i = 1, \ldots, n$. The results of applying classic and our modified version of ROC are presented in Figure 6.8. We can see, that AUC_M is slightly larger than classic AUC. This is caused by the fact, that the misclassified protocols PR_{13} and PR_{19} have low validity rate (and hence low quality of data) - and thus the misclassification based on the available information is not seen as a serious mistake (notice that $q_{13} = 0$) in these two cases.

Remark: Publication **IV** also discusses a "*don't know principle*" we have formulated in connection with the design of expert knowledge based classifiers for practice. It can be formulated as a rule of thumb that suggests, that a rule concerning the diagnosis (assigned class) for data with low quality (validity) should be present in the classifier (in our case from previous section in the rule base). The main idea is that if we have an instance of data of really low quality (the meaning of "very low" has to be specified), it is better to assign an uncertain output by the classifier (say 0.5 if the possible outputs are in $[0, 1]$) regardless of the information characterising the data instance, than to let the classifier assign a class (output either 1 or 0). It can be seen as a sort of data preprocessing procedure implemented into the classification process. The usual approach in classification is to discard data

instances of low quality. If these can not be discarded, it seems reasonable to use the *don't know principle* instead.

Publication **IV** provides a detailed analysis of the method and its functioning on artificial data. The main advantage of the proposed modification is the ability of a classifier performance measure to reflect the quality of data. From the linguistic modelling point of view, the mathematical model (e.g. classifier introduced for the MMPI-2 example) is a direct reflection of the expert knowledge provided by a skilled diagnostician. We can now see the modified AUC_M not only as a classifier performance measure, but as *a measure of the diagnostic abilities of the diagnostician*. If we need to evaluate his/her performance (his correct interpretation of the outputs of MMPI-2 in the context of conversion symptoms presence identification), we are in principle more interested in his/her success rate on the data that are not distorted and as such tell us something about the person who provided it. Wrong classification based on data instances with low quality is understandable in this case - the misclassification is a result of the fact that although the data suggest a specific diagnosis, the person who provided them might not be well described by the data in the first place. The modified ROC might be a first step to test the quality of expert knowledge. If an expert intuitively found a pattern in the data that identifies a given diagnosis better than current means, we can now test this new diagnostics lead and see, if it improves the success rate of the diagnostics process or not. And if it does, we might propose to use it in a larger scale. Linguistic modelling allows us here to test such intuitive approaches to data interpretation (as long as we are able to model them formally by e.g. bases of linguistic fuzzy rules) on large data sets. This may prove useful in updating or improving diagnostics strategies in psychology or medicine, or to eliminate less successful approaches to data interpretation.

CHAPTER VII

Discussion and future prospects

The thesis aimed to answer two research questions posed at its very beginning: "*What needs to be done and ensured to provide decision makers with mathematical models they can safely use to facilitate their decisions, without compromising the authority of the decision maker in the process of reaching decisions? And how to ensure that the responsibility for the consequences of the decisions rests mainly with the decision maker and the decision making process is entirely under his/her control?*" In Chapter 2 we have outlined an approach to linguistic modelling that requires the linguistic and mathematical level of the models to remain mutually connected throughout the whole process of building the mathematical model. This is especially needed when linguistic outputs are required. Although there are many approaches to the retranslation of mathematical outputs of a model into the natural language, many of these can result in the loss of information or in a too complicated linguistic description. The potential of linguistic modelling in decision support is in our opinion fully utilized when the outputs of the model provided in the linguistic level are self-explanatory and intuitive and contain all the information necessary for qualified decision (note that this does not mean that the decision support models should provide decisions as such).

The first research question is answered by the guidelines for linguistic decision support models outlined in Chapter 2. In a very condensed form, the answer can be summarized here as *"The decision maker has to be taken into account in the process of designing mathematical model - his/her needs, understanding of the outputs we are able to provide, vocabulary, meanings of linguistic terms. In general the model should be intuitive for the decision maker."* This usually means that linguistic or graphical outputs are provided, if the tools of linguistic modelling are used, the model has to be customized (adjusted) for the decision maker and the decision maker should be able to identify "what went wrong" if the results he/she obtains are odd or counterintuitive. To answer the second research question means to turn our attention to the modeling itself and to the person building the model. The more we (model builders) use representations of outputs that are convenient for us, but possibly confusing or not well suited for the user of the model, the larger is the possibility of misinterpretation. In the opinion of the author, unless we do everything that is possible to prevent misinterpretations, we are responsible (at least partially) for their consequences. There is no "I told him not to use it that way" excuse for us, if there was a better (but perhaps less elegant or more demanding) way of presenting the results - better in the sense that it would be easier for the user to comprehend and understand. This of course means that *we should keep looking for more appropriate ways of presenting mathematical result to laymen* - ways that would retain all the

information, but that would be able to convey it to people without mathematical background. Or we (that is all those who design mathematical models) can also openly accept partial responsibility for the possible consequences of the use of our models.

Although our guidelines for designing linguistic models came first in the thesis, they are based on the author's experience (and the experience of those more experienced around him) for the last five years. Publications **I** to **XII**, the practical applications and theoretical results presented in them were crucial in formulating the author's attitude to linguistic modelling presented in this thesis. We hope that the guidelines and the outputs of the discussion on problematic and also ethical issues of linguistic modelling throughout the thesis can be considered an answer to the two research questions. Hence we consider the objective *"to propose conditions or guidelines for the design of linguistic decision support models that allow the decision maker to remain responsible and in control in the decision making process (that is providing support for qualified decision making)"* achieved.

The need for well understood outputs has been showcased and discussed by the EMRS decision support system (Chapter 3) presented also in Publication **I**. Disaster management decision support systems (when a disaster actually occurs) leave no space for errors or misinterpretations - each delay and each mistake can mean losses of lives. Linguistic fuzzy modelling has been able to provide applicable results (see also Publication **I**) that are comprehensible to EMRS operators. In the context of HR management (staff evaluation - Chapter 5), where the understandability of the evaluation process as well as of its results is of crucial importance (see Publications **II**, **VIII** and **IX**), misinterpretations do not mean losses of lives, but the consequences of e.g. loosing a job based on a possible flaw in the evaluation methodology or a misunderstanding of the outputs of the evaluation tool are still significant. We have not only proposed a linguistic fuzzy rule based evaluation methodology with a linguistic description of the outputs as well as of the evaluation mechanism, but also a way of resolving the issue of responsibility. The results are provided in such a form that discourages their "direct use" - that is decisions as such are not provided - *decisions need to be made based on the outputs of the model*. This way the evaluator has to be active in the evaluation process and accept the responsibility for his/her final decision. He/she knows how the aggregation works, how partial evaluations are constructed, all the evaluation data are available on all levels of aggregation - a qualified decision can therefore be made. But to reach it, sometimes more information may be required.

We have also set the objective *"to discuss possible ways of outputs representation that provide the decision maker with as much information (not necessarily precise) on the outputs of the models as possible (and to develop new approaches to outputs representation in practical applications)"*. The discussion on possible and desirable ways of presenting outputs of linguistic decision support models is presented in Section 2.4. Part II discusses 5 practical applications of the linguistic modeling to decision support. The mathematical models in all these applications naturally deal with the issue of presenting results. Each of the solutions is customized to meet the specific requirements of the respective problem. Hence e.g. for EMRS decision support, a linguistic label roughly describing the number of reinforcements is provided (see Chapter 3), while for the purposes of academic faculty evaluation, color bars and linguistic descriptions of performance are provided. The design of the models follows the guidelines set in Chapter 2 and showcases various possibilities of presenting outputs of the models to decision makers so that the process is as natural as possible for the decision makers. In our opinion this objective has been achieved. On the other hand it is obvious that research in this area will have to continue to keep up with the demands of practice.

Chapter 2 also comprises the basic ideas of ordinal decision making, fuzzy linguistic modelling and

computing with words and perceptions, thus satisfying the objective *"to briefly summarize the state of the linguistic modeling for decision support in mathematics, with particular focus on the area of systems where human factor is involved"* and *"to suggest a unifying general view on the linguistic models for decision support, their design and connected issues including the ethical ones"*.

We have also set out *"to contribute to the mathematical theory of linguistic (fuzzy) modelling for decision support and to suggest a unifying general view on the linguistic models for decision support, their design and connected issues including the ethical ones"*. A significant part of the thesis (Part II) deals with practical applications of linguistic decision support models. We have provided several instances of decision support models in various areas ranging from medical rescue services decision support through HR management and staff evaluation, evaluation of arts and R&D outcomes to applications in humanities including psychological diagnostics. Each of these areas of practice has enabled interesting insights into the mathematical theory of linguistic fuzzy modelling and new concepts and methods have been developed to meet the requirements on linguistic decision support models in these fields. To summarize:

- The α*-degree upper bound of a fuzzy number* (Chapter 3) has been proposed to enable comparisons of fuzzy and crisp numbers and to provide a means of expressing the tolerance of the decision maker to violations of conditions.
- The concept of *weak consistency* in Saaty's AHP and the overall modification of AHP for its use on large matrices of preference intensities were proposed in Chapter 4.
- We have discussed the importance of avoiding discrepancies in the linguistic and mathematical level of a linguistic model in the context of Saaty's AHP in Chapter 4 - the linguistic level of the fundamental scale has been analyzed, some of its apparent weaknesses have been pointed out and possible methods of improving the quality and functioning of the linguistic level of the AHP (also using the weak consistency condition) have been proposed (see Publications **III**, **VII** and **XI**). This enabled the evaluation methodology and the underlying mathematical model for creative work outcomes of Czech art colleges and faculties to be created. The linguistic modelling approach to AHP made the evaluation of works of art possible and its outputs widely acceptable.
- A *fuzzy inference mechanism providing easily interpretable outputs for HR management* has been summarized in Chapter 5. The inference mechanism is designed to provide outputs that can be easily interpreted using a linguistic scale or converted into colour representation.
- A *modification of the ROC analysis* to be able *to reflect the quality of data in the process of classifier performance assessment* has been proposed in Chapter 6.

In our opinion, we can say that the goal of contributing to the theory of linguistic (fuzzy) modelling for decision support has been achieved. What remains to reflect is the objective *"to demonstrate the usability of linguistic modelling in real-life applications and decision making situations by presenting several working applications of linguistic models in various areas - ranging from the evaluation of works of art through disaster management to psychological diagnostics"*. We consider this objective achieved - this claim is backed up by Chapters 3 to 6 and Publications **I** to **XII**.

Several of the presented models that have been developed with a significant contribution of the author of this thesis have already reached a national impact - the Registry of Artistic Performances

(RUV) has been included in the national methodology for funding universities in the Czech Republic, the information system for academic faculty evaluation (IS HAP) is being implemented on several universities in the Czech Republic. We have also presented an R&D evaluation methodology for scientific monographs, which has been used at the Faculty of Science, Palacký University, Olomouc in 2013.

It is the opinion of the author that the concept of linguistic modelling with a strong accent on the connection between its two crucial levels outlined in the first part of the chapter on linguistic fuzzy modelling has proven successful in practical applications. Several issues remain opened in the area of proper representation of meaning, of retranslation procedures and providing intuitive and easy to understand results to practitioners - and their necessary mathematical basis. These will be in the scope of the author in the future. Linguistic mathematical models strive to represent the meaning of expressions from the natural language - yet much of the research remains purely in the field of mathematics, computer science and partially also linguistics. In our opinion an interdisciplinary cooperation with experts and professionals from the field of humanities is also necessary. Making "soft" mathematics without the cooperation with human sciences seems not only illogical, but it also seems to be missing the opportunity of achieving interesting synergical effects.

The issue of ethics in mathematical modeling, responsibility and other important matters brought back into focus by the behavioral operations research seem worth exploring, clarifying and pursuing in the future. We are convinced that there is need for linguistic fuzzy modelling, we have identified a large demand for linguistic fuzzy modelling and we are sure there is much potential in this approach to mathematical modelling. It is our hope that this field of applied mathematics will continue its effort in describing the world around us in a way that is comprehensible - thus providing tools for representing expert knowledge and possibly also for transferring knowledge to others. There are many possibilities for further development of this branch of mathematics, and the author sincerely hopes to have the opportunity to participate in this effort in the future.

ACKNOWLEDGEMENTS

The research presented in this thesis was supported by the following grants and projects:

- Grants PrF_2010_08, PrF_2011_022, PrF_2012_017 and PrF_2013_013 of the Internal Grant Agency of Palacký University in Olomouc,
- Grant $GA\ 14-02424S$ of the Grant Agency of the Czech Republic,
- Centralized Developmental Project C41 entitled *Evaluating Creative Work Outcomes Pilot Project* (financed by the Czech Ministry of Education),
- Educational policy fund - indicator F (Registry of Artistic Performances - RUV) - financed from the state budget of the Czech Republic
- and indirectly also by the Research Foundation of Lappeenranta University of Technology.

I would like to express my thanks and gratitude for all the support.

[1] B. Arfi. *Linguistic fuzzy logic methods in social sciences*. Springer-Verlag, Berlin Heidelberg, 2010.

[2] R. E. Bellman and L. A. Zadeh. Decision-making in a fuzzy environment. *Management science*, 17(4):141–164, 1970.

[3] U. Bodenhofer. A general framework for ordering fuzzy sets. In B. Bouchon-Meunier, J. Gutiérrez-Ríos, M. Luis, and R. R. Yager, editors, *Technologies for Constructing Intelligent Systems 1*, chapter A general, pages 213–224. Physica-Verlag HD, 2002.

[4] U. Bodenhofer. Orderings of fuzzy sets based on fuzzy orderings. part I: the basic approach. *Mathware & soft computing*, 15:201–218, 2008.

[5] U. Bodenhofer and P. Bauer. Interpretability of linguistic variables: a formal account. *Kybernetika*, 41(2):227–248, 2005.

[6] M. Brunelli, L. Canal, and M. Fedrizzi. Inconsistency indices for pairwise comparison matrices: a numerical study. *Annals of Operations Research*, 211(1):493–509, February 2013.

[7] M. Brunelli and J. Mezei. How different are ranking methods for fuzzy numbers? A numerical study. *International Journal of Approximate Reasoning*, 54(5):627–639, 2013.

[8] G. Canfora and L. Troiano. Fuzzy ordering of fuzzy numbers. *Proceedings of the IEEE International Conference of Fuzzy Systems*, 1(2):669–674, 2004.

[9] C. Carlsson and R. Fullér. On fuzzy screening systems. In *Proceedings of EUFIT 95 Conference*, pages 1–6, 1995.

[10] C. Carlsson and R. Fullér. Benchmarking in linguistic importance weighted aggregations. *Fuzzy Sets and Systems*, 114(1):35–41, 2000.

[11] C. Carlsson and R. Fullér. *Fuzzy Reasoning in Decision Making and Optimization*, volume 82 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag HD, Heidelberg, 2002.

[12] C. Carlsson, R. Fullér, and J. Mezei. An Approximate Reasoning Approach to Rank the Results of Fuzzy Queries. In *Proceedings of the 2nd International Conference on Applied Operational Research*, pages 382–387, 2010.

[13] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, editors. *Accuracy improvements in linguistic fuzzy modeling*. Springer-Verlag, Berlin Heidelberg GmbH, 2003.

[14] G. Chen and T. T. Pham. *Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems*. Springer-Verlag, Boca Raton, London, New York, Washington, 2001.

[15] J. de Boer. *Order in chaos: modelling medical disaster management*. Free University Hospital, Netherland, Amsterdam, 1999.

[16] R. Degani and G. Bortolan. The problem of linguistic approximation in clinical decision making. *International Journal of Approximate Reasoning*, 2(2):143–162, April 1988.

[17] M. Delgado, F. Herrera, E. Herrera-Viedma, and L. Martinez. Combining numerical and linguistic information in group decision making. *Information Sciences*, 107:177–194, 1998.

[18] M. Delgado, M. D. Ruiz, D. Sánchez, and M. A. Vila. Fuzzy quantification: a state of the art. *Fuzzy Sets and Systems*, 242(1):1–30, May 2014.

[19] M. Delgado, J. L. Verdegay, and M. A. Vila. On aggregation operations of linguistic labels. *International Journal of Intelligent Systems*, 8(3):351–370, 1993.

[20] D. Dubois and H. Prade. *Fuzzy sets and systems: theory and applications*. Number Nf. Academic Press, New York, 1980.

[21] D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy sets and systems*, 84:169–185, 1996.

[22] D. Dubois and H. Prade, editors. *Fundamentals of Fuzzy Sets*. Kluwer Academic Publishers, Massachusetts, 2000.

[23] J. P. Egan. *Signal Detection Theory and ROC analysis*. Academic, New York, 1975.

[24] M. Enea and T. Piazza. Project Selection by Constrained Fuzzy AHP. *Fuzzy Optimization and Decision Making*, 3(1):39–62, March 2004.

[25] T. Fawcett. ROC graphs: Notes and practical considerations for researchers. Technical report, HP Labs, HPL-2003-4, 2004.

[26] M. Fedrizzi. Distance-Based Characterization of Inconsistency in Pairwise Comparisons. In S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager, editors, *Advances in Computational Intelligence*, volume 300, pages 30–36. Springer, Berlin Heidelberg, 2012.

[27] M. Fedrizzi and S. Giove. Optimal sequencing in incomplete pairwise comparisons for large-dimensional problems. *International Journal of General Systems*, 42(4):366–375, 2013.

[28] J. Fodor. Left-continuous t-norms in fuzzy logic: An overview. *Acta Polytechnica Hungarica*, 1(2):35–47, 2004.

[29] C. Franco, J. T. Rodríguez, and J. Montero. An ordinal approach to computing with words and the preference-aversion model. *Information Sciences*, 258:239–248, February 2014.

[30] D. M. Green and J. A. Swets. *Signal detection theory and psychophysics*. Wiley, New York, 1966.

[31] R. L. Greene. *The MMPI-2: An Interpretive Manual*. Allyn and Bacon, Needham Heights, 2000.

[32] R. P. Hämäläinen, J. Luoma, and E. Saarinen. On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research*, 228(3):623–634, August 2013.

[33] D. J. Hand. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, 77(1):103–123, June 2009.

[34] T. Hasuike, H. Katagiri, H. Tsubaki, and H. Tsuda. Constructing membership function based on fuzzy shannon entropy and human's interval estimation. In *Proceedings of IEEE International Conference on Fuzzy Systems (WCCI 2012)*, number 22700233, pages 1–6, 2012.

[35] S. R. Hathaway and J. C. Mckinley. A Multiphasic Personality Schedule (Minnesota) : I. Construction of the Schedule. *The Journal of Psychology*, 10(2):249–254, October 1940.

[36] F. Herrera, S. Alonso, F. Chiclana, and E. Herrera-Viedma. Computing with words in decision making: foundations, trends and prospects. *Fuzzy Optimization and Decision Making*, 8(4):337–364, September 2009.

[37] F. Herrera and E. Herrera-Viedma. Aggregation operators for linguistic weighted information. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 27(5):646–656, 1997.

[38] F. Herrera and E. Herrera-Viedma. Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, 115(1):67–82, October 2000.

[39] F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 8(6):746–752, 2000.

[40] P. Holeček, J. Talašová, and J. Stoklasa. Fuzzy classification systems and their applications. In *Proceedings of the 29th International Conference on Mathematical Methods in Economics - part I*, pages 266–271. University of Economics, Prague, Faculty of Informatics and Statistics, 2011.

[41] D. H. Hong. Note on the expected value of a function of a fuzzy variable. *Journal of applied mathematics and informatics*, 27(3-4):773–778, 2009.

[42] V. N. Huynh, Y. Nakamori, T. B. Ho, and G. Resconi. A Context Model for Constructing Membership Functions of Fuzzy Concepts Based on Modal Logic. In T. Eiter and K.-D. Schewe, editors, *Foundations of Information and Knowledge Systems*, pages 93–104. 2002.

[43] J. Jantzen. *Foundations of Fuzzy Control*. John Wiley & Sons, Chichester, UK, August 2007.

[44] J. Kacprzyk and S. Zadrozny. Linguistically quantified propositions for consensus reaching support. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 1135–1140, 2004.

[45] J. Kacprzyk and S. Zadrozny. Computing with words is an implementable paradigm: fuzzy queries, linguistic data summaries, and natural-language generation. *Fuzzy Systems, IEEE Transactions on Fuzzy Systems*, 18(3):461–472, 2010.

[46] J. Kacprzyk and S. Zadrozny. Linguistic Data Summarization: A High Scalability through the Use of Natural Language? In A. Laurent and M.-J. Lesot, editors, *Scalable Fuzzy Algorithms for Data Management and Analysis: Methods and Design*, pages 214–237. IGI Global, 2010.

[47] J. Kacprzyk and S. Zadrozny. Comprehensiveness and interpretability of linguistic data summaries: A natural language focused perspective. In *2013 IEEE Symposium on Computational Intelligence for Human-like Intelligence (CIHLI)*, pages 33–40. Ieee, April 2013.

[48] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Springer Science+Business Media, Dordrecht, 2000.

[49] G. J. Klir. Fuzzy arithmetic with requisite constraints. *Fuzzy Sets and Systems*, 91(2):165–175, 1997.

[50] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Aplications*. Prentice Hall, New Jersey, 1995.

[51] J. Kluska. *Analytical methods in fuzzy modeling and control*. Springer-Verlag, Berlin Heidelberg, 2009.

[52] Z. Kovacic and S. Bogdan. *Fuzzy controller design: theory and applications*. CRC Press, Taylor & Francis group, Boca Raton, 2006.

[53] J. Krejčí and J. Talašová. A proper fuzzification of Saaty's scale and an improved method for computing fuzzy weights in fuzzy AHP. In *Proceedings of the 31st International Conference on Mathematical Methods in Economics - part II*, pages 452–457. College of Polytechnics Jihlava, 2013.

[54] L. I. Kuncheva. *Fuzzy Classifier Design*. Springer Physica Verlag, Heidelberg, New York, 2000.

[55] G. Lakoff. Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2(4):458–508, 1973.

[56] W. V. Leekwijck and E. E. Kerre. Defuzzification: criteria and classification. 108(2):159–178, 1999.

[57] E. H. Mamdani. Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis. *IEEE Transactions on Computers*, C-26(12):1182–1191, December 1977.

[58] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Human-Computer Studies*, 51(2):135–147, 1999.

[59] J. G. Marín-Blázquez, G. M. Pérez, and M. G. Pérez. A linguistic fuzzy-XCS classifier system. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 1–6, 2007.

[60] L. Martínez and F. Herrera. An overview on the 2-tuple linguistic model for computing with words in decision making: Extensions, applications and challenges. *Information Sciences*, 207(1):1–18, November 2012.

[61] L. Martínez and F. Herrera. Challenges of computing with words in decision making. *Information Sciences*, 258:218–219, February 2014.

[62] S. Massanet, J. V. Riera, J. Torrens, and E. Herrera-Viedma. A new linguistic computational model based on discrete fuzzy numbers for computing with words. *Information Sciences*, 258:277–290, February 2014.

[63] J. M. Mendel. Computing with words and its relationships with fuzzistics. *Information Sciences*, 177(4):988–1006, February 2007.

[64] J. Mezei and R. Wikström. Aggregation operators and interval-valued fuzzy numbers in decision making. In Á. Rocha, A. M. Correia, T. Wilson, and K. A. Stroetmann, editors, *Advances in Information Systems and Technologies*, volume 206 of *Advances in Intelligent Systems and Computing*, pages 535–544. Springer-Verlag, Berlin, Heidelberg, 2013.

[65] V. Novák. *Fuzzy množiny a jejich aplikace*. SNTL, Praha, 1986.

[66] V. Novák, I. Perfilieva, and J. MočkoÅ. *Mathematical Principles of Fuzzy Logic*. Kluwer Academic Publishers, Massachusetts, 1999.

[67] T. Nuguyen, N. R. Prasad, C. L. Walker, and E. A. Walker. *A first course in fuzzy and neural control*. Chapman & Hall/CRC, Boca Raton, 2003.

[68] C. E. Osgood. *The Measurement of Meaning*. University of Illinois Press, Illinois, 1957.

[69] S. Ovchinnikov. Similarity relations, fuzzy partitions, and fuzzy orderings. *Fuzzy Sets and Systems*, 40(1):107–126, March 1991.

[70] S. V. Ovchinnikov. General negations in fuzzy set theory. *Journal of Mathematical Analysis and Applications*, 92(1):234–239, 1983.

[71] R. Parasuraman, A. J. Masalonis, and P. Hancock. Fuzzy Signal Detection Theory: Basic Postulates and Formulas for Analyzing Human and Machine Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(4):636–659, December 2000.

[72] C. Parker. On measuring the performance of binary classifiers. *Knowledge and Information Systems*, 35(1):131–152, September 2013.

[73] W. Pedrycz and S. H. Rubin. Data compactification and computing with words. *Engineering Applications of Artificial Intelligence*, 23(3):346–356, April 2010.

[74] J. Ramík. Consistency of Pair-wise Comparison Matrix with Fuzzy Elements. In *IFSA/EUSFLAT Conf.*, pages 98–101, 2009.

[75] J. Ramík. Measuring transitivity of fuzzy pairwise comparison matrix. In *Proceedings of the 30th International Conference Mathematical Methods in Economics*, pages 751–756, 2012.

[76] J. Ramík and M. Vlach. *Generalized Concavity in Fuzzy Optimization and Decision Analysis*, volume 41 of *International Series in Operations Research & Management Science*. Springer US, Boston, MA, 2002.

[77] J. Ramík and M. Vlach. Measuring consistency and inconsistency of pair comparison systems. *Kybernetika*, 49(3):465–486, 2013.

[78] E. H. Ruspini. A new approach to clustering. *Information and control*, 15(1):22–32, 1969.

[79] T. L. Saaty. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281, June 1977.

[80] T. L. Saaty. *Fundamentals of Decision Making and Priority Theory With the Analytic Hierarchy Process*. RWS Publishers, 2000.

[81] T. L. Saaty. Deriving the AHP 1-9 scale from first principles. In *ISAHP 2001 proceedings, Bern, Switzerland*, pages 397–402, 2001.

[82] T. L. Saaty. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1):83–98, 2008.

[83] T. L. Saaty. Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 102(2):251–318, September 2008.

[84] T. L. Saaty and L. T. Tran. On the invalidity of fuzzifying numerical judgments in the Analytic Hierarchy Process. *Mathematical and Computer Modelling*, 46(7-8):962–975, October 2007.

[85] T. L. Saaty and L. G. Vargas. *Decision Making with the Analytic Network Process: Economic, Political, Social and Technological Applications with Benefits, Opportunities, Costs and Risks*. Springer, New York, 2006.

[86] S. Safarzadegan Gilan, M. H. Sebt, and V. Shahhosseini. Computing with words for hierarchical competency based selection of personnel in construction companies. *Applied Soft Computing*, 12(2):860–871, February 2012.

[87] A. Sancho-Royo and J. L. Verdegay. Methods for the Construction of Membership Functions. *International Journal of Intelligent Systems*, 14(12):1213–1230, December 1999.

[88] S. Schockaert. Construction of membership functions for fuzzy time periods. In *Proc. ESSLLI 2005 Student Session*, volume 281, pages 305–317, 2005.

[89] M. Smithson and J. Verkuilen. *Fuzzy set theory: applications in the social sciences*. Sage Publications, Thousand Oaks, London, New Delhi, 2006.

[90] J. Stoklasa. Aplikace teorie chaosu a její použití pri rešení krizových situací, unpublished Bachelor thesis, Silesian University in Opava, 2007.

[91] J. Stoklasa. Klasické a fuzzy modely hodnocení efektivnosti, unpublished Masters thesis, Palacký University in Olomouc, 2009.

[92] J. Stoklasa, P. Holeček, and J. Talašová. A holistic approach to academic staff performance evaluation - a way to the fuzzy logic based evaluation. In *Peer Reviewed Full Papers of the 8th International Conference on Evaluation for Practice: "Evaluation as a Tool for Research, Learning and Making Things Better"*, number June, pages 121–131. University of Tampere, School of Humanities and Sciences, Unit at University Consortium of Pori, 2012.

[93] M. Sugeno and K. Murakami. Fuzzy parking control of model car. In *Proceedings of the 23rd conference on Decision and Control, Las Vegas*, pages 902–904, 1984.

[94] M. Sugeno and T. Yasukawa. A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, 1(1):7–31, February 1993.

[95] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15(1):116–132, 1985.

[96] J. Talašová. *Fuzzy metody vícekriteriálního hodnocení a rozhodování*. Palacký University in Olomouc, Olomouc, 2003.

[97] J. Talašová and P. Holeček. Multiple-Criteria Fuzzy Evaluation : The FuzzME Software Package. In *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference*, pages 681–686, 2009.

[98] J. Talašová and O. Pavlačka. Academic Staff Evaluation Model Design for the Faculty of Science (in Czech) - research report. Technical report, Palacký University in Olomouc, Olomouc, 2006.

[99] R. M. Tong and P. P. Bonissone. A linguistic approach to decisionmaking with fuzzy sets. *IEEE Transactions on Systems, Man, and Cybernetics*, 10(11):716–723, 1980.

[100] E. Trillas. On the use of words and fuzzy sets. *Information Sciences*, 176(11):1463–1487, June 2006.

[101] I. B. Türksen. Type 2 representation and reasoning for CWW. *Fuzzy Sets and Systems*, 127(1):17–36, 2002.

[102] S. Valliappan and T. D. Pham. Constructing the membership function of a fuzzy set with objective and subjective information. *Computer-Aided Civil and Infrastructure Engineering*, 8(1):75–82, 1993.

[103] W. Voxman. Canonical representations of discrete fuzzy numbers. *Fuzzy Sets and Systems*, 118(3):457–466, March 2001.

[104] F. Wenstø p. Quantitative analysis with linguistic values. *Fuzzy Sets and Systems*, 4(2):99–115, 1980.

[105] D. Wu, J. M. Mendel, and S. Coupland. Enhanced Interval Approach for Encoding Words Into Interval Type-2 Fuzzy Sets and Its Convergence Analysis. *IEEE Transactions on Fuzzy Systems*, 20(3):499–513, 2012.

[106] Z. Xu. A method based on linguistic aggregation operators for group decision making with linguistic preference relations. *Information Sciences*, 166(1):19–30, October 2004.

[107] R. R. Yager. A new methodology for ordinal multiobjective decisions based on fuzzy sets. *Decision Sciences*, 12(4):589–600, 1981.

[108] R. R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision-making. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190, 1988.

[109] R. R. Yager. Non-numeric multi-criteria multi-person decision making. *Group Decision and Negotiation*, 2(1):81–93, March 1993.

[110] R. R. Yager. An approach to ordinal decision making. *International Journal of Approximate Reasoning*, 12(3-4):237–261, 1995.

[111] R. R. Yager. Knowledge-based defuzzification. *Fuzzy Sets and Systems*, 80(2):177–185, 1996.

[112] R. R. Yager. On the retranslation process in Zadeh's paradigm of computing with words. *IEEE transactions on systems, man, and cybernetics. Part B: Cybernetics*, 34(2):1184–1195, April 2004.

[113] R. R. Yager and J. Kacprzyk, editors. *The Ordered Weighted Averaging Operators: Theory and Applications*. Springer Science+Business Media, New York, 1997.

[114] L. A. Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[115] L. A. Zadeh. Quantitative fuzzy semantics. *Information sciences*, 3(2):159–176, 1971.

[116] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information sciences*, 3(2):177–200, 1971.

[117] L. A. Zadeh. A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges. *Journal of Cybernetics*, 2(3):4–34, January 1972.

[118] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man and Cybernetics*, 3(1):28–44, 1973.

[119] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, 8(3):199–249, January 1975.

[120] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-II. *Information sciences*, 8(4):301–357, 1975.

[121] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-III. *Information sciences*, 9(1):43–80, 1975.

[122] L. A. Zadeh. The linguistic approach and its application to decision analysis. In Y. C. Ho and S. K. Mitter, editors, *Directions in large-scale systems*, pages 339–361. 1976.

[123] L. A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computers & Mathematics with Applications*, 9(1):149–184, 1983.

[124] L. A. Zadeh. Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, 4(2):103–111, 1996.

[125] L. A. Zadeh. From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications*, 45(1):105–119, April 1999.

[126] L. A. Zadeh. A Note on Z-numbers. *Information Sciences*, 181(14):2923–2932, 2011.

[127] L. A. Zadeh. *Computing with words: Principal concepts and ideas*. Springer, Heidelberg New York Dordrecht London, 2012.

[128] L. A. Zadeh and J. Kacprzyk, editors. *Computing with words in Information/Intelligent systems 1: Foundations*. Springer-Verlag, Berlin Heidelberg, 1999.

[129] L. A. Zadeh and J. Kacprzyk, editors. *Computing with words in information/intelligent systems 2: Applications*. Springer-Verlag, Berlin Heidelberg GmbH, 1999.

[130] A.-X. Zhu, L. Yang, B. Li, C. Qin, T. Pei, and B. Liu. Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma*, 155(3-4):164–174, March 2010.

[131] H.-J. Zimmermann. *Fuzzy Set Theory and Its Applications*. Kluwer Academic Publishers, Boston/Dordrecht/London, 2001.

PART III: PUBLICATIONS

PUBL. I

Stoklasa, J., *A Fuzzy Approach to Disaster Modeling: Decision Making Support and Disaster Management Tool for Emergency Medical Rescue Services*. IN Mago, V. K. and Bhatia, N. (eds.) Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies, IGI Global, 2012. DOI: 10.4018/978-1-61350-429-1.ch028

Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition:

Advancing Technologies

Vijay Kumar Mago
Simon Fraser University, Canada

Nitin Bhatia
DAV College, India

A volume in the Advances in Computational Intelligence and Robotics (ACIR) Book Series

Managing Director: Lindsay Johnston
Senior Editorial Director: Heather Probst
Book Production Manager: Sean Woznicki
Development Manager: Joel Gamon
Development Editor: Mike Killian
Acquisitions Editor: Erika Gallagher
Typesetters: Lisandro Gonzalez
Cover Design: Nick Newcomer, Lisandro Gonzalez

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Cross-disciplinary applications of artificial intelligence and pattern recognition : advancing technologies / Vijay Kumar Mago and Nitin Bhatia, editors.
p. cm.
Summary: “This book provides a common platform for researchers to present theoretical and applied research findings for enhancing and developing intelligent systems, discussing advances in and applications of pattern recognition technologies and artificial intelligence”-- Provided by publisher.
Includes bibliographical references and index.
ISBN 978-1-61350-429-1 (hardcover) -- ISBN 978-1-61350-430-7 (ebook) -- ISBN 978-1-61350-431-4 (print & perpetual access) 1. Pattern recognition systems. 2. Artificial intelligence. I. Mago, V. K. II. Bhatia, Nitin, 1978-
TK7882.P3C66 2012
006.3--dc23
2011046541

This book is published in the IGI Global book series Advances in Computational Intelligence and Robotics (ACIR) Book Series (ISSN: 2327-0411; eISSN: 2327-042X)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 28
A Fuzzy Approach to Disaster Modeling:
Decision Making Support and Disaster Management Tool for Emergency Medical Rescue Services

Jan Stoklasa
Palacky University in Olomouc, Czech Republic

ABSTRACT

The decision making process of the Emergency Medical Rescue Services (EMRS) operations centre during disasters involves a significant amount of uncertainty. Decisions need to be made quickly, and no mistakes are tolerable, particularly in the case of disasters resulting in a large number of injured people. A multiphase linguistic fuzzy model is introduced to assist the operator during the initial phase of the medical disaster response. Based on uncertain input data, estimating the severity of the disaster, the number of injured people, and the amount of forces and resources needed to successfully deal with the situation is possible. The need for reinforcements is also considered. Fuzzy numbers, linguistic variables and fuzzy rule bases are applied to deal with the uncertainty. Outputs provided by the model (severity of the disaster, number of reinforcements needed etc.) are available both as fuzzy sets (for the purposes of disaster planning) and linguistic terms (for emergency call evaluation purposes).

INTRODUCTION

Disaster can be defined as an event threatening human life, health, property or environment, with an unusually extensive impact on the society. Such situations usually require a change of attitude and value system revision to be successfully dealt with.

For the purpose of this chapter only disasters that result in a large (significantly surpassing the usual) number of injured people with prevailing mechanical injuries will be considered. There are many disaster classifications. We can distinguish between man-made disasters (traffic accidents, industrial accidents etc.) and god-made (or natural) disasters, such as earthquakes, tsunamis etc.

DOI: 10.4018/978-1-61350-429-1.ch028

It makes sense to use this approach to disaster classification from the perspective of general disaster response. The case of medical rescue services response to disasters requires a different classification approach. A useful classification should be based on the prevailing type (source) of injuries. Among the typical types on injuries, mechanical injuries are the most frequent. We can also consider chemicals, radiation, biological agents or explosions to be possible sources of injuries. In this chapter, we focus on the mechanical type of injuries and all disasters resulting in this type of injuries.

Should such a life or health threatening event occur in our lives, we count on the Emergency Medical Rescue Services (EMRS) to provide us with assistance. Their forces are trained to be able to cope with almost any every-day health threatening event that might happen. But when a more serious event – disaster – occurs, classical problem solving strategies seize to work. EMRS staff needs to think and act differently, procedures they know and do well no longer suffice (Boer, 1999).

The amount of forces and resources needed to successfully cope with such situations may be difficult to determine or even to estimate. The key role in the rescue process is played by the EMRS operator, who evaluates the emergency call. Disasters occur quickly, suddenly and with an unusual impact on the environment and people. This implies that whatever decisions need to be made can not be postponed, and every mistake can result in damage to property or health or even in casualties. Moreover, people reporting these events to the EMRS Operations center may be affected by the disaster themselves. This can make their evaluation of the severity of the disaster inappropriate (both under and over-estimated).

In order to make the decision as correctly and quickly as possible, every available piece of information needs to be taken into consideration, regardless of its uncertainty. Decision making support in the form of a mathematical model can mean a substantial simplification of EMRS operator's work as well as a means of mistake elimination. And in this context mistakes can mean life losses.

The main purpose of this chapter is to show that linguistic fuzzy modeling can prove itself useful even in the context of medical disaster response, where mainly during the initial time period uncertainty is inevitable and has to be dealt with. The emergency call usually comprises rough information describing the event. The exact location, number of casualties, severity of injuries etc. is not available. We usually deal with guesses of the person that is reporting the disaster. This is however the only piece of information concerning the disaster itself and its impact (Stoklasa, 2009) that is available during the first minutes of disaster response. Decisions need to be made even in situations when there is lack of information, or the precision of data is low. We may even need to deal with contradictory information. Any tool that can help us use this kind of data effectively, to verify it somehow and to draw valid conclusions is most welcome. We need to speed up the decision making process and eliminate possible mistakes when lives are at stake. Fuzzy logic and linguistic fuzzy modeling may provide such a tool.

The stress under which the operators of EMRS operation centers work reaches critical levels during disasters. It is surprising that we still do not have available a sufficiently well working system for these purposes (at least in the Czech Republic) – not even now in the 21st century. The need for such a decision making support tool has been at least recognized during the last few years. This chapter describes how we deal with this challenge. Based on the practical experience of professional EMRS workers and operators, as well as hospital representatives and interviews with them, the linguistic fuzzy model described later in this chapter was developed.

BACKGROUND

The use of operations research (OR) in disaster management is nothing new. According to Altay & Green (2006) there is an increasing need for OR study in disaster management. It is also true to say that some areas of disaster management need yet to be convinced that the use of OR may bring benefits. Wide use of mathematical programming in the context of disaster management is understandable. Statistics are another branch of OR that is widely used in disaster management (according to Altay & Green, 2006). Interesting applications of soft computing methods in connection with disaster management can be also found. Cret et al. (1993) use the fuzzy approach for earthquake damage estimation; seismic hazard analysis is considered by Dong, Chiang & Shah (1987). Applications to flood control are presented by Esogbue, Theologidu & Guo (1992) and some more general recommendations given in Esogbue (1996). It is interesting to see that similar approaches are applied in the field of medicine. Fuzzy rule bases help with medical diagnostics in the field of internal medicine (Rotshtein, 1999) and even to control drug infusion during anesthesia (Abbod & Linkens, 1998). Almost all the fields of medicine seem to profit (at least potentially) from the use of fuzzy logic (see Abbod et al. (2001) for more detailed information on the use of fuzzy methods in medicine). Disaster medicine however seems to resist this trend. We can find applications of fuzzy approach even in the form of a mobile Triage decision support (San Pedro et al., 2004) – a useful tool that helps hospital nurses perform injury severity classification, but the area of EMRS remains almost untouched by OR. This is surprising, as without functional and well working medical rescue services, there is usually nobody to be classified and treated in hospitals during disasters. This chapter describes an application of fuzzy methods on EMRS decision making process and provides a mathematical description of the disaster from the medical rescue services' point of view. We reflect the specific needs of the EMRS (in the Czech Republic) and provide them with a custom made decision support tool.

The chain of medical care can be divided, according to Štětina (2000) and Boer (1999), into four phases. The first phase (omitted from the chain of medical care by some authors) comprises first aid activities performed by laymen directly at the disaster site before the arrival of the EMRS (including the emergency call). The second phase consists of professional first aid administration by medical rescue workers and doctors (still at the disaster site). The third phase can be described as the transportation phase – the main goal is to transport all the injured people from the disaster site to appropriate medical facilities (specialized hospitals etc.) within given time limits. After that the fourth phase follows, comprising all the medical care provided to the injured people by hospital staff. All these four phases are linked and problems in any of them can negatively influence the whole chain of medical care.

In some countries, the third part of the chain of medical care is not performed by the EMRS. As this model was developed mainly for the use by Czech EMRS, where patients need to be both treated and transported to hospitals by the emergency medical rescue services, the second and third phase will play an important role in the design of the model. We will try to determine the appropriate amount of forces and resources needed to enable the emergency medical rescue services to provide medical treatment to all people affected by the disaster and to transport them appropriate medical facilities. However the first phase (mainly the emergency call) and the fourth phase (mainly in terms of hospital treatment capacity (HTC) – the number of patients that can be treated per hour, for more details see Boer (1999)) – need to be reflected in the model.

During the construction of the model we realized, that all four phases of the chain of medical care should be reflected in any decision making support model for the previously described pur-

pose. Only then we have a tool that can provide us with valuable information. In this chapter we present a mathematical model of the EMRS operator decision making process during disasters that result in a large number of mechanical injuries. The model uses linguistic variables, fuzzy rule bases and performs optimization tasks under fuzzy constraints in order to provide outputs useful in the decision making process. In the first phase it determines the capacity of a current EMRS center (number of patients that can be treated by available EMRS teams per hour). Then, based on the emergency call, it estimates the severity of the disaster (number of people affected by the disaster, number of people injured) in the second phase. In the third phase it determines which medical facilities need to be alerted to start preparations to receive patients. Finally the fourth phase provides the number of ambulances (EMRS teams) needed to successfully deal with the disaster and assesses the need for reinforcements. The model has to be able to deal with uncertainty in all these phases. Fuzzy set theory is used to meet this requirement and to allow the use of linguistic labels.

PRELIMINARIES

Fundamentals of the fuzzy set theory (introduced by Zadeh (1965)) are described in detail, e.g., in Dubois & Prade (2000). Let U be a nonempty set (the universe). A fuzzy set A on U is defined by the mapping $A: U \rightarrow [0,1]$. For each $x \in U$ the value $A(x)$ is called the membership degree of the element x in the fuzzy set A and A(•)is called the membership function of the fuzzy set A. The height of a fuzzy set A is the real number $hgt(A) = \sup_{x \in U}\{A(x)\}$. Other important concepts related to fuzzy sets are: (a) the kernel of A, $Ker(A) = \{x \in U \,|A(x) = 1\}$, (b) the support of A, $Supp(A) = \{x \in U \,|A(x) > 0\}$ and c) the α-cut of A, $A_\alpha = \{x \in U \,|A(x) \geq \alpha\}$, for $\alpha \in [0,1]$.

A function T:$[0,1]^2 \rightarrow [0,1]$ is called a triangular norm or t-norm if for all $\alpha, \beta, \gamma, \delta \in [0,1]$ it satisfies the following four properties: (1) commutativity: $T(\alpha,\beta) = T(\beta,\alpha)$, (2) associativity: $T(\alpha,T(\beta,\gamma)) = T(T(\alpha,\beta),\gamma)$, (3) monotonicity: if $\alpha \leq \gamma, \beta \leq, \delta$, then it holds that $T(\alpha,\beta) \leq T(\gamma,\delta)$, and (4) boundary condition: $T(\alpha,1) = \alpha$.

A function S: $[0,1]^2 \rightarrow [0,1]$ is called a triangular conorm or t-conorm if for all $\alpha, \beta, \gamma, \delta, \in [0,1]$ it satisfies the properties (1)–(3) from the previous definition and (4) the boundary condition: $S(\alpha,0) = \alpha$.

A function N: [0,1,]→[0,1] satisfying conditions: (a) $N(0) = 1$ and $N(1) = 0$, (b) N is strictly decreasing, (c) N is continuous and 4) $N(N(x)) = x$ for all $x \in [0,1]$ (N is involutive), is called a strong negation. For the purposes of this paper we consider the following strong negation: $N(x) = 1\text{-}x$, where $x \in [0,1]$.

If $T(x,y) = N(S(N(x),N(y)))$ for all $x,y \in [0,1]$, we call S the N-dual t-conorm to T. Triangular norms and conoroms are used to define the intersection and union of fuzzy sets respectively. Let A and B be fuzzy sets on U. The intersection of A and B is a fuzzy set $(A \cap_T B)$ on U given by $(A \cap_T B)(x) = T(A(x),B(x))$ for all $x \in U$, where T is a t-norm. The union of A and B on U is a fuzzy set $\left(A \cup_S B\right)$ on U given by

$$\left(A \cup_S B\right)(x) = S(A(x), B(x))$$

for all $x \in U$, where S is a t-conorm N-dual to T, for more details see Dubois & Prade (2000). Let A be a fuzzy set on U and B be a fuzzy set on V. Then the Cartesian product of A and B is a fuzzy set $A \times_T B$ on $U \times V$ given by $(A \times_T B)(x,y) = T(A(x),B(y))$ for all $(x,y) \in U \times V$. See Dubois & Prade (2000) for more details on triangular norms and conorms. A binary fuzzy relation is any fuzzy set P on $U \times V$.

In this paper we will use the minimum t-norm ($T(\alpha,\beta) = \min\{\alpha,\beta\}$, for all $\alpha,\beta \in [0,1]$) and the maximum t-conorm ($S(\alpha,\beta) = \max\{\alpha,\beta\}$, for all $\alpha,\beta \in [0,1]$). For the union, intersection and Cartesian product of fuzzy sets A and B based on this

t-norm and t-conorm we use the following notation: $\left(A \cup B\right)$, ($A \cap B$) and ($A \times B$) respectively.

Let P be a fuzzy relation on $U \times V$ and Q be a fuzzy set on $V \times W$. The composition of these two fuzzy relations is a fuzzy set $P \circ Q$ on $U \times W$ with a membership function defined for all $(x,z) \in U \times W$ by the formula $(P \circ Q)(x,z) = \sup_{y \in V} \{\min\{P(x,y), Q(y,z)\}\}$.

Let R denote the set of all real numbers. Fuzzy set C on R is called fuzzy number if it satisfies three conditions: (1) the kernel of C, Ker(C), is a nonempty set, (2) the α-cuts of C, C_α, are closed intervals for all $\alpha \in (0,1]$, and (3) the support of C, Supp(C), is bounded. The symbol $F_N(\mathrm{R})$ denotes the family of all fuzzy numbers on R. If Supp(C)$\subseteq$[a,b], we call C a fuzzy number on the interval [a,b]. The family of all fuzzy numbers on the interval [a,b] will be denoted by $F_N([a,b])$.

Let $A_1, A_2, ..., A_n \in F_N([a,b])$, we say that $A_1, A_2, ..., A_n$ form a fuzzy scale on [a,b] if these fuzzy numbers form a Ruspini fuzzy partition (see Ruspini, 1969 or Codara, D'Antona & Marra, 2009) on [a,b] (i.e. $\sum_{i=1}^{n} A_i(x) = 1$, for all $x \in [a,b]$) and are numbered in accordance with their ordering.

The basics of linguistic fuzzy modelling were introduced by Zadeh (1975). A linguistic variable is the quintuple (X, $T(X)$, U, M, G) where X is the name of the linguistic variable, $T(X)$ is the set of its linguistic values (linguistic terms), U is the universe, $U = [a,b] \subseteq \mathsf{R}$, which the mathematical meanings (fuzzy numbers) of the linguistic terms are defined on, G is a syntactic rule (grammar) for generating linguistic terms from $T(X)$ and M is a semantic rule (meaning), that assigns to every linguistic term $\mathcal{A} \in T(X)$ its meaning $A = M(\mathcal{A})$ as a fuzzy number on U. Linguistic terms and fuzzy numbers representing their meanings will be distinguished in the text by different fonts (calligraphic letters for linguistic terms and standard capital letters for their respective meanings - fuzzy numbers on U).

The linguistic variable (X, $T(X)$, U, M, G), $T(X) = \{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_s\}$, $M\left(\mathcal{T}_p\right) = T_p$, $T_p \in F_N\left(U\right)$ for $p = 1,\ldots,s$, defines a linguistic scale on U, if the fuzzy numbers $T_1, T_2, \ldots, T_s$ form a fuzzy scale on U.

Let (X_j, $T(X_j)$, U_j, M_j, G_j), j=1,...,m, and (Y, $T(Y)$, V, M, G) be linguistic variables. Let $\mathcal{A}_{ij} \in T(X_j)$ and $M_j(\mathcal{A}_{ij}) = A_{ij} \in F_N(U_j)$, $i = 1,\ldots,n, j = 1,\ldots,m$. Let $\mathcal{B}_i \in T(Y)$ and $M(\mathcal{B}_i) = B_i \in F_N(V)$, $i = 1,\ldots,n$. Then the following scheme is called a linguistically defined function (a base of fuzzy rules, see Zadeh (1975)):

If X_1 is $\mathcal{A}_{11}$ and … and X_m is $\mathcal{A}_{1m}$ then Y is $\mathcal{B}_1$.
If X_1 is $\mathcal{A}_{21}$ and … and X_m is $\mathcal{A}_{2m}$ then Y is $\mathcal{B}_2$.
……
If X_1 is $\mathcal{A}_{n1}$ and … and X_m is $\mathcal{A}_{nm}$ then Y is $\mathcal{B}_n$.

(1)

The process of calculating an output for current input values by means of such a rule base is called approximate reasoning (or fuzzy inference). In the model we use approximate reasoning mechanisms based on Takagi & Sugeno's fuzzy inference (Takagi & Sugeno, 1985) and Generalized Sugeno fuzzy inference (for more details see Sugeno & Yasukawa (1993) or Talašová (2003)).

Let us have the base of fuzzy rules defined previously. Let us have an observation in the form of X_1 is $\mathcal{A}_1'$ and … and X_m is $\mathcal{A}_m'$. Then by entering these observed values into the linguistically defined function, we get the output

$B^S = \left(\sum_{i=1}^{n} h_i B_i\right) / \left(\sum_{i=1}^{n} h_i\right)$, where

$h_i = \min\left\{\mathrm{hgt}\left(A_{i1} \cap A_1'\right), ..., \mathrm{hgt}\left(A_{im} \cap A_m'\right)\right\}$. For the classical Sugeno's fuzzy inference, B_i is a real number for all i=1,...,n. If we use the generalized Sugeno's fuzzy inference

introduced by Talašová (2003), B_i is a fuzzy number for all i=1,...,n.

Mamdani & Asilian (1975) introduced another approach to fuzzy inference. Let us consider the rule base (1). Each rule is modeled by the fuzzy relation

$$R_i = A_{i1} \times_T A_{i2} \times_T \ldots \times_T A_{im} \times_T B_i,\ i=1,\ldots,n$$

The whole rule base is represented by the union of all these fuzzy relations $R = \bigcup_{i=1}^{n} R_i$. Let $(A'_1, A'_2, \ldots, A'_m)$ be an m-tuple of fuzzy inputs. The output B^M is then calculated as

$$B^M = (A'_1 \times_T A'_2 \times_T \ldots \times_T A'_m) \circ R .$$

Mamdani & Asilian's approach (1975) preserves information regarding the uncertainty of input values. This is important particularly when the inputs are highly uncertain. On the other hand, the output of Mamdani's fuzzy model is usually not a fuzzy number. To interpret the Mamdani output linguistically may prove problematic. The center of gravity method is usually used to defuzzify the Mamdani & Asilian's output B^M. This way we get a crisp output

$$b^M = \int_{y \in V} B^M(y) \cdot y \, dy \,/ \int_{y \in V} B^M(y) \, dy).$$

The asymmetry of fuzzy numbers can negatively influence the output of the defuzzification process and thus reduce the interpretation possibilities of such an output. A proper linguistic approximation of B^M may, on the other hand, be too uncertain to provide the desired amount of information. The generalized Sugeno inference (Talašová, 2003 or Sugeno & Yasukawa, 1993) provides fuzzy numbers as outputs (if the inputs are fuzzy numbers) which makes the output B^S easier to interpret linguistically. Partial loss of information concerning the uncertainty of input values is compensated by easier interpretability of the output. As linguistic terms (linguistic approximations of B^S) are easier to interpret for the EMRS operators than general fuzzy set outputs, we use the Sugeno's fuzzy inference (Takagi & Sugeno, 1985) and the generalized Sugeno's fuzzy inference (Talašová, 2003) in the model.

PROBLEM SPECIFICATION

A decision making support for the EMRS Operations Center can be considered beneficial, if it (1) is able to process any input data (uncertain or precise), (2) provides an estimation of the number of people moderately and severely wounded, (3) determines to which medical facilities should the injured people be transported (4) is able to assess the sufficiency of any provincial EMRS center's forces and resources and (5) performs these operations in real-time.

The relatively low frequency rate of disasters limits the amount of data available to construct the mathematical model. Some parts of the model therefore need to be based on expert and professional medical rescue worker's experience. The model has to be able to communicate with the EMRS operator in an easy to understand way. Outputs must be "intuitively clear" to the operator. A linguistic form of outputs is therefore preferred. A multiphase mathematical model designed to meet these requirements will be described.

In the whole chapter we consider only disasters resulting in mechanical injuries (which is the case of many traffic accidents, but also earthquakes etc.). There is not enough experience and data available for other sources of injuries, such as chemicals, radiation, explosions, fire, biological agents etc. to develop similar model. This fact should be addressed in future research.

MATHEMATICAL MODEL

To be able to deal with the uncertainty of input data and to model the expert-based knowledge of some parts of this problem, the tools of fuzzy mathematics described above are used. Namely fuzzy sets, fuzzy numbers, linguistic scales and linguistic fuzzy models are the basis of the described solution. For a more detailed explanation see Talašová (2003) or Dubois & Prade (2000). The fuzzy number α- degree upper bound had to be defined as a new way of fuzzy and crisp number comparison in order to meet the model requirements.

Phase 1: Rescue Capacity Determination

One of the key pieces of information we are interested in at the moment any disaster occurs (resulting in many injured people) is the amount of people we are able to provide care to within a time unit (hour). Such quantity is, according to Boer (1999), called the Medical Rescue Capacity (*MRC*).

Each EMRS center has a certain number of medical rescue teams on duty. There does not necessarily have to be a doctor in all these teams. Let us assume that teams have 2 to 4 members and that each team can be treating no more than one patient at a time. We can assume that only teams with doctors will be taking care of seriously injured people and both teams with and without doctors will take care of people with moderate injuries. People that are just slightly injured do not interest us now, as during disasters all the forces and resources of EMRS need to concentrate on people that are unable to transport themselves to the hospital.

We consider two categories of injured people: seriously injured (T1 – for "triage group 1") and moderately injured (T2). T1 and T2 patients differ mainly in the time needed to treat them so that they are able to withstand the transport to hospital – T1 patients need more treatment and therefore more time (Boer (1999) suggests a method of mean treatment time determination). We also have many different teams with different experience and skills. The *MRC* should therefore be determined for each team separately and of course we need to distinguish between the *MRC* for T1 patients and the MRC for T2 patients.

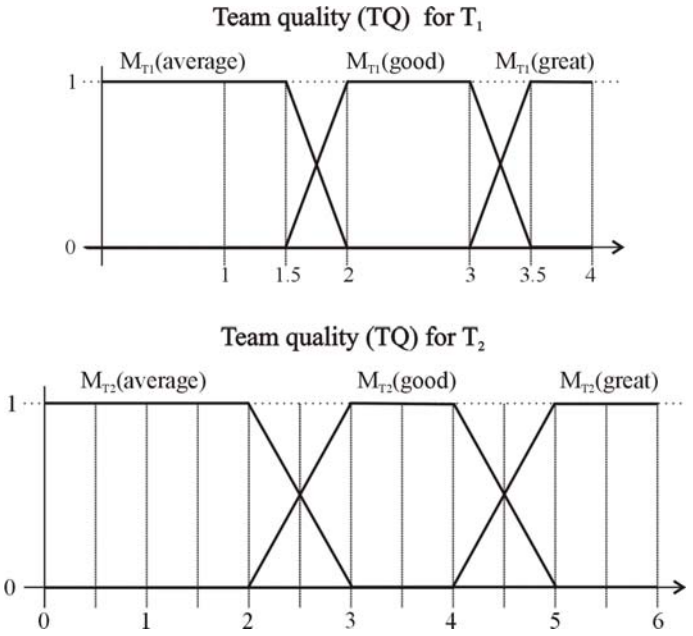
We consider the team quality (*TQ*), its time on duty (*ToD*) and the weather at the disaster site (*WE*) to be the most important factors influencing the actual medical rescue capacity of each team. Each of these variables is represented by a linguistic scale in the model. We use linear membership functions to define the meanings of the respective linguistic terms on the respective universes (example for *TQ* can be seen in Figure 1).

As the relationship among these three factors and the resulting *MRC* can only be described linguistically (based on interviews with experienced rescue workers – no other information is available at this time) both inputs (*TQ,ToD,WE*) and output of this phase (*MRC*) are modeled by linguistic scales. The relationship is then described by a fuzzy rule base consisting of 48 rules such as:

"If TQ is average and ToD is short and WE is ideal, the MRC is average."

We use the generalized Sugeno approximate reasoning mechanism (Talašová, 2003) to obtain results for any combination of inputs. Thus we get a fuzzy number describing the current *MRC* for a particular team. The use of linguistic variables allows us to have a single rule base for *MRC* determination for the case of T1 and T2 patients, thus simplifying the model significantly. We can use the same linguistic terms for team quality description, only their meanings (fuzzy numbers assigned to them) will differ. For T1, we have the universe [0,4] (in terms of patients treated per hour) and for T2 we have the universe [0,6] again in terms

Figure 1. Differences in the meanings of linguistic terms for team quality description for T1 and T2 patients

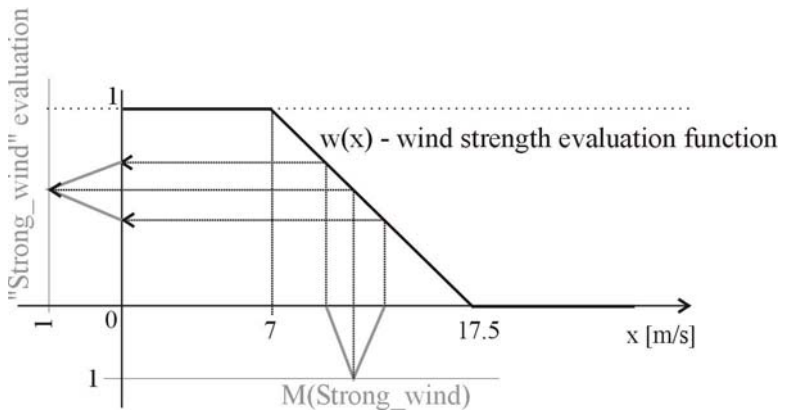


of patients treated per hour. For the differences in meanings of the linguistic terms see Figure 1.

The fuzzy approach also allows users to input *TQ* in linguistic terms (using the predefined linguistic variable for team quality description) as well as in precise numbers (supposing a team quality evaluation mechanism is available).

The exact time of duty is easy to obtain and input as a precise number. To assess the weather quality (*WE*), we define an evaluation function (a membership function of a fuzzy set that describes conditions appropriate for the rescue operation) for each of the three important weather characteristics – temperature (*t*), wind strength (*w*) and rainfall (*r*). These evaluation functions map the real weather characteristics values into the closed interval [0,1] (see Figure 2) and using the extension principle (see Dubois & Prade, 2000) they

Figure 2. Evaluation function for one of the important weather characteristics – wind. Evaluation is determined for a fuzzy (uncertain) input as a fuzzy number using the extension principle.



also map the fuzzy sets describing the estimates of current weather characteristics (defined on the universe of the respective weather characteristics) into fuzzy sets on [0,1]. The resulting evaluation 1 means our complete satisfaction and evaluation 0 our complete dissatisfaction or even dangerousness of the weather. The overall weather quality evaluation is then computed by aggregating the above mentioned important weather characteristics evaluations (*T* for temperature, *W* for wind strength and *R* for rainfall) using a fuzzified geometric mean:

$$(F)\left(T \cdot W \cdot R\right)^{\frac{1}{3}}(y) = \sup\{\inf\{T(x_1), W(x_2), R(x_3)\} \mid y = (x_1 \cdot x_2 \cdot x_3)^{\frac{1}{3}}; x_1, x_2, x_3 \in [0,1]\} \quad (2)$$

The geometric mean (its fuzzification) was chosen because of its suitable properties: 1) if any of the values is completely dissatisfactory, then the weather evaluation is 0 (completely dissatisfactory), 2) if any of the weather characteristics evaluations is low, the overall weather evaluation will be lowered accordingly (only weather with all evaluations really close to 1 will be evaluated as satisfactory), 3) it is easy enough to use and to be understood by laymen.

The output of the first rule base, that describes the relationship between the inputs (*TQ, ToD, WE*) and output (*MRC*) is the medical rescue capacity for a current team (*MRC*). The time of day (*TDy*) can also be taken into account in order to model reaction time prolonging, fatigue and drowsiness of the staff. For this purpose a second fuzzy rule base is defined to determine how to correct the *MRC* according to the time of the day. There are 4 linguistic fuzzy rules available for this purpose and classical Sugeno approximate reasoning is applied (Sugeno & Yasukawa, 1993):

- *If TDy is night, then the correction is 1.*
- *If TDy is morning, then the correction is 0.*
- *If TDy is afternoon, then the correction is 0.3.*
- *If TDy is evening, then the correction is 0.6.*

The result of this rule base is shifting the *MRC* lower or leaving it unchanged. *TDy* was left out of the first rule base to maintain simplicity of the model and to preserve the information concerning potential (maximal) *MRC* for the purposes of disaster planning.

Applying this process on every team on duty and adding up the resulting fuzzy numbers we get the total *MRC* (*TMRC*) for a particular provincial EMRS center: $TMRC = \sum_{i=1}^{t} MRC_i$, where *t* is the number of teams available at this center. An average *MRC (AMRC)* is easy to determine as well: $AMRC = \sum_{i=1}^{t} MRC_i / t$. Fuzzy number arithmetic used to carry out the calculations is based on interval arithmetic (see Dubois & Prade, 2000).

As the location of the disaster is not known in advance, some calculations can be done no sooner than the emergency call is answered (particularly those requiring specific weather data from the disaster site). To obtain the results of phase one – the *MRC* for each team and the *TMRC* and *AMRC* – we need to consider all the data concerning the conditions at the disaster site and its location from the emergency call (precise or not). We can supply additional data from other sources if possible (e.g. weather conditions data from meteorological stations). Outputs of phase 1 are therefore highly relevant for the emergency medical rescue services decision making, for they reflect much of the disasters context.

Phase 2: Severity of the Disaster Determination

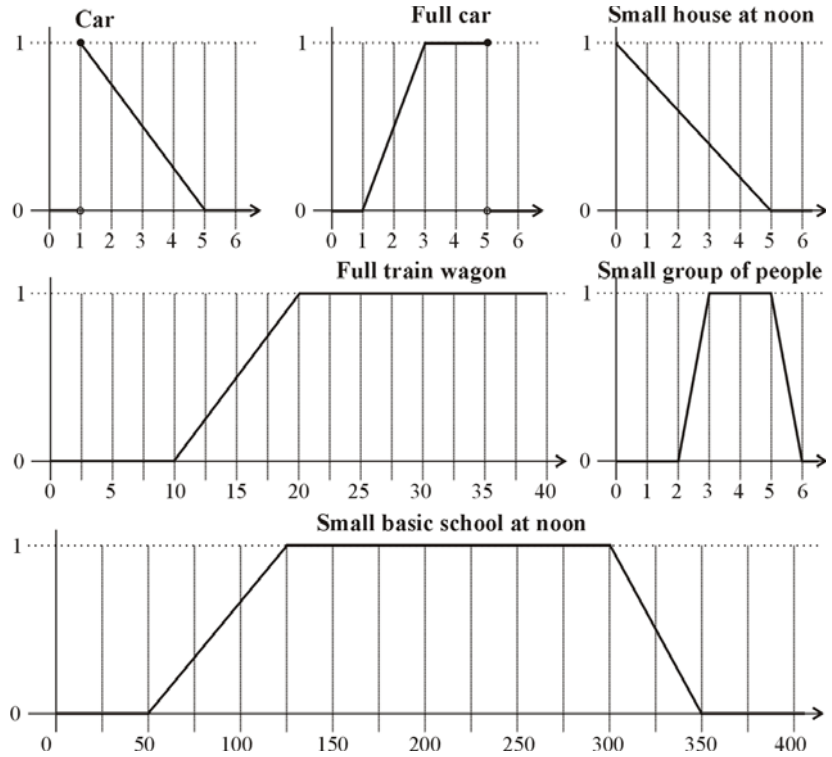
This part of the model is partly dependent on the data obtained during the emergency call and partly on the so called common knowledge. Both these

things have to be combined to derive an output that really provides a decision making support. The disaster has to be described using all the data available, regardless of the uncertainty. We need to assess the disaster severity – to be able to estimate the number of people injured. It is even more important to determine how many people belong to T1 group and how many to T2.

First we define a linguistic variable describing the disaster severity with the set of linguistic terms {*small traffic accident, medium traffic accident, serious traffic accident, small disaster, medium disaster and serious disaster*}, their respective meanings being fuzzy numbers on the interval [0;10000]. These fuzzy numbers should be defined with respect to the needs of the EMRS and disaster management of the country our model will be used in. Thus the EMRS operator gets a means of roughly describing the disaster (linguistically), when almost no other information is available. We also allow the operator to input exact numbers of injured people (in case the precise number is available), as well as his guesses in the form of fuzzy numbers (a way of customizing/refining the meanings of the previously mentioned predefined linguistic terms).

However the largest benefit of our proposed fuzzy approach in this phase is the possibility to input linguistic terms (Figure 3). We predefine the meanings of such terms as a "full car", "small group of people", "full train wagon", "small basic school" as fuzzy numbers. The information from an emergency call: "A car crashed into a small group of people and then collided with a bus" can then be inputted by simply choosing the buttons "bus", "small group of people" and "car" on a touch screen in the EMRS operations centre. The computer can add up the meanings of these linguis-

Figure 3. Examples of the meanings of some predefined linguistic terms describing the basic units of possible disasters in terms of the amount of people affected by the disaster



tic terms and thus give the operator the expected amount of people affected by the disaster (*AFF*) again as a fuzzy number or in linguistic terms.

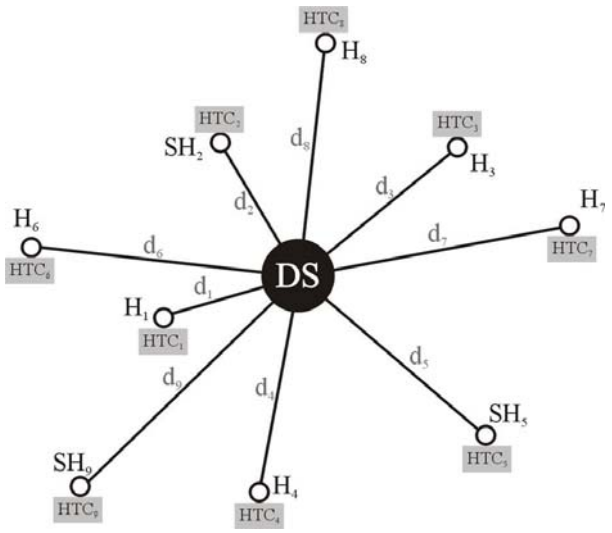
When mechanical injuries prevail (which is the case of most disasters), the relative amount of severely wounded is about 10% and the relative amount of moderate injuries is about 20% (Boer, 1999). To estimate how many people have been injured, we define two fuzzy numbers – *M*("about 10%") and *M*("about 20%"). The kernels of these fuzzy numbers will be 0.1 and 0.2 respectively and their supports (0,0.2) and (0.1,0.3) respectively. Using fuzzy number arithmetic (Talašová, 2003, Dubois & Prade, 2000) we can obtain the estimate of the amount of both severely injured (*NT*1, where *NT*1=*AFF*·*M*("about 10%")) and moderately injured (*NT*2, where *NT*2=*AFF*·*M*("about 20%")) people as fuzzy numbers.

Phase 3: Determination of Medical Facilities Involved

At this point we know how many patients we are able to treat per hour and how many of them we can expect at the disaster site. Since the location of the disaster is known, we also know how far the nearest hospitals are and what their treatment capacity (*HTC* – number of patients that can receive appropriate treatment per hour, should be a piece of information available in advance) is. We need to consider time limitations – for the severely injured must be treated in appropriate medical facilities within an hour and moderately injured within 6 hours (Boer, 1999; Stoklasa, 2009).

Let us consider two types of hospitals: regular hospitals (*H*) and specialized hospitals with emergency departments (*SH*). We suppose that T1 patients can be treated only in specialized hospitals and we need to transport them there within one hour (so called "golden hour" – see Boer (1999) or Stoklasa (2009)). T2 patients can be treated either in regular or in specialized hospitals and we have 6 hours to transport them there (so called Friedrich's time). Figure 4 shows the situation. Let $I = \{1, 2, ..., n\}$ be the set of indices of all the hospitals close to the disaster site (indexed according to the distance from the disaster site, index 1 meaning the hospital closest to the disaster site). The last index n is dependent on the time

Figure 4. Determining which medical facilities will be involved. (DS=disaster site, HTC=hospital treatment capacities)



limits we set for the transport of the patients to medical facilities. Let us denote I_{SH} the set of indices of all specialized hospitals, $I_{SH} \subseteq I$, $I_{SH}=\{i_1, i_2, ..., i_r\}$ and I_H the set of indices of all regular hospitals, $I_H \subseteq I$, $I_H=\{j_1, j_2, ..., j_s\}$. It holds that $I_{SH} \cup I_H = I$ and $I_{SH} \cap I_H = \varnothing$. Conditions that need to be fulfilled are as follows.

We need to employ so many specialized hospitals, so that all the T1 patients can receive appropriate care within the first hour

$$\sum_{i \in I_{SH}} HTC_i \geq NT1 \tag{3}$$

For this purpose we suppose that no T2 patients will be transported within the first hour.

We need to employ so many regular and specialized hospitals, so that all the T2 patients can receive appropriate care within the following five hours

$$\sum_{i \in I_{SH}} HTC_i + \sum_{j \in I_H} HTC_j \geq \frac{NT2}{5} \tag{4}$$

As the sums of hospital treatment capacities are real numbers, whereas *NT*1 and *NT*2 are fuzzy numbers, we need to define the meaning of the symbol "≥". For this purpose we introduce the fuzzy number α- degree upper bound.

Definition 1. Let *h* be a real number on [*a,b*] and *A* be a fuzzy number with a non-zero membership function on [*a,b*] and zero membership function outside this interval. Then we say that *h* is the α- *degree upper bound for the fuzzy number A*, α∈ [0,1], if and only if

$$\alpha = \frac{\int_a^h A(t)dt}{\int_a^b A(t)dt} \tag{5}$$

Definition 1 describes how to compare a fuzzy number with a crisp one. Parameter α describes the amount of uncertainty we are willing to tolerate. If α is set to 1, we need *h* to be larger or equal than the largest number from Supp(*A*), in other words we do not tolerate any uncertainty (*h* has to be larger or equal than any value with a non-zero membership degree to *A*). For a fixed α the condition (3) is fulfilled if and only if $\sum_{i \in I_{SH}} HTC_i$ is a β- degree upper bound of *NT*1, and β ≥ α. By choosing α we can influence the degree of satisfaction of the constraints. According to (1-α) (the degree to which the total *HTC* of all the hospitals can be exceeded) we determine to which hospitals the injured people need to be transported to receive appropriate care. If we choose α= 1, we employ as many hospitals as it is needed to provide care to the maximum estimated number of injured people (for T1 category it is the supremum of Supp(*NT*1). We try to determine the least amount of *SH* hospitals so that the condition (3) is fulfilled for a given α. This way we get the set of all employed specialized hospitals in the first hour $I_{SH_1} = \{i_1, i_2, ..., i_{r_1}\}$. We do the same for the condition (4), but this time for any type of hospitals – *H* and *SH*.

Phase 4: Estimation of the Amount of Forces and Resources Needed

Once we know which hospitals need to be alerted to start preparing themselves to receive patients, the number of ambulances needed to transport the patients to the hospitals has to be determined. Let us for the i-th employed hospital *i*=1, 2, ..., n, define

$$trav_i = \frac{v}{2 \cdot d_i}, \tag{6}$$

where *v* is the supposed average traveling speed of ambulances (say 50 km/h) and d_i is the distance from the i-th hospital (regular or specialized) to

the disaster site. $Trav_i$ then describes the number of journeys from the disaster site to the hospital and back that a single ambulance traveling at an average speed of 50 km/h is able to make within one hour. Let us for now consider the first hour of the disaster response (and therefore only the transportation of T1 patients to *SH*). Let us denote x_i the number of ambulances needed to transport patients to the i-th specialized hospital. The following conditions need to be fulfilled.

1. The total number of ambulances should be as low as possible

$$\sum_{k=1}^{r_1} x_{i_k} \rightarrow \min \tag{7}$$

2. All T1 patients need to be transported to hospitals

$$\sum_{k=1}^{r_1} trav_{i_k} \cdot x_{i_k} \geq NT1 \tag{8}$$

3. We do not allow any of the hospital treatment capacities to be exceeded

$$trav_{i_k} \cdot x_{i_k} \leq HTC_{i_k}, \quad k = 1,2,...,r_1 \tag{9}$$

As can be easily seen, we have a fuzzy linear programming problem to solve. Condition (8) has a fuzzy number constraint *NT*1. We can again use the fuzzy number α- degree upper bound to deal with this condition.

We have devised a heuristic algorithm based partially on the α- degree fuzzy number upper bound to determine how many ambulances will be needed to successfully cope with the situation. The previous selection of involved medical facilities was performed in a way that can be exploited now. We still consider just the T1 category of injuries. T2 category can be treated analogously. If we take all the hospitals with indices belonging into $I_{SH_1} = \{i_1, i_2, ..., i_{r_1}\}$, we have enough hospital treatment capacity to treat all the T1 patients within the first hour. This is ensured by the hospital selection process introduced earlier in this chapter. The minimum number of ambulances needed to transport all the T1 patients (*NT*1 in total) to these hospitals remains to be determined. The main idea of our heuristic approach is to start with the hospital closest to the disaster site (just a reminder - for T1 patients we need only specialized hospitals) and determine the minimum number of ambulances needed to fill this hospital's treatment capacity within one hour and then continue the same with the second closest hospital. Such a number is now easy to obtain for the i-th hospital,

$$i \in I_{SH_1} - \{i_{r_1}\} = \{i_1, i_2, ..., i_{r_1-1}\},$$

as $x_i = HTC_i / trav_i$. For calculations involving T1 patients we may round $trav_i$, for if the fractional part of $trav_i$ is greater or equal to 0.5, we can assume that the ambulance will manage to transport the last patient to the hospital within the time limit (in this case we round $trav_i$ up), while if the fractional part of $trav_i$ is less than 0.5, the last patient will not arrive in the hospital in time (so we round $trav_i$ down in this case).

We shall round up the resulting x_i to the closest positive integer (to avoid confusions during the decision making process). However doing so we risk exceeding the hospital treatment capacity of the particular hospital (should all x_i ambulances transport patients for the whole hour) by $trav_i$ -1 in the worst case. This is in our opinion a low price to pay for a simpler computation.

The treatment capacity of the most remote hospital might not be fully exploited. To determine the number of ambulances needed to transport patients to this hospital $x_{i_{r_1}}$, we first add up all the previously determined numbers

of ambulances needed $\sum_{k=1}^{r_1-1} x_{i_k}$. We choose an αand set $x_{i_{r_1}} = 1$. We then increase $x_{i_{r_1}}$ by 1, until

$$\sum_{k=1}^{r_1} trav_{i_k} \cdot x_{i_k} \geq NT1,$$

in other words until $\sum_{k=1}^{r_1} trav_{i_k} \cdot x_{i_k}$ is a β- degree upper bound of *NT*1, and β≥α. The total amount of ambulances needed is then $AMB = \sum_{k=1}^{r_1} x_{i_k}$.

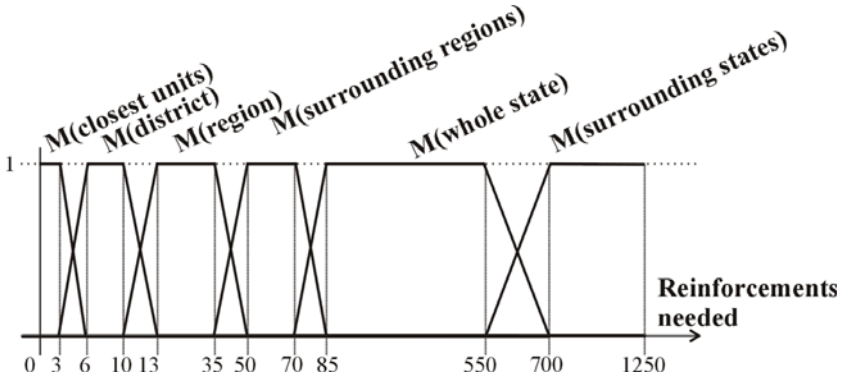
Knowing the number of ambulances needed to transport patients away from the disaster site is however not yet sufficient to be able to assess if the current (provincial) EMRS center has enough forces and resources to deal with the disaster. In order to do so, we need to determine how many teams will be needed at the disaster site to provide care to all the severely injured within the first hour (phases 1 and 2) first.

We expect *NT*1 severely injured people at the disaster site (output of phase 2) and we can determine the average medical rescue capacity (*AMRC*) as *TMRC*/*N*, where *N* is the number of teams used to compute *TMRC* (output of phase 1). *TMRC* is a sum of fuzzy numbers. Generalized Sugeno deduction is used to determine current medical rescue capacities. The output of this process for each team – the current *MRC* – is a weighted average of fuzzy numbers and therefore a fuzzy number. It can be easily seen that the average medical rescue capacity will be a fuzzy number. If we divide *NT*1 by *AMRC*, we obtain the amount of EMRS teams needed at the disaster site as a fuzzy number (let us denote this fuzzy number *TEAMS*).

Once we know the required amount of forces and resources needed both at the disaster site and to transport patients to hospitals, we can compare it with the amount that is available at the current provincial EMRS centre. We can assume that a provincial EMRS center can handle the situation without any assistance from other EMRS centers, if (*TEAMS* + *AMB*) ≤ *Number of teams available*. We can again use the fuzzy number α- degree upper bound to interpret the symbol ≤. Should the resources of a current provincial EMRS center prove insufficient, we can linguistically approximate the fuzzy number (*TEAMS* + *AMB*) - *Number of teams available* using the linguistic variable "Reinforcements", its linguistic terms being for example "closest provincial EMRS units", "whole country EMRS stations", "surrounding countries EMRS stations", etc. (Figure 5). The meanings of these linguistic terms need to be set up according to the real numbers of reinforcements available.

Similarly we can run the whole procedure for the moderately injured people. This way an easy to understand piece of information concerning

Figure 5. Meanings of the linguistic terms of the linguistic variable "Reinforcements"



the need of reinforcements is available to the EMRS operator.

NUMERICAL EXAMPLE

Let us consider the following situation. It is September, we are in the Czech Republic at 10:30. An emergency call is registered by an EMRS operator: "*A train collided with a bridge construction. All 10 wagons! I am in XZ.*" This is all the information we have. How serious is the event? How many people are injured? How many ambulances should we send? Do we have enough sources and resources?

Let us suppose we have 7 teams with a doctor (all considered "good") and 14 teams without a doctor (7 considered "good" and 7 considered "great") available. Let us suppose that it is 17°C, the wind is 2 m/s and it is not raining at the disaster site and this information is available to us.

Let us consider the following hospital structure near the disaster site (*SH* = specialized hospitals capable of treating T1 and T2, *H* = ordinary hospitals, capable of treating T2):

- H_1d = 5km HTC = 5
- SH_2d = 10 km HTC = 20
- SH_3d = 15km HTC = 15
- H_4d = 15 km HTC = 10
- H_5d = 20km HTC = 10
- SH_6d = 25 km HTC = 30,

where d is the distance between the current hospital and the disaster site, HTC is the hospital treatment capacity of the respective hospital (number of people the hospital is able to treat per hour). We also consider the ambulances to be moving at the average speed of 50 km/h.

Operator has to put into the model two facts: "10 full train wagons" are involved and that the disaster site is XZ. The model then provides the following outputs (only the most important ones for the decision are listed).

1. 39 ambulances are needed during the first hour (25 to transport patients to hospitals, 14 teams to provide care at the disaster site)
2. 11 ambulances will be needed during the following 5 hours (7 to transport patients to hospitals, 4 to provide care at the disaster site)
3. "You need reinforcements for the first hour. Mobilise the region!"

Outputs provided by the model are easily interpretable. It is still the operator who has to decide, what to do. He can now compare his or her guess (for he/she has no time for anything more sophisticated than guesses) with the output of the model and then act. More outputs are available from the model – estimated number of injured people, distribution of ambulances (patients) among the hospitals etc.

FUTURE RESEARCH DIRECTIONS

Until now not much attention has been paid to the decision making processes of the medical rescue services and the uncertainty they involve. If the first decisions are wrong or late, we can not expect the whole system to work well. This is one of the reasons why this model has been developed – to show the importance of first decisions and the role of the vagueness that is carried through the decision making process.

During the work on the model, we encountered some problems that may be addressed in the future. The first one is the lack of experience and even information concerning disasters and disaster procedures thanks to the relatively low frequency of disaster appearance. Further on no EMRS quality assessment tool was found in the Czech Republic, although the rescue capacity of each team is an important piece of information both for the disaster response and planning, even for EMRS management.

We also realized that a functional data exchange system for the EMRS, medical facilities and other rescue services would simplify the decision making process significantly, by reducing the uncertainty of some data by confirming them from multiple sources and by sharing the decision making process outputs in real time with all the units involved in the disaster response.

CONCLUSION

The model described above was designed primarily as a decision making support for the EMRS operators, should a disaster resulting in a large number of casualties occur. The main advantage of the proposed model and the mathematical tools used is the ability to accept and also provide linguistic values and this way deal with uncertainty. The model estimates the current total medical rescue capacity (number of people treated per hour) for a current provincial EMRS centre. Based on the emergency call it estimates the severity of the disaster and then determines which hospitals will be included into the rescue process and the number of ambulances needed. The number of ambulances needed is then compared with the number of forces and resources available and the required number of reinforcements is suggested in linguistic terms. At present we are gathering more data to field-test the model and to adapt it to the needs of the disaster response system in the Czech Republic.

The use of this model can not substitute the human factor completely – EMRS operators still play the key role in the disaster response. This model is intended to provide information to compare the EMRS operator's opinion with, this way eliminating mistakes and speeding up the decision making process. The second field of use for this model is disaster planning. The model provides insights into EMRS activity during disasters, describing this activity in an easy to understand way. Many potential disaster scenarios can be explored using the model and the need of sources and resources can be determined. This way an optimal distribution of EMRS centers and resource stores can be determined.

In such situations as the first minutes of the disaster response, when little information is available and the uncertainty level is high, the above mentioned approach can provide useful results. A model that is able to deal with uncertainty when the uncertainty is inherent to the situation modeled is most appreciated for it can help us identify the blind spots within the system. Once we know how uncertain the decisions may be, we can start working on measures to reduce the level of uncertainty.

REFERENCES

Abbod, M. F., & Linkens, D. A. (1998). Anaesthesia monitoring and control using fuzzy logic fusion. *J. Biomed. Eng. – Appl. Basis Commun. Special Issue on Control Methods in Anaesthesia*, *10*(4), 225–235.

Abbod, M. F., von Keyserlingk, D. G., Linkens, D. A., & Mahfouf, M. (2001). Survey of utilisation of fuzzy technology in medicine and healthcare. *Fuzzy Sets and Systems*, *120*, 331–349. doi:10.1016/S0165-0114(99)00148-7

Altay, N., & Green, W. G. III. (2006). OR/MS research in disaster operations management. *European Journal of Operational Research*, *175*, 475–493. doi:10.1016/j.ejor.2005.05.016

Boer, J. D. (1999). *Order in chaos*. Amsterdam, The Netherlands: Free University Hospital.

Codara, P., D'Antona, O. M., & Marra, V. (2009). An analysis of Ruspini partitions in Gödel logic. *International Journal of Approximate Reasoning*, *50*, 825–836. doi:10.1016/j.ijar.2009.02.007

Cret, L., Yamazaki, F., Nagata, S., & Katayama, T. (1993). Earthquake damage estimation and decision-analysis for emergency shutoff of city gas network using fuzzy set theory. *Structural Safety*, *12*(1), 1–19. doi:10.1016/0167-4730(93)90015-S

Dong, W. M., Chiang, W. L., & Shah, H. C. (1987). Fuzzy information processing in seismic hazard analysis and decision making. *Soil Dynamics and Earthquake Engineering*, *6*(4), 220–226. doi:10.1016/0267-7261(87)90003-0

Dubois, D., & Prade, H. (2000). *Fundamentals of fuzzy sets*. Boston, MA: Kluwer Academic Publishers.

Esogbue, A. O. (1996). Fuzzy sets modeling and optimization for disaster control systems planning. *Fuzzy Sets and Systems*, *81*(1), 169–183. doi:10.1016/0165-0114(95)00248-0

Esogbue, A. O., Theologidu, M., & Guo, K. J. (1992). On the application of fuzzy sets theory to the optimal flood control problem arising in water resources systems. *Fuzzy Sets and Systems*, *48*(2), 155–172. doi:10.1016/0165-0114(92)90330-7

Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-machine Studies*, *7*, 1–13. doi:10.1016/S0020-7373(75)80002-2

Rotshtein, A. (1999). Design and tuning of fuzzy rule-based systems for medical diagnosis . In Teodorescu, H. N., Kandel, A., & Jain, L. C. (Eds.), *Fuzzy and neuro-fuzzy systems in medicine*. Boca Raton, FL: CRC Press.

Ruspini, E. (1969). A new approach to clustering. *Information and Control*, *15*, 22–32. doi:10.1016/S0019-9958(69)90591-9

San Pedro, J., Burstein, F., Churilov, L., Wassertheil, J., & Cao, P. (2004). Mobile decision support system for triage in emergency departments. *Decision Support in Uncertain and Complex World: The IFIP TC8/WG8.3 International Conference,* (pp. 714-723).

Štětina, J. (Ed.). (2000). *Disaster medicine*. Praha, Czech Republic: Grada Publishing. (in Czech)

Stoklasa, J. (2009). *Classical and fuzzy models for efficiency assessment* (in Czech). Unpublished Master's thesis, Palacky University, Olomouc, Czech Republic

Sugeno, M., & Yasukawa, T. (1993). A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, *1*(1), 7–31. doi:10.1109/TFUZZ.1993.390281

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, *1*(15), 116–132.

Talašová, J. (2003). *Fuzzy methods of multiple-criteria evaluation and decision making*. Olomouc, Czech Republic: VUP. (in Czech)

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353. doi:10.1016/S0019-9958(65)90241-X

Zadeh, L. A. (1975). The concept of linguistic variable and its application to approximate reasoning. *Information Sciences, 8,* 199-249; 301-357; *9*, 43-80.

ADDITIONAL READING

Altay, N., & Green, W. G. III OR/MS research in disaster operations management. (2006). *European Journal of Operational Research*, *16*(175), 475-493. doi:10.1016/j.ejor.2005.05.016

Avouris, N. M. (1995). Cooperating knowledge-based systems for environmental decision support. *Knowledge-Based Systems*, *1*(8), 39–54. doi:10.1016/0950-7051(94)00289-U

Avouris, N. M., & Finotti, S. (1993). User interface design to expert systems based on hierarchical spatial representations. *Expert Systems with Applications*, *2*(6), 109–118. doi:10.1016/0957-4174(93)90001-M

Chongfu, H. (1996). Fuzzy risk assessment of urban natural hazards. *Fuzzy Sets and Systems*, *2*(83), 271–282. doi:10.1016/0165-0114(95)00382-7

Gray, J. (1981). Characteristic patterns of and variations in community response to acute chemical emergencies. *Journal of Hazardous Materials*, *4*(4), 357–365. doi:10.1016/0304-3894(81)87006-9

Guohua, Ch. & Xinmei, Zhang. (2009). Fuzzy-based methodology for performance assessment of emergency planning and its application. *Journal of Loss Prevention in the Process Industries*, *2*(22), 125–132.

Hogan, D. E., & Burstein, J. L. (Eds.). (2007). *Disaster medicine*. Philadelphia: Lippincott Williams & Wilkins.

Huang, Ch., & Inoue, H. (2007). Soft risk maps of natural disasters and their applications to decision-making. *Information Sciences*, *7*(177), 1583–1592. doi:10.1016/j.ins.2006.07.033

Jiuh-Biing, S. (2007). An emergency logistics distribution approach for quick response to urgent relief demand in disasters. *Transportation Research Part E, Logistics and Transportation Review*, *6*(43), 687–709.

Karimi, I. & Hüllermeier, E. (2007). Risk assessment system of natural hazards: A new approach based on fuzzy probability. *Fuzzy Sets and Systems*, *9*(158), 987–999. doi:10.1016/j.fss.2006.12.013

Lillibridge, S. L., Noji, K. & Burkle, F. M. Jr. (1993). Disaster assessment: The emergency health evaluation of a population affected by a disaster. *Annals of Emergency Medicine*, *11*(22), 1715–1720. doi:10.1016/S0196-0644(05)81311-3

Masár, O., Štorek, J., Brenner, M., Turečková, H., Sysel, D., & Belejová, H. (2010). *Selected chapters from disaster medicine*. Bratislava: Faculty of Medicine, Komenský University in Bratislava. (in Czech)

McCaughrin, W. C., & Mattammal, M. (2003). Perfect storm: Organizational management of patient care under natural disaster conditions. *Journal of Healthcare Management*, (48): 295–308.

Pokorný, J., & Štorek, J. (2003). Aktuelle Entwicklungen im tschechischen Rettungsdienst. Beericht anslässlich der ersten Koferenz zu den Terroranschlägen des 11.September 2001. *Notfall&Rettungsmedizin*, *2*(6), 107–108.

Rosenthal, U., & Kouzmin, A. (1997). Crises and crisis management: Toward comprehensive government decision making. *Journal of Public Administration Research and Theory: J-PART*, *2*(7), 277–304.

Quelch, J., & Cameron, I. T. (1994). Uncertainty representation and propagation in quantified risk assessment using fuzzy sets. *Journal of Loss Prevention in the Process Industries*, *6*(7), 463–473. doi:10.1016/0950-4230(94)80004-9

Son, J., Aziz, Z., & Pena-Mora, Z. (2007). Supporting disaster response and recovery through improved situation awareness. *Structural Survey*, *5*(26), 411–425. doi:10.1108/02630800810922757

Stoklasa, J., & Štorek, J. (2008). *Disaster Medicine – first aid principles. (in Czech, study materials)*. Opava, Mathematical Institute in Opava, Silesian University in Opava.

Štorek, J. (2001). Traumatological planning model of a emergency medical rescue services centre. (in Czech). *Urgentní medicína*, 4(4), 6-8.

Štorek, J. (2004) Disaster management of the hospital care provider. (in Czech). *Urgentní medicína*, 2(7), 4 – 9.

Štorek, J. (2005) Public health and National security system – preparedness of the department to face disasters and crisis situations – documentation area. (in Czech). *Urgentní medicína*, 1(8), 4-6.

Tufekci, S. (1995). An integrated emergency management decision support system for hurricane emergencies. *Safety Science*, *1*(20), 39–48. doi:10.1016/0925-7535(94)00065-B

Wallace, W. A. & De Balogh, F. (1985). Decision support systems for disaster management. *Public Administration Review,* Special issue: Emergency Management: A Challenge for Public Administration (45), 134-146.

Yang, L., Jones, B. F., & Yang, S.-H. (2007). A fuzzy multi-objective programming for optimization of fire station locations through genetic algorithms. *European Journal of Operational Research*, *2*(181), 903. doi:10.1016/j.ejor.2006.07.003

Yang, L., Prasanna, R., & King, M. (2009). On-site information system design for emergency first responders. *Journal of Information Technology Theory and Application*, *1*(10), 5–27.

Zerger, A., & Smith, D. I. (2003). Impediments to using GIS for real-time disaster decision support. *Computers, Environment and Urban Systems*, *2*(27), 123–141. doi:10.1016/S0198-9715(01)00021-7

KEY TERMS AND DEFINITIONS

Decision Making Support: A tool to help the decision maker achieve the desired decision by eliminating possible mistakes, carrying out some difficult computations and speeding up the process.

Disaster: An unexpected and devastating event with unusual impact on health, lives and/or property of people or the environment.

Emergency Medical Rescue Services: A component of the disaster response system, whose task it is (in the Czech Republic) to provide medical care to those people that are injured during the disaster and to transport them into proper medical facilities to receive further care.

Fuzzy Number α-Degree Upper Bound: A way of comparing a fuzzy number with a real number introduced in this chapter.

Linguistic Fuzzy Modeling: A two level mathematical modeling tool, with the first level described linguistically (using linguistic terms, linguistically defined functions) intended to mediate the communication between the model and the user of the model and the second level dealing with the meanings of these linguistic terms (using fuzzy sets, fuzzy numbers and approximate reasoning), computations are carried out within the second level and the outputs then linguistically approximated – this way transferred into the first level.

Linguistic Scale: A linguistic variable with a special structure of the meanings of its linguistic terms, such that the belonging of each element of the universal set is divided completely among the fuzzy numbers representing the meanings of the linguistic terms. For each element of the universal set the sum of the degrees of membership of this element to all the fuzzy numbers is equal to one.

Uncertainty: The lack of precise or desired information.

PUBL. II

Stoklasa, J., Talašová, J. and Holeček, P., Academic Staff Performance Evaluation - Variants of Models. *Acta Polytechnica Hungarica*, 8(3), 91-111, 2011.

Available online at *http://www.uni-obuda.hu/journal/Stoklasa_Talasova_Holecek_29.pdf*.

Academic Staff Performance Evaluation – Variants of Models

Jan Stoklasa, Jana Talašová, Pavel Holeček

Department of Mathematical Analysis and Applications of Mathematics,
Faculty of Science, Palacký University
17. listopadu 1192/12, 771 46 Olomouc, Czech Republic
jan.stoklasa@upol.cz, talasova@inf.upol.cz, holecekp@inf.upol.cz

Abstract: In the paper we describe the development process of the academic staff performance evaluation model for Palacky University in Olomouc (Czech Republic). Various alternatives of the mathematical solution are discussed. All the models share the same basic idea – we evaluate the staff member's performance in the area of Pedagogical Activities and in the area of Research and Development. The input data for the models is obtained from structured forms containing information about all the activities performed by a current staff member in the respective year. We require an aggregated piece of information concerning the yearly performance of a particular staff member at a current work position (achievement of standard performance, achievement of excellence, etc.). In the first part of the paper we analyse a group of models that share the algorithm for normalized partial evaluations in both areas of interest (Pedagogical Activities, Research and Development); the partial evaluation normalization function is determined by the scores for standard and excellent performance (defined by the evaluator for different work positions and for both areas of interest separately). Models within this group differ by the aggregation operator used to calculate the overall performance evaluation – weighted arithmetic average (WA), OWA, and WOWA. The second part of the paper presents a model where partial evaluations are determined simply as multiples of standard score for the current work position and area of interest, but the aggregation of these partial evaluations is performed by a fuzzy-rule-based system. This fuzzy model is currently being implemented at Palacky University.

Keywords: evaluation; academic staff; aggregation; fuzzy model

1 Introduction

The general requirements on the model to be developed and used at Palacky University were as follows: It should a) include, if possible, every aspect of academic staff activity; b) use only easily verifiable and objective data; and c) be easy to work with. Other requirements were for the final evaluation: d) to

maximally reflect staff benefit to the Faculty; and e) not to be a simple average of partial evaluations in separate areas of activity, but to be able to appreciate excellent performance in any of the two evaluated areas (Pedagogical Activities - PA, Research and Development - RD).

The main objective of the model is to globally assess the performance and overall work load of each academic staff member in regular time intervals (annually). To achieve this, detailed information in a unified form concerning particular activities and outcomes of a particular academic staff member will be gathered. Aggregated overall evaluation information will also be available (at different levels of aggregation). As far as the aggregated evaluation is concerned, the desired output of the model was neither to arrange members of academic staff in order of their performance, nor to obtain crisp numerical evaluations interpretable only with difficulty. A rough piece of information concerning the performance of a particular academic staff member is sufficient for staff management. If more detailed information is needed, evaluations on lower levels of aggregation are available (i.e. multiples of standard score for each area of interest).

To be able to design a model with the desired properties, we studied general problems of quality assessment in high education institutions (see [1] for the Czech Republic and [2] for the EU), and fundamentals of human resource management (see [3]). At the same time we were looking for appropriate mathematical tools for these purposes (see [4, 5, 6, 7]). Various academic staff evaluation models currently used in the USA (see e.g. [8]), Canada ([9]), and Australia ([10, 11]) were subjected to a detailed analysis. Later, even the models recently designed at various Czech universities (see [12, 13, 14]) were analysed. Models of performance assessment of whole departments were also studied (see [Babak Daneshvar Rouyendegh, Serpil Erol]) as well as business models of performance assessment (see [Lívia Róka-Madarász]). The analysis concentrated on both the contents and mathematical aspects of these evaluation models and resulted in the design of several academic staff evaluation models (see [15, 16]). The models described later in the paper differ both in the manner of how members of academic staff are evaluated in separate areas of their activity and in the aggregation method for these partial evaluations (Weighted average, OWA, and WOWA operators were used; for the theory of aggregation operators see [5, 6]; we also considered fuzzy expert systems as a means of aggregation [17, 18]).

2 Preliminaries

The fundamentals of the fuzzy set theory (introduced in 1965 by Zadeh [19]) are described in detail, e.g., in [4]. Let U be a nonempty set (the universe). A fuzzy set A on U is defined by the mapping $A: U \to [0,1]$. For each $x \in U$ the value $A(x)$

is called the membership degree of the element x in the fuzzy set A and $A(\cdot)$ is called the membership function of the fuzzy set A. The height of a fuzzy set A is the real number $\mathrm{hgt}(A) = \sup_{x \in U} \{A(x)\}$. Other important concepts related to fuzzy sets are: a) the kernel of A, $\mathrm{Ker}(A) = \{x \in U \mid A(x) = 1\}$, b) the support of A, $\mathrm{Supp}(A) = \{x \in U \mid A(x) > 0\}$ and c) the α-cut of A, $A_\alpha = \{x \in U \mid A(x) \geq \alpha\}$, for $\alpha \in [0,1]$.

A function $T : [0,1]^2 \to [0,1]$ is called a triangular norm or t-norm if for all $\alpha, \beta, \gamma, \delta \in [0,1]$ it satisfies the following four properties: 1) commutativity: $T(\alpha, \beta) = T(\beta, \alpha)$, 2) associativity: $T(\alpha, T(\beta, \gamma)) = T(T(\alpha, \beta), \gamma)$, 3) monotonicity: if $\alpha \leq \gamma$, $\beta \leq \delta$, then it holds that $T(\alpha, \beta) \leq T(\gamma, \delta)$, and 4) boundary condition: $T(\alpha, 1) = \alpha$.

A function $S : [0,1]^2 \to [0,1]$ is called a triangular conorm or t-conorm if for all $\alpha, \beta, \gamma, \delta \in [0,1]$ it satisfies the properties 1) - 3) from the previous definition and 4) the boundary condition: $S(\alpha, 0) = \alpha$.

A function $N : [0,1] \to [0,1]$ satisfying conditions: a) $N(0) = 1$ and $N(1) = 0$, b) N is strictly decreasing, c) N is continuous and 4) $N(N(x)) = x$ for all $x \in [0,1]$ (N is involutive), is called a strong negation. For the purposes of this paper we consider the following strong negation: $N(x) = 1 - x$, where $x \in [0,1]$.

If $T(x, y) = N(S(N(x), N(y)))$ for all $x, y \in [0,1]$, we call S the N-dual t-conorm to T. Triangular norms and conoroms are used to define the intersection and union of fuzzy sets respectively. Let A and B be fuzzy sets on U. The intersection of A and B is a fuzzy set $(A \cap_T B)$ on U given by $(A \cap_T B)(x) = T(A(x), B(x))$ for all $x \in U$, where T is a t-norm. The union of A and B on U is a fuzzy set $(A \cup_S B)$ on U given by $(A \cup_S B)(x) = S(A(x), B(x))$ for all $x \in U$, where S is a t-conorm N-dual to T, for more details see [4]. Let A be a fuzzy set on U and B be a fuzzy set on V. Then the Cartesian product of A and B is a fuzzy set $A \times_T B$ on $U \times V$ given by $(A \times_T B)(x, y) = T(A(x), B(y))$ for all $(x, y) \in U \times V$. See [4] for more details on triangular norms and conorms. A binary fuzzy relation is any fuzzy set P on $U \times V$.

In this paper we will use the product t-norm ($T(\alpha, \beta) = \alpha \cdot \beta$, for all $\alpha, \beta \in [0,1]$) and the probabilistic sum t-conorm ($S(\alpha, \beta) = \alpha + \beta - \alpha \cdot \beta$, for all $\alpha, \beta \in [0,1]$). For the union, intersection and Cartesian product of fuzzy sets A and B based on

this t-norm and t-conorm we use the following notation: $(A \cup B)$, $(A \cap B)$ and $(A \times B)$ respectively.

Let R denote the set of all real numbers. Fuzzy set C on R is called fuzzy number if it satisfies three conditions: 1) the kernel of C, Ker(C), is a nonempty set, 2) the α-cuts of C, C_α, are closed intervals for all $\alpha \in (0,1]$, and 3) the support of C, Supp(C), is bounded. The symbol $F_N(\mathsf{R})$ denotes the family of all fuzzy numbers on R. If $\mathrm{Supp}(C) \subseteq [a,b]$, we call C a fuzzy number on the interval $[a,b]$. The family of all fuzzy numbers on the interval $[a,b]$ will be denoted by $F_N([a,b])$.

Let $A_1, A_2, ..., A_n \in F_N([a,b])$, we say that $A_1, A_2, ..., A_n$ form a fuzzy scale on $[a,b]$ if these fuzzy numbers form a Ruspini fuzzy partition (see [20, 21]) on $[a,b]$ (i.e. $\sum_{i=1}^{n} A_i(x) = 1$, for all $x \in [a,b]$) and are numbered in accordance with their ordering.

The basics of linguistic fuzzy modelling were introduced by Zadeh in [22]. A linguistic variable is the quintuple (X, $T(X)$, U, M, G) where X is the name of the linguistic variable, $T(X)$ is the set of its linguistic values (linguistic terms), U is the universe, $U = [a,b] \subseteq \mathsf{R}$, which the mathematical meanings (fuzzy numbers) of the linguistic terms are defined on, G is a syntactic rule (grammar) for generating linguistic terms from $T(X)$ and M is a semantic rule (meaning), that assigns to every linguistic term $A \in T(X)$ its meaning $A = M(\mathcal{A})$ as a fuzzy number on U. Linguistic terms and fuzzy numbers representing their meanings will be distinguished in the text by different fonts (calligraphic letters for linguistic terms and standard capital letters for their respective meanings - fuzzy numbers on U).

The linguistic variable (X, $T(X)$, U, M, G), $T(X) = \{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_s\}$, $M(\mathcal{T}_p) = T_p$, $T_p \in F_N(U)$ for $p = 1,...,s$, defines a linguistic scale on U, if the fuzzy numbers $T_1, T_2, ..., T_s$ form a fuzzy scale on U.

Let (X_j, $T(X_j)$, U_j, M_j, G_j), j=1,...,m, and (Y, $T(Y)$, V, M, G) be linguistic variables. Let $A_{ij} \in T(X_j)$ and $M_j(\mathcal{A}_{ij}) = A_{ij} \in F_N(U_j)$, $i = 1,...,n$, $j = 1,...,m$. Let $\mathcal{B}_i \in T(Y)$ and $M(\mathcal{B}_i) = B_i \in F_N(V)$, $i = 1,...,n$. Then the following scheme is called a linguistically defined function (a base of fuzzy rules, see [22]):

$$\begin{aligned}
&\text{If } X_1 \text{ is } \mathcal{A}_{11} \text{ and ... and } X_m \text{ is } \mathcal{A}_{1m} \text{ then } Y \text{ is } \mathcal{B}_1.\\
&\text{If } X_1 \text{ is } \mathcal{A}_{21} \text{ and ... and } X_m \text{ is } \mathcal{A}_{2m} \text{ then } Y \text{ is } \mathcal{B}_2. \qquad (1)\\
&\text{.....}\\
&\text{If } X_1 \text{ is } \mathcal{A}_{n1} \text{ and ... and } X_m \text{ is } \mathcal{A}_{nm} \text{ then } Y \text{ is } \mathcal{B}_n.
\end{aligned}$$

Mamdani & Assilian [17] introduced the following approach to fuzzy control. Let us consider the rule base (1). Each rule is modeled by the fuzzy relation $R_i = A_{i1} \times_T A_{i2} \times_T ... \times_T A_{im} \times_T B_i$, i =1,...,n. The whole rule base is represented by the union of all these fuzzy relations $R = \bigcup_{i=1}^{n} R_i$. Let ($a_1$, a_2, ... , a_m) be an m-tuple of crisp inputs. The output of the i-th Mamdani-Assilian fuzzy rule B_i^M is then calculated (according to [17]) as $B_i^M(y) = \min\{\min\{A_{i1}(a_1), A_{i2}(a_2), ..., A_{im}(a_m)\}, B_i(y)\}$ for all $y \in V$. The overall output of Mamdani-Assilian fuzzy controller is $B^M(y) = \max_{i=1,...,n}\{B_i^M(y)\}$ for all $y \in V$. A crisp output b^M can be then obtained using the center of gravity method: $b^M = \int_{y \in V} B^M(y) \cdot y \, dy / \int_{y \in V} B^M(y) \, dy$.

The approach of Takagi & Sugeno [23] considers a rule base in the form of (2).

$$\begin{aligned} &\text{If } X_1 \text{ is } A_{11} \text{ and ... and } X_m \text{ is } A_{1m} \text{ then } Y = g_1(X_1, ..., X_m). \\ &\text{If } X_2 \text{ is } A_{21} \text{ and ... and } X_m \text{ is } A_{2m} \text{ then } Y = g_2(X_1, ..., X_m). \\ &\qquad ..... \\ &\text{If } X_1 \text{ is } A_{n1} \text{ and ... and } X_m \text{ is } A_{nm} \text{ then } Y = g_n(X_1, ..., X_m). \end{aligned} \tag{2}$$

Here $X_1, X_2, ..., X_m$ are the input variables, $A_{i1}, A_{i2}, ..., A_{im}$ are fuzzy sets with linear membership functions that are identical to the meanings of $A_{i1}, A_{i2}, ..., A_{im}$ used in (1) for all $i = 1,...,n$ and $Y = g_i(X_1, ..., X_m)$ describes the control function for the i-th rule. Let us consider an m-tuple of crisp input values $a_1, a_2, ..., a_m$, $a_j \in U_j$, $U_j \subseteq \mathbb{R}$ is the universal set of A_{ij} for all $i = 1,...,n$ and $j = 1,...,m$. The output of Takagi & Sugeno's fuzzy controller is computed as $y^{TS} = \sum_{i=1}^{n}\left(t_i \cdot g_i(a_1, a_2, ..., a_m)\right) / \sum_{i=1}^{n} t_i$, $t_i = \min\{A_{i1}(a_1), A_{i2}(a_2), ..., A_{im}(a_m)\}$ for all $i = 1,...,n$. Sugeno's approach (see [23]) is a special case of this approach, where $Y = b_i$, $b_i \in \mathbb{R}$. If we consider Sugeno's approach, the output (control action) is determined as $y^S = \sum_{i=1}^{n}\left(t_i \cdot b_i\right) / \sum_{i=1}^{n} t_i$. Takagi & Sugeno's approach and particularly the one of Sugeno are based on practical experience with control – a control function or a control action is suggested for all fuzzy conditions. If we choose to model the Cartesian product using the same t-norm and if B_i are fuzzy singletons for all $i = 1,...,n$, Sugeno's fuzzy controller becomes a special case of Mamdani's fuzzy controller.

Using the approach to fuzzy control of Sugeno & Yasukawa [24], we assume the rule base (1) and an m-tuple of crisp input values (a_1, a_2, ... , a_m). By entering these observed values into the linguistically defined fuzzy relation, we get the output $b^S = \left(\sum_{i=1}^{n} h_i \cdot b_i\right) / \left(\sum_{i=1}^{n} h_i\right)$, where $h_i = A_{i1}(a_1) \cdot A_{i2}(a_2) \cdot ... \cdot A_{im}(a_m)$, $i = 1,...,n$.

The output of this so called qualitative model [23] is the weighted average of b_i with respect to h_i, where b_i is calculated as the center of gravity of B_i, for all $i = 1,...,n$, using the formula $b_i = \int_{y \in V} B_i(y) \cdot y \, dy / \int_{y \in V} B_i(y) \, dy$. This approach is in fact a special case of Takagi & Sugeno's approach presented in [23], where the consequent parts of the rules are modeled by constant functions. In [24] the constants b_i are real-valued characteristics of the fuzzy numbers B_i that represent the meanings of linguistic terms B_i, $i = 1,...,n$.

If we compare the previously mentioned approaches to fuzzy control, the main advantage of Mamdani's approach is that it provides information regarding the uncertainty of output values. This is important particularly when the inputs are uncertain. On the other hand, the output of Mamdani's fuzzy model is usually not a fuzzy number. To interpret the Mamdani output linguistically may prove problematic (so the center of gravity method is usually used). The asymmetry of fuzzy numbers can negatively influence the output of the defuzzification process and thus reduce the interpretation possibilities of such an output. A proper linguistic approximation may be too uncertain to provide the desired amount of information. As interpretability plays an important role in staff evaluation, we have based our evaluation model on Sugeno & Yasukawa's approach.

The approach of Sugeno & Yasukawa [24] deals with the rule base differently. The rules are defined linguistically but, for computational purposes, the fuzzy sets on the right sides of the rules are replaced by their centers of gravity and the classical Sugeno's fuzzy controller procedure is applied. Fuzzy sets B_i are then used for the interpretation of crisp outputs of this procedure. In this paper we use Sugeno & Yasukawa's approach [24] in a slightly modified form. Instead of the centers of gravity we use the elements of kernels of triangular fuzzy numbers. These triangular fuzzy numbers form a fuzzy scale on the domain of the output variable. This allows us to perform a fuzzy classification (see section 3.3 for more details). We also use the product t-norm. The approach used in our model is computationally simple and the input-output function meets all the requirements on the model (see Section 3.3).

3 Academic Staff Performance Evaluation Models

There are many reasons for staff performance evaluation. From the viewpoint of chief executives, the identification of strengths and weaknesses of staff (staff-member focus) may be important. The evaluation may serve as a basis for funds allocation and work assignment. On the other hand, the staff can also benefit from an objective evaluation tool. Such a tool can provide an academic staff member with an overview of all the work performed by him or – her – in this way the outputs of the evaluation process become a valuable document for various

purposes, i.e. future job applications and interviews. Faculty or University management can set up the evaluation function to enable staff specialisation or to encourage people to be more active in the area that is currently most important.

In the following sections we will introduce two families of academic staff evaluation models: the family of models using WA, OWA, and WOWA operators to aggregate partial evaluations and a "new" family of models – models of academic staff performance evaluation where the evaluation function is described by a fuzzy rule base.

3.1 Common Features of the Models

The performance of each member of academic staff is evaluated in both pedagogical (PA), and research and development (RD) areas of activities. Input data are acquired from a form filled in by the staff where particular activities are assigned scores according to their importance and time requirements. Three areas are taken into consideration for pedagogical performance evaluation: a) lecturing, b) the supervision of students, and c) work associated with the development of fields of study. The research and development activity evaluation is based on the methodology valid for the evaluation of R&D results in the Czech Republic (papers in important journals, books, patents, etc. are evaluated highly [25]) but other important activities (grant project management, editorial board memberships, etc.) are also included in the model.

Both pedagogical and RD areas are assigned standard scores (different for senior assistant professors, associate professors, and professors). For example, the standard score for all academic staff members in PA is 800; 40 points are assigned to the worker annually for each hour of lecturing per week and 1 point for every examined student. For RD, the standard scores default values are 14, 28, 56 for assistant professors, associate professors, and full professors respectively, where e.g. 8 points are assigned for a proceedings paper in Scopus. Standard scores can of course be modified to maximally reflect the needs of the evaluator and department. A partial evaluation of a staff member in both evaluated areas is determined using these standard scores. Such partial evaluation represents, in the simplest case, a multiple of the standard score for the current work position. The process of aggregating these partial evaluations divides the mathematical models into two groups.

3.2 The Use of WA, OWA, and WOWA for the Aggregation of Partial Evaluations

For the use of weighted average (WA), ordered weighted average (OWA), or weighted ordered weighted average (WOWA) operators to aggregate partial evaluations, we need to ensure that the values of partial evaluations are defined on

the same scale. This, however, has to be done with respect to the meanings of these partial evaluations. It is natural to determine the partial evaluations for PA and RD in terms of standard score multiples. While the evaluation in PA is based mainly on the time consumption of the activities (number of lectures, seminars, examined students), the RD area is scored according to the importance of the outcome (paper, book, invited lecture at a conference, …). The RD scores also reflect the current methodology for R&D assessment in the Czech Republic, which emphasizes excellence of the outcomes.

If the evaluation is based mainly on time consumption, the performance of a particular staff member increases more or less linearly depending on the time consumed by the activities (the increase is limited by a maximum time capacity – say two times the standard working hours). The more work he or she performs, the higher the evaluation (raising the performance twice results in an evaluation twice as high). Natural limits exist, as it is impossible to work more than 16 hours a day (for a longer period). If we base the evaluation on the current R&D assessment methodology (valid in the Czech Republic), the evaluation increases exponentially as we move towards the top journals in the particular field. In case of papers published in impacted journals, the evaluation is determined as $J_{imp} = 10 + 295 \cdot Factor$, where $Factor = (1-N)/(1+(N/0.057))$. N is the normalized ranking of the journal, $N = (P-1)/(P_{\max}-1)$, where P is the rank of the journal in the current field according to the Journal Citation Report and $P_{\max}$ is the total number of journals in the field according to the Journal Citation Report (for more details see [25]).

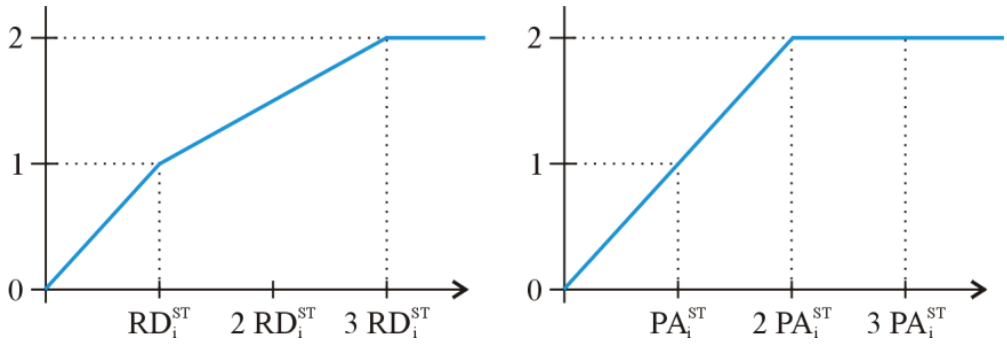


Figure 1

Research and Development partial evaluation normalization function (left) and Pedagogical Activities partial evaluation normalization function (right), both for the *i*-th work position

For example, it is possible to achieve ten times the standard score (performance) in the R&D area. To be able to aggregate the evaluations of PA and R&D, normalization is needed. We transform the evaluations using a normalization function to [0,2]. Different functions are used for PA and R&D (see Figure 1).

The normalization function for RD partial evaluations can be defined as follows (see [15]):

$$PE_i^{RD}(x_i) = \begin{cases} \dfrac{x_i}{RD_i^{ST}} & \text{for} \quad x_i \in \left[0, RD_i^{ST}\right) \\ 0.5 \cdot \dfrac{x_i}{RD_i^{ST}} + 0.5 & \text{for } x_i \in \left[RD_i^{ST}, 3 \cdot RD_i^{ST}\right) \\ 2 & \text{for} \quad x_i \geq 3 \cdot RD_i^{ST}, \end{cases} \tag{3}$$

where:

RD_i^{ST} is the standard score in Research and Development assigned to the *i*-th work position (*i*=1 for assistant professor, *i*=2 for associate professor, *i*=3 for professor);

x_i is the total score in Research and Development obtained by a current staff member in the *i*-th work position by filling in the form;

PE_i^{RD} is the normalized RD partial evaluation of a current staff member (in the *i*-th work position).

Any performance better than $3 \cdot RD_i^{ST}$ will be assigned the value 2, meaning an excellent performance. This is no problem as we do not intend to rank staff members in order of their performance (if we wanted to do so, there is still the "raw" not-normalized score available for this purpose). We have chosen this type of normalization (with normalized values from [0,2]) so that standard performance is always assigned the value 1 (in order to maintain a high level of comprehensibility for the people using these models). Our goal is not to identify the best staff member of the faculty. A rough classification of academic staff members into categories such as "close to standard", "worthy of appreciation" and, of course, the determination of "problematic" staff members is more important. If distinguishing among people evaluated as excellent is needed, it should be based on their particular outcomes and scientific achievements. From managerial point of view, having excellent people is enough and there is no need to say who is "more excellent" than others. Analogously to (3), we may define the normalization function for PA as follows: $PE_i^{PA}(x_i) = x_i / PA_i^{ST}$ for all $x_i \in \left[0,\ 2 \cdot PA_i^{ST}\right)$ and $PE_i^{PA}(x_i) = 2$ for all $x_i \geq 2 \cdot PA_i^{ST}$.

Figure 1 shows the normalization functions for RD ("excellent" means three times the standard score or better) and PA ("excellent" means two times the standard score or better) partial evaluations. We can now apply the WA, OWA, and WOWA on the normalized partial evaluations.

3.2.1 Weighted Average (WA)

Let $w_1, w_2, \ldots, w_m$ be real numbers, $w_i \geq 0$, $i = 1, 2, \ldots, m$, $\sum_{i=1}^{m} w_i = 1$. We will call $w_1, w_2, \ldots, w_m$ normalized real weights.

Let $w_1, w_2, \ldots, w_m$ be normalized real weights. Let $a_1, a_2, \ldots, a_m$ be real numbers. The mapping $WA: \mathrm{R}^m \rightarrow \mathrm{R}$ is called the Weighted Averaging operator (WA), if $WA(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{m} w_i \cdot a_i$; see [4 or 5].

In our case $w_1, w_2, \ldots, w_m$ are the weights of the areas of interest and $a_1, a_2, \ldots, a_m$ are the corresponding (normalized) partial evaluations $PE_1, PE_2, \ldots, PE_m$. This aggregation operator is fairly easy to use and compute. That is why WA is the most commonly used aggregation operator in the existing academic staff evaluation models. However, using this operator, we are unable to appreciate excellent performance and to penalize unsatisfactory performance (see Figure 2).

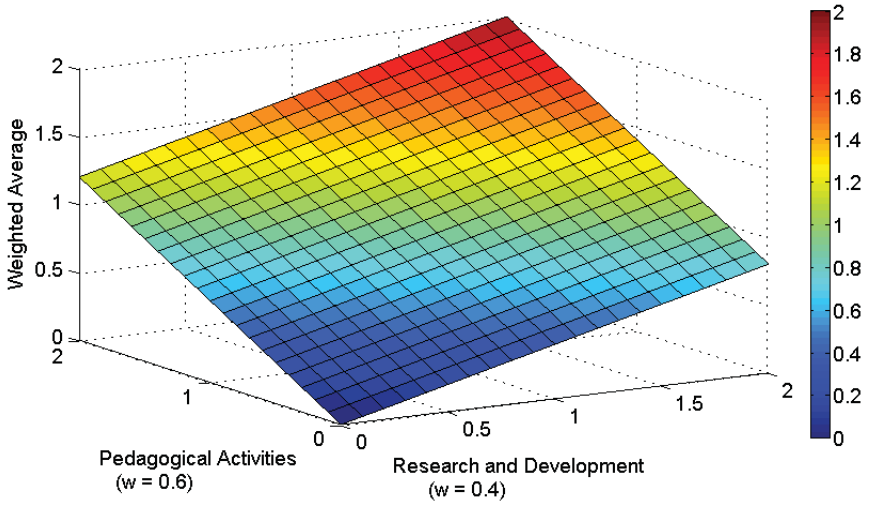


Figure 2
Weighted Average

Fixed weights for both areas of interest that are the same for all staff members do not allow us to assess people according to their focus (the area they are good in). Such an evaluation approach motivates people to concentrate on the area with greater assigned weight. (Let us say PA has the weight w=0.6 and RD has the weight w=0.4. Balanced performance represented by the standardized score of 1 for both areas results in the overall evaluation of 1. However if the normalized partial evaluation in PA is 0 and 2 in RD, the overall performance in this case is 0.8. Thus we can see that excellent performance in the activity with lower weight is unable to outweigh balanced performance (with scores 1 and 1) in both areas of activities.)

3.2.2 Ordered Weighted Average (OWA)

Let $w_1, w_2, \ldots, w_m$ be normalized real weights. Let $a_1, a_2, \ldots, a_m$ be real numbers. The mapping $OWA: \mathrm{R}^m \to \mathrm{R}$ is called the Ordered Weighted Averaging operator (OWA), if $OWA(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{m} w_i \cdot a_{\sigma(i)}$, where $\{\sigma_{(1)}, \sigma_{(2)}, \ldots, \sigma_{(m)}\}$ is a permutation of $\{1, 2, \ldots, m\}$ such that $a_{\sigma(1)} \geq a_{\sigma(2)} \geq \ldots \geq a_{\sigma(m)}$; see [6].

Again, $a_1, a_2, \ldots, a_m$ correspond to the normalized partial evaluations $PE_1, PE_2, \ldots, PE_m$ for all the areas of interest. According to the OWA definition, for any $i \in \{1, 2, ..., m\}$ w_i is the weight assigned to the *i*-th largest normalized partial evaluation. For our model it holds that $w_1 \geq w_2 \geq \ldots \geq w_m$, because we want to reflect (promote) the specialization of academic staff members. As can be easily seen (Figure 3), this approach penalizes balanced performance.

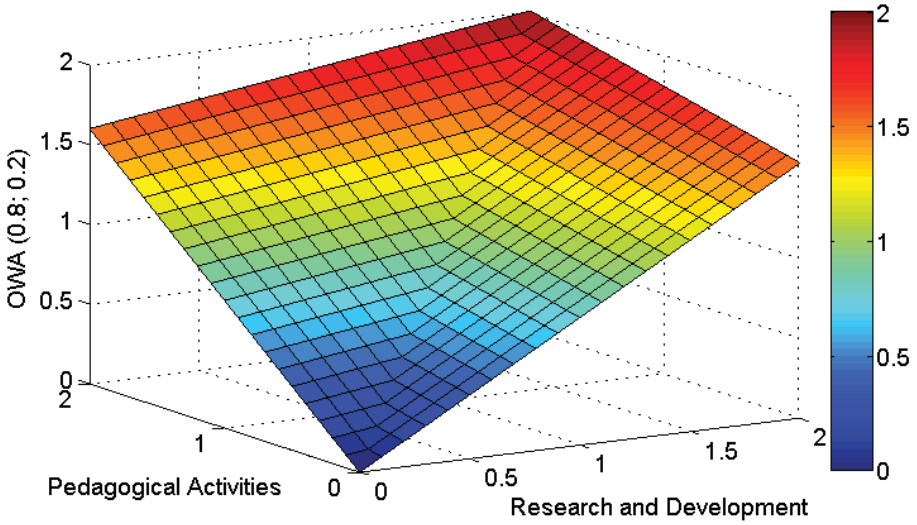


Figure 3
Ordered Weighted Average

Using this aggregation operator we motivate people to specialize but they are free to choose the area (in contrast with the WA, where only specialization in the area with greater weight, assigned by the evaluator, results in better overall evaluation). If all the staff members wished (and had the skills) to excel in RD, they could all get a good overall evaluation even if there was nobody teaching students and the university failed in one of the key areas.

3.2.3 Weighted Ordered Weighted Average (WOWA)

We can combine both previously mentioned aggregation operators into one – the Weighted Ordered Weighted Average (see Figure 4).

Let us consider two sets of normalized real weights $w_1, w_2, \ldots, w_m$ and $p_1, p_2, \ldots, p_m$. Let $a_1, a_2, \ldots, a_m$ be real numbers. The mapping $WOWA: \mathrm{R}^m \to \mathrm{R}$ is called the Weighted Ordered Weighted Averaging operator (WOWA), if $WOWA(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{m} \omega_i \cdot a_{\sigma(i)}$, where $\{\sigma_{(1)}, \sigma_{(2)}, \ldots, \sigma_{(m)}\}$ is a permutation of $\{1, 2, \ldots, m\}$ such that $a_{\sigma(1)} \geq a_{\sigma(2)} \geq \ldots \geq a_{\sigma(m)}$ and ω_i are defined as $\omega_i = w^*\left(\sum_{j \leq i} p_{\sigma(j)}\right) - w^*\left(\sum_{j<i} p_{\sigma(j)}\right)$ with w^* being a nondecreasing function that interpolates the points $\left\{\left(i/m, \sum_{j \leq i} w_j\right)\right\}, i = 1, 2, \ldots, m$, together with the point (0,0); see [26].

Using this approach we have two sets of weights available – OWA weights to reflect staff specialisation (again we use $w_1 \geq w_2 \geq \ldots \geq w_m$ to appreciate staff specialization) and fixed WA weights $p_1, p_2, \ldots, p_m$ assigned to the areas of interest according to their importance for the success of the university or faculty. Such aggregation of partial evaluations, however, proves to be too complicated to be understood by the people using the model (executives, heads of departments etc.) and by the academic staff members as well.

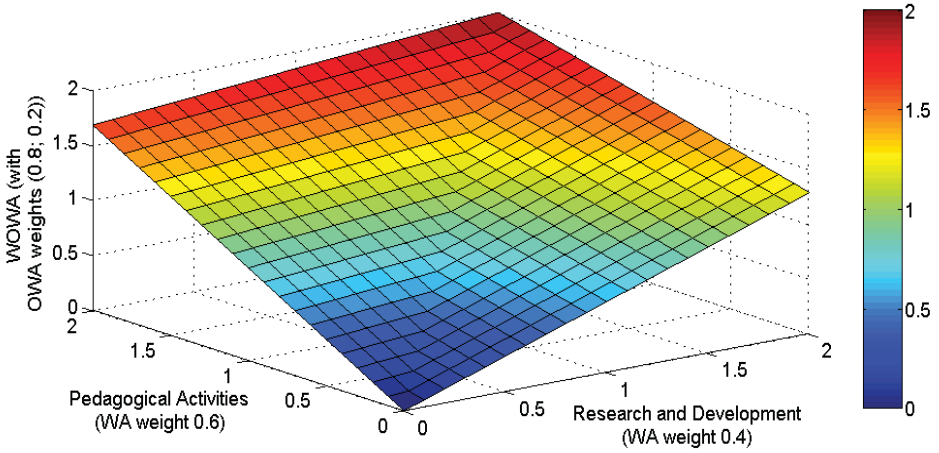


Figure 4
Weighted Ordered Weighted Average

Models using WOWA appear "unpredictable" to practitioners as they transform two sets of weights into one, the values of which sometimes surprise the user of the model – we may say that it is not considered "intuitive enough" by the evaluators. The penalization of balanced performance is not removed as well.

3.3 Aggregation of Partial Evaluations by Means of a Fuzzy-Rule-Base System (FRBS)

In order to avoid the penalization of balanced performance, as well as to be able to appreciate excellence on one hand and penalize unsatisfactory performance on the other, a model based on fuzzy linguistic modelling was developed. Another asset of the approach that will be mentioned later in this paper is its comprehensibility, as all the relations between inputs and outputs are described linguistically.

Let us assume that we have available the partial evaluations of PA and RD in terms of multiples of standard scores (for the particular area of interest and work position). Using the tools of linguistic fuzzy modelling, we can now construct a user/evaluator based model – first in a purely linguistic form. Then we assign proper mathematical objects and methods whenever needed using the following algorithm:

1) We define the set of linguistic terms for the following linguistic variables

   - *PA* (input1): *T(PA)* = {V*ery_Low*, *Low*, *Standard*, *High*, *Extreme*},
   - *RD* (input2): *T(RD)* = {V*ery_Low*, *Low*, *Standard*, *High*, *Extreme*},
   - *Overall* (output): *T(Overall)* = {*Unsatisfactory*, *Substandard*, *Standard*, *Very_Good*, *Excellent*}.

   T(PA), *T(RD)* and *T(Overall)* are naturally ordered according to the meanings of the linguistic terms.

2) We define the expected (linguistic) output for each combination of input values (linguistic), thus forming a linguistic rule base containing *k* rules (25 in our case), such as:

   …

   If *PA* is *Standard* and *RD* is *Standard* then *Overall* is *Standard*.

   If *PA* is *Standard* and *RD* is *High* then *Overall* is *Very_Good*.

   If *PA* is *High* and *RD* is *Standard* then *Overall* is *Very_Good*.

   If *PA* is *High* and *RD* is *High* then *Overall* is *Excellent*.

   …

3) Now we need to specify both input variables regarding the mathematical level of description of their values. As both inputs are mathematically expressed in terms of standard score multiples, the domains for *PA* and *RD* are [0,BB] and [0,CC] respectively, where BB and CC are sufficiently high real numbers not to be exceeded by any actual *PA* and *RD* partial evaluation respectively.

We define the "most typical" real value of the partial evaluation (in terms of standard score multiples) for each linguistic term of all the inputs defined in step 1):

- most typical values for *PA* linguistic terms: {0, 0.5, 1, 1.5, 2};
- most typical values for *RD* linguistic terms: {0, 0.5, 1, 2, 3}.

For the output linguistic variable *Overall* we define the universe to be [0,2]. We need to define the most typical values of its linguistic terms as well. These values serve here as category labels. We may see the evaluation process as a classification problem. The information that a staff member is *Unsatisfactory* in the degree of 0, *Substandard* in the degree of 0, *Standard* in the degree of 0.4, *Very_Good* in the degree of 0.6 and *Excellent* in the degree of 0 is sufficient. We need to perform a fuzzy classification. To achieve this we assign the key output linguistic terms the values of an ordinal scale: 0 for *Unsatisfactory*, 1 for *Standard*, and 2 for *Excellent*. Meanings of the remaining two linguistic terms are 0.5 for *Substandard* and 1.5 for *Very_Good*.

- Most typical values for *Overall* linguistic terms: {0, 0.5, 1, 1.5, 2}

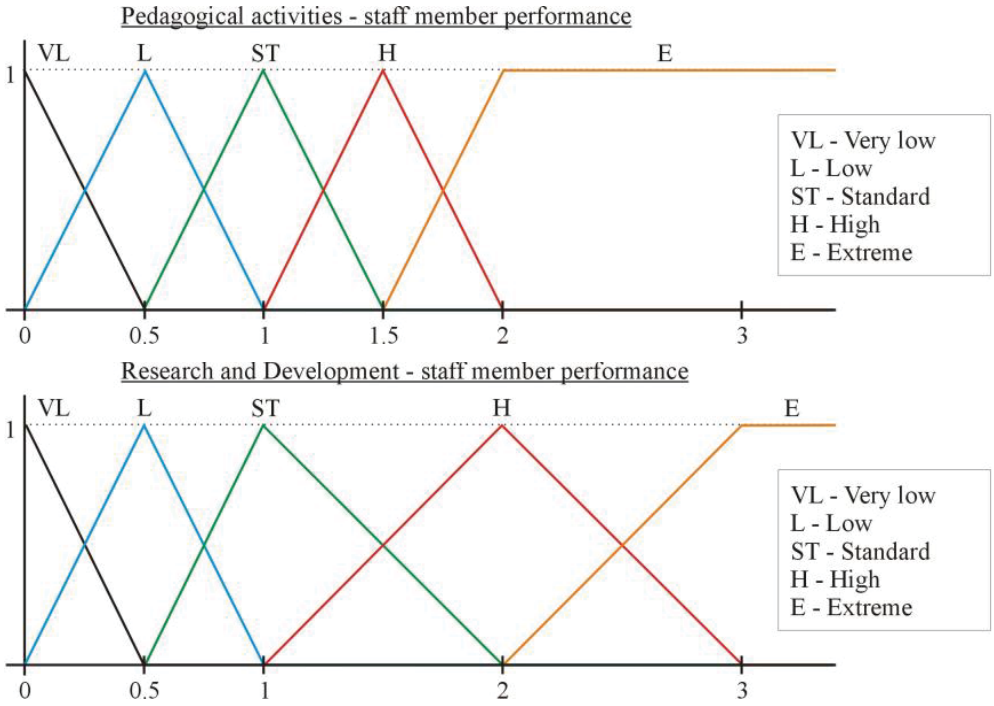


Figure 5
Linguistic scales

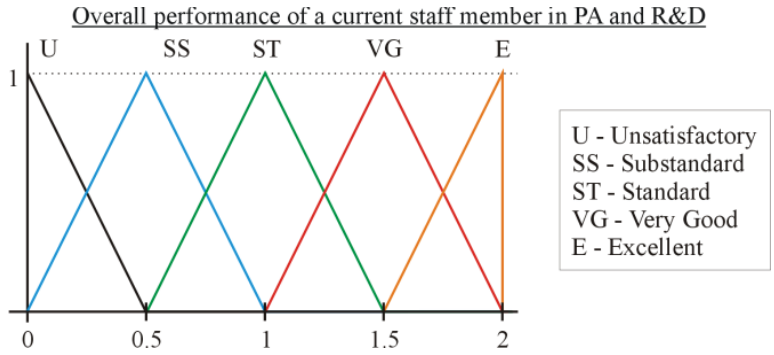


Figure 6
Fuzzy scale describing the overall performance in PA and RD of a current staff member

4) For the input variables *PA* and *RD* and for the output variable *Overall* we construct (on the respective universes) fuzzy scales using the already defined linguistic terms. The "most typical" values lie in the kernels of these fuzzy numbers (Figures 5 and 6). This way we get

- $(PA, T(PA) = \{Very\_Low, Low, Standard, High, Extreme\}, [0, BB], M_{PA})$
- $(RD, T(RD) = \{Very\_Low, Low, Standard, High, Extreme\}, [0, CC], M_{RD})$
- $(Overall, T(Overall) = \{Unsatisfactory, Substandard, Standard, Very\_Good, Excellent\}, [0, 2], M_{Overall})$.

The definition of $M_{PA}(Extreme)$ and $M_{RD}(Extreme)$ corresponds with the normalization process described previously (see Figure 1).

5) For any pair of real inputs $pa \in [0,BB]$ and $rd \in [0,CC]$ we can now compute the output (real) value

$$eval(pa, rd) = \frac{\sum_{j=1}^{k} A_j(pa) \cdot B_j(rd) \cdot ev_j}{\sum_{j=1}^{k} A_j(pa) \cdot B_j(rd)} = \sum_{j=1}^{k} A_j(pa) \cdot B_j(rd) \cdot ev_j \,, \qquad (4)$$

where

- *Aj* is the fuzzy number representing the meaning of the linguistic term describing *PA* in rule *j*, *j=1*, ..., *k;*
- B_j is the fuzzy number representing the meaning of the linguistic term describing *RD* in rule *j*, *j=1*, ..., *k;*
- ev_j is the real number representing the most typical value of the linguistic term describing the Overall in rule *j* , *j*=1, ..., *k*; ev_j lies in the kernel of the respective triangular fuzzy number.

As we are using linguistic scales and have only two crisp inputs, no more than 4 rules can be called for at the same time. It is easy to prove that $\sum_{j=1}^{k} A_j(pa) \cdot B_j(rd) = 1$. Let $A_1(pa) = a \neq 0$ and $B_1(rd) = b \neq 0$, $a, b \in [0,1]$, which means that the truth value of this rule is $a \cdot b$. We can find no more than three other rules with non zero truth values, namely: $(1-a) \cdot b$, $a \cdot (1-b)$ and $(1-a) \cdot (1-b)$. The sum of these four truth values is equal to 1.

Formula (4) interpolates the overall evaluation function *eval*(*pa, rd*) defined by a finite amount of known values (25 in this case) as shown in Figure 7. The result is a piece-wise bilinear function. Moreover for all $x_1 \leq x_2$, $x_1, x_2 \in [0, \mathrm{BB}]$, and $y_1 \leq y_2$, $y_1, y_2 \in [0, \mathrm{CC}]$, it holds that $eval(x_1, y_1) \leq eval(x_2, y_2)$. As we have assured that *eval* is nondecreasing in both arguments for the 25 typical combinations of values (defined in steps 2 and 3), the interpolated function is nondecreasing in both arguments as well.

To linguistically interpret the crisp output *eval* of step 5, we use the linguistic fuzzy scale *Overall.* The output can now be interpreted in terms of membership degrees to the fuzzy numbers that represent the meanings of linguistic terms from *T*(*Overall)*. For example, the overall evaluation 1.2 will be interpreted as 0.6 "Standard" and 0.4 "Very_Good". This way the fuzzy classification is complete. The result of the algorithm is a description of a current staff member's performance that uses the predefined five linguistic terms (labels of the categories) and specifies the membership degree of the staff member to each category. Such description is easy to understand and still provides a valuable piece of information.

The linguistic rule base constructed in step 2 describes the aggregation of PA and RD partial evaluations much more transparently than all the previously mentioned models (particularly for laymen). By the use of linguistic fuzzy modelling we have constructed an evaluation tool that is easy to understand, easy to use and even easy to modify for various purposes. Due to the chosen approximate reasoning mechanism, it is computationally undemanding as well. Figure 7 shows the shape of the aggregation function described by the fuzzy rule base. It meets all the requirements concerning excellence appreciation and unsatisfactory performance penalization mentioned in the introduction. The outputs are available as real numbers as well as their linguistic descriptions.

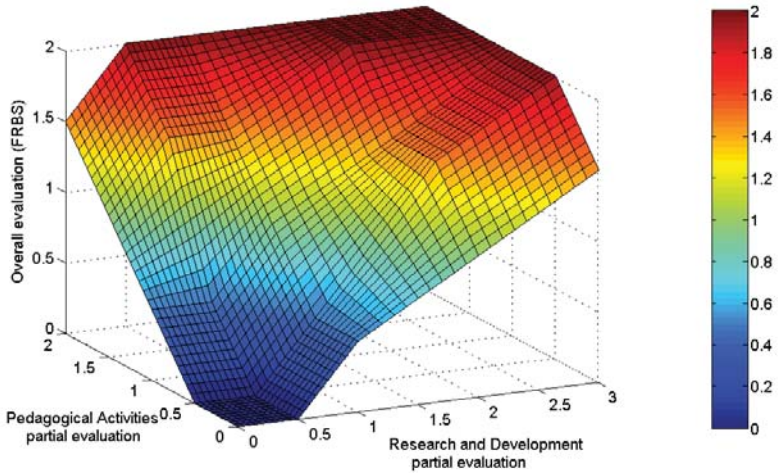


Figure 7
The shape of the linguistically defined aggregation function for PA and RD partial evaluations

3.4 Numerical Example

Let us consider six academic staff members (SM_1,…,SM_6). For each of them we have the partial evaluations in terms of multiples of the appropriate standard scores (see Tab. 1). We calculate the normalized partial evaluations as described earlier in the paper (setting excellence at three times the standard score for RD and twice the standard score for PA). All these normalized partial evaluations lie in [0,2], where 1 corresponds to a standard performance and 2 to an excellent performance. To aggregate these partial evaluations, WA, OWA, WOWA, and the fuzzy-rule-base model introduced in this paper were applied.

Staff member 1, who is standard in both areas, is always evaluated as standard regardless of the aggregation method used. Using the WA, SM_2 and SM_6 are evaluated worse than the "standard" SM_1, even though they show excellent performance in RD. By comparing the WA evaluation of SM_1, SM_2 and SM_6, it is obvious that specialization in RD is discouraged, as excellence in RD is unable to outweigh the low performance in PA. Due to the fixed weights, the use of WA can result in classifying people excellent in one of the areas of interest as standard or worse.

This is not the case with the OWA operator, which is able to reflect and appreciate staff members' specialization, as Tab. 1 clearly illustrates. However, there is no way for the executives to influence the area of specialisation of their staff (by the use of the evaluation model). The WOWA operator solves even this problem but the results of combining two sets of weights defined by the evaluator are not well accepted by laymen. If the partial evaluation in the area with the larger fixed weight is larger than evaluation in the other area (SM_3, SM_5), the resulting aggregated evaluation is larger than those obtained by the use of WA and OWA.

The use of the fuzzy-rule-base evaluation model described in this paper results in a linguistic description of each staff member's performance. The numerical value of the function *eval* that results from step 5 in Section 3.3 is also available to the evaluator (its values are 1 for SM_1, 1.8 for SM_2, 1.5 for SM_3, 0.4 for SM_4, 1.6 for SM_5, 1.6 for SM_6). Results provided by the fuzzy-rule-base model (the fuzzy classification of a staff member according to his/her performance in PA and RD) are easy to understand and need no further explanation. The evaluation process is described linguistically, and therefore even staff members themselves can see how the evaluation works.

Conclusions

We have described several mathematical tools that can be used in academic staff performance evaluation for the aggregation of partial evaluations. Having identified the weak spots of the previously discussed aggregation operators, we have suggested a new model that is based on fuzzy-rule-base systems. The main advantage of the proposed model is that it is easy to understand, easy to use and easy to modify to meet the specific requirements of the evaluator. Outputs (evaluations) are available on different levels of aggregation, thus giving an overall picture of a staff member's performance in a graphical form with linguistic labels, as well as detailed information concerning the performance in all the areas relevant for evaluation. This makes the proposed model, which is currently being implemented at Palacky University, a multipurpose performance assessment tool.

The developed performance evaluation system is beneficial to academic staff members as well – it serves as a record of their activities for their own needs. It provides feedback on their performance (and how the employer sees this performance). Aggregated information available in an easy to understand form is an important management tool for the executives, namely the heads of departments. The long-term use of the model offers the opportunity to observe the dynamics of staff member performance over time, which can be seen as another valuable asset of our model.

Acknowledgement

The research was supported by the grant PrF_2010_08 of the Internal Grant Agency of Palacky University in Olomouc.

References

[1] Chvátalová, A., Kohoutek, J., Šebková, H. (eds.): Quality Assurance in Czech Higher Education (in Czech). Aleš Čeněk, Plzeň 2008, ISBN 978-80-7380-154-0

[2] European Association for Quality Assurance in Higher Education [online] <http://www.enqa.eu/>

Table 1
Overview of the results for WA, OWA, WOWA, and FRBS aggregation of partial evaluations

		Staff member 1		Staff member 2		Staff member 3		Staff member 4		Staff member 5		Staff member 6	
		PA	RD	PA	RD	PA	RD	PA	RD	PA	RD	PA	RD
Standard score multiples		1	1	0.3	3	1.5	0.8	0.6	0.8	1.5	1.2	0.1	5
Normalized partial evaluations		1	1	0.3	2	1.5	0.8	0.6	0.8	1.5	1.1	0.1	2
WA	$w_{PA}=0.6$ $w_{RD}=0.4$	1		0.98		1.22		0.68		1.34		0.86	
OWA	$w_{max}=0.6$ $w_{min}=0.4$	1		1.32		1.22		0.72		1.34		1.24	
OWA	$w_{max}=0.8$ $w_{min}=0.2$	1		1.66		1.36		0.76		1.42		1.62	
WOWA	$w_{PA}=0.6$ $w_{RD}=0.4$ $w_{max}=0.8$ $w_{min}=0.2$	1		1.39		1.39		0.73		1.44		1.32	
Proposed rule based model		Standard…1		Very Good…0.4 Excellent…0.6		Very Good…1		Unsatisfactory…0.2 Substandard…0.8		Very Good…0.8 Excellent…0.2		Very Good…0.8 Excellent…0.2	

[3] Matheson, W., Van Dyk, C., Millar, K. I.: Performance Evaluation in the Human Services. The Haworth Press, New York-London, 1995. ISBN 1-56024-379-1

[4] Dubois, D., Prade, H. (Eds.): Fundamentals of Fuzzy Sets. The Handbook of Fuzzy Sets Series. Kluwer Academic Publishers, Boston-London-Dordrecht. 2000. ISBN 0-7923-7732-X

[5] Torra, V., Narukawa, Y.: Modeling Decisions. Springer, Heidelberg, 2007. ISBN 978-3-540-68789-4

[6] Yager, R. R.: On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making. *IEEE Trans. On Systems, Man and Cyberneics* 3 (1) 1988, pp. 183-190

[7] Talašová, J.: Fuzzy Methods of Multiple Criteria Evaluation and Decision Making (in Czech) Palacky University, Olomouc, 2003, ISBN 80-244-0614-4

[8] 2009-10 Guidelines for Evaluation of Academic Staff [online] Wayne State University [cited 30. 5. 2010] <http://www.aaupaft.org/pdf/AcStaffguidelines_2009-10.pdf>

[9] Academic Performance Evaluation [online] c2010, last revision May 15, 2010, McGill University [cited 30. 5. 2010] <http://www.mcgill.ca/medicine-academic/performance/>

[10] Performance Management [online] c2009, University of Technology Sydney [cited 30. 5. 2010] <http://www.hru.uts.edu.au/performance /reviewing/rating.html>

[11] Performance Management [online] Flinders University [cited 30. 5. 2010] <http://www.flinders.edu.au/ppmanual/review.html>

[12] Determination of Criteria for Pedagogical and Other Activities Evaluation (in Czech) [online] Brno, Masaryk University, Faculty of Law [cited 30. 5. 2010]. <http://www.law.muni.cz/dokumenty/7601>

[13] Academic Staff Evaluation Criteria for Personal Extra Pay Distribution (in Czech) [online] Ústí nad Labem, Jana Evangelista Purkyne University, Faculty of Environment [cited 30. 5. 2010] <http://fzp.ujep.cz/dokumenty/kritosoh.pdf>

[14] Pedagogical and Creative Activities Evaluation (in Czech) [online] Zlín, Tomas Bata University in Zlín, Faculty of Applied Informatics [cited 30. 5. 2010] <http://web.fai.utb.cz/cs/docs/SD_09_09.pdf>

[15] Talašová, J., Pavlačka, O.: Academic Staff Evaluation Model Design for the Faculty of Science, Palacky University in Olomouc (in Czech) Research report. Faculty of Science, Palacky University, Olomouc 2006

[16] Talašová, J., Stoklasa, J., Pavlačka, O., Holeček, P.: New Academic Staff Evaluation Model Design for the Faculty of Science, Palacky University in Olomouc (in Czech) Research report. Faculty of Science, Palacky University, Olomouc 2009

[17] Mamdani, E. H., Assilian, S.: An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, *Int. J. Man-Machine Studies*, Vol. 7, 1975, pp. 1-13

[18] Sugeno, M.: An Introductory Survey on Fuzzy Control. *Information Sciences*, 36, 1985, pp. 59-83

[19] Zadeh, L. A.: Fuzzy Sets. *Inform. Control*, 8, 1965, pp. 338-353

[20] Ruspini, E.: A New Approach to Clustering. *Inform. Control*, 15, 1969, pp. 22-32

[21] Codara, P., D'Antona, O. M., Marra, V.: An Analysis of Ruspini Partitions in Gödel Logic. *International Journal of Approximate Reasoning*, 50, 2009, pp. 825-836

[22] Zadeh, L. A.: The Concept of Linguistic Variable and its Application to Approximate Reasoning. *Information Sciences*, Part 1: 8, 1975, pp. 199-249, Part 2: 8 1975, pp. 301-357, Part 3: 9 1975, pp. 43-80

[23] Takagi, T., Sugeno, M.: Fuzzy Identification of Systems and its Application to Modeling and Control. *IEEE Transactions on Systems, Man and Cybernetics*, 1 (15), 1985, pp. 116-132

[24] Sugeno, M., Yasukawa, T.: A Fuzzy-Logic-based Approach to Qualitative Modeling. *IEEE Transactions on Fuzzy Systems*, 1 (1), 1993, pp. 7-31

[25] Methodology for Research and Development Outcomes Evaluation (in Czech) [online] Research and Development in the Czech Republic [cited 30.5. 2010], http://www.vyzkum.cz/storage/att/CDDC542199F1640B59A7D1E84 1B7151C/Metodika%202009_schv%c3%a1leno.pdf

[26] Torra, V.: The Weighted OWA Operator. International Journal of Intelligent Systems. 2 (12), 1997, pp. 153-166

[27] Babak Daneshvar Rouyendegh, Serpil Erol: The DEA – FUZZY ANP Department Ranking Model Applied in Iran Amirkabir University, in Acta Polytechnica Hungarica, Vol. 7, No. 4, 2010, pp. 103-114

[28] Lívia Róka-Madarász: Performance Measurement for Maintenance Management of Real Estate, in Acta Polytechnica Hungarica, Vol. 8, No. 1, 2011, pp. 161-172

PUBL. III

Stoklasa, J., Jandová, V. and Talašová, J., Weak consistency in Saaty's AHP - evaluating creative work outcomes of Czech Art Colleges. *Neural Network World*, 23(1), 61-77, 2013.

WEAK CONSISTENCY IN SAATY'S AHP – EVALUATING CREATIVE WORK OUTCOMES OF CZECH ART COLLEGES

*Jan Stoklasa, Věra Jandová, Jana Talašová**

Abstract: The full consistency of Saaty's matrix of preference intensities used in AHP is practically unachievable for a large number of objects being compared. There are many procedures and methods published in the literature that describe how to assess whether Saaty's matrix is "consistent enough". Consistency is in these cases measured for an already defined matrix (i.e. ex-post). In this paper we present a procedure that guarantees that an acceptable level of consistency of expert information concerning preferences will be achieved. The proposed method is based on dividing the process of inputting Saaty's matrix into two steps. First, the ordering of the compared objects with respect to their significance is determined using the pairwise comparison method. Second, the intensities of preferences are defined for the objects numbered in accordance with their ordering (resulting from the first step). In this paper the weak consistency of Saaty's matrix is defined, which is easy to check during the process of inputting the preference intensities. Several propositions concerning the properties of weakly consistent Saaty's matrices are proven in the paper. We show on an example that the weak consistency, which represents a very natural requirement on Saaty's matrix of preference intensities, is not achieved for some matrices, which are considered "consistent enough" according to the criteria published in the literature.
The proposed method of setting Saaty's matrix of preference intensities was used in the model for determining scores for particular categories of artistic production, which is an integral part of the Registry of Artistic Results (RUV) currently being developed in the Czech Republic. The Registry contains data on works of art originating from creative activities of Czech art colleges and faculties. Based on the total scores achieved by these institutions, a part of the state budget subsidy is being allocated among them.

Key words: *Multiple criteria evaluation, AHP, weak consistency, work of art, Registry of Artistic Results*

Received: July 15, 2012
Revised and accepted: November 1, 2012

*Jan Stoklasa, Věra Jandová, Jana Talašová
Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacky University in Olomouc, E-mail: jan.stoklasa@upol.cz, vera.jandova01@upol.cz, jana.talasova@upol.cz

1. Introduction

When designing mathematical models for such purposes as evaluation of works of art, the evaluators' experience and background needs to be taken into account. Suitable mathematical tools have to be chosen so that the resulting model is not only mathematically sound, but also possible to implement in real setting. Particularly when dealing with abstract categories and large amounts of pairwise comparisons and experts not closely related to the field of mathematics, it is important to find an appropriate way of inputting the data. During each step we might need to go back and correct some partial inconsistencies, but the result should be a reasonably consistent mathematical representation of experts' knowledge concerning their preferencies on the given set of objects. Tools enabling the evaluators to check the consistency of inputted preferences for pairs of objects (even during the process of data input) and guidelines for such purposes can be the key to success in such application areas. We are going to present here a real-life problem and our solution to it. The task was to develop a mathematical model for the evaluation of works of art, which required cooperation with experts from the field of artistic production.

In the second section of the paper, we start with the description of the Registry of Artistic Results (RUV – from Czech "Registr Uměleckých Výstupů"), its purpose and structure. We also introduce the evaluation criteria and the resulting categories of works of art in this section. Section 3 describes the two-step mathematical model used to obtain scores for each category, and the respective evaluation methodology. As consistency is a great issue when using Saaty's matrices of large dimensions, Section 4 discusses various measures of inconsistency in Saaty's matrix and presents a short overview of the relevant research. We introduce here a new concept of weak consistency and prove several properties of weakly consistent Saaty's matrices. Section 5 presents two methods for determining the scores of categories of works of art by Saaty's matrix of preference intensities – the eigenvector method and the logarithmic least squares method. We discuss here the possibility of seeing the data in Saaty's matrix as repeated measurements of relative information concerning the importances in the set of categories of works of art and hence dealing with it as with compositional data. All the results presented in this paper are then summarized and discussed in Section 6.

2. Classification of Works of Art

The Registry of Artistic Results has been developed and is currently being pilot tested in the Czech Republic. It contains information on works of art originating from creative activities of art colleges and faculties (see [14]). The RUV is conceived as an analogy to the register of R&D outcomes where information on outcomes of research institutions (including universities) has been collected for some years already. In both registers the outcomes are stored under several categories. These categories are assigned scores. The sum of scores of all the outcomes of a given university is considered an indicator of its performance in the area of creative activity. These numerical values can then be used in decisions regarding one part of the total money to be allocated among universities from the state budget. The

structure of the evaluated categories of works of art used in the Registry of Artistic Results was inspired, to some extent, by the artistic categories in the Slovak Republic (see [12]). However, the mathematical model used to determine scores for each category in Slovakia is quite different.

For the purposes of registration of works of art originating from creative activities of the Czech art colleges and faculties, the whole area of artistic production is divided into seven fields: architecture, design, film, fine arts, literature, music and theatre. Each piece of art, regardless of the field, is categorized according to the following three criteria:

- *Relevance or significance of the piece*;
- *Extent of the piece*;
- *Institutional and media reception/impact of the piece.*

In each criterion, three different levels are distinguished (denoted by capital letters for easier handling):

- The criterion *Relevance or significance of the piece*:

  A – a new piece of art or a performance of crucial significance;

  B – a new piece of art or a performance containing numerous important innovations;

  C – a new piece of art or a performance pushing forward modern trends.

- The criterion *Extent of the piece*:

  K – a piece of art or a performance of large extent;

  L – a piece of art or a performance of medium extent;

  M – a piece of art or a performance of limited extent.

- The criterion *Institutional and media reception/impact of the piece*:

  X – international reception/impact;

  Y – national reception/impact;

  Z – regional reception/impact.

The resulting category for a piece of art is given by a combination of three capital letters – e.g. AKX, BKY, or CLZ. There are 27 categories altogether that are assigned a score. The decision concerning the relevance or significance of the piece (choice of A, B or C) rests upon expert assessment; the experts have at their disposal general definitions of each category and specific real-life (historical) examples of works of art in each category for all 7 fields of artistic production and these examples assist them in the decision process. (Gathering real-life representatives of all the categories for all the fields of arts was also important to confirm a common understanding of the categories and to ensure that corresponding categories are really comparable in terms of evaluation across all the fields of arts.) As for the extent of the piece (levels K, L, M), all the classes are clearly specified for

all the fields of art. As for the institutional and media reception/impact, lists of institutions corresponding to categories X, Y, Z are available for all fields.

Our task was to develop a mathematical model to determine the scores for such categories of pieces of art (each described by a triplet of capital letters). We have decided to solve this problem by applying Saaty's AHP method (introduced in [10]). We need to realize that there are interactions among the three mentioned criteria. For example the first one (expertly defined *Relevance or significance of the piece of art*) and the third one (*Institutional and media reception/impact of the piece*) partly overlap. It was, therefore, not possible to use the approach, where first we would determine the weights of the criteria and scores for their individual levels, and then set the scores of categories as respective weighted averages. We did not choose the ANP method either (see [11] for more details) which is able to solve tasks with mutually dependent criteria. It was because deriving information concerning the links among the criteria has proven to be extremely difficult for the experts in the field of arts. We have decided to compare directly the 27 works of art categories. In the case of such a large number of objects (categories), Saaty suggests (see [8]) to split the problem into several smaller ones and then apply the AHP on these. If we chose to do so, we would have to define relative significances for abstract supercategories. This is a difficult task for the experts from the field of arts. The difficulties resulting from a large dimension of the matrix of preference intensities were considered small compared to the difficulties resulting from the use of other ways of solving the problem. For a large number of mutually compared objects the issue of obtaining a Saaty's matrix that is consistent enough arises. Our solution to this problem will be described and further discussed in the following sections.

3. Determining Scores for Particular Categories of Artistic Production

Saaty's method (see [8, 6, 9]) served as a basis for determination of scores for all 27 categories of artistic production. No matter how obvious it was that this mathematical tool is the most appropriate for this task, certain challenges concerning its use were also clearly apparent: 1) difficulty for a team of experts to express preferences with respect to abstract categories; 2) reaching consensus within the group of experts (professional guarantors of particular fields of art); and most importantly 3) difficulty to reach acceptable consistency of Saaty's matrix for such a large number of categories (Section 4 deals with this issue). Admittedly, expressing one's opinion on intensities of preferences with respect to abstract categories is difficult. Experts – professional guarantors of particular artistic fields – were first asked to provide examples of works of art in all categories in their field (see Section 2). Next, professional guarantors of each field of art set their preferences concerning pairs of categories, while using the representatives (examples) as an aid in their decision making. Although it would be possible for each of these experts to express their preferences separately, and only then to derive the collective preferences (from the individual ones), we used a different approach. The collective preferences were set directly at a team meeting of experts. The reason was that

art-college experts are not used to work with mathematical models and individual inputting of required data could prove difficult for them. Achieving consensus was also intentionally preferred over averaging different opinions.
Great effort was made to find the best way of converting expert preferences concerning the 27 categories of artistic production (represented in each field of art by specific examples) into a mathematical model. Such model is required to be a consistent representation of experts' preferences and to allow calculating the scores of all the categories of works of art. To facilitate the process of inputting required data by the experts, to achieve the necessary consistency of this input and to obtain consensual scores for all the categories of works of art, the following two-step procedure was performed. First, a pairwise comparison method was used to determine the order of importance of the 27 categories (their quasi-ordering). Second, a Saaty's matrix was constructed with categories numbered according to this quasi-ordering. Such matrix of preference intensities was then used to determine scores for the categories.

3.1 Matrix of preferences and indifferences

In the first step, we have determined the order of importance of the categories by the Pairwise Comparison Method (see [6, 13]). This method employs a matrix of preferences and indifferences $P = \{p_{ij}\}_{i,j=1,\dots,27}$. For its elements it holds that:

$p_{ij} = 1$, if the i^{th} category is more important than the j^{th} category;

$p_{ij} = 0.5$, if the i^{th} category is equally important as the j^{th} category;

$p_{ij} = 0$, if the j^{th} category is more important than the i^{th} category.

It is sufficient for the experts to fill in the upper right triangle of the matrix, that is, the elements p_{ij}, $i < j$, as $p_{ii} = 0.5$ and $p_{ji} = 1 - p_{ij}$. The row sums $R_i = \sum_{j=1}^{27} p_{ij}$, $i = 1, \dots, 27$, are used in this method to determine the order of the mutually compared objects according to their significance. To be able to accept the results of this method, we need to be sure that the matrix P defined by experts contains consistent information on their preferences on the set of objects. The matrix P, therefore, has to represent a quasi-ordering of objects, i.e. a complete and transitive relation (a relation that can be described as a linear ordering of classes of indifferent objects). The completeness of this relation is ensured by the process of inputting of matrix P ($p_{ji} = 1 - p_{ij}$); the transitivity in the terms of matrix P can be expressed by the following condition:

$$p_{ik} = \max\{p_{ij}, p_{jk}\}, \text{ for all } p_{ij}, p_{jk} \geq 0.5, \ i, j, k = 1, \dots, 27. \tag{1}$$

If the matrix does not satisfy the condition (1), we make the minimum amount of changes necessary for it to become so. These changes are then consulted with the team of experts and if they are approved of, we can proceed. All the changes actually made while solving our problem are summarized in Fig. 1.

3.2 Saaty's matrix of preference intensities

Saaty's matrix of preference intensities for n mutually compared objects is a square matrix $S = \{s_{ij}\}_{i,j=1}^{n}$, that is reciprocal (i.e. $s_{ij} = 1/s_{ji}$ for all $i, j = 1, 2, \dots, n$)

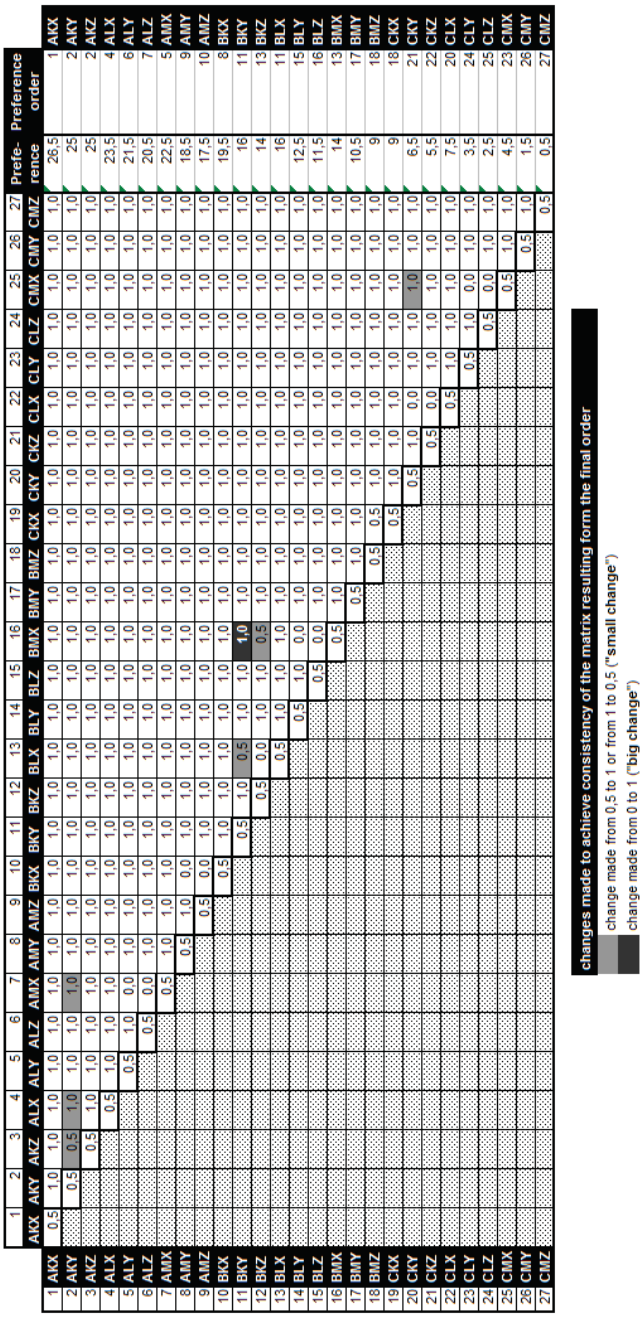


Fig. 1 *Pairwise comparison matrix for 27 categories. Necessary changes are highlighted.*

and for an object i that is more or equally preferred than object j the element $s_{ij} \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Tab. I provides the linguistic descriptors of these numerical values for Saaty's scale. If, for example, $s_{ij} = 3$, it can be interpreted that the object i is 3 times as important as the object j. From Perron-Frobenius theorem it follows that Saaty' matrix always has a maximum eigenvalue (spectral radius – see [6]). A fully consistent Saaty's matrix has a single nonzero eigenvalue, which is equal to the order of the matrix.

In the second step of our method, Saaty's matrix of preference intensities $S = \{s_{ij}\}_{i,j=1}^{27}$ was constructed for categories numbered in ascending order according to their significance determined in the previous step. Again, it was in this case sufficient to fill in the upper right triangle of the matrix S, as S is reciprocal. The elements s_{ij}, $i < j$, were set using Saaty's scale presented in Tab. I. Before we could proceed with calculating the scores of the categories, the consistency of the information provided by experts through the matrix S had to be checked. It is well known that the full consistency defined by Saaty:

$$s_{ik} = s_{ij} \cdot s_{jk}, \text{ for all } i, j, k = 1, 2, \ldots, n, \tag{2}$$

is basically unachievable even for not too large sets of mutually compared objects. Various authors, including Saaty (see [7]), therefore define for the practical use of Saaty's matrix various criteria to decide, whether a Saaty's matrix that is not fully consistent is at least consistent enough to represent expert knowledge concerning the relative preferences on a set of objects that are being compared. Hence, these authors allow some tolerance in the fulfillment of condition (2). We have approached this problem differently. We have defined directly the notion of weak consistency of Saaty's matrix. This natural condition has the advantage that already during the process of inputting data into Saaty's matrix that is constructed for categories ordered in accordance with their significance, the experts can easily check its fulfillment. The details concerning the issue of consistency of Saaty's matrix are given in the following section.

s_{ij}	linguistic meaning
1	i^{th} object is **equally important as** j^{th} object
3	i^{th} object is **slightly/moderately more important than** j^{th} object
5	i^{th} object is **strongly more important than** j^{th} object
7	i^{th} object is **very strongly more important than** j^{th} object
9	i^{th} object is **extremely/absolutely more important than** j^{th} object
2,4,6,8	correspond with the respective intermediate linguistic meanings

Tab. I *Saaty's scale.*

4. Consistency and Weak Consistency of Saaty's Matrix

In this section, we are going to deal with a general Saaty's matrix $S = \{s_{ij}\}_{i,j=1}^{n}$, which represents the information concerning preference intensities among n ob-

jects (in our application categories of works of art) given by experts. In the sense of the previously mentioned definition of Saaty's matrix, this means that $s_{ij} \in \{1/9, 1/8, 1/7, \ldots, 1/2, 1, 2, \ldots, 8, 9\}$ and the matrix is reciprocal, i.e. $s_{ij} = 1/s_{ji}$ for all $i, j = 1, \ldots, n$. We also require the matrix to be consistent enough to be able to use Saaty's method to calculate the relative importances of the objects.

The full consistency condition of Saaty's matrix is expressed by (2). Such a full consistency is, however, unachievable in real situations. Consider, for example four arbitrary objects linearly ordered according to their importance. If each of them is just slightly more important than the following one, then in the case of full consistency, the first object would have to be 27 times more important as the fourth one. But we have no value greater than 9 available on Saaty's 9 point scale to express our preferences (Tab. I). Saaty [7], therefore, proposes an inconsistency (Saaty introduced it as consistency index) index (CI) based on the spectral radius ($\lambda_{\max}$) of the pairwise comparison matrix S:

$$CI(S) = \frac{\lambda_{\max} - n}{n - 1}, \tag{3}$$

where n is the dimension of the matrix S. For any Saaty's matrix it holds that $CI(S) \geq 0$. $CI(S)$ was defined by Saaty to introduce an inconsistency measure for Saaty's matrices that would be independent of the dimension of the matrix. The average $CI(S)$ of randomly generated Saaty's matrices, however, proved to be growing as the dimension of the matrix grows. Saaty, therefore, introduced the inconsistency ratio $CR(S)$:

$$CR(S) = \frac{CI(S)}{RI(n)}, \tag{4}$$

where $RI(n)$ is the so-called random inconsistency index that represents the inconsistency of a randomly generated reciprocal pairwise comparison matrix of dimension n. $RI(n)$ is calculated as an average of indices $CI(S)$ calculated for a set of randomly generated reciprocal pairwise comparison matrices of dimension n. Matrix S for which $CR(S) < 0.1$ is then considered consistent enough.

4.1 Other approaches to determining satisfactory level of consistency of Saaty's matrix

Various authors have been trying to construct alternative measures of inconsistency of the matrix S. Alonso & Lamata [2] pointed out that the use of randomly generated reciprocal pairwise comparison matrices S of dimension n to determine the $RI(n)$ may result in slightly different indices depending on the number of such matrices used to compute the $RI(n)$. As they try to lower the computational complexity of determining $RI(n)$ for larger matrices, they realize that the growth of an average largest eigenvalue $\overline{\lambda}_{\max}(n)$ is easier to predict than the $RI(n)$ as the dimension n of the matrix S grows. For $\overline{\lambda}_{\max}(n)$, the expression $\overline{\lambda}_{\max}(n) = 2.7740n - 4.3513$ obtained by the least square method proves to be very accurate and easy to compute. Using $\overline{\lambda}_{\max}(n)$, they compute the random inconsistency index $RI_\lambda(n) = (\overline{\lambda}_{max} - n)/(n - 1)$. These authors, therefore, propose to

compute $CR(S)$ using the following formula:

$$CR(S) = \frac{CI(S)}{RI_\lambda(n)} = \frac{\lambda_{max} - n}{\overline{\lambda}_{max} - n}. \tag{5}$$

Analogically to Saaty's approach, for $CR(S) < 0.1$ is the matrix S considered consistent enough.

Lamata & Pelaez [5] propose an inconsistency index CI^* for a matrix S of type $n \times n$ using determinants:

$$CI^*(S) = \begin{cases} 0 & \text{if } n < 3, \\ det(S) & \text{if } n = 3, \\ \frac{1}{NT(S)} \sum\limits_{i=1}^{NT(S)} CI^*(\sigma_i) & \text{if } n > 3, \end{cases} \tag{6}$$

where σ is a submatrix of S of type 3×3 consisting of the rows and columns $i, j, k \in \{1, \ldots, n\}$, $i \neq j \neq k$, and $NT(S)$ is the number of such submatrices, i.e.

$$NT(S) = \begin{cases} 0 & \text{if } n < 3 \\ \frac{n!}{(n-3)!3!} & \text{otherwise.} \end{cases}$$

Next they generate 10 000 random Saaty's matrices of type $n \times n$. For this data they determine the p-value (e.g. 0.05) and for this p-value a critical value CR^* is calculated. If $CI^*(S) > CR^*$, then the matrix S is considered inconsistent.

Ji & Jiang [4] find Saaty's scale to be problematic – this paper deals with the transitivity on the linguistic and numerical parts of the scales most commonly used in AHP and with various types of inconsistency causes inherent to the scales used in AHP. They propose a scale that is transitive both in linguistic and in numerical part. The numerical part of this scale consists of the following set of values: $\{0, +0.5, -0.5, +1, -1, +1.5, \ldots, -3.5, +4, -4\}$. A common Saaty's matrix S can be transformed into a matrix D using this new so-called derived transitive scale in the following way:

$$d_{ij} = \begin{cases} (s_{ij} - 1)/2 & \text{if } s_{ij} \geq 1 \\ -\left(\frac{1}{s_{ij}} - 1\right)/2 & \text{if } s_{ij} < 1. \end{cases} \tag{7}$$

The matrix D is absolutely consistent, if

$$d_{ij} = \frac{1}{n}\sum_{k=1}^{n} d_{ik} + \frac{1}{n}\sum_{k=1}^{n} d_{kj} = \frac{1}{n}\sum_{k=1}^{n} d_{ik} - \frac{1}{n}\sum_{k=1}^{n} d_{jk} \tag{8}$$

i.e. if $d_{ij} = \overline{d}_i - \overline{d}_j$, where $\overline{d}_i = \frac{1}{n}\sum_{k=1}^{n} d_{ik}$ and $\overline{d}_j = \frac{1}{n}\sum_{k=1}^{n} d_{jk}$. An average grade deviation per comparison is then determined:

$$\epsilon = \sqrt{\frac{\sum\limits_{i=1}^{n-1}\sum\limits_{j=i+1}^{n} [d_{ij} - (\overline{d}_i - \overline{d}_j)]^2}{n(n-1)/2}}. \tag{9}$$

The decision maker sets up an acceptable level of deviation ald, and if $\epsilon < ald$ the matrix is considered consistent.

All of these approaches to assessing the consistency of the matrix of preference intensities are mathematically sound. These approaches, however, forget about the experts that input the data. It is almost impossible for the experts to check whether they are consistent (or consistent enough) in their preferences during the process of inputting data. Which is a major drawback when these experts are far from the field of mathematics and the dimension of the matrix is large. If we obtain from the experts a matrix of preference intensities that is not consistent enough, the usual advice is to start from the beginning and fill the matrix in again. This approach, however, does not make much sense as it does not guarantee that the new matrix of preference intensities will be better (more consistent) than the previous one. We need to find a way of obtaining a better result. Our solution to this problem is presented in the following subsection.

4.2 Weak consistency

Unlike most of the authors that start with full consistency and try to determine an acceptable level of its violation (see [2, 5, 4]), we have chosen a different approach. We define directly a weak consistency that is based on the properties that should intuitively hold, and we require these properties to be fully met. This is of great use when we need the experts that are expressing their preference intensities to check the consistency themselves during the process of inputting. If we utilize the linguistic meanings of the elements of Saaty's scale, we can define weak consistency such that for example if an object A is slightly more important than object B and object B is strongly more important than object C, we need at least the larger of the preference intensities to hold between A and C (which means that a stronger preference than the larger one of these two is also acceptable). For situations, when two objects are equally important, such as if A is equally as important as B and B is very strongly more important than C, it is reasonable to require A to be very strongly more important than C (the preference between the two objects that are not indifferent should hold between A and C). This understanding of weak consistency is summarized in Definition 1.

Definition 1: Let $S = \{s_{ij}\}_{i,j=1}^{n}$ be Saaty's matrix of preference intensities. We say that S is *weakly consistent*, if for all $i, j, k \in \{1, 2, \ldots, n\}$ the following holds:

$$s_{ij} > 1 \wedge s_{jk} > 1 \quad \Longrightarrow \quad s_{ik} \geq \max\{s_{ij}, s_{jk}\}; \tag{10}$$

$$(s_{ij} = 1 \wedge s_{jk} \geq 1) \vee (s_{ij} \geq 1 \wedge s_{jk} = 1) \quad \Longrightarrow \quad s_{ik} = \max\{s_{ij}, s_{jk}\}. \tag{11}$$

The property $s_{ik} \geq \max\{s_{ij}, s_{jk}\}$ can be found as max-max transitivity in the literature [3].

If we order the objects (alternatives) being compared according to their importance from the most important to the least, we get $s_{ij} \geq 1$ for all $i, j = 1, 2, \ldots, n$ such that $i < j$; $s_{ii} = 1$ for all $i = 1, 2, \ldots, n$. The upper triangle of the matrix S then consists only of numbers from $\{1, 2, \ldots, 9\}$. In such case, according to the Definition 1, it is sufficient to check whether conditions (10) and (11) are fulfilled for the elements in the upper triangle of S to assess the weak consistency of S.

It is evident that for the weak consistency condition to hold, the elements in the upper triangle of Saaty's matrix S have to be nondecreasing from left to right in the rows and from the bottom up in the columns. This property was used by the experts to continuously check the weak consistency while entering the data into Saaty's matrix of preference intensities for categories of works of art.

Analogically, we could define weak consistency using elements that are lower than 1 by minimum. This is summarized in the following proposition.

Proposition 1: Let $S = \{s_{ij}\}_{i,j=1}^{n}$ be Saaty's matrix of preference intensities. Then, S is weakly consistent if and only if the following holds for all $i, j, k \in \{1, 2, \ldots, n\}$:

$$s_{ij} < 1 \wedge s_{jk} < 1 \quad \Longrightarrow \quad s_{ik} \leq \min\{s_{ij}, s_{jk}\}; \tag{12}$$

$$(s_{ij} = 1 \wedge s_{jk} \leq 1) \vee (s_{ij} \leq 1 \wedge s_{jk} = 1) \quad \Longrightarrow \quad s_{ik} = \min\{s_{ij}, s_{jk}\}. \tag{13}$$

Proof:

1. First we prove that weak consistency implies conditions (12) and (13):
   (a) Let $s_{ij} < 1$ and $s_{jk} < 1$, then from the reciprocity of S we get $s_{ji} > 1$ and $s_{kj} > 1$. The weak consistency implies that $s_{ki} \geq \max\{s_{ji}, s_{kj}\}$, i.e. $s_{ik} \leq \frac{1}{\max\{s_{kj}, s_{ji}\}}$. Hence, $s_{ik} \leq \frac{1}{s_{kj}} = s_{jk}$ and $s_{ik} \leq \frac{1}{s_{ji}} = s_{ij}$, in other words $s_{ik} \leq \min\{s_{jk}, s_{ij}\}$.
   (b) Let $s_{ij} = 1$ and $s_{jk} \leq 1$. Then, $s_{ji} = 1$ and $s_{kj} \geq 1$, weak consistency implies that $s_{ki} = s_{kj}$, from reciprocity $s_{ik} = s_{jk} = \min\{s_{ij}, s_{jk}\}$.
   (c) Let $s_{ij} \leq 1$ and $s_{jk} = 1$. Then, $s_{ji} \geq 1$ and $s_{kj} = 1$, weak consistency implies $s_{ki} = s_{ji}$, from reciprocity $s_{ik} = s_{ij} = \min\{s_{ij}, s_{jk}\}$.
2. Now let us suppose that S fulfills (12) and (13). We will prove that such matrix S is weakly consistent:
   (a) Let $s_{ji} > 1$ and $s_{kj} > 1$. Reciprocity implies $s_{ij} < 1$ a $s_{jk} < 1$. From (12) we get $s_{ik} \leq \min\{s_{ij}, s_{jk}\}$. Then, $s_{ik} \leq s_{ij}$ and $s_{ik} \leq s_{jk}$. From reciprocity we get $s_{ki} \geq s_{ji}$ and $s_{ki} \geq s_{kj}$, i.e. $s_{ki} \geq \max\{s_{kj}, s_{ji}\}$.
   (b) Let $s_{ji} = 1$ and $s_{kj} \geq 1$. Reciprocity implies $s_{ij} = 1$ and $s_{jk} \leq 1$. From (13) we get $s_{ik} = s_{jk} = \min\{s_{ij}, s_{jk}\}$. Thus, from reciprocity $s_{ki} = s_{kj} = \max\{s_{kj}, s_{ji}\}$.
   (c) Let $s_{ji} \geq 1$ and $s_{kj} = 1$. Reciprocity implies $s_{ij} \leq 1$ and $s_{jk} = 1$. From (13) we get $s_{ik} = s_{ij} = \min\{s_{ij}, s_{jk}\}$. Thus, from reciprocity $s_{ki} = s_{ji} = \max\{s_{kj}, s_{ji}\}$. □

Relations between elements greater than 1 and lower than 1 can be described by propositions 2 and 3.

Proposition 2: Let $S = \{s_{ij}\}_{i,j=1}^{n}$ be a weakly consistent Saaty's matrix of preference intensities. If for $i, j, k \in \{1, 2, \ldots, n\}$ it holds that $s_{ij} > 1$ and $s_{jk} < 1$, then the following holds for s_{ik}:

$$1 < s_{ik} \leq s_{ij}, \text{ if } s_{ij} > \frac{1}{s_{jk}} = s_{kj}; \tag{14}$$

$$1 > s_{ik} \geq s_{jk}, \text{ if } s_{ij} < s_{kj}; \tag{15}$$

$$s_{ji} \leq s_{ik} \leq s_{ij}, \text{ if } s_{ij} = s_{kj}. \tag{16}$$

Proof:
Considering the relationship between s_{ij} and s_{kj} we need to deal with the following 3 situations separately:

1. Let us consider $s_{ij} > s_{kj}$.
   (a) Let us suppose that $s_{ik} < 1$. Reciprocity then implies $s_{ki} > 1$. From weak consistency we get $(s_{ki} > 1 \wedge s_{ij} > 1) \Longrightarrow (s_{kj} \geq \max\{s_{ij}, s_{ki}\})$, which is a contradiction to the assumption that $s_{ij} > s_{kj}$.
   (b) Let us suppose that $s_{ik} = 1$. As $s_{kj} > 1$, we get from weak consistency that $s_{ij} = \max\{s_{kj}, s_{ik}\} = s_{kj}$, which is again a contradiction to the assumption that $s_{ij} > s_{kj}$.
   (c) Consequently, $s_{ik} > 1$ must hold. As $s_{kj} > 1$, weak consistency implies that $s_{ij} \geq \max\{s_{ik}, s_{kj}\}$. Thus, $s_{ij} \geq s_{ik} > 1$ holds.
2. Now let $s_{ij} < s_{kj}$.
   (a) Let $s_{ik} > 1$. As in 1c) we get $s_{ij} \geq s_{kj}$, which is a contradiction to the assumption that $s_{ij} < s_{kj}$.
   (b) Let $s_{ik} = 1$. As in 1b) we get $s_{ij} = s_{kj}$, which is again a contradiction to the assumption that $s_{ij} < s_{kj}$.
   (c) Consequently, $s_{ik} < 1$ must hold. Analogically to 1a), we now get $1 > s_{ik} \geq s_{jk}$.
3. Let $s_{ij} = s_{kj}$. As S is weakly consistent, one of the following situations may occur:
   (a) Let $s_{ik} > 1$. Then, as $s_{kj} > 1$, we get from the weak consistency $s_{ij} \geq \max\{s_{ik}, s_{kj}\}$. As $s_{ij} = s_{kj}$, to fulfill the implication (10) it has to hold that $s_{ij} \geq s_{ik} > 1$.
   (b) Now let $s_{ik} < 1$. Then, $s_{ki} > 1$ and as $s_{ij} > 1$, the weak consistency implies $s_{kj} \geq \max\{s_{ij}, s_{ki}\}$. As $s_{ij} = s_{kj}$, to fulfill the implication (10) it has to hold that $s_{ij} \geq s_{ki}$, i.e. $s_{ji} \leq s_{ik} < 1$.
   (c) The last situation we need to check is $s_{ik} = 1$. As $s_{kj} > 1$, the weak consistency implies that $s_{ij} = s_{kj}$. As this equation holds, a situation when $s_{ik} = 1$ can occur.

   When we put together 3a) – 3c), we get $s_{ji} \leq s_{ik} \leq s_{ij}$. □

Proposition 3: Let $S = \{s_{ij}\}_{i,j=1}^{n}$ be a weakly consistent Saaty's matrix of preference intensities. If for $i, j \in \{1, 2, \ldots, n\}$ it holds that $s_{ij} < 1$ and $s_{jk} > 1$, then the following holds for s_{ik}:

$$1 < s_{jk} \leq s_{ik}, \text{ if } s_{jk} > \frac{1}{s_{ij}} = s_{ji}; \tag{17}$$

$$s_{ij} \leq s_{ik} < 1, \text{ if } s_{jk} < s_{ji}; \tag{18}$$

$$s_{kj} \leq s_{ik} \leq s_{jk}, \text{ if } s_{jk} = s_{ji}. \tag{19}$$

Proof:
The proof is analogical to the proof of Proposition 2 – to obtain (17), (18) and (19), we again distinguish among three cases: $s_{jk} > s_{ji}$, $s_{jk} < s_{ji}$ and $s_{jk} = s_{ji}$ and for each of them we investigate $s_{ik} > 1$, $s_{ik} < 1$ and $s_{ik} = 1$. □

The concept of weak consistency (10), (11) represents a weakening of the concept of consistency (2). This is summarized in the following proposition.

Proposition 4: Let a Saaty's matrix of preference intensities $S = \{s_{ij}\}_{i,j=1}^{n}$ be consistent, i.e. $s_{ik} = s_{ij} \cdot s_{jk}$ for all $i, j, k = 1, 2, \ldots, n$. Then, S is also weakly consistent.

Proof:
Let S be consistent (i.e. consistency condition (2) is fulfilled). Then, $s_{ij} > 1$ and $s_{jk} > 1$ imply $s_{ik} = s_{ij} \cdot s_{jk} > \max\{s_{ij}, s_{jk}\}$, which means that the first condition of weak consistency (10) is fulfilled. Next, if $s_{ij} = 1$, then, $s_{ik} = s_{jk} = \max\{s_{ij}, s_{jk}\}$ and if $s_{jk} = 1$, then $s_{ik} = s_{ij} = \max\{s_{ij}, s_{jk}\}$. The second condition of weak consistency (11) is also fulfilled. □

The implication in the Proposition 4 holds only for the consistency defined by (2). On the other hand, it is naturally not true that a matrix that is deemed "consistent enough" according to some of the criteria found in literature has to necessarily fulfill the weak consistency conditions. For example, according to $CR(S)$ even such matrix S may be considered consistent enough, where the decision maker did not manage to keep the preference ordering of the alternatives – at some place he prefers alternative C to alternative D and at the same time he inputs information that D is preferred to C. This situation will be illustrated by the following numerical example.

4.3 Numerical example

Let us consider the following Saaty's matrix.

$$S = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 1 & 2 & 2 & 2 \\ 1/2 & 1 & 3 & 2 \\ 1/2 & 1/3 & 1 & 2 \\ 1/2 & 1/2 & 1/2 & 1 \end{pmatrix} \end{array}$$

Its maximum eigenvalue is $\lambda_{\max} = 4.2153$. The inconsistency index $CI = \frac{4.2153-4}{4-1} = 0.0718$ and the inconsistency ratio is $CR(S) = \frac{0.0718}{0.89} = 0.0807$. If we use Saaty's weakening of the consistency condition, this matrix will be considered consistent enough as $CR(S) < 0.1$.

From the second row of the matrix S it follows that B is preferred to C and D. As the intensity of preference of B to C is larger than B to D, we reasonably conclude that D is preferred to C. While according to the third row C is preferred to D. The preference ordering of the alternatives is clearly violated and still the matrix is considered consistent enough if we use Saaty's inconsistency ratio and the threshold 0.1. We can easily see that S is not weakly consistent: for $s_{23} = 3$ and $s_{34} = 2$ we would need $s_{24} \geq \max\{s_{23}, s_{34}\} = 3$. However, $s_{24} = 2$ which violates condition (10) of the weak consistency.

5. Determining Scores of the Categories

The weak consistency of Saaty's matrix can be easily checked during the process of entering data into the matrix. In our case, the experts have decided to additionally (after having completed Saaty's matrix) re-divide the classes of indifferent categories that resulted from the Pairwise Comparison Method. The experts defined the intensities of preferences between pairs of previously indifferent categories and then compared the new categories with the others so that Saaty's matrix remained weakly consistent. Fig. 2 illustrates final Saaty's matrix after re-dividing the pairs of indifferent categories.

If S is close to the ideally consistent matrix, the scores of 27 categories, representing their relative importance, can be calculated by Saaty's method as components of the eigenvector corresponding to the largest eigenvalue of Saaty's matrix S.

We can obtain the scores of artistic categories from S also in a different way. The columns of S can be interpreted as repeated measurements of the relative importances of the 27 categories. These measurements are performed by the team of experts who compare all the categories with the first one, then the second one, and so on until the 27^{th} one. From the point of view of mathematical statistics, these are compositional data, i.e. data bearing only relative information (see [1]). Information contained in these data can be expressed by estimating their mean value. A proper estimator of the mean value of this kind of data is a vector whose components are geometric means of the corresponding components of vectors representing single measurements. The relative scores of all 27 categories can be also obtained by computing geometric means of the rows of Saaty's matrix (this calculation method is known as the Logarithmic Least Squares Method, see [6]). If the experts satisfy the condition of weak consistency of the matrix of preference intensities throughout the input process, we can expect the individual measurements and the estimate of the mean value of the compositions to be better.

Fig. 3 compares the scores determined by Saaty's matrix eigenvector method with those determined as geometric means of the rows. The scores are normalized so that the maximum is 305 (analogy to R&D outcomes evaluation). It is easy to see that the differences between the results of these two methods are not large. Saaty's matrix eigenvector method was used in testing the model on the first real dataset, gathered by Czech art colleges and faculties for the years 2008 to 2010.

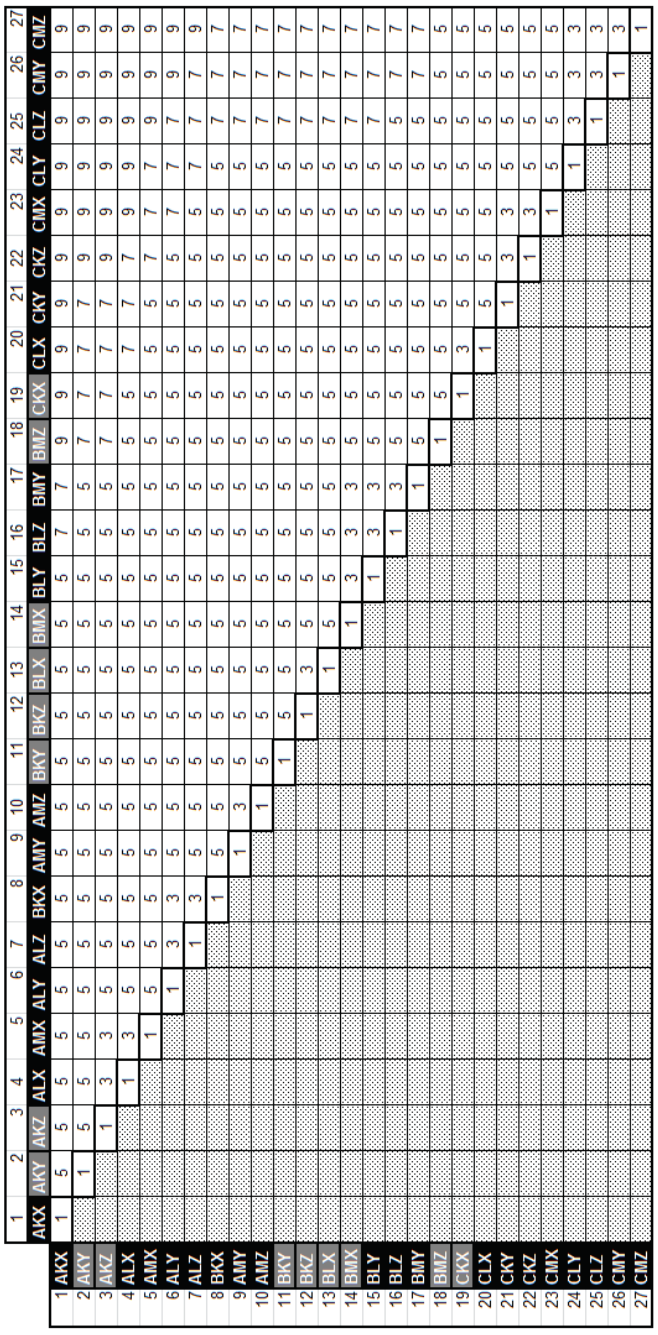


Fig. 2 *Saaty's matrix of preference intensities for 27 categories ordered according to their significance. The re-devided categories are highlighted.*

Category	Relevance or significance	Extent	Institutional reception	Eigenvector method	Geom. means method
AKX	crucial significance	large	international	305	305
AKY	crucial significance	large	national	259	254
AKZ	crucial significance	large	regional	210	217
ALX	crucial significance	medium	international	191	194
AMX	crucial significance	limited	international	174	171
ALY	crucial significance	medium	national	138	138
ALZ	crucial significance	medium	regional	127	124
BKX	containing numerous important innovations	large	international	117	112
AMY	crucial significance	limited	national	97	94
AMZ	crucial significance	limited	regional	90	87
BKY	containing numerous important innovations	large	national	79	75
BKZ	containing numerous important innovations	large	regional	66	66
BLX	containing numerous important innovations	medium	international	62	61
BMX	containing numerous important innovations	limited	international	48	50
BLY	containing numerous important innovations	medium	national	44	46
BLZ	containing numerous important innovations	medium	regional	40	41
BMY	containing numerous important innovations	limited	national	37	38
BMZ	containing numerous important innovations	limited	regional	31	30
CKX	pushing forward modern trends	large	international	26	26
CLX	pushing forward modern trends	medium	international	24	24
CKY	pushing forward modern trends	large	national	19	20
CKZ	pushing forward modern trends	large	regional	17	18
CMX	pushing forward modern trends	limited	international	16	16
CLY	pushing forward modern trends	medium	national	12	13
CLZ	pushing forward modern trends	medium	regional	10	11
CMY	pushing forward modern trends	limited	national	9	9
CMZ	pushing forward modern trends	limited	regional	8	9

Fig. 3 *Comparison of the results of eigenvector method and the logarithmic least squares method (Geom. means method).*

6. Conclusion

The paper describes a multiple criteria evaluation model for the works of art resulting from the creative activities of Czech art colleges and faculties. The evaluation model is an integral part of the Registry of Artistic Results (RUV), where information concerning these works of art is stored. The results of this evaluation model have been used as a basis for allocating a part of the state-budget subsidy among art colleges in the Czech Republic since 2012.

For the purpose of determining scores for 27 categories of works of art a two-step procedure is proposed. It was developed in an effort to achieve the best possible conversion of preferences of the expert team into scores for different categories of artistic production. It is based on Saaty's method. Due to the large number of compared objects, our effort was focused on the consistency of Saaty's matrix. Various criteria of sufficient consistency of Saaty's matrix published in the literature were studied and consequently a new notion of weak consistency of Saaty's matrix has been introduced in this paper. For objects descending ordered in accordance with their importance (obtained e.g. by the Pairwise Comparison Method) the weak consistency is easy to check even during the process of entering data into Saaty's matrix of preference intensities. It also constitutes a natural requirement

for the consistency of information provided by experts concerning their preferences on a set of objects.

The paper shows on a practical application how much effort is needed to obtain information as consistent as possible from a group of experts in a field far away from mathematics (in this case arts).

Acknowledgement

This research is conducted with the support of the Centralized Developmental Project C41 entitled *Evaluating Creative Work Outcomes Pilot Project* and financed by the Czech Ministry of Education.

References

[1] Aitchison J.: The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London, 1986.

[2] Alonso J. A., Lamata M. T.: Consistency in the Analytic Hierarchy Process: A New Approach. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **14**, 4, 2006, pp. 445-459.

[3] Herrera-Viedma E., Herrera F., Chiclana F., Luque M.: Some issues on consistency of fuzzy preference relations. European Journal of Operational Research, 154, 2004, pp. 98-109.

[4] Ji P., Jiang R.: Scale Transitivity in the AHP. The Journal of the Operational Research Society, **54**, 8, 2003, pp. 896-905.

[5] Lamata M. T., Pelaez J. I.: A Method for Improving the Consistency Judgements. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **10**, 6, 2002, pp. 677-686.

[6] Ramík J.: Analytický hierarchický process (AHP) a jeho využití v malém a středním podnikání. Slezská univerzita v Opavě, Karviná, 2000.

[7] Saaty T. L.: Relative Measurement and its Generalization in Decision Making, Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors – The Analytic Hierarchy/Network Process. RACSAM, **102**, 2, 2008, pp. 251-318.

[8] Saaty T. L.: The Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process. Vol. VI of the AHP Series, RWS Publ., 2000.

[9] Saaty T. L.: The Brain: Unraveling the Mystery of How it Works, The Neural Network Process. RWS Publ., 1999.

[10] Saaty T. L.: A Scaling Method for Priorities in Hierarchical Structures. Journal of Mathematical Psychology, 15, 1977, pp. 57-68.

[11] Saaty T. L., Vargas L. G.: Decision Making with the Analytic Network Process: Economic, Political, Social and Technological Applications with Benefits, Opportunities, Costs and Risks. Springer, New York, 2006.

[12] Smernica č. 13/2008-R zo 16. októbra 2008 o bibliografickej registrácii a kategorizácii publikačnej činnosti, umeleckej činnosti a ohlasov. Ministerstvo školstva Slovenskej republiky, 2008.

[13] Talašová J.: Fuzzy metody vícekriteriálního hodnocení a rozhodování. Vydavatelství Univerzity Palackého, Olomouc, 2003.

[14] Zelinský M. (ed.): Registr uměleckých výkonů. Akademie múzických umění v Praze, Praha, 2010.

PUBL. IV

Stoklasa, J. and Luukka, P., Receiver operating characteristics and the quality of data.

Submitted to *Psychometrika*, Springer. (May 2014)

RECEIVER OPERATING CHARACTERISTICS AND THE QUALITY OF DATA

Jan Stoklasa[1]

palacký university in olomouc

Pasi Luukka

lappeenranta university of technology

[1]This research was supported by the grant *GA 14-02424S* of the Grant Agency of the Czech Republic.
Corresponding author's full address: Jan Stoklasa, Palacký University in Olomouc, Faculty of Science, Dept. of Math. Analysis and Applications of Mathematics, 17. listopadu 1192/12, 77146, Olomouc, Czech Republic

Correspondence should be sent to

E-Mail: jan.stoklasa@upol.cz
Phone: +420-585-634107
Fax: +420-585-634002
Website: www.upol.cz

RECEIVER OPERATING CHARACTERISTICS AND THE QUALITY OF DATA

Abstract

We propose an approach to classifier performance assessment that regards misclassifications of instances with high quality of data as more serious than misclassifications of instances with lower data quality. We choose the receiver operating characteristics (ROC) and the area under the ROC curve and its fuzzification as a starting point and show how such classifier performance measures can be modified to reflect data quality. We analyze the performance of the modified ROC method on artificial data and on a real-life case from psychological diagnostics. Data quality incorporation into the classification is discussed on the example of the "don't know" principle.

Key words:

classification, receiver operating characteristic, ROC, area under curve, AUC, sensitivity, specificity, fuzzy, diagnostics, psychology.

1. Introduction

Classification is an important task in decision making. Classifiers are therefore an important tool for designing decision support systems. In this paper we will use examples from humanities. The reason to do so is that in many cases the decision making in humanities has to be based on data with varying quality. This is of no surprise when we consider human beings as being the source of the data. Various measures of data quality (deliberate distortions of data), lie scores and other instruments aiming on identifying data with problematic interpretability have been developed in psychology, sociology and other disciplines and embedded into the tools these sciences use to gather data. Data quality is therefore not only a well known issue in these sciences, there are also some tools for the assessment of the quality of data.

Numerous decision support systems and classifiers have been developed in the past decades for both psychological and medical diagnostics purposes (see Miller (1994); Sim et al. (2001) for an overview), some of them even using fuzzy set theory (Godil et al. (2011) provide a summary of applications of fuzzy logic in medicine, neuroscience, psychiatry and psychology), methods reflecting some specific needs of diagnostics in various fields have been developed (e.g. Parasuraman et al. (2000)). The diagnostic situation is in fact a classification problem. We know the definitions of the categories (the diagnoses - given by World Health Organisation (2006) or diagnostics manuals of particular test methods) and need to assign a proper category label to the patient. Many statistical methods are available for classification purposes and frequently used (also in medicine and related fields) such as decision trees (Markey et al., 2003), random forests (Ramírez et al., 2010; Smith et al., 2010) and their fuzzification (Bonissone et al., 2010), discriminant analysis (Smith et al., 2010) and logistic regression (Bielza et al., 2011; Zhou et al., 2004). Bayesian network classifiers (Robles et al., 2004) and neural network classifiers (Bhatikar et al., 2005; De Gaetano et al., 2009) are also commonly used in this area. Fuzzy logic and linguistic fuzzy modeling can be used to construct mathematical classifiers as well (see e.g. Kuncheva (2000) for an overview). Software packages providing implementations of fuzzy methods suitable for such applications are also available (see i.e. Holeček and Talašová (2010); Mathworks Inc. (2011); Talašová and Holeček (2009); von Altrock (1995)).

Regardless of the mathematical tool we use to build a classifier, we either model expert knowledge (by e.g. fuzzy rule bases) or "teach" the classifier to perform well on some reliable data set (or combine these approaches). Either way we design a tool to perform well on data with high quality. This is understandable, as we do not want to make mistakes when we are dealing with unambiguous, high-quality and well interpretable data. The question is how our classifiers perform on real life data. We can find various classifier performance measures (see e.g. Parker (2013) for a comparison of several most

commonly used classifier performance measure; a recently proposed *H measure* introduced in Hand (2009)). The quality of data on which the classification is based (and hence the interpretability or reliability of the classification assigned) is, as far as we know, not directly considered in the process of performance assessment of a classifier. This way we suppose that all the data we use as inputs are of the same quality and hence misclassification (or successful classification) of any instance provides the same amount of information concerning the performance of the classifier.

In this paper we intend to show, that this might result in a distorted image of the classifier performance, particularly when there are data of various quality in our data set. Usually we avoid such distortions by discarding all the instances with low quality of data (and as such with low information value) before we assess the performance of the classifier. This way we however change the validation set. If the classifier is intended to be used in a real-life setting where data of varying quality are common, such approach makes in fact no sense. We therefore propose in Section 5 to assess the performance of a classifier on the whole set of data (no data omitting due to low-quality) - but to reflect the quality of the data in the assessment process. The main idea of the approach proposed in this paper is simple: a classifier performs well if it makes as little misclassifications on high-quality data as possible. As the quality of data decreases, each misclassification (as well as successful classification) tells us less about the classifier up to the point of zero quality, where any outcome of the classifier tell absolutely nothing about its performance. We present this idea on the example of the receiver operating characteristics (ROC) curves and the area under these curves (AUC) as classifier performance measures. In Section 2 we recall some necessary mathematical notions that will be necessary for our analysis and for the modification of the classic ROC presented in Section 3. In Section 4 we briefly summarize the fuzzified ROC. The modified ROC approach is introduced in Section 5 and its similarities to the classic and fuzzy ROC are discussed here. Finally in Section 6 we compared the results of the proposed modification of ROC with the classic approach on artificial data and also on a real-life data set from psychological diagnostics. We also discuss in this section how the quality of data can be reflected in the classification process by discussing the "don't know principle" which could be summarized by the rule "if the quality of data is low, then any of the two classes might be assigned, therefore assign something in the middle".

2. Used mathematical apparatus

The quality of data will be considered a characteristic, that each instance of data can have to a various degree (that is discriminating just between quality data and poor quality data will not be considered sufficient). We will therefore use the notion of fuzzy sets as

introduced by Zadeh (1965), a more detailed description of the related concepts can be found for example in Dubois and Prade (2000).

Let U be a nonempty set (the universe). A *fuzzy set* A on U is defined as the mapping $A : U \to [0, 1]$. For each $x \in U$ the value $A(x)$ is called a *membership degree* of the element x in the fuzzy set A and $A(.)$ is called a *membership function* of the fuzzy set A. The symbol $F(U)$ denotes the set of all fuzzy sets on U.

The *height of* a fuzzy set A is a real number $\mathrm{hgt}(A) = \sup_{x \in U}\{A(x)\}$. Other important concepts related to fuzzy sets are: a) the *kernel* of A, $\mathrm{Ker}(A) = \{x \in U \mid A(x) = 1\}$; b) the *support* of A, $\mathrm{Supp}(A) = \{x \in U \mid A(x) > 0\}$; and c) for $\alpha \in [0, 1]$ the α-*cut* of A, $[A]_\alpha = \{x \in U \mid A(x) \geq \alpha\}$. If the support of A is a finite set, $\mathrm{Supp}(A) = \{x_1, \ldots, x_k\}$, then the fuzzy set A will be described as $A = \{{}^{A(x_1)}\!/_{x_1}, \ldots, {}^{A(x_k)}\!/_{x_k}\}$. The cardinality of such a fuzzy set A is given by $Card(A) = \sum_{i=1}^{k} A(x_i)$. Any crisp set $\{x_1, \ldots, x_k\}$ can be represented by the fuzzy set $\{{}^{1}\!/_{x_1}, \ldots, {}^{1}\!/_{x_k}\}$.

A *union* of fuzzy sets A and B on U is a fuzzy set $A \cup B$ on U with the membership function for all $x \in U$ given by $(A \cup B)(x) = \max\{A(x), B(x)\}$. An *intersection* of fuzzy sets A and B on U is a fuzzy set $A \cap B$ on U with the membership function for all $x \in U$ given by $(A \cap B)(x) = \min\{A(x), B(x)\}$. Let A be a fuzzy set on U and B be a fuzzy set on V. Then the Cartesian product of A and B is the fuzzy set $A \times B$ on $U \times V$ with the membership function defined for all $(x, y) \in U \times V$ by $(A \times B)(x, y) = \min\{A(x), B(y)\}$.

Let x be a vector in an n-dimensional real space $\mathbb{R}^n$ and let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ be a set of class labels. A (crisp) *classifier* is any mapping

$$D : \mathbb{R}^n \to \Omega. \tag{1}$$

If Ω is a closed interval, we call D a *continuous classifier*.

As we do not aim to propose any particular type of classifier here and we intend to focus on the performance of classifiers in general in the context of data quality, we leave out a more detailed discussion of various classifiers. In this paper and in the presented real-life example, we consider $\Omega = [0, 1]$. This way the diagnostics process in medicine or psychology is still seen as trying to decide whether a person is healthy (classifier output 0) or not (classifier output 1). But we also allow the classifier to assign values among 0 and 1 to instances. This is in accordance with the fact that in many cases in medicine and in psychology, although typical symptoms can be defined, a real life person who has these symptoms fully and exactly is not a frequent sight. In fact much of the diagnostics and decision making in human sciences is based on partially met criteria and the closest match is sought (instead of exact math with the description of the category prototype). Nevertheless at the end a decision has to be made whether a diagnosis should be assigned or whether the person is healthy or not. This is usually done by introducing some sort of

threshold - crisp (see the crisp ROC analysis in the following section) or fuzzy (see the section on fuzzy ROC). The quality of data is another issue to be considered and reflected in the mathematical models for decision support or classification.

In accordance with Kuncheva (2000) we can define a fuzzy classifier for the purposes of this paper in the following way. Let x be a vector in an n-dimensional real space $\mathbb{R}^n$ and let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ be a set of class labels (crisp, linguistic or fuzzy). A *fuzzy classifier* is an if-then inference system (a fuzzy rule based system) IS which either

a) yields a single class label (crisp, linguistic or fuzzy) for x:

$$IS : \mathbb{R}^n \to \Omega, \tag{2}$$

b) or for a discrete Ω maps $\mathbb{R}^n$ into a fuzzy set on Ω :

$$IS : \mathbb{R}^n \to F(\Omega) = \{\Omega_x \mid x \in \mathbb{R}^n\} \tag{3}$$

such that for all $x \in \mathbb{R}^n$ it holds that $\sum_{i=1}^{c} \Omega_x(\omega_i) = 1$. Thus IS distributes the full membership of x among the classes. The fuzzy set Ω_x can be interpreted as "appropriate class label for x".

The next sections briefly summarize one of the tools commonly used for classifier performance assessment - the receiver operating characteristics and the AUC - area under the ROC curve. We briefly summarize the classic version of ROC in Section 3, its fuzzification in Section 4 and propose a modification of ROC able to reflect quality of the data on which the classification is based in Section 5.

3. Receiver operating characteristic (ROC)

Receiver operating characteristics (ROC) analysis is a method originally developed in signal detection theory (SDT) and further extended in psychology by Green and Swets (1966) to graphically represent the performance of classifiers (see Egan (1975) for more details). Its use has been recently extended to other fields of science - there are numerous medical applications of ROC for diagnostics systems performance (see Fawcett (2004); Zou et al. (2007) for reviews of ROC applications in this field). There are many applications of fuzzy SDT in psychology as well (see Parasuraman et al. (2000)). The use of fuzzy methods and their combination with classical approaches such as the SDT is also frequent (see Godil et al. (2011); Zolghadri and Mansoori (2007)). The use of fuzzy logic in classifier fusion and construction of fuzzy ROC curves is discussed by Evangelista et al. (2005a,b).

The signal detection theory (Egan, 1975; Fawcett, 2004; Parasuraman et al., 2000) assumes there are two states - signal and noise. Our goal is to find reliable tools for signal detection. Such tools (classifiers) result in a Positive (Pos) and Negative (Neg) judgement (Pos meaning the signal is believed to be present and Neg meaning the signal is believed

to be absent, only noise is present). These classifiers can by their nature be binary (which directly determine one of the two values - 0 for "negative" or 1 for "positive") or continuous, which assign to each instance $x_i \in \{x_1, \ldots, x_n\}$ a value $cl_i \in [0, 1]$. In order to obtain a crisp classification with a continuous classifier a threshold $t \in [0, 1]$ has to be determined. Instances with values above this threshold ($cl_i \geqq t$) are then considered *Pos* instances and instances with values below the threshold ($cl_i < t$) are considered *Neg*. An optimal threshold can to be determined which guarantees the best performance of the classifier. ROC graphs are commonly used for this purpose (see Fawcett (2004)).

For each given threshold $t \in [0, 1]$ the classifier divides the set of all instances into four subsets:

1. TP_t (true positive) - set of instances where signal was classified as positive (*Pos*),
2. FP_t (false positive) - set of instances where noise was classified as positive (*Pos*),
3. FN_t (false negatives) - set of instances where signal was classified as negative (*Neg*),
4. TN_t (true negatives) - set of instances where noise was classified as negative (*Neg*).

In our context, signal represents the real (confirmed) presence of the disease, noise will represent the absence of the disease (also confirmed). Let now $S \subseteq \{x_1, \ldots, x_n\}$ be a set of signal instances and $N \subseteq \{x_1, \ldots, x_n\}$ be a set of noise instances. For the crisp ROC it holds, that $S \cap N = \varnothing$ and $S \cup N = \{x_1, \ldots, x_n\}$. Let $Card(X)$ be generally the cardinality of a set X, i.e. the number of its elements. The performance of a classifier can for any threshold $t \in [0, 1]$ be described using the following characteristics (see e.g. Fawcett (2004) for more details):

$$TP\_rate_t = \frac{Card(TP_t)}{Card(S)} = \frac{Card(TP_t)}{Card(TP_t \cup FN_t)}, \tag{4}$$

$$FP\_rate_t = \frac{Card(FP_t)}{Card(N)} = \frac{Card(FP_t)}{Card(TN_t \cup FP_t)}, \tag{5}$$

$$FN\_rate_t = \frac{Card(FN_t)}{Card(S)} = \frac{Card(FN_t)}{Card(TP_t \cup FN_t)}, \tag{6}$$

$$TN\_rate_t = \frac{Card(TN_t)}{Card(N)} = \frac{Card(TN_t)}{Card(TN_t \cup FP_t)}. \tag{7}$$

All the characteristics (4) - (7) are real numbers from $[0, 1]$ for any thresholds $t \in [0, 1]$. *Sensitivity* can be defined as the TP_rate_t, *specificity* as $1 - FP\_rate_t$. The performance of the classifier described by these characteristics can be graphically represented by the ROC graph, which visualises the combinations of sensitivity and (1-specificity) (TP_rate_t is depicted on the vertical axis, FP_rate_t on the horizontal axis) for all possible classification

thresholds. Discrete classifiers produce a single point in the ROC space, for continuous classifiers an ROC curve can be plotted, where each point corresponds to a certain threshold value for distinguishing between signal and noise, that is $ROC\_curve = \{(x, y) \mid x = FP\_rate_t, y = TP\_rate_t, \text{for all } t \in [0, 1]\}$.

Generally the point $(0, 0)$ represents a classifier that never assigns *Pos*. Such classifier makes no false positive errors, but also results in no true positives. The point $(1, 1)$ represents a classifier that always issues *Pos* classification. The point $(0, 1)$ describes a perfect classifier. For continuous classifiers the area under the ROC curve (AUC) can be computed. For an ideal classifier it holds that $AUC = 1$; classifiers with $AUC > 0.5$ may be considered better than random classification (see for example Fawcett (2004); Krzanowski and Hand (2009) for more details). The AUC can be interpreted as the probability that a randomly chosen signal instance will be assigned greater value by the classifier, than a randomly chosen noise instance.

This approach however does not reflect the quality of the data based on which a classification was assigned by the classifier. Information concerning the performance of our classifier may therefore be misleading particularly when we expect it to be used on data with varying quality. The main idea behind the modification of the classic ROC we propose in Section 5 is, that if a classifier performs well on high-quality data and makes mistakes on low-quality data it does not mean that it works poorly. Given the fact that most classifiers are taught on data sets that are reliable (and hence the input data for the classifier are of high quality), the performance measure on the validation set should take the quality of the data into account. Even more so, if the data in the real life situation are expected to be of varying quality. That is if we have means for data quality assessment, we can include the information on data quality into the classifier assessment process. If a classifier does not make mistakes on high-quality data, it can be considered to work well. On the other hand, if the classifier classifies correctly mainly instances with low-quality data, something is wrong and this should be reflected by the classifier performance assessment tool (in our case by the modified ROC analysis).

It should also be mentioned here, that the use of AUC for inter-classifier comparison (as a classifier performance measure) has been questioned recently (see Flach et al. (2011); Hand (2009); Parker (2013)). According to Hand (2009) the AUC is incoherent in terms of misclassification costs, a modified performance measure is proposed - the *H measure*. We refer the reader to Flach et al. (2011) or Hand (2009) for a discussion concerning this issue and to Parker (2013) for comparisons of *H measure* with AUC and other frequently used classifier performance measures, such as the *average precision*, *ϕ-coefficient* and the *area under the Cohen's κ curve*. The argument against AUC in Hand (2009) is based on the dependence of this measure on the classifiers' score distribution. As we do not compare

different classifiers in this paper, but our goal is to find a measure of classifiers performance that is able to reflect the quality of data, we can proceed with modifications of AUC. We will use the AUC as a simple enough example of possible advantages of reflecting data quality in the assessment of classifier performance.

4. Fuzzification of ROC

So far we have supposed that the set of *signal* instances and the set of *noise* instances are crisp. That is we have supposed that we know the actual diagnosis for each object from the training set with certainty. We have also paid no attention to the quality of the data that characterize each instance to be classified.

Parasuraman et al. (2000) introduce a fuzzification of the signal detection theory for fuzzy signals or responses. They assume that both the signal and the response (the decision made by the classifier) can be fuzzy (in our context it could mean certain to a given degree within [0,1]). Let us consider a set of instances $\{x_1, \ldots, x_k\}$. Let $r_i \in [0,1]$ be the output of the classifier for an instance x_i describing the degree to which a *Pos* conclusion was suggested by the classifier for x_i, $i = 1, \ldots, k$. This way we assume that the interval $[0,1]$ of classifier outputs is in fact a continuum representing the degree to which a given instance corresponds with the prototype of signal (meets all the criteria for a given diagnosis). At the same time $(1 - r_i)$ describes the degree of correspondence of the given instance with the prototype of noise.

Let s_i be a value from $[0,1]$ describing the degree to which an event (instance) x_i is a signal for all $i = 1, \ldots, k$ (in the case of medical or psychological diagnostics s_i would represent the reliability or sureness of the actual diagnosis - that is the sureness of the value with which we are going to be comparing the classifier output; obviously $(1 - s_i)$ is the degree to which x_i is considered noise). The outcome of the classification can be described by the membership degree of the input x_i into the categories TP, FP, FN, TN with analogical interpretation as those in the classical SDT . The membership degrees are defined for any x_i, $i = 1, \ldots, k$ in the following way:

$$TP(x_i) = \min(s_i, r_i), \tag{8}$$

$$FP(x_i) = \max(r_i - s_i, 0), \tag{9}$$

$$FN(x_i) = \max(s_i - r_i, 0), \tag{10}$$

$$TN(x_i) = \min(1 - s_i, 1 - r_i). \tag{11}$$

This way we can define a fuzzy set of true positive instances as $\widetilde{TP} =$

$\{^{TP(x_1)}/_{x_1}, \ldots, ^{TP(x_k)}/_{x_k}\}$. Analogically, we can define the fuzzy set of false positives, false negatives and true negatives respectively: $\widetilde{FP} = \{^{FP(x_1)}/_{x_1}, \ldots, ^{FP(x_k)}/_{x_k}\}$, $\widetilde{FN} = \{^{FN(x_1)}/_{x_1}, \ldots, ^{FN(x_k)}/_{x_k}\}$ and $\widetilde{TN} = \{^{TN(x_1)}/_{x_1}, \ldots, ^{TN(x_k)}/_{x_k}\}$. Obviously each instance can partially belong to more than one these fuzzy sets. The fuzzy set of signal instances can be defined as $\tilde{S} = \{^{s_1}/_{x_1}, \ldots, ^{s_n}/_{x_k}\}$ and the fuzzy set of all noise instances as $\tilde{N} = \{^{1-s_1}/_{x_1}, \ldots, ^{1-s_n}/_{x_k}\}$. We can observe, that since r_i and s_i are interpreted as membership degrees, there is no need to set a threshold explicitly. The choice of some (fuzzy) threshold is implicitly present in determining r_i, that is embedded in the classifier itself. In accordance with Parasuraman et al. (2000) we can now generalize (4) - (7) using (8) - (11) to define the following classifier performance characteristics:

$$TP\_rate = \frac{Card(\widetilde{TP})}{Card(\widetilde{TP} \cup \widetilde{FN})} = \frac{\sum_{i=1}^{k} TP(x_i)}{\sum_{i=1}^{k} s_i}, \tag{12}$$

$$FP\_rate = \frac{Card(\widetilde{FP})}{Card(\widetilde{TN} \cup \widetilde{FP})} = \frac{\sum_{i=1}^{k} FP(x_i)}{\sum_{i=1}^{k} (1 - s_i)}, \tag{13}$$

$$FN\_rate = \frac{Card(\widetilde{FN})}{Card(\widetilde{TP} \cup \widetilde{FN})} = \frac{\sum_{i=1}^{k} FN(x_i)}{\sum_{i=1}^{k} s_i}, \tag{14}$$

$$TN\_rate = \frac{Card(\widetilde{TN})}{Card(\widetilde{TN} \cup \widetilde{FP})} = \frac{\sum_{i=1}^{k} TN(x_i)}{\sum_{i=1}^{k} (1 - s_i)}. \tag{15}$$

This approach however does not allow us to construct a ROC curve, as we have only a single combination (FP_rate, TP_rate) available. Let us also remark, that in case that $r_i \in \{0, 1\}$ and $s_i \in \{0, 1\}$ this approach is reduced to the classic ROC. This way classic ROC is a special case of this approach.

In accordance with Castanho et al. (2007) we can construct a fuzzy ROC curve by making the threshold explicit. Let us consider the same set of instances $\{x_1, \ldots, x_k\}$. Let again $r_i \in [0, 1]$ be the output of the classifier for an instance x_i. Let $\tilde{t} \in F([0, 1])$ be a fuzzy set on $[0, 1]$ describing a positive outcome of the classifier and let us suppose, that $\mathrm{hgt}(\tilde{t}) = 1$. This way we have generalized the crisp threshold t used in classic ROC into a fuzzy threshold, that for each value r_i describes a degree $\tilde{t}(r_i)$ to which given x_i should be considered *Pos* for all $i = 1, \ldots, k$. Analogically to the previous case $(1 - \tilde{t}(r_i))$ describes the degree to which an x_i should be considered *Neg*. Let us keep the meaning of $s_i \in [0, 1]$ describing the degree to which an instance x_i is a signal for all $i = 1, \ldots, k$.

The outcome of the classification can now be described for any $\tilde{t} \in F([0, 1])$ by the membership degree of the input x_i into the categories $TP_{\tilde{t}}$, $FP_{\tilde{t}}$, $FN_{\tilde{t}}$, $TN_{\tilde{t}}$:

$$TP_{\tilde{t}}(x_i) = \min(s_i, \tilde{t}(r_i)), \tag{16}$$

$$FP_{\tilde{t}}(x_i) = \max\left(\tilde{t}(r_i) - s_i, 0\right), \tag{17}$$

$$FN_{\tilde{t}}(x_i) = \max\left(s_i - \tilde{t}(r_i), 0\right), \tag{18}$$

$$TN_{\tilde{t}}(x_i) = \min\left(1 - s_i, 1 - \tilde{t}(r_i)\right). \tag{19}$$

Having defined the fuzzy sets of signal, noise, true positive, false positive, false negative and true negative instances analogically to the previous case, we obtain the following classifier performance characteristics:

$$TP\_rate_{\tilde{t}} = \frac{Card(\widetilde{TP_{\tilde{t}}})}{Card(\widetilde{TP_{\tilde{t}}} \cup \widetilde{FN_{\tilde{t}}})} = \frac{\sum_{i=1}^{k} TP_{\tilde{t}}(x_i)}{\sum_{i=1}^{k} s_i}, \tag{20}$$

$$FP\_rate_{\tilde{t}} = \frac{Card(\widetilde{FP_{\tilde{t}}})}{Card(\widetilde{TN_{\tilde{t}}} \cup \widetilde{FP_{\tilde{t}}})} = \frac{\sum_{i=1}^{k} FP_{\tilde{t}}(x_i)}{\sum_{i=1}^{k} (1 - s_i)}, \tag{21}$$

$$FN\_rate_{\tilde{t}} = \frac{Card(\widetilde{FN_{\tilde{t}}})}{Card(\widetilde{TP_{\tilde{t}}} \cup \widetilde{FN_{\tilde{t}}})} = \frac{\sum_{i=1}^{k} FN_{\tilde{t}}(x_i)}{\sum_{i=1}^{k} s_i}, \tag{22}$$

$$TN\_rate_{\tilde{t}} = \frac{Card(\widetilde{TN_{\tilde{t}}})}{Card(\widetilde{TN_{\tilde{t}}} \cup \widetilde{FP_{\tilde{t}}})} = \frac{\sum_{i=1}^{k} TN_{\tilde{t}}(x_i)}{\sum_{i=1}^{k} (1 - s_i)}. \tag{23}$$

The ROC curve can now be constructed as a plot of $TP\_rate_{\tilde{t}}$ versus $FP\_rate_{\tilde{t}}$ for all possible (and reasonable) values of the fuzzy threshold $\tilde{t} \in F([0,1])$. This is, however a much more complicated task as we need to vary the membership function of $\tilde{t}$. The area under the ROC curve (AUC) can again be computed as a measure of classifier performance.

To apply the fuzzified ROC approach, training sets with known certainty of diagnoses would have to be available. And still the issue of comparability of the certainty level of training instances and the output of the classifier needs to be dealt with. We also need to point out, that the quality of data is still not directly reflected in this approach. However, low quality of data can be taken into account (for example in a fuzzy rule-based classifier by adding rules that would ensure a low value of classifier's output for instances with low data quality). But incorporating data quality into the diagnostics process might not be an easy task. In Section 5 we introduce and discuss a "Don't know principle" and its possible use in reflecting data quality in the classification process.

5. Modification of the classic ROC approach

Let us now propose a modification of the classic ROC analysis, that aims to embed data quality into the performance assessment of classifiers using ROC and AUC. The

quality of data on which the classification is based has great influence on the performance of the classifier. We can expect that only data of high quality will be used during the training of the classifier. If however the real-life data with which the classifier is supposed to work are of variable quality (and particularly when we know that we can not discard instances with lower quality of data) we should use appropriate measures of classifier performance to assess its functionality. Since a classifier is usually trained on high-quality data, it should perform well on data with high quality. Any misclassification of an instance described by high-quality data should be reflected by lowering the "performance score" of the classifier. On the other hand misclassification of instances described by data with poor quality should not be reflected too much in the performance score of the classifier. This is the idea that on unreliable or weird data we can not expect a good classifier to be flawless - in fact it is understandable that the classifier error rate will increase as the quality of data decreases, this however does not make the classifier worse or less functional. To be frank we expect such behavior. We aim to propose here a measure of classifier performance, that will reflect the quality of the data in a way, that punishes misclassifications of instances described by high-quality data and regards misclassifications of instances with low data quality as less serious.

We will consider the *quality of the data* to be a property of the data that the data (each object/instance we want to classify) possess in various extent. We will treat the data-quality as a continuous attribute that can be measured. In fact measurability of quality would be a quite restrictive condition, as quality measurement is not an easy issue. In this paper it would suffice that there is some means of quality assessment - be it by expert assessment, validity scales (as used in psychological questionnaires see e.g. Greene (2000); Hathaway and McKinley (1940); Stoklasa and Talašová (2011)), lie scores, measures of intentional distortion of data by the respondent or similar tools. Let us also suppose that the outputs of these quality measures can be linearly transformed onto a [0,1] interval (1 meaning top quality - good information value; 0 meaning poor quality - interpretation almost impossible). We will call this transformed quality measure $q_i \in [0,1]$ a *quality rate* of the data instance i. This way each instance or object $x_i \in \{x_1, \ldots, x_n\}$ we want to classify will be assigned a quality rate q_i. We can define a fuzzy set of quality data instances as $\widetilde{QDI} = \{{}^{q_1}\!/_{x_1}, \ldots, {}^{q_n}\!/_{x_n}\}$. This fuzzy set will play an important role in the proposed modification of the classic ROC analysis.

Let us now suppose that we have the same continuous classifier as was considered in Section 3. This classifier assigns to each instance $x_i \in \{x_1, \ldots, x_n\}$ a value $cl_i \in [0,1]$. As we already have the set of quality data instances defined as a fuzzy set, we will define the sets of signal, noise, true positive, false positive, false negative and true negative instances respectively as fuzzy sets as well in the following way. For any value $t \in [0,1]$:

1. $\widetilde{S} = \{{}^{s_1}/_{x_1}, \ldots, {}^{s_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $s_i = 1$ iff x_i is a signal instance and $s_i = 0$ otherwise,
2. $\widetilde{N} = \{{}^{1-s_1}/_{x_1}, \ldots, {}^{1-s_n}/_{x_n}\}$,
3. $\widetilde{TP_t} = \{{}^{\alpha_1}/_{x_1}, \ldots, {}^{\alpha_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $\alpha_i = 1$ iff $[(cl_i \geqq t) \wedge (s_i = 1)]$ and $\alpha_i = 0$ otherwise
4. $\widetilde{FP_t} = \{{}^{\beta_1}/_{x_1}, \ldots, {}^{\beta_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $\beta_i = 1$ iff $[(cl_i \geqq t) \wedge (s_i = 0)]$ and $\beta_i = 0$ otherwise
5. $\widetilde{FN_t} = \{{}^{\gamma_1}/_{x_1}, \ldots, {}^{\gamma_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $\gamma_i = 1$ iff $[(cl_i < t) \wedge (s_i = 1)]$ and $\gamma_i = 0$ otherwise,
6. $\widetilde{TN_t} = \{{}^{\delta_1}/_{x_1}, \ldots, {}^{\delta_n}/_{x_n}\}$, where for any $i = 1, \ldots, n$, $\delta_i = 1$ iff $[(cl_i < t) \wedge (s_i = 0)]$ and $\delta_i = 0$ otherwise.

Obviously $\alpha_i + \beta_i + \gamma_i + \delta_i = 1$ for all $i = 1, \ldots, n$. Now we can define the modified characteristics of the classifier using the above mentioned definitions and $\widetilde{QDI}$. For any value of threshold $t \in [0, 1]$ we now propose:

$$TP\_rate_t = \frac{Card(\widetilde{TP_t} \cap \widetilde{QDI})}{Card(\widetilde{S} \cap \widetilde{QDI})} = \frac{Card(\widetilde{TP_t} \cap \widetilde{QDI})}{Card[(\widetilde{TP_t} \cup \widetilde{FN_t}) \cap \widetilde{QDI}]}, \tag{24}$$

$$FP\_rate_t = \frac{Card(\widetilde{FP_t} \cap \widetilde{QDI})}{Card(\widetilde{N} \cap \widetilde{QDI})} = \frac{Card(\widetilde{FP_t} \cap \widetilde{QDI})}{Card[(\widetilde{TN_t} \cup \widetilde{FP_t}) \cap \widetilde{QDI}]}, \tag{25}$$

$$FN\_rate_t = \frac{Card(\widetilde{FN_t} \cap \widetilde{QDI})}{Card(\widetilde{S} \cap \widetilde{QDI})} = \frac{Card(\widetilde{FN_t} \cap \widetilde{QDI})}{Card[(\widetilde{TP_t} \cup \widetilde{FN_t}) \cap \widetilde{QDI}]}, \tag{26}$$

$$TN\_rate_t = \frac{Card(\widetilde{TN_t} \cap \widetilde{QDI})}{Card(\widetilde{N} \cap \widetilde{QDI})} = \frac{Card(\widetilde{TN_t} \cap \widetilde{QDI})}{Card[(\widetilde{TN_t} \cup \widetilde{FP_t}) \cap \widetilde{QDI}]}. \tag{27}$$

We can now construct the ROC curve analogically as in the crisp case and compute the AUC. As a result of this modification of the ROC, high-quality data influence the classifier performance measure (here represented by the AUC) more than low-quality data. The proposed approach is a generalization of the non-fuzzy approach described by Fawcett (2004). The crisp ROC approach is a special case of our approach, where $q_1 = q_2 = \cdots = q_n$. In the following section, we will discuss the performance of the proposed modification of the ROC analysis in comparison with the classical one and provide a real-life data example of its performance.

6. Results

In this section, we present a numerical study of the proposed modification of ROC on artificial data. We also comment on the "Don't know principle" - a rule of thumb that concerns data quality and classification and we conclude with a real-life example from psychological diagnostics. The fuzzified ROC will not be considered here for as we aim to discuss the role of data quality in classifier performance assessment and its effect will can be better demonstrated under non-fuzzy signals. We therefore refer to Castanho et al. (2007) for examples of the fuzzified ROC.

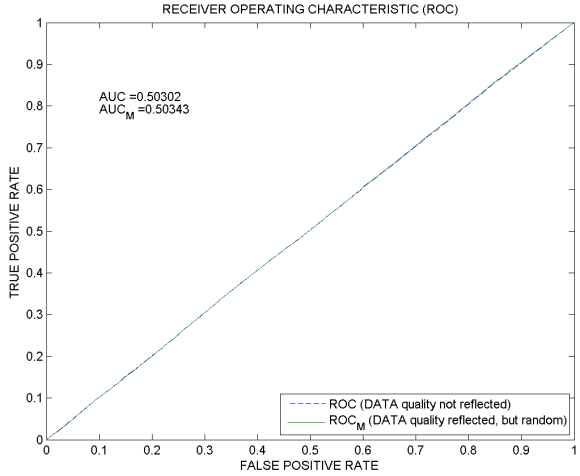


FIGURE 1.
ROC curves for the classic and modified ROC analysis for artificial data (randomly assigned diagnoses and data-quality).

6.1. Numerical study

In the numerical example, we will consider the modified ROC and the classical one. To show how the modified ROC works on data and to compare it with the classic ROC, we have randomly generated 100 000 instances of data $\{x_1, \ldots, x_{100000}\}$. Each instance was randomly assigned a class (either $s_i = 0$ for Noise, or $s_i = 1$ for a Signal, $i = 1, \ldots, 100000$). Each instance was randomly assigned a hypothetical classifier output $cl_i \in [0, 1]$ and a quality measure $q_i \in [0, 1]$ for all $i = 1, \ldots, 100000$. We assume, that all cl_i are the outputs of a classifier that has been trained on a data set containing high-quality data. The type of classifier plays no role. The data set $\{x_1, \ldots, x_{100000}\}$ is now considered as a testing set to assess the performance of the classifier. We aim to stress some of the important characteristics of the proposed modification of ROC - such a setting was therefore chosen

mainly for the sake of clarity and easy interpretability of the results. As all the data are generated randomly (from a uniform distribution), we would expect the classifier to perform "randomly" that is, the area under the ROC curve should both in the case of classic ROC (AUC) and in the modified ROC (AUC_M; M stands here for "modified to reflect the quality of data") be close to 0.5. In accordance with this expectation Figure 1 shows, that according to both measures - that is AUC and AUC_M the classifier performs randomly.

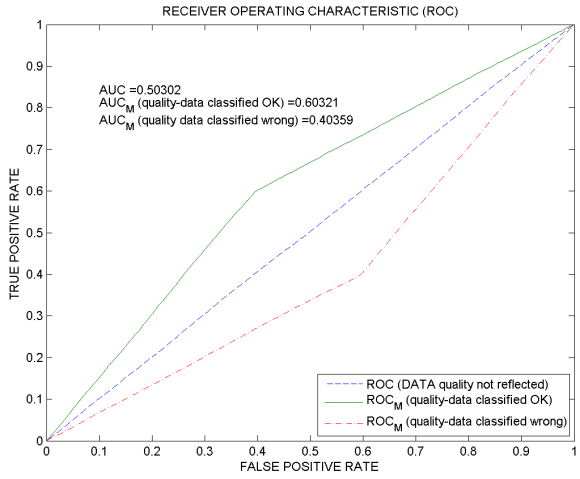


FIGURE 2.
ROC curves for the classic and modified ROC analysis for artificial data. Classic ROC and AUC do not reflect data quality. Two modified ROC curves and the respective AUC_M curves reflect the quality of the data.

Let us now take the quality of the data into account. As we have discussed in previous sections, we can expect that any classifier we will need to assess will be trained on high-quality data and its performance will be optimized on the training set. Let us now therefore suppose that on the generated set the classifier performs well on high-quality data and makes mistakes on the data with low quality. To model this, we need to alter our example in the following way. We will keep the values cl_i and the classification into Noise and Signal unchanged. We will however alter the q_i for each instance in such a way, that if for a given instance x_i the classifier output is close to the proper classification (cl_i is close to s_i) we set q_i higher, if (cl_i is far from s_i) we set q_i lower. Modifications of q_i were performed in the following way in our example: if $|cl_i - s_i| < 0.5$ then $q_i = \min\{1, 0.5 + (0.5 * rand_{[0,1]} - 0.3 * rand_{[0,1]})\}$, otherwise $q_i = \max\{0, 0.5 - (0.5 * rand_{[0,1]} - 0.3 * rand_{[0,1]})\}$, where $rand_{[0,1]}$ is a random number generated from $[0, 1]$ under uniform distribution.

We can also consider another situation, where the classifier performs well on low-

quality data and makes mistakes on high-quality data instances. For this analogical modification of q_i was suggested - for instances where cl_i is close to s_i the q_i was set low and for instances where cl_i is far from s_i the q_i was se high (analogical modifications as described in the previous paragraph were performed). Figure 2 summarizes the results. The three ROC curves in Figure 2 represent the same classifier and its performance on the testing set, the three cases differ by the fact on what data the classifier makes mistakes. We can easily see, that if we disregard the quality of the data, the performance of the classifier can be considered random (AUC close to 0.5). However when we add the information concerning data quality, we can now distinguish between a classifier, that makes mistakes mainly on low-quality data (AUC close to 0.6) and a classifier, that misclassifies mainly high-quality data (AUC close to 0.4). Misclassification of high-quality data is seen as worse that random performance. These are the conclusions that are expected and intuitive given the fact that we want to have an idea about the functioning of our classifier. The modified ROC analysis that takes into account the quality of the data, seems to reflect well the fact that we do not mind making mistakes on low-quality data - it is the quality data that have to be classified properly without any question. The classic ROC has no means of distinguishing between "understandable misclassifications" and "intolerable ones".

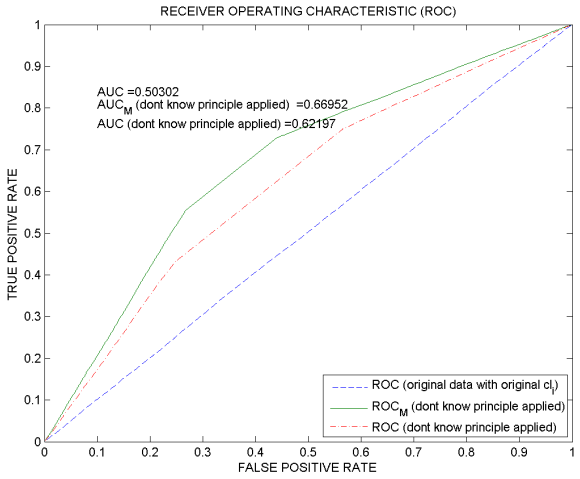


FIGURE 3.
ROC curve and AUC for the original data compared to the ROC and AUC for data after the application of the "don't know principle". The modified ROC and AUC_M are also provided for the case after application of the "don't know principle" (that is after setting $cl_i = 0.5$ for all instances where $q_i < 0.4$).

6.2. Don't know principle

Here we would like to explore how the quality of data can be used in the classification process and hence captured at least partially in the classic ROC. We have already mentioned, that the quality of data may be reflected even sooner than in the phase of classifier performance assessment. In fact the output of the classifier (the diagnosis) can already reflect the quality of the data. When designing classifiers, it is common practice to discard low quality data as not reliable and not to work with these at all. There might, however, be situations, when we can not afford to discard some instances simply on behalf of their low reliability (low quality of the data). In these cases it might be useful to utilize what we call the "don't know principle". This intuitive rule of thumb suggests, that if we need to work with low-quality data, we should not add such rules in the classification process (either by modifying parameters of the classifier, functions rules in a rule base) that suggest either one outcome or another. We should rather incorporate rules in the form *if the quality of data is low, then we do not know which class to assign*. In binary classification, this might mean assigning low-quality data instances such a value, that lies directly in between the two values typical for each class. If signal is represented by 1 and noise by 0, low-quality data instances might be assigned 0.5. Figure 3 shows the results of

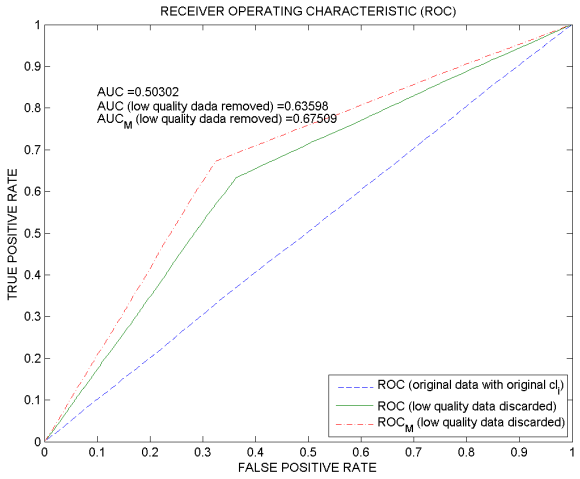


FIGURE 4.
ROC curve and AUC for the original data compared to the ROC and AUC for data after discarding data instances with low quality. The modified ROC and AUC_M are also provided for the after discarding data instances with low quality (that is after removing all data instances for which $q_i < 0.4$).

utilizing this rule of thumb on the artificial data set presented in previous subsection. For simplicity low quality was defined here as $q_i < 0.4$. Both classic and modified ROC curves

are plotted for the data where the don't know principle was applied and the AUC and AUC_M is computed. The same is plotted and computed for the same data, but without the use of the don't know principle. The case, where classifier performs better on low-quality data is not considered here - we suppose from now on that after proper training of the classifier it should perform well on high-quality data. That is the figure depicts a situation where well classified instances are expected to have high quality of data. We can see, that both measured by the AUC or the AUC_M, the performance of the classifier gets better if we use the don't know principle - we have assured that $|cl_i - s_i| = 0.5$ for all instances of data with low quality (as data with low quality were common source of misclassifications, we have reduced the distance $|cl_i - s_i|$ for many low-quality data instances from a higher value to 0.5). The AUC_M is larger than the AUC as using the "don't know principle", because even though all the instances of low-quality data we assigned a $cl = 0.5$ they still retain the low quality.

We should note here, that the suggested rule of thumb alters the outputs of the classifier. The AUC computed from the changed data is then a measure of performance of a modified classifier - one that comprises the "don't know principle" and assigns classes accordingly for low-quality data.

In Figure 4 we can see (on the same set of 100000 testing data instances) what is the effect of discarding instances with low quality. Obviously both the AUC and the AUC_M got bigger as we have removed instances with low quality of data (since these were mainly associated with misclassification). As we can see, there are no large differences between the respective AUC and AUC_M depicted in Figures 3 and 4 for our chosen definition of low-quality $q_i < 0.4$. This would suggest, that eliminating instances with low validity can be replaced by the use of the "don't know principle" while maintaining the whole set of data.

Table 1. Fuzzy rule base applied in the classifier

Rule number	*lar*	*sar*	Output (symptoms are)
1	High	High	Present
2	High	Average	Present
3	High	Low	Possibly present
4	Average	High	Present
5	Average	Average	Possibly present
6	Average	Low	Not present
7	Low	High	Possibly present
8	Low	Average	Not present
9	Low	Low	Not present

6.3. Example from psychological diagnostics

So far, we have explored the properties of the proposed modification of ROC analysis on artificial data. As the need for such a modification came from the field of psychological diagnostics, we will now present a short and simplified case from this area. In Stoklasa and Talašová (2011) a fuzzy rule based classifier was suggested to be used for the interpretation of outputs of a psychological questionnaire to decide, whether converse symptoms are present in a patient or not. This questionnaire was the Minnesota Multiphasic Personality Inventory (more precisely its second revision MMPI-2, see e.g. Greene (2000); Hathaway and McKinley (1940) for more details on this method). We will use this method as a basis

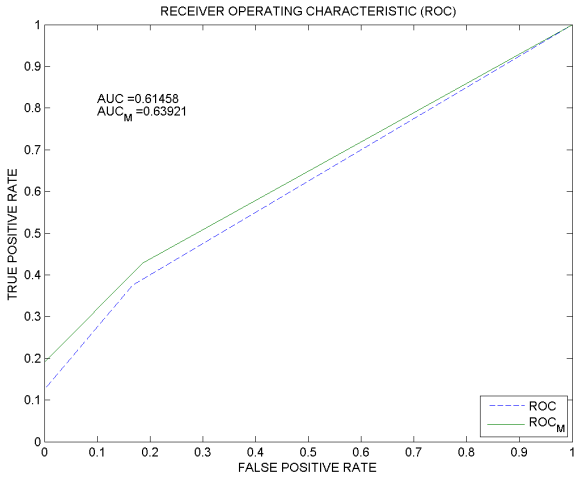


FIGURE 5.
MMPI-2 diagnostics: ROC curve and AUC for the classic ROC analysis compared to the results (ROC curve and AUC_M) computed using the modified ROC analysis suggested in this paper on real-life data.

for our example, as it provides several measures for data validity, ranging from the number of unanswered questions throw the consistency of answers to similar items to deliberate distortion of answers and tendency to look more healthy or ill. Each of these distortions of data is checked by a validity scale - that is certain measures of these distortions exist. Seven validity scales were chosen by Stoklasa and Talašová (2011) to represent the overall validity of the data provided by the patient through the MMPI-2. For each validity scale a fuzzy set representing its acceptable scores was defined. A fuzzy set representing a *prototype of an overall valid MMPI-2 protocol* was then defined as a Cartesian product of the seven fuzzy sets representing acceptable scores of each validity scale. For each protocol (an instance of data, based on which we want to assign a diagnosis to the given patient) a validity rate vr was computed as the membership degree of the 7-tuple of the validity

scale scores to the fuzzy set describing the prototype of an overall valid MMPI-2 protocol. In the example presented here, we will consider 20 MMPI-2 protocols as a validation set. The validity rate of protocol i ($vr_i \in [0,1]$) can be interpreted as a measure of the quality of the data q_i we need to be able to use the modified ROC analysis, that is we set $q_i = vr_i$ for all $i = 1, \ldots, 20$.

Based on each protocol, two other measures were computed for the diagnostics purposes - the location appropriateness rate $lar_i \in [0,1]$ and the shape appropriateness rate $sar_i \in [0,1]$ for all $i = 1, \ldots, 20$ (see Stoklasa and Talašová (2011) for more details). A linguistic fuzzy rule base reflecting expert diagnostician's knowledge and experience with the method was used to classify the protocols (it presented in Table 1). The outputs of the classifier cl_i are summarized in Table 2 as well as the respective values of $vr_i = q_i$, lar_i and sar_i.

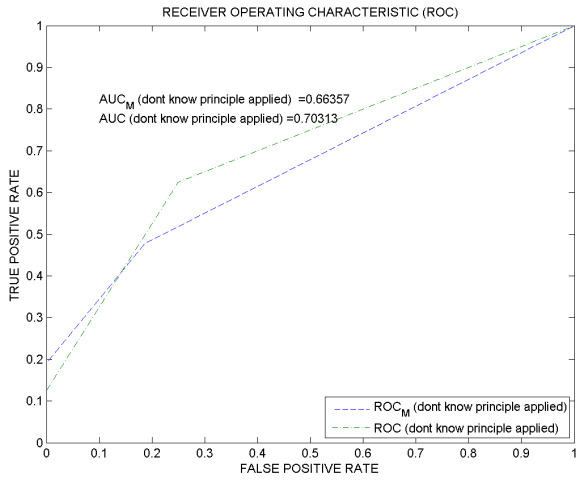


FIGURE 6.
MMPI-2 diagnostics: ROC curve and AUC for data after the application of the "don't know principle" compared to the modified ROC curve and AUC_M also for the case after application of the "don't know principle" (that is after setting $cl_i = 0.5$ for all instances where $q_i < 0.4$).

What interests us here is how to assess the performance of the proposed classifier on the validation set of 20 protocols, where quality rates of the data instances are known. Classic and modified ROC curves were constructed and AUC and AUC_M were computed for this purpose - results are summarized in Figure 5. We can see, that AUC_M is slightly larger than classic AUC. This can be attributed mainly to the fact, that the misclassified protocols PR_{13} and PR_{19} have low validity rate (and hence low quality of data). These misclassifications are therefore seen in the modified version of the ROC suggested in this paper not as an essential failure of the classifier, but as a natural result of the low quality

of the data. Due to the variable quality of the data, the modified ROC suggests a bit better assessment of the performance of the classifier. Another reason why to reflect the

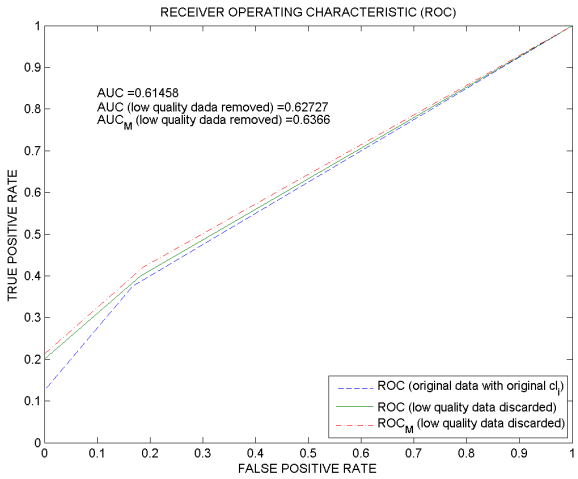


FIGURE 7.
MMPI-2 diagnostics: ROC curve and AUC for the original data compared to the ROC and AUC for data after discarding data instances with low quality ($q_i < 0.4$). Modified ROC and AUC_M are also provided.

quality of the data in the performance assessment process of classifiers is the possibility of testing expert knowledge (here in the area of diagnostics) on data. The classifier presented in Stoklasa and Talašová (2011) can be either approached as a mathematical model for determining into which class each object belongs, or alternatively as a model of everyday practice of an expert. If the second view is adopted, we need to accept that low-quality instances of data cannot be discarded - to each patient a diagnosis has to be assigned. The proposed modification of the ROC analysis is in our opinion a step towards more realistic assessment of performance under everyday (that is not idealized) conditions.

As the "don't know principle" was discussed in the previous subsection, we can now explore its effect on real life data. The results (again with low validity defined by $q_i < 0.4$) are summarized in Figure 6. We can see that in our particular example there is no substantial improvement of the performance of the classification (even though the AUC and AUC_M suggest that utilizing the "don't know principle" the performance of the classifier has improved slightly). What is interesting here is the fact, that the modified ROC suggests a smaller improvement of the performance of the classifier than the classic ROC. This is caused by the fact that only three outputs of the classifier were actually changed: cl_4 from 0 to 0.5 ($q_4 = 0$), cl_{13} from 0 to 0.5 ($q_{13} = 0$) and cl_{19} from 0 to 0.5 ($q_{19} = 0.25$). Due to the fact that protocols 4 and 13 are considered completely unreliable, any change

in cl_4 or cl_{13} will not affect the performance assessment of the classifier using the modified ROC at all (the change of cl_{19} will also have only a small impact on AUC_M considering $q_{14} = 0.25$ is quite low). This is why the AUC_M measure registered a lower increase in performance quality. In our opinion this is an intuitive behavior of a performance measure. Claiming, that changing the classification of an object based on completely unreliable piece of data improves the performance of a classifier seems illogical.

7. Conclusion

In this paper, we have presented a modification of the classical ROC analysis as a tool for performance assessment of classifiers. We have done so in the context of classic ROC analysis and its fuzzification, summarizing each of these approaches and pointing out possible shortcomings when the quality of data varies. We have argued, that the during the assessment of classifier performance information concerning the reliability (quality) of inputs leading to misclassifications should be reflected. As classifiers are usually designed and trained on high-quality data, bad performance on high-quality data is intolerable. This is captured by the classic ROC curve and the AUC as a measure of classifier performance. On the other hand, if misclassifications happen based on low-quality inputs, is should not be regarded as equally severe - classifiers are developed using high-quality data after all. This is something that the classic nor the fuzzified ROC do not reflect.

In real-life applications, the quality of data may vary - there are situations where low-quality data are the only inputs we have for some instances we need to classify. Under these circumstances we can not discard data instances of lower quality - this could mean for example that a doctor refuses to assign a diagnosis to a patient. We use the tools developed for high-quality data (and we need these as we usually can not afford to make mistakes on high-quality data) but we need to interpret the results of these methods in context of the quality of the inputs. Although this might be a challenging task for designing decision support models and systems for practice, it is a path we need to follow in order to keep in touch in the demands of everyday problems.

We have identified humanities as the main field where quality of data may play an important role and hence the main field of application of the proposed modified ROC analysis. This is however mainly due to the fact that the quality of data is a frequent issue in these sciences (mainly due to the fact that human beings are the source of data) and at least some measures for data quality exist here. This does not restrict the possible use of the proposed modification outside humanities. It seems reasonable to reflect the quality of inputs when assessing the performance of any input-output system. Reliable measures of the quality of data (understood as measures of possible distortions, or identifiers of lower information value) will in our opinion play an important role in the development of

suitable models and methods in operations research and decision support system design, as more and more focus is on big data problems and if fact more and more data is available.

The modified approach suggested in Section 5 includes information concerning the quality of the data on which the classifier works. As such the proposed modification regards misclassifications of low-quality data instances as less severe than misclassifications of high-quality data instances. This is, in our opinion, an intuitive and fair way of assessing the performance of classifiers. We have chosen ROC curves and the area under these curves as an example of a measures of classifier performance. We have shown on simulated and real life data, what are the consequences of incorporating data quality into the classifier performance assessment process. We have also shortly discussed, how the quality of data can be incorporated into the classifier itself via the "don't know principle". We the paper we have presented some evidence that the quality of data deserves the attention of researchers, as it is something that the practitioners need to be able to deal with. The presented results of the numerical study as well as the practical example give support to our claim that the proposed modification of ROC might be one of the first steps on this way.

Discarding data with low quality ($q_i < 0.4$) in such a small validation set is not very reasonable. Since we have pointed this out, we can at least plot the result of such a step in Figure 7. Four instances of data were removed from the data set (protocols 4, 10, 13, 19). We can see, that the performance of the classifier got slightly better both according to the AUC and AUC_M. Again removing invalid data from the data set does not seem to improve the performance of the classifier more than utilizing the "don't know principle".

Table 2. Fuzzy classifier outputs summary

Protocol	Diagnosis	*vr*	*lar*	*sar*	*cl*
PR_1	N	0.500	0.0	0.000	0.000
PR_2	N	0.500	0.0	0.000	0.000
PR_3	N	0.500	0.0	0.000	0.000
PR_4	N	0.000	0.0	0.000	0.000
PR_5	S	0.960	1.0	0.000	0.500
PR_6	N	1.000	0.0	1.000	0.500
PR_7	N	0.705	0.0	0.000	0.000
PR_8	S	0.980	1.0	0.375	0.875
PR_9	N	0.720	0.0	0.000	0.000
PR_{10}	S	0.262	1.0	0.000	0.500
PR_{11}	N	0.910	0.0	0.000	0.000
PR_{12}	N	1.000	0.0	0.000	0.000
PR_{13}	S	0.000	0.0	0.000	0.000
PR_{14}	S	0.960	0.2	0.000	0.000
PR_{15}	N	1.000	0.0	0.000	0.000
PR_{16}	N	0.541	0.0	1.000	0.500
PR_{17}	N	1.000	0.0	0.000	0.000
PR_{18}	S	0.720	0.0	0.000	0.000
PR_{19}	S	0.250	0.0	0.000	0.000
PR_{20}	S	1.000	0.2	0.000	0.000

References

Bhatikar, S. R. , DeGrof, C. & Mahajan, R. L. (2005). A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics. *Artificial Intelligence in Medicine*, 33(3), 251-260.

Bielza, C. , Robles, V. & Larrañaga, P. (2011). Regularized logistic regression without a penalty term: An application to cancer classification with microarray data. *Expert Systems with Applications*, 38(5), 5110-5118.

Bonissone, P., Cadenas, J. M., Garrido, M. C. & Daz-Valladares, R. A. (2010). A fuzzy random forest. *International Journal of Approximate Reasoning*, 51(7), 729-747.

Castanho, M. J. P., Barros, L. C., Yamakami, A. & Vendite, L. L. (2007). Fuzzy receiver operating characteristic curve: an option to evaluate diagnostic tests. *IEEE Transactions on Information Technology in Biomedicine*, 11(3), 244-250.

De Gaetano, A., Panunzi, S., Rinaldi, F., Risi, A. & Sciandrone, M. (2009). A patient adaptable ECG beat classifier based on neural networks. *Applied Mathematics and Computation*, 213(1), 243-249.

Dubois, D. & Prade, H. (2000). *Fundamentals of fuzzy sets*. Kluwer Academic Publishers, Dordrecht.

Egan, J. (1975). *Signal detection theory and ROC analysis.* Academic, New York.

Evangelista, P. F., Bonnisone, P., Embrechts, M. J. & Szymanski, B. K. (2005a). Fuzzy ROC Curves for Unsupervised Nonparametric Ensemble Techniques. In *Proc. International Joint Conference on Neural Networks 2005, Montreal,* 3040-3045.

Evangelista, P. F., Bonnisone, P., Embrechts, M. J. & Szymanski, B. K. (2005b). Fuzzy ROC curves for the 1 class SVM: application to intrusion detection. In *Proc. 13th European Sympozium on Artificial Neural Networks ESANN05, Burges, 2005*, 345-350.

Fawcett, T. (2004). *ROC graphs: notes and practical considerations for researchers*. Technical report HPL-2003-4, HP Labs.

Flach, P., Hernández-Orallo, J. & Ferri, C. (2011). A Coherent Interpretation of AUC as a Measure of Aggregated Classification Performance. In *Proceedings of the 28th International Conference on Machine Learning, Bellevue,WA, USA, 2011*.

Godil, S. S., Shamim, M. S., Enam, S. A. & Qidwai, U. (2011). Fuzzy logic: A "simple" solution for complexities in neurosciences? *Surgical neurology international*, 24(2), 109-117. http://surgicalneurologyint.com/article.asp?issn=2152-7806;year=2011;volume=2;issue=1;spage=24;epage=24;aulast=Godil. Accessed 8 May 2014.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics.* Wiley, New York.

Greene, R. L. (2000). *The MMPI-2: An interpretive manual.* Allyn and Bacon, Boston.

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, 77(1), 103-123.

Hathaway, S. R. & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of psychology*, 10(2), 249-254.

Holeček, P. & Talašová, J. (2010). Designing Fuzzy Models of Multiple-Criteria Evaluation in FuzzME Software. In *Proceedings of the 28th International Conference on Mathematical Methods in Economics,* 2010, 250-256.

Krzanowski, W. J. & Hand, D. J. (2009). *ROC curves for continuous data.* London: Chapman and Hall.

Kuncheva, L. I. (2000). *Fuzzy classifier design.* Physica Verlag, Heidelberg; New York.

Markey, M. K., Tourassi, G. D. & Floyd, C. E. (2003). Decision tree classification of proteins identified by mass spectrometry of blood serum samples from people with and without lung cancer. *Proteomics*, 9(3), 1678-1679.

Mathworks Inc. (2011). Fuzzy logic toolbox 2.2.13 The Mathworks Inc.

Miller, R. A. (1994). Medical Diagnostic Decision Support Systems - Past, Present, and Future: A Threaded Bibliography and Brief Commentary. *JAMIA*, 1(1), 8-27.

Parasuraman, R., Masalonis, A. J. & Hancock, P. A. (2000). Fuzzy signal detection theory: basic postulates and formulas for analyzing human and machine performance. *Human factors,* 42(4), 636-659.

Parker, C. (2013). On measuring the performance of binary classifiers. *Knowledge Information Systems*, 35(1), 131-152.

Ramírez, J., Górriz, J. M., Segovia, F., Chaves, R., Salas-Gonzalez, D., López, M., Alvarez, I. & Padilla, P. (2010). Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification. *Neuroscience Letters*, 472(2), 99-103.

Robles, V., Larrañaga, P., Peña, J. M., Menasalvas, E., Pérez, M. S., Herves, V. & Wasilewska, A. (2004). Bayesian network multi-classifiers for protein secondary structure prediction. *Artificial Intelligence in Medicine*, 31(2), 117-136

Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehman, H. & Tang, P. C. (2001). Clinical Decision Support Systems for the Practice of Evidence-based Medicine. *JAMIA*, 8(6), 527-534.

Smith, A., Sterba-Boatwright, B. & Mott, J. (2010). Novel application of a statistical technique, Random Forests, in a bacterial source tracking study. *Water Research*, 44(14), 4067-4076.

Stoklasa, J. & Talašová, J. (2011). Using linguistic fuzzy modeling for MMPI-2 data interpretation. In *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011 - part II, 2011, Praha, Czech Republic*, 653-658.

Talašová, J. & Holeček, P. (2009). Multiple-Criteria Fuzzy Evaluation: The FuzzME Software Package. *Proceedings of 2009 IFSA World Congress, July 20-24*, 681-686.

von Altrock, C. (1995). *Fuzzy logic and neurofuzzy applications explained.* Prentice-Hall, New York.

World Health Organisation. (2006). *Mezinárodní klasifikace nemocí - 10. revize: Duševní poruchy a poruchy chování (3rd edition).* Psychiatrické centrum Praha, Praha. [International classification of diseases - (10^{th} revision: Mental and behavioral disorders (3^{rd} edition)]

Zadeh, L. A. (1965). Fuzzy sets. *Inform. Control*, 8(3), 338-353.

Zolghadri, M. J. & Mansoori, A. G. (2007). Weighting fuzzy classification rules using receiver operating characteristics (ROC) analysis. *Information sciences*, 117(11), 2296-2307.

Zhou, X., Kuang-Yu Liu & Wong, S. T.C. (2004). Cancer classification and prediction using logistic regression with Bayesian gene selection. *Journal of Biomedical Informatics*, 37(4), 249-259.

Zou, K. H., O'Malley, A. J. & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115(5), 654-657.

PUBL. V

Krejčí, J. and Stoklasa, J., Fuzzified AHP in the evaluation of R&D results.

Submitted to *Central European Journal of Operations Research*, Springer. (April 2014)

Noname manuscript No.
(will be inserted by the editor)

Fuzzified AHP in the evaluation of R&D results

Jana Krejčí · Jan Stoklasa

Received: date / Accepted: date

Abstract Fuzzification of the Analytic Hierarchy Process (AHP) is of great interest to researchers since it is a frequently used method for coping with complex decision making problems. There have been many attempts to fuzzify the AHP. In this paper we focus particularly on the construction of fuzzy pairwise comparison matrices and on obtaining fuzzy weights of objects from them subsequently. We review the fuzzification of the geometric mean method for obtaining fuzzy weights of objects from fuzzy pairwise comparison matrices. We illustrate here the usefulness of the fuzzified AHP on a real-life problem of the evaluation of quality of R&D results in university environment. The benefits of the presented evaluation methodology and its suitability for quality assessment of R&D results are discussed. When the task of quality assessment in R&D is considered, an important role is played by peer-review evaluation. Evaluations provided by experts in the peer-review process have a high level of subjectivity and can be expected in a linguistic form. New decision-support methods (or adaptations of classic methods) well suited to deal with such inputs, to capture the consistency of experts' preferences and to restrict the subjectivity to an acceptable level are necessary. A new consistency condition is therefore defined here to be used for expertly defined fuzzy pairwise comparison matrices.

Keywords Fuzzy Analytic Hierarchy Process · Fuzzy pairwise comparison matrix · Constrained fuzzy arithmetic · Triangular fuzzy numbers · Evaluation · R&D

Jana Krejčí
Department of Industrial Engineering University of Trento, 38123 Povo, Via Sommarive, Trento, Italy
E-mail: jana.krejci@unitn.it

Jan Stoklasa
Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic
E-mail: jan.stoklasa@upol.cz

1 Introduction

The Analytic Hierarchy process (AHP), developed by T. L. Saaty (1980), is a methodology that supports decision making in a multiple criteria environment. It is an effective tool based on structuring a complex problem into a hierarchy and making pairwise comparisons of objects in one level of the hierarchy against an object in the upper level of the hierarchy. The simplest type of hierachy is a 3-level hierarchy where the goal of a decision making problem is in the first level of the hierarchy, criteria describing the problem are in the second level, and alternatives of the decision making problem are in the third level of the hierarchy. In this simple hierarchy, pairwise comparison matrices of alternatives against each criterion and a pairwise comparison matrix of criteria against the goal of the decision making problem are constructed using elements from Saaty's scale. Weights of objects (alternatives, criteria or categories) are obtained from pairwise comparison matrices and aggregated into overall weights of alternatives afterwards.

The elements from Saaty's scale are real numbers to which linguistic terms expressing intensity of preference of one object over another one are assigned. However, because linguistic terms are vague, fuzzy numbers are more suitable to describe them. Therefore, fuzzification of the AHP has been proposed by many authors. The most well known approaches were published by Van Laarhoven and Pedrycz (1983), Buckley (1985), Cheng and Mon (1994), Chang (1996) and Xu (2000).

In this paper, we will focus only on one part of the AHP which is obtaining weights of objects from pairwise comparison matrices. Aggregation of intermediate weights of criteria and alternatives into overall weights of alternatives will not be dealt with here. We will review the geometric mean method for obtaining weights of objects from pairwise comparison matrices originally proposed by Buckley (1985) and fuzzify weak consistency proposed in Stoklasa et al. (2013). The proposed fuzzification of the AHP will be then applied on a real-life example - an evaluation of scientific monographs. Since quality assessment in R&D is becoming an important issue in many countries, we show what benefits the use of the fuzzified AHP may have in this context. The presented evaluation methodology describes a natural integration of the peer-review component into the evaluation process. The main benefit of the proposed approach is its ability to incorporate peer-review-based (and hence subjective) evaluations. We show how intervals of possible evaluations and default evaluations of objects can be determined using the fuzzified AHP. This provides a good starting point for the subsequent peer-review assessment and provides limits within which the peer-review can affect the evaluation.

The paper is organized as follows: In Section 2, triangular fuzzy numbers and operation with them are defined. In Section 3, the classic AHP is described. Section 4 deals with the proper fuzzification of the AHP, and in Section 5, the proposed method is applied to a real life problem - the evaluation of R&D outcomes (scientific monographs) at the Palacky University in Olomouc, Czech Republic.

2 Triangular fuzzy numbers and operations with them

In this Section, triangular fuzzy numbers and operations with them will be defined.

A fuzzy set $\tilde{a}$ on a universe U is characterised by its membership function $\tilde{a} : U \to [0,1]$. By $Core\,\tilde{a}$ we denote a core of $\tilde{a}$, i.e. $Core\,\tilde{a} := \{u \in U \mid \tilde{a}(u) = 1\}$, and by $Supp\,\tilde{a}$ we denote a support of $\tilde{a}$, i.e. $Supp\,\tilde{a} := \{u \in U \mid a(u) > 0\}$. For any $\alpha \in (0,1]$, $\tilde{a}_\alpha$ means an α-cut of $\tilde{a}$, i.e. $\tilde{a}_\alpha := \{u \in U \mid \tilde{a}(u) \geq \alpha\}$.

A fuzzy number $\tilde{c}$ is a fuzzy set defined on the set of real numbers $\mathbb{R}$ with the following properties: a) $Core\,\tilde{c} \neq \emptyset$, b) for all $\alpha \in (0,1]$, $\tilde{c}_\alpha$ is a closed interval, c) $Supp\,\tilde{c}$ is bounded. Fuzzy number $\tilde{c}$ is said to be positive, when $Supp\,\tilde{c} \subset (0,\infty)$.

A triangular fuzzy number $\tilde{c}$ is a fuzzy number whose membership function is determined by a triplet of real numbers $c_1 \leq c_2 \leq c_3$ in the form

$$\tilde{c}\,(x) = \begin{cases} \frac{x-c_l}{c_2-c_1},\ c_1 \leq x \leq c_2, \\ \frac{c_3-x}{c_3-c_2},\ c_2 < x \leq c_3, \\ 0, \quad \text{otherwise}, \end{cases} \tag{1}$$

where $Core\,\tilde{c} = \{c_2\}$ and $Supp\,\tilde{c} = (c_1, c_3)$. For a triangular fuzzy number $\tilde{c}$, whose membership function is given by (1), the notation $\tilde{c} = (c_1, c_2, c_3)$ will be used hereafter. The real numbers c_1, c_2, and c_3 will be called the lower, the middle, and the upper significant values of the triangular fuzzy number $\tilde{c}$.

A real number $c \in \mathbb{R}$ can be regarded a triangular fuzzy number $\tilde{c}$ where $\tilde{c} = (c, c, c)$. According to this, the classic AHP can be seen as a particular case of the fuzzified AHP where the fuzzification is done by triangular fuzzy numbers.

Mathematics of fuzzy numbers is based on the extension principle introduced by Zadeh (1975). It is known that not even the result of a simple arithmetic operation with triangular fuzzy numbers (e.g. multiplication) has to be a triangular fuzzy number. Similarly it is for the computation of the value of a function of n variables with entries given by triangular fuzzy numbers. In real applications, however, it is common to approximate also these vague outputs by triangular fuzzy numbers. Therefore, we will restrict ourselves to triangular fuzzy numbers in this paper. The supports and the cores of resulting fuzzy numbers will be determined correctly and left and right membership functions of the fuzzy numbers will be approximated by linear functions. According to this we will define simplified arithmetic of triangular fuzzy numbers. Because elements of fuzzy pairwise comparison matrices are positive fuzzy numbers, we will restrict ourselves only to them in definitions of arithmetic operations.

Let $\tilde{c} = (c_1, c_2, c_3)$ and $\tilde{d} = (d_1, d_2, d_3)$ be positive triangular fuzzy numbers. Then we define addition, multiplication and division of the triangular fuzzy numbers as $\tilde{c} + \tilde{d} = (c_1 + d_1, c_2 + d_2, c_3 + d_3)$, $\tilde{c} \cdot \tilde{d} = (c_1 \cdot d_1, c_2 \cdot d_2, c_3 \cdot d_3)$ and $\tilde{c}/\tilde{d} = (c_1/d_3, c_2/d_2, c_3/d_1)$ respectively. The reciprocal of a triangular fuzzy number $\tilde{c} = (c_1, c_2, c_3)$ is defined as $1/\tilde{c} = (1/c_3, 1/c_2, 1/c_1)$.

More generally we can define a simplified version of the extension principle for a continuous function and its entries expressed by triangular fuzzy numbers. Let f be a continuous function, $f : \mathbb{R}^n \to \mathbb{R}$, and let $\tilde{c}_i = (c_{i1}, c_{i2}, c_{i3}), i = 1, \ldots, n$, be triangular fuzzy numbers. Then $f(\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_3)$ is a triangular fuzzy number $\tilde{c} = (c_1, c_2, c_3)$ whose significant values are given in the form

$$c_1 = \min\{f(x_1, \ldots, x_n); x_i \in [c_{i1}, c_{i3}], i = 1 \ldots, n\},$$
$$c_2 = f(c_{12}, \ldots, c_{n2}),$$
$$c_3 = \max\{f(x_1, \ldots, x_n); x_i \in [c_{i1}, c_{i3}], i = 1 \ldots, n\}.$$

In the special case of a non decreasing function f (e.g. addition), the computation of the significant values of the resulting triangular fuzzy number is extremely simple. Let f be a continuous non decreasing function, $f : \mathbb{R}^n \to \mathbb{R}$, and let $\tilde{c}_i = (c_{i1}, c_{i2}, c_{i3}), i = 1, \ldots, n$, be triangular fuzzy numbers defined on $\mathbb{R}$. Then $\tilde{c} = (c_1, c_2, c_3)$ is a triangular fuzzy number defined on $\mathbb{R}$ whose significant values are given as follows:

$$c_1 = f(c_{11}, \ldots, c_{n1}),$$
$$c_2 = f(c_{12}, \ldots, c_{n2}),$$
$$c_3 = f(c_{13}, \ldots, c_{n3}).$$

In case that there are some interactions among the input variables $x_1, \ldots, x_n$ described by a subset $D \subseteq \mathbb{R}^n$, the resulting triangular fuzzy number is described in the following way. Let f be a continuous function, $f : \mathbb{R}^n \to \mathbb{R}$, and let $\tilde{c}_i = (c_{i1}, c_{i2}, c_{i3}), i = 1, \ldots, n$, be triangular fuzzy numbers. Let D be a relation in $\mathbb{R}^n$ describing interactions among the variables. Then $\tilde{c} = (c_1, c_2, c_3)$ is a triangular fuzzy number whose significant values are given as follows:

$$c_1 = \min\{f(x_1, \ldots, x_n); (x_1, \ldots, x_n) \in D \cap [c_{11}, c_{13}] \times \cdots \times [c_{n1}, c_{n3}]\},$$
$$c_2 = f(c_{12}, \ldots, c_{n2}),$$
$$c_3 = \max\{f(x_1, \ldots, x_n); (x_1, \ldots, x_n) \in D \cap [c_{11}, c_{13}] \times \cdots \times [c_{n1}, c_{n3}]\}.$$

This constrained extention principle will be applied on the arithmetic operations with triangular fuzzy numbers throughout this paper.

Triangular fuzzy numbers $\tilde{c}_i \in [0, 1], i = 1, \ldots, n$, are called normalized fuzzy weights if

$$\forall c_i \in \tilde{c}_{i\alpha}\ \exists\ c_j \in \tilde{c}_{j\alpha}, j = 1, \ldots, n, j \neq i : c_i + \sum_{j=1, j\neq i}^{n} c_j = 1, \tag{2}$$

for all $\alpha \in (0, 1]$. A set of triangular fuzzy numbers $\tilde{c}_i = (c_{i1}, c_{i2}, c_{i3}), i = 1, \ldots, n$, satisfies (2) and hence can be called normalized fuzzy weights if

$$\sum_{j=1}^{n} c_{j2} = 1, \qquad c_{j1} + \sum_{i=1, i\neq j} c_{i3} \geq 1, \qquad c_{j3} + \sum_{i=1, i\neq j} c_{i1} \leq 1 \tag{3}$$

hold for each $j \in \{1, \ldots, n\}$.

Table 1 Saaty's 5-point scale.

Intensity of preference	Linguistic term
1	equal preference
3	moderate preference
5	strong preference
7	very strong preference
9	extreme preference

3 Pairwise comparison matrices

In this section, Saaty's scale, construction of pairwise comparison matrices and verifying their consistency, and methods for obtaining weights of objects from pairwise comparison matrices will be revised.

To make pairwise comparisons of objects (criteria, alternatives or categories), Saaty (1980) used a scale of integer numbers from 1 to 9. To each number from the scale a linguistic term describing the preference of one object over another one was assigned. Usually only a 5-point scale of numbers $1, 3, 5, 7$ and 9 given in Tab. 1 is used to make pairwise comparisons of objects. In case that more detailed comparisons are needed a 9-point scale with intermediate values $2, 4, 6$, and 8 can be used. The intermediate values are expressed using the neighbouring linguistic terms and connecting them with the word "'between"'.

Using any of these scales, each two objects (criteria, alternatives or categories) in the same level of the hierarchy are compared with respect to each object in the upper level of the hierarchy. Generally, when we have n objects $x_1, x_2, \ldots, x_n$ in one level of the hierarchy and want to compare them with respect to an object from the upper level of the hierarchy, a square matrix $A = \{a_{ij}\}_{i,j=1}^{n}$ of pairwise comparisons is constructed. An element a_{ij} expresses the intensity of a preference of an object x_i over an object x_j with respect to a particular object from the upper level of the hierarchy. It is obvious that there are ones on the main diagonal as we compare always an object with itself. Further, if the object x_i is a_{ij}-times more important than the object x_j, then obviously the object x_j takes only the $\frac{1}{a_{ij}}$-th of the preference of the object x_j. Therefore, the pairwise comparison matrix $A = \{a_{ij}\}_{i,j=1}^{n}$ must be reciprocal.

The reciprocity is not the only requirement on pairwise comparison matrices. If we want to obtain reasonable weights of objects from a pairwise comparison matrix, the pairwise comparisons have to be entered in the matrix in a reasonable way. It means decision makers should be able to make reasonable pairwise comparisons (they should be consistent in their decisions). According to this, another property of pairwise comparison matrices called consistency was defined.

According to Saaty (1980), a pairwise comparison matrix $A = \{a_{ij}\}_{i=1}^{n}$ is said to be consistent if $a_{ik} = a_{ij} \cdot a_{jk}$, for all $i, j, k \in \{1, \ldots, n\}$. However, the requirement of consistency is very strong and is not reachable for all pairwise

Table 2 Random index RI.

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.52	0.89	1.11	1.25	1.35	1.4	1.45	1.49

comparison matrices. That is caused especially by the fact that Saaty's scale contains only the numbers up to 9. When for example $a_{ij} = 7$ and $a_{jk} = 9$, the pairwise comparison a_{ik} should be 63 to keep the consistency. However, only the numbers up to 9 are allowed. Therefore, using the elements from Saaty's scale, the consistency is not reachable in most cases and the problem of acceptable inconsistency has to be addressed.

Saaty (1980) defined a consistency index CI of a pairwise comparison matrix $A = \{a_{ij}\}_{i,j=1}^{n}$ in the form

$$CI = \frac{\lambda - n}{n - 1}, \tag{4}$$

where n is the number of compared objects and λ is the maximal eigenvalue of the pairwise comparison matrix. It was shown by Ramík (2000) that if a decision maker is absolutely consistent in his judgements, i.e. if the pairwise comparison matrix is consistent, then $\lambda = n$ and, therefore, $CI = 0$. However, as was already mentioned above, meeting the condition $a_{ik} = a_{ij} \cdot a_{jk}$ for all $i, j, k \in \{1, \ldots, n\}$ is usually impossible with limited Saaty's scale. Hence, some degree of inconsistency is allowed.

It is required CI to be close to 0. However, it is difficult to determine a value of CI to which the matrix is regarded as consistent and over which is regarded as inconsistent. Therefore, Saaty (1980) defined a consistency ratio CR in the form

$$CR = \frac{CI}{RI}, \tag{5}$$

where RI is a random consistency index, which is computed separately for each order of the matrix as an average value of consistency indexes of reciprocal matrices that are randomly generated from elements of Saaty's scale. Tab. 2 shows values of RI for matrix orders up to $n = 10$. According to Saaty, a pairwise comparison matrix is regarded as consistent if $CR < 0.1$. Otherwise, the decision maker's judgements about the preferences should be reconsidered. Nevertheless, even the condition $CR < 0.1$ is unreachable for most pairwise comparison matrices of larger dimensions. Therefore, other indexes measuring inconsistency of pairwise comparison matrices were introduced in the literature. An overview of some idexes is given for example by Brunelli et al. (2013).

In this paper, we apply the weak consistency condition introduced by Stoklasa et al. (2013). According to this condition a fuzzy pairwise comparison matrix $A = \{a_{ij}\}_{i,j=1}^{n}$ is said to be weakly consistent if for all $i, j, k \in \{1, 2, \ldots, n\}$ the following holds:

$$a_{ij} > 1 \ \wedge \ a_{jk} > 1 \quad \Rightarrow \quad a_{ik} \geq \max\{a_{ij}, a_{jk}\}, \tag{6}$$

$$(a_{ij} = 1 \wedge a_{jk} \geq 1) \ \vee \ (a_{ij} \geq 1 \wedge a_{jk} = 1) \quad \Rightarrow \quad a_{ik} = \max\{a_{ij}, a_{jk}\}. \quad (7)$$

Using the linguistic terms of the elements from Saaty's scale to make pairwise comparisons of objects, the weak consistency condition is a minimum consistency requirement which should be met to consider a pairwise comparison matrix to be reasonably consistent. In this paper, we combine both approaches; we require all paiwise comparison matrices to be weakly consistent and to meet the condition $CR < 0.1$.

After verifying the consistency of a pairwise comparison matrix, weights of objects can be computed. Originally, Saaty proposed the eigenvalue method for deriving weights of objects from a pairwise comparison matrix. According to this method, weights of objects are calculated as components of the eigenvector corresponding to the largest eigenvalue of the pairwise comparison matrix, i.e. the vector $w = (w_1, \ldots, w_n)$ of weights of n objects is the solution of the equation $Aw = \lambda w$, where A is a pairwise comparison matrix of n objects and λ is a maximal eigenvalue of the matrix A.

The weights of objects can be also obtained by the logarithmic least squares method (LLSM), see Crawford and Williams (1985). By normalizing the weights obtained by the eigenvalue method and by the LLSM very similar but not identical weights are obtained. The result of the LLSM can be also obtained by computing geometric means of the rows of Saaty's pairwise comparison matrix. This method is called the geometric means method and the formula for computing a weight w_i of an i-th object from a pairwise comparison matrix $A = \{a_{ij}\}_{i,j=1}^{n}$ is given as follows:

$$w_i = \sqrt[n]{\prod_{j=1}^{n} a_{ij}}, \qquad i = 1, \ldots, n. \quad (8)$$

Using the weights $w_i, i = 1, \ldots, n$, an absolutely consistent pairwise comparison matrix W can be reconstructed in the form

$$W = \left\{ \frac{w_i}{w_j} \right\}_{i,j=1}^{n}. \quad (9)$$

This pairwise comparison matrix in general differs from the original pairwise comparison matrix used for computing the weights $w_i, i = 1, \ldots, n$.

The weights w_i, $i = 1, \ldots, p$, are usually normalized so that the sum of the resulting weights is equal to 1:

$$v_i = \frac{w_i}{\sum_{j=1}^{n} w_j}, \qquad v_i \geq 0, \ i = 1, \ldots, n, \ \sum_{i=1}^{n} v_i = 1. \quad (10)$$

Table 3 Fuzzified Saaty's 5-point scale.

Fuzzy number	Membership function	Linguistic term
$\widetilde{1}$	$\left(\frac{1}{3}, 1, 3\right)$	equal preference
$\widetilde{3}$	$(1, 3, 5)$	moderate preference
$\widetilde{5}$	$(3, 5, 7)$	strong preference
$\widetilde{7}$	$(5, 7, 9)$	very strong preference
$\widetilde{9}$	$(7, 9, 9)$	extreme preference

Table 4 Fuzzified Saaty's 9-point scale.

Fuzzy number	Membership function	Linguistic term
$\widetilde{1}$	$\left(\frac{1}{2}, 1, 2\right)$	equal preference
$\widetilde{2}$	$(1, 2, 3)$	between equal and moderate preference
$\widetilde{3}$	$(2, 3, 4)$	moderate preference
$\widetilde{4}$	$(3, 4, 5)$	between moderate and strong preference
$\widetilde{5}$	$(4, 5, 6)$	strong preference
$\widetilde{6}$	$(5, 6, 7)$	between strong and very strong preference
$\widetilde{7}$	$(7, 8, 9)$	very strong preference
$\widetilde{8}$	$(7, 8, 9)$	between very strong and extreme preference
$\widetilde{9}$	$(8, 9, 9)$	extreme preference

4 Fuzzy pairwise comparison matrices

As was described in the previous section, Saaty defined a scale of integer numbers with assigned linguistic terms to make pairwise comparisons of objects. However, the meanings of these linguistic terms are vague, and therefore, fuzzification of real numbers from the scale is appropriate. For the fuzzification of Saaty's scale triangular fuzzy numbers are usually used.

Many different approaches to the fuzzification of Saaty's scale have been proposed in literature. We will work with properly fuzzified Saaty's 5-point scale and 9-point scale given in Tab. 3 and Tab. 4 proposed by Enea and Piazza (2004) and Krejčí et al. (2013). Using the elements from these fuzzified Saaty's scales fuzzy pairwise comparison matrices of objects are constructed. Because pairwise comparison matrices are required to be reciprocal in the classic AHP, the reciprocity should be preserved also during the fuzzification.

In the following, a fuzzy pairwise comparison matrix is a reciprocal matrix $\tilde{A} = \{\tilde{a}_{ij}\}_{i,j=1}^{n}$ whose elements $\tilde{a}_{ij}, i \neq j$, are triangular fuzzy numbers from fuzzified Saaty's scales shown in Tab. 3 or Tab. 4 in case that the i-th object is more important than the j-th object, otherwise $\tilde{a}_{ij} = \frac{1}{\tilde{a}_{ji}}$. The elements $\tilde{a}_{ii}, i = 1, \ldots, n$, on the main diagonal are equal to 1 because we always compare an object with itself, and therefore, there is no fuzziness in the comparisons.

In the fuzzified AHP, similarly as in the classic AHP, consistency of fuzzy pairwise comparison matrices has to be verified. However, in many applications of the fuzzified AHP the problem of verifying consistency is neglected, see e.g.

Cheng et al. (2009); Güngör et al. (2009); Kwong and Bai (2002). In Pan (2008); Tesfamariam and Sadiq (2006); Vahidnia et al. (2009), consistency of fuzzy pairwise comparison matrices is verified using Saaty's CR computed from the matrix of middle significant values of fuzzy numbers.

Apart from the sufficient consistency in Saaty's sense (that is CR<0.1), we will also use here a fuzzification of the weak consistency (6) and (7). This way, at least a minimum requirement on consistency represented by the fuzzified weak consistency condition can be checked during the construction of a fuzzy pairwise comparison matrix.

We say that a fuzzy pairwise comparison matrix $\tilde{A} = \{\tilde{a}_{ij}\}_{i,j=1}^{n}$, $\tilde{a}_{ij} = (a_{ij1}, a_{ij2}, a_{ij3})$, is weakly consistent if for all $i, j, k \in \{1, 2, \ldots, n\}$ the following holds:

$$a_{ij2} > 1 \ \wedge \ a_{jk2} > 1 \quad \Rightarrow \quad a_{ik2} \geq \max\{a_{ij2}, a_{jk2}\}, \tag{11}$$

$$(a_{ij2} = 1 \wedge a_{jk2} \geq 1) \ \vee \ (a_{ij2} \geq 1 \wedge a_{jk2} = 1) \quad \Rightarrow \quad a_{ik2} = \max\{a_{ij2}, a_{jk2}\}. \tag{12}$$

Once the consistency of a fuzzy pairwise comparison matrix is checked, fuzzy weights of objects are computed from the matrix. Fuzzification of the geometric mean method will be used here. As was mentioned in the Introduction, the geometric mean method for obtaining fuzzy weights of objects from fuzzy pairwise comparison matrices was for the first time fuzzified by Buckley (1985). However, he did not take into account the reciprocity of pairwise comparison matrices in his formulas. Thus, the resulting fuzzy weights were too vague and did not represent actual fuzzy preferences of objects. Enea and Piazza (2004) introduced proper formulas for obtaining significant values of triangular fuzzy weights $\tilde{v}_i = (v_{i1}, v_{i2}, v_{i3})$, $i = 1, \ldots, n$, of objects from a fuzzy pairwise comparison matrix using the geometric mean method in this form:

$$v_{i1} = \min \left\{ \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj}}} \; ; \begin{array}{l} a_{kj} \in [a_{kj1}, a_{kj3}], \forall j > k, \\ a_{jk} = \dfrac{1}{a_{kj}}, \forall j < k \\ a_{jj} = 1, \forall j \end{array} \right\}, \tag{13}$$

$$v_{i2} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij2}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj2}}}, \tag{14}$$

$$v_{i3} = \max \left\{ \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj}}} \; ; \begin{array}{l} a_{kj} \in [a_{kj1}, a_{kj3}], \forall j > k, \\ a_{jk} = \dfrac{1}{a_{kj}}, \forall j < k \\ a_{jj} = 1, \forall j \end{array} \right\}. \tag{15}$$

These fuzzy weights are less vague than the fuzzy weights obtained by formulas proposed by Buckley and represent actual fuzzy preferences of objects (Krejčí et al., 2013). However, the formulas (13) and (15) for obtaining the lower and the upper significant values of a fuzzy weight $\tilde{v}_i$ are computationally demanding as we are looking for a minimum, resp. maximum, of a function of $\frac{n(n-1)}{2}$ variables. (The matrix $\tilde{A}$ has n^2 elements from which n on the main diagonal are equal to 1 and the elements below the main diagonal are the reciprocals of the corresponding elements above the main diagonal.) As was shown by Krejčí et al. (2013), the formulas (13) and (15) can be simplified so that the optimization is done over $\frac{n(n-1)}{2} - (n-1)$ variables using the following formulas:

$$v_{i1} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij1}}}{\sqrt[n]{\prod_{j=1}^{n} a_{ij1}} + \max\left\{ \begin{array}{ll} \sum_{\substack{k=1 \\ k\neq i}}^{n} \sqrt[n]{a_{ki3} \prod_{\substack{l=1 \\ l\neq i}}^{k-1} \frac{1}{a_{lk}} \prod_{\substack{l=k+1 \\ l\neq i}}^{n} a_{kl}} & a_{kl} \in [a_{kl1}, a_{kl3}], \\ & k,l = 1,\ldots,n, \\ & k,l \neq i,\ k > l \end{array} \right\}}, \tag{16}$$

$$v_{i3} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij3}}}{\sqrt[n]{\prod_{j=1}^{n} a_{ij3}} + \min\left\{ \begin{array}{ll} \sum_{\substack{k=1 \\ k\neq i}}^{n} \sqrt[n]{a_{ki1} \prod_{\substack{l=1 \\ l\neq i}}^{k-1} \frac{1}{a_{lk}} \prod_{\substack{l=k+1 \\ l\neq i}}^{n} a_{kl}} & a_{kl} \in [a_{kl1}, a_{kl3}], \\ & k,l = 1,\ldots,n, \\ & k,l \neq i,\ k > l \end{array} \right\}}. \tag{17}$$

These formulas will be used for obtaining fuzzy weights of categories of scientific monographs in the following section.

5 Evaluation of R&D outcomes - a numerical example

In this section, we aim to show how fuzzy pairwise comparison matrices can be successfully used in a real-life application. We will discuss a practical R&D outcomes evaluation problem and propose a solution to it using the formulas (14), (16) and (17). We will show how fuzzy pairwise comparison matrices can be used to analyze and visualize preferences of the experts (evaluators). If a part of an evaluation process is a classification task (objects are sorted into classes according to some criteria and these classes can be ordered according to the preferences of the experts), then fuzzy evaluations of these classes computed by (14), (16) can be interpreted as intervals of possible evaluations of each object from the respective class. The middle significant values (17) of the

corresponding fuzzy evaluations can be interpreted as "default" evaluations of a typical representative of each evaluation class. Such an approach allows the evaluation process to incorporate peer-review evaluation with its supposed subjectivity, but still restricting the effects of subjectivity by providing evaluation intervals and default evaluations for objects belonging to each class. This will be illustrated by a practical application presented in this section.

First, let us define the context of the real-life example that will be presented here. A practical need to revise the evaluation of certain R&D outcomes at the Palacky University in Olomouc (UP), Faculty of Science, was identified (see Krejčí et al. (2012)). The national R&D evaluation methodology (RVVI, 2012) (which was used in the Czech Republic to evaluate R&D outcomes published in the year 2012) assigned to all scientific monographs a fixed amount of points. The points were the same for all the monographs regardless of their quality or content; the same national evaluation methodology, however, distinguished well among scientific papers based on ranking of a journal according to its impact factor in a given field. This way, no distinction could be made among the monographs based on the national R&D evaluation at the Faculty of Science. Since quality assessment and promotion was (and still is) seen as crucial in higher education institutions, such an evaluation was seen as too restrictive. Motivation to publish quality monographs was missing. Formal criteria (the length of the monograph, the scientific board of the publisher and so on) were used to decide whether to assign the fixed amount of points or not to a monograph; no quality assessment component was identifiable in the evaluation of monographs. A different evaluation methodology was searched for, that would be able to reflect expertly assessed quality of monographs, while still retaining a substantial amount of measurability. The outputs of the evaluation were also required to be easy to use for funds distribution purposes at the Faculty of Science.

To formalize the problem, let us consider we are to evaluate a set of objects (scientific monographs in the case of this paper) based on a set of criteria. For simplicity, let us assume that we consider two criteria. The first one is categorical and represents the reputation of the publisher of the monograph. The second criterion is the quality of the respective monograph assessed through the peer-review. The evaluation based on the publishers' reputation (the first criterion) is required to provide a good starting point and necessary restrictions for the following peer-review assessment. The peer-review then can account for the fact that not all the monographs published by the same publisher (or a category of publishers) are necessary of the same scientific quality - very good scientific quality of monographs can be emphasized in the peer-review process, and some monographs of poorer scientific quality can be identified and evaluated accordingly.

In Krejčí et al. (2012), four categories of publishers were identified. The categories reflected the scientific reputation of the publisher in those fields of science relevant for the Faculty of Science - *Category 1* being of the highest scientific reputation and *Category 4* being of the lowest scientific reputation. Defining the categories was an iterative and heuristic process, in which the

scientific quality of the monographs published in the publishing houses, the editorial boards and the reputation of the publishing houses in the scientific community were considered. A general description of each category was provided in the end as well as a list of several typical publishers from each category. All the publishing houses in which the evaluated monographs were published in 2012 were then classified into these four categories.

Based on the general description of the categories and on the criteria used to define these categories, the board of experts responsible for the evaluation of monographs at the Faculty of Science of UP provided intuitive information concerning the expected evaluation of the monographs published by publishers from each group. This information was provided as intervals of scores for each category of publishers in the following form:

$$\begin{array}{ll} \text{Category 1:} & 50-75 \text{ points,} \\ \text{Category 2:} & 30-40 \text{ points,} \\ \text{Category 3:} & 15-20 \text{ points,} \\ \text{Category 4:} & 5-10 \text{ points.} \end{array} \tag{18}$$

Although the intervals of scores (18) were presented as a consensual suggestion of the whole board of experts, the justification and the meaning of these suggested evaluation intervals were missing. For this reason the academic senate requested an analysis of these intervals to have a clearer interpretation of their meaning before an evaluation methodology can be based on them. The classic AHP was used to represent the preference structure of the board of experts. Each category was assigned an evaluation equal to the mean score from the suggested interval (see (19)).

$$\begin{array}{lll} \text{Category 1:} & 62.5 & \text{points,} \\ \text{Category 2:} & 35 & \text{points,} \\ \text{Category 3:} & 17.5 & \text{points,} \\ \text{Category 4:} & 7.5 & \text{points.} \end{array} \tag{19}$$

Crisp Saaty's matrix of preference intensities S was reconstructed from the evaluations of the four categories (19) using the formula (9), and its elements were subsequently rounded to integers (see (20)). This allowed us to use Saaty's 9-point scale (see Section 3) to interpret the intuitive preference structure initially expressed by (18).

$$S = \begin{pmatrix} 1 & 1.79 & 3.57 & 8.34 \\ \frac{1}{1.79} & 1 & 2.00 & 4.67 \\ \frac{1}{3.57} & \frac{1}{2.00} & 1 & 2.33 \\ \frac{1}{8.34} & \frac{1}{4.67} & \frac{1}{2.33} & 1 \end{pmatrix} \quad \rightarrow \quad \text{rounded } S = \begin{pmatrix} 1 & 2 & 4 & 8 \\ \frac{1}{2} & 1 & 2 & 5 \\ \frac{1}{4} & \frac{1}{2} & 1 & 2 \\ \frac{1}{8} & \frac{1}{5} & \frac{1}{2} & 1 \end{pmatrix} \tag{20}$$

The consistency ratio of the rounded S is 0.0023. The information presented in the rounded S above the main diagonal can be summarized in the following

way - one book in the higher (more preferred) category can be compensated by two books from the neighboring lower category. Right above the main diagonal of the rounded S we can see only the lowest integer values associated with preference. Using the linguistic terms suggested by Saaty for the 9-point scale, we can describe the relationship of neighboring categories as "each higher category is between equally important and moderately more important than the lower category". Although the original matrix S in fact expresses that the difference in preferences between neighboring categories decreases slightly with the increasing preference of the categories ($s_{34} > s_{23} > s_{12}$), the board of experts confirmed that the rounded S is a good approximation of their preferences.

In the next step, we applied fuzzified Saaty's scale presented in Table 4 to fuzzify the rounded S. We have obtained $\tilde{S}$ as presented in (21). The idea behind this step is that if we asked the experts to provide their preferences using linguistic Saaty's scale, it would be reasonable (for the reasons already presented in previous sections of this paper) to reflect the uncertainty of the linguistic description of their preferences.

$$\text{fuzzified S:} \qquad \tilde{S} = \begin{pmatrix} 1 & (1,2,3) & (3,4,5) & (7,8,9) \\ \left(\frac{1}{3},\frac{1}{2},1\right) & 1 & (1,2,3) & (4,5,6) \\ \left(\frac{1}{5},\frac{1}{4},\frac{1}{3}\right) & \left(\frac{1}{3},\frac{1}{2},1\right) & 1 & (1,2,3) \\ \left(\frac{1}{9},\frac{1}{8},\frac{1}{7}\right) & \left(\frac{1}{6},\frac{1}{5},\frac{1}{4}\right) & \left(\frac{1}{3},\frac{1}{2},1\right) & 1 \end{pmatrix} \tag{21}$$

From $\tilde{S}$ we can compute the evaluations of the four categories of publishers (22) using the formulas (14), (16) and (17) and multiply each characteristic value by 100 to obtain integer evaluations in the form of (23).

$$\begin{aligned} \tilde{H}_1 &= (0.41, 0.53, 0.61) \\ \tilde{H}_2 &= (0.19, 0.28, 0.40) \\ \tilde{H}_3 &= (0.09, 0.13, 0.20) \\ \tilde{H}_4 &= (0.05, 0.06, 0.09) \end{aligned} \tag{22}$$

$$\begin{aligned} \tilde{h}_1 &= (41, 53, 61) \\ \tilde{h}_2 &= (19, 28, 40) \\ \tilde{h}_3 &= (9, 13, 20) \\ \tilde{h}_4 &= (5, 6, 9) \end{aligned} \tag{23}$$

As the evaluations (22) are normalized fuzzy weights according to (3) and the initially suggested evaluation intervals (18) are not, we can not compare (18) and (23) directly. We can however observe, that in (23) the neighboring evaluation intervals defined by the respective left and right significant values of the fuzzy numbers $\tilde{h}_1, \tilde{h}_2, \tilde{h}_3$ and $\tilde{h}_4$ lie close to each other and for $\tilde{h}_2$ and $\tilde{h}_3$ even overlap. This is a significant difference from the initial idea of four evaluation intervals that are not only disjunctive, but also distinctly separated (see (18)).

In discussions with the board of experts (evaluators), we have not found a firm rationale for the evaluation intervals to be as distinctly separated as in

(18). The intensity of preference was confirmed to be the same between each two neighboring categories and could be identified with the lowest linguistic label still expressing preference. If we consider this to be the only initial information available from (and confirmed by) the experts, we can construct initial Saaty's matrix of preference intensities $\tilde{Z}_I$ using fuzzified Saaty's 5-point scale presented in Table 3 in the form of (24). We do not consider the 9-point scale now as the intermediate linguistic values were confirmed by the experts not to be intuitive, particularly the preference expressed as being "between equally preferred and moderately preferred to"; hence the lowest expressible preference we decided to use is "moderate" with an associated meaning $(1, 3, 5)$.

$$\tilde{Z}_I = \begin{pmatrix} 1 & (1,3,5) & ? & ? \\ \left(\frac{1}{5}, \frac{1}{3}, 1\right) & 1 & (1,3,5) & ? \\ ? & \left(\frac{1}{5}, \frac{1}{3}, 1\right) & 1 & (1,3,5) \\ ? & ? & \left(\frac{1}{5}, \frac{1}{3}, 1\right) & 1 \end{pmatrix} \tag{24}$$

Based on the initial information provided in the form of $\tilde{Z}_I$, we can fill in the missing values (using fuzzified Saaty's 5-point scale) so that $\tilde{Z}_I$ remains as consistent as possible (that is, its consistency ratio is below 0.1 and as close to zero as possible, and the matrix is weakly consistent). Thus, we obtain $\tilde{Z}$ in the form of (25). The consistency ratio of $\tilde{Z}$ is 0.0286 and it is weakly consistent according to (11) and (12).

$$\tilde{Z} = \begin{pmatrix} 1 & (1,3,5) & (3,5,7) & (7,9,9) \\ \left(\frac{1}{5}, \frac{1}{3}, 1\right) & 1 & (1,3,5) & (3,5,7) \\ \left(\frac{1}{7}, \frac{1}{5}, \frac{1}{3}\right) & \left(\frac{1}{5}, \frac{1}{3}, 1\right) & 1 & (1,3,5) \\ \left(\frac{1}{9}, \frac{1}{9}, \frac{1}{7}\right) & \left(\frac{1}{7}, \frac{1}{5}, \frac{1}{3}\right) & \left(\frac{1}{5}, \frac{1}{3}, 1\right) & 1 \end{pmatrix} \tag{25}$$

The evaluations of the four categories of publishers computed from $\tilde{Z}$ using the formulas (14), (16) and (17) and multiplied by 100 are expressed in (26).

$$\begin{aligned} \tilde{l}_1 &= (38, 58, 69) \\ \tilde{l}_2 &= (14, 25, 45) \\ \tilde{l}_3 &= (6, 11, 23) \\ \tilde{l}_4 &= (4, 5, 10) \end{aligned} \tag{26}$$

The fuzzy numbers $\tilde{l}_1, \tilde{l}_2, \tilde{l}_3$ and $\tilde{l}_4$ representing the evaluations of categories of publishers computed from (25) can be interpreted in the following way. For each category, the left and the right significant values of the fuzzy numbers $l_1, \tilde{l}_2, \tilde{l}_3$ and $\tilde{l}_4$ define the interval of possible scores any monograph published by a publisher from the given category can be assigned. The middle value of any of these fuzzy numbers can be interpreted as a "default" evaluation of a typical book published by a typical publisher in the given category. We can see, that under this interpretation we obtain evaluation intervals for the four categories of publishers that are significantly overlapping. This seems to contradict the initial information provided by the board of experts through (18). The difference between these two pieces of information lies mainly in

the fact, that nothing more than intuition was used to provide the initial information. After its formalization, the shift of the evaluation intervals from not overlapping at all to significantly overlapping followed clearly from the information provided by the board of experts. This way, the fuzzified AHP served as a tool for visualising and interpreting the preferences of experts. The final results (26) were well accepted by the board of experts as representing their intentions and preferences concerning the reputation of publishers.

The next step of the evaluation of scientific monographs can be summarized in the following way. After the classification of each monograph according to its publisher, we can asses its quality in a peer-review process. Based on the reputation of the publishing house (and hence the category it belongs to), we have an interval of all possible scores that can be assigned to any monograph published there. We also have a default evaluation available (the middle significant value of the fuzzy number representing the evaluation of the publisher category), which can be used as a starting point for the peer-review process. If the board of experts considers the monograph to be better than an "average" monograph published by publishers from the respective category, it can shift its evaluation higher up to the upper bound of the evaluation interval. Analogical approach is at hand for monographs of "lower than average" quality. The fact that the evaluation intervals overlap allows a "really good" monograph published by a publisher from a lower category to be assigned more points than a "bad" monograph published by a publisher from a higher category.

The evaluation intervals provide limits and a starting point for the peer-review evaluation. Each shift from the default value needs to be justified (its rationale provided) and must not result in an evaluation outside the respective evaluation interval. This way, peer-review is implemented into the evaluation as a quality assessment tool and yet at least some of its subjectivity is removed by providing the evaluation intervals within which the evaluation of a particular monograph must lie.

6 Conclusion

In this paper, we have reviewed the fuzzification of Saaty's scale for making pairwise comparisons of objects and of the geometric mean method for obtaining fuzzy weights of objects from fuzzy pairwise comparison matrices. Moreover, fuzzification of the weak consistency defined by Stoklasa et al. (2013) was proposed here to reflect the linguistic level of the intensities of preferences used in Saaty's scale. The fuzzified weak consistency was combined with Saaty's CR in order to obtain reasonably consistent fuzzy pairwise comparison matrices. The proposed method was then applied to a real-life problem which was the evaluation of R&D outcomes at the Faculty of Science at the Palacky University, Olomouc, Czech Republic.

The fuzzified AHP was used to determine the evaluations of four categories of publishers based on their reputation. This way, evaluations of all four categories were computed as triangular fuzzy numbers. The supports of these

fuzzy numbers were interpreted as intervals of possible scores of any monograph from the given category. The number from the core of each of these fuzzy numbers was interpreted as an evaluation that a typical book published by a publisher from the given category might be assigned. This interpretation provides a very good starting point for the subsequent peer-review assessment of the quality of each monograph. Although it is presented here to showcase a practical application of the use of the fuzzified AHP, it can also be viewed as a practical example of a general evaluation methodology incorporating a peer-review component. An important feature of the presented approach lies in its ability to limit the effects of the peer-review evaluation while still retaining its subjectivity. Expert evaluation this way becomes an integral part of the evaluation methodology and its inherent subjectivity is no longer a problem. This makes the presented evaluation methodology suitable for quality assessment issues, where peer-review (or other forms of expert assessment of quality) is required.

References

Brunelli M, Canal L, Fedrizzi M (2013) Inconsistency indices for pairwise comparison matrices: a numerical study. Ann Oper Res 211, 493-509

Buckley JJ (1985) Fuzzy hierarchical analysis. Fuzzy Set Syst 17, 233-247

Chang DY (1996) Applications of the extent analysis method on fuzzy AHP. Eur J Operat Res 95, 649-655

Cheng CH, Mon DL (1994) Evaluating weapon system by analytical hierarchy process based on fuzzy scales. Fuzzy Set Syst 63, 1-10

Cheng JH, Lee CM, Tang CH (2009) An application of fuzzy delphi and fuzzy AHP on evaluating wafer supplier in semiconductor industry. Wseas Transactions on Inf Sci Appl 6, 756-767

Crawford GB (1987) The geometric mean procedure for estimating the scale of a judgment matrix. Math Model 9, 327-334

Crawford G, Williams CA (1985) A note on the analysis of subjective judgment matrices. J Math Psychol 29, 387-405

Enea M, Piazza T (2004) Project selection by constrained fuzzy AHP. Fuzzy Optim Decis Ma 3, 39-62

Güngör Z, Serhadlioglu G, Kesen SE (2009) A fuzzy AHP approach to personnel selection problem. Appl Soft Comput 9, 641-646

Krejčí J, Pavlačka O, Talašová J (2013) On the fuzzification of the Analytic Hierarchy Process. submitted to Inf Sci

Krejčí J, Jandová V, Stoklasa J, Talašová J (2012) Bodové hodnocení knih [Evaluation of monographs - in Czech]. research report, Palacky University, Olomouc.

Kwong CK, Bai H (2002) A fuzzy AHP approach to the determination of importance weights of customer requirements in quality function deployment. J Intell Manuf 13, 367-377

Pan NF (2008) Fuzzy AHP appproach for selecting the suitable bridge construction method. Automat Constr 17, 958-965

Pan Y, Yuan B (1997) Bayesian Inference of Fuzzy Probabilities. Int J Gen Syst 26, 73-90

Ramík J (2000) Analytický hierarchický proces (AHP) a jeho využití v malém a středním podnikání [Analytic Hierarchical Process (AHP) and its applications in small and medium business - in Czech]. Slezská univerzita, OPF Karviná, ISBN 80-7248-088-X

RVVI (2012) Metodika hodnocení výsledků výzkumných organizací a hodnocení ukončených programů (platná pro léta 2010, 2011 a rok 2012) [Research, Development and Innovation Council: Methodology for the evaluation of outcomes of research organisations and the evaluation of finished programmes (valid for 2010, 2011 and 2012) - in Czech] [online], available from: http://www.vyzkum.cz/FrontClanek.aspx?idsekce=650022, [cited 2014-09-03]

Saaty T L (1980) The Analytic Hierarchy Process. McGraw Hill, New York

Stoklasa J, Jandová V, Talašová J (2013) Weak consistency in Saaty's AHP - evaluating creative work outcomes of Czech Art Colleges. Neural netw world 23, 61-77

Tesfamariam S, Sadiq R (2006) Risk-based environmental decision-making fuzzy analytic hierarchy process (F-AHP). Stoch Env Res Risk A 21, 35-50

Vahidnia MH, Alesheikh AA, Alimohammadi A (2009) Hospital site selection using fuzzy AHP and its derivatives. J Environ Manage 90, 3048-3056

Van Laarhoven PJM, Pedrycz W (1983) A fuzzy extension of Saaty's priority theory. Fuzzy Set Syst 11, 199-227

Xu R (2000) Fuzzy least-square priority method in the analytic hierarchy process. Fuzzy Set Syst 112, 395-404

Zadeh LA (1975) Concept of a linguistic variable and its appliation to approximate reasoning I,II. Inf Sci 8, 199-249, 301-357; III. Inf Sci 9, 43-80

PUBL. VI

Stoklasa, J., Talášek, T. and Musilová, J.,Fuzzy approach - a new chapter in the methodology of psychology? *Human Affairs*, 24(2), 189–203, 2014.

VERSITA

HUMAN AFFAIRS *24*, 189–203, 2014
DOI: 10.2478/s13374-014-0219-8

FUZZY APPROACH – A NEW CHAPTER IN THE METHODOLOGY OF PSYCHOLOGY?[1]

JAN STOKLASA, TOMÁŠ TALÁŠEK AND JANA MUSILOVÁ

Abstract: This paper aims to briefly introduce the main idea behind the fuzzy approach and to identify the areas and problems encountered in the humanities that might profit from using this approach. Based on a short overview of selected applications of fuzzy in psychology we identify key areas in which the fuzzy approach has already been applied, and propose a list of general types of problems that the fuzzy approach may provide solutions for in psychology and the humanities in general. These types of problems are illustrated using practical examples. The benefits and possible shortcomings of using the fuzzy approach compared to classical approaches in use today are discussed.

The goal of this paper is to indicate areas in research and practice in the humanities, where modern mathematical tools—in this case linguistic fuzzy modelling—have already been used or might prove promising.

Keywords: methodology; fuzzy; linguistic modelling; decision support; diagnostics.

Introduction

The goal of every science can be formulated like this: to describe, explain, and predict the world, or more specifically the behaviour of the object of study. In psychology, the object is the human mind. However, it is not an object that is easy to access. There are not many ways in which the human mind or specific mental processes can be directly assessed or measured.

Psychology uses methods and formal models developed in other sciences for other purposes (mathematics, physics, medicine and others) as well as methods developed directly for psychology. Many of these originate from other sciences and use their tools. Of all these formal tools, statistics has an important role to play (especially in quantitative methodology). It is one of the few mathematical tools that all psychology majors meet during their studies and as far as we can say from our experience, the only one that psychology students in the Czech Republic are really required to be familiar with. It is used in psychological diagnostics to define the norm, to assess the validity and reliability of psychological tests and methods,

[1] The research presented in this paper was supported by grant PrF 2013 013 Mathematical models of the Internal Grant Agency of Palacky University in Olomouc.

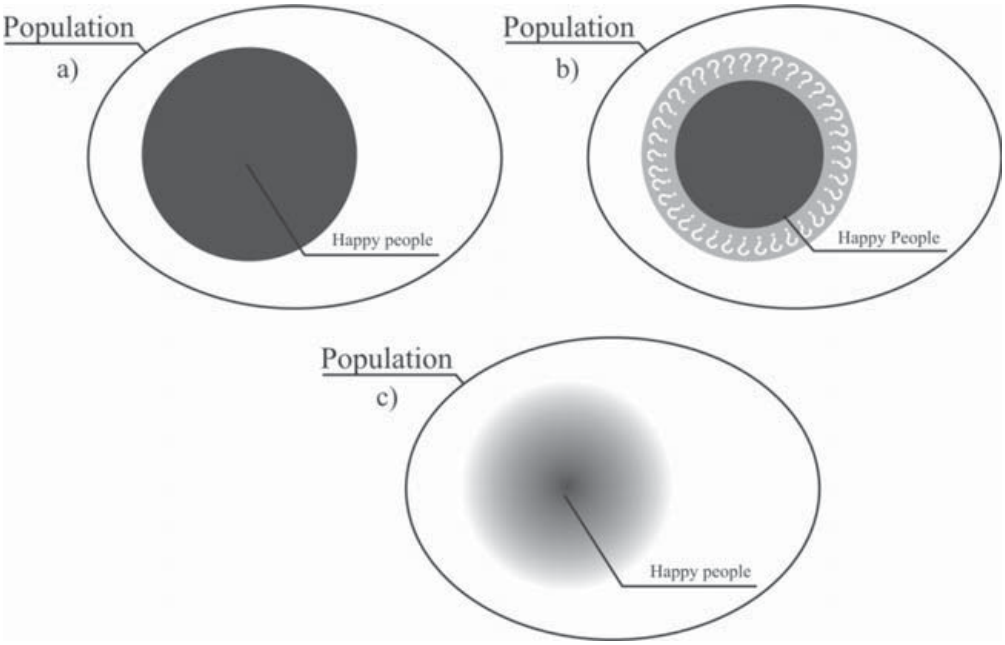


Figure 1. The concept of a fuzzy set: a) crisp set of happy people in the population—people are either happy or not happy; b) crisp set with borderline cases (grey area with question marks) where we cannot decide whether these people are happy or not; c) fuzzy set of happy people—people can be happy to various extents—in the centre the people are completely happy, the further away from the centre they are, the less happy the people are.

to test hypotheses—its uses are numerous and in many cases the use of statistics is not only apt, but beneficial to the psychological understanding of the world (or at least of a part of it).

We might question how much statistics can be of service if we really want to concentrate on uniqueness, if we want to capture what it is that makes every human being different from other human beings. The fact that qualitative methodologies have been introduced into psychology (if introduction is the correct term for ideas that have always been implicitly present in psychology, although perhaps not sufficiently methodologically and formally grounded) means that the answer this question is a clear "not enough".

In this paper we would like to point out that if we create a psychological methodology based mainly on statistics, we might sooner or later find that there is a hole in it. And for all the problems that fall into this hole, statistics and other mathematical tools commonly used in psychology (scaling, optimisation, etc) might not be able to provide satisfactory models. The hole might not be visible from a distance—only when we encounter a problem lying really close to the hole or even directly inside it do we realize that new tools are necessary and that a different approach to building formal models is required. So it is quite possible that many psychologists will not get closer to the problems near this hole during their whole professional career. But if they eventually do, they need to have tools to deal with them appropriately. Representing human knowledge, working with linguistic descriptions of reality or mental processes (self-reports), dealing with uncertain information or describing human decision-making are issues that form just a subset of the problems that might fall into this

"hole in methodology". In our opinion we encounter problems from this area quite frequently in psychology, but we either treat them with methods ill-suited to these problems or the data they produce, or we ignore them owing to the lack of appropriate tools.

If we consider some of the most typical sources of information in psychology—interviews, observations and similar methods—we usually obtain a linguistic description of the problem or process. This description is based on a self-report by a particular human being, and as such can be understood only as precisely as the words and language allow. The meaning of the words is, however, not certain—some of the linguistic expressions we normally use partially overlap, and their meanings are context dependent and may even differ from person to person. If uncertainty is inherent to linguistic description (due to the process whereby one person codes ideas into words and then they are decoded back into ideas—that is, a second person—the psychologist—assigns meaning to the words), then classical methods not equipped to deal with uncertainty may produce incorrect results when applied to model situations or systems that are described linguistically.

We aim to briefly introduce the basic concept of fuzzy approach in the following section. Using a list of a number of successful applications of fuzzy in a psychological context, we identify several prototypical issues which typically lead to the use of fuzzy tools (or at least suggest that the use of fuzzy might be considered). We discuss several implications and areas that typically encounter several of these issues. Finally, we provide two practical applications of fuzzy in the humanities context to show how the prototypical issues can be dealt with in real life.

Fuzzy approach in a nutshell

The fuzzy approach is based on the idea that, in some cases, it is not reasonable to say that an object either has a property or it does not (the fuzzy approach in fact assumes that the logical law of the excluded middle does not hold). Objects or people may exhibit some properties only partially—to a certain extent. This becomes even more apparent when the properties are described in common language—by words. Let us for example consider happiness. If we would like to select all the happy people from the population, we would have to be able to define a strict threshold between "happiness" and "not happiness" —that is, we would have to be able to decide whether each person is happy or not (see Figure 1, subfigure a). This approach is, however, counterintuitive. In this case, we would probably be able to select those who are "definitely happy" and those who are "definitely not happy". But there would be a certain amount of people for whom we would not be able to decide with certainty (see Figure 1, subfigure b). This is usually used in diagnostics for borderline values of scores or indicators. If we obtain values close to the threshold, we interpret them with more caution (for example as being inconclusive).

If we consider happiness then there are people that are "very happy", some of them may even be "manic", there may also be people that are "a bit happy", "somewhat unhappy" and so on. It would therefore seem that happiness is an emotion that people experience to different extents (Figure 1, subfigure c) describes a fuzzy set of happy people—the darker the colour, the higher the level of happiness). We can view the characteristic property of a set as a linguistic label of a set as well and the degree to which a member belongs to this

set (usually a number between 0 and 1) can be interpreted as the level of compatibility of the member with this linguistic label—the extent to which the linguistic label describes the member well. This can be of course interpreted also in a logical sense—statements in fuzzy logic can be true, false or everything between these two extremes—this means a statement can hold only partially.

To refer to the concept of fuzzy modelling and fuzzy logic as a new branch of mathematics would not be appropriate. Fuzzy sets were introduced as far back as in 1965 by Zadeh and he outlined the possibility that fuzzy sets could be used to model the meanings of certain linguistic terms ten years later (Zadeh, 1975). There is a considerable amount of literature on fuzzy logic, fuzzy set theory and linguistic fuzzy modelling and it is not within the scope of this paper to provide theoretical insights into this area (interested readers can see for example Klir & Yuan (1995) or Dubois & Prade (2000)).

Applying fuzzy in psychology and social sciences

Since 1965, there has been a fair amount of development in the field of fuzzy, both in the theory and applications. Surprisingly, fuzzy set theory has received more attention in the technical sciences and heavy industry than in the humanities. There are a number of books and book chapters on fuzzy methods in the social sciences and psychology—for example, Smithson (1986), Zétényi (1988), Smithson & Oden (1999), Ragin (2000), Smithson & Verkuilen (2006) and Arfi (2010). Most of these authors expect that the fuzzy approach will attract greater attention in the humanities soon. It would not be correct to say that there are no cases of fuzzy mathematics or linguistic fuzzy modelling being applied so far—some interesting psychological results can be found, such as:

- fuzzy logical model of perception (Oden & Massaro, 1978)
- fuzzy set based theory of memory (Massaro et al., 1991)
- approach to depression as a fuzzy concept (Horowitz & Malle, 1993)
- fuzzy burnout syndrome concept (Burisch, 1993)
- fuzzy scaling and various fuzzifications of Likert scales
- fuzzy coding in qualitative research
- fuzzy developmental stages theories (overlapping stages)

Researchers have also focused on the use of linguistic fuzzy modelling in psychological diagnostics (focus on the MMPI-2 interpretation)—see Bebčáková et al. (2010) or Stoklasa & Talašová (2011) for an example of MMPI-2 (a psychological personality inventory) interpretation tools using fuzzy concepts and linguistic modelling.

There are also numerous applications of fuzzy methods in formal mathematical theory of group and multiple criteria decision-making (which are very close to psychology) and fuzzy data analysis methods. The use of fuzzy methods in HR management in companies has been discussed in Zemková & Talašová (2011); Stoklasa et al. (2011, 2013) describe potential uses of fuzzy rule bases in HR management at tertiary education institutions.

Fuzzy concepts have also been covered in fuzzy linguistics. The linguistic modelling approach also provides valuable insights into classical decision support methods. It can be used even in the evaluation of arts—for example an evaluation model for the creative work outcomes of Czech art colleges and faculties (described in Stoklasa et al., 2013, Stoklasa

& Talašová, 2013) shows how a linguistically described condition on consistency of expert preferences can prove useful in large evaluation problems.

Prototypical issues: where human sciences can benefit from the fuzzy approach

These applications of fuzzy in the humanities all share some common features that can be extracted to produce a list of typical cases of when one might consider using the fuzzy approach. All the examples address issues that cannot be sufficiently reflected upon and dealt with in the formal models in psychology using the classical crisp approach. These include:

- **inadequacy of crisp boundaries and "grey zones"**—a typical example of this issue is deciding whether a particular observation, test score etc., is within the norm or not. It is not reasonable to assume that the shift from being one unit below the threshold (can be defined numerically or linguistically) to being one unit above the threshold means a transition from being "normal" to being "beyond the norm". In diagnostics, setting scores and observations around the threshold can be treated as "inconclusive" or "borderline". But this does not solve the problem as we still need to decide what is "normal" and when it becomes "borderline". The fuzzy approach can provide tools that enable the continuous transition from one state to another, allowing an observation to be partially normal and partially above the norm.
- **ill-defined and overlapping categories**—in many cases we need to classify people or objects into classes. These classes are usually defined by their characteristic feature (this can be a measurable quality or a purely qualitative feature). Classical approaches operate under the assumption that an object cannot belong to more than one class at the same time. The fuzzy approach makes it possible for an object to distribute its membership among several categories, as well as to belong fully to several categories at the same time. This includes also diagnostics situations, testing, management decisions and so on.
- **continuity of transformation between stages**—many theories operating with stages might again benefit from the possibility of modelling continuous transitions between stages. Not only developmental stages as mentioned in the previous section—evaluation is also a good example of this problem (an improving performance means a person gradually ceases to be "average" and begins to be "good").
- **linguistic data**—when we deal with information provided in words, we need to be able to account for the uncertainty inherent in such data. Since a concept can mean different things to two different people, formal models should be able to reflect these differences. Also the fact that the same linguistic term can equally well describe various actual objects or situations (a "long sleep" can be something between 6 and 12 hours for me) should be modelled adequately. A single object might even be described using several words (to various degrees of compatibility). It may be necessary to allow a description to be partially compatible with an object. A fuzzy approach can provide tools to represent linguistic data.
- **measurement/assessment with linguistically labelled scales**—all assessment and measurement instruments that use linguistic labels or scales (for example: never—sometimes—always) may encounter problems with the uncertainty of the words used and the different meanings of these words among different people. When subjective

differences in meaning become an issue, appropriate tools to model the meanings of words are welcome. The issues of meaning might also arise when only numerical scales are used.

- **partial validity of statements or data**—in the humanities, where human beings provide a great amount of the data and where observation and interpretation play an important part in the methodology, we cannot rely on the fact that the data we work with are completely valid (some instruments even provide tools for the validity assessment of the data obtained from people). Human knowledge of the world can be contradictory, incomplete or uncertain. If we have no more objective means of obtaining data than self-assessment, we need to be able to reflect the different validities of our findings, and the varying importance of the rules we use to describe the behaviour of the system. Fuzzy can not only provide tools to represent the partial validity of statements and data, it can also provide the means for assessing the methods we already use in the context of partially valid data.

We do not claim that the fuzzy approach will solve all these problems. The fuzzy approach also has its limits, which are usually defined by people's ability to express the meaning of words, the issue of the context dependency of the meaning and the inconsistency of expert knowledge of the systems. Fuzzy methodology was developed to deal with uncertainty and as such might provide at least some level of assistance for these issues. However, we need to admit that the continued collaboration between fuzzy set theoreticians, psychologists, linguists and sociologists is required to find even more appropriate ways of capturing the meaning of words in ordinary language.

Using these prototypical issues identified above, we can generate several possible areas in which the fuzzy approach can be used in the humanities. Combining the ability to deal with uncertainty (and hence to model some aspects of language descriptions of reality) and allowing the partial validity of statements, we can build powerful tools for the humanities that could be used for example in expert knowledge representation, knowledge transfer and provide assistance in difficult decision problems (such as diagnostics in psychology).

Since language is our main tool for communication, being able to build models using words (narrative descriptions) that reflect knowledge of the systems we are interested in seems to be the natural course of research in the humanities. The uncertainty inherent in words is the key to the relative simplicity and effectiveness of our communication. Providing precise descriptions is not only unnatural to human beings, in many cases it is also impossible (we do not know exactly what "fast" is in km/h, we do not have a precise representation of "a while"), but we still understand each other well enough. And the models that fit "well enough" remain relatively simple and understandable and are the main domain of fuzzy mathematics and linguistic fuzzy modelling.

Once we have a model of expert knowledge, we can easily distribute it to others. This might be an interesting feature in the context of education. Let us consider that we are able to model the diagnostics process of a skilled diagnostician, his work using the diagnostics method, his way of dealing with the data and interpreting results. Linguistic fuzzy modelling can provide us with a formal (mathematical) level and an attached linguistic description level (see also the next section for more information on this). That way if we input the expert knowledge into a computer, we obtain a good training tool for students—future

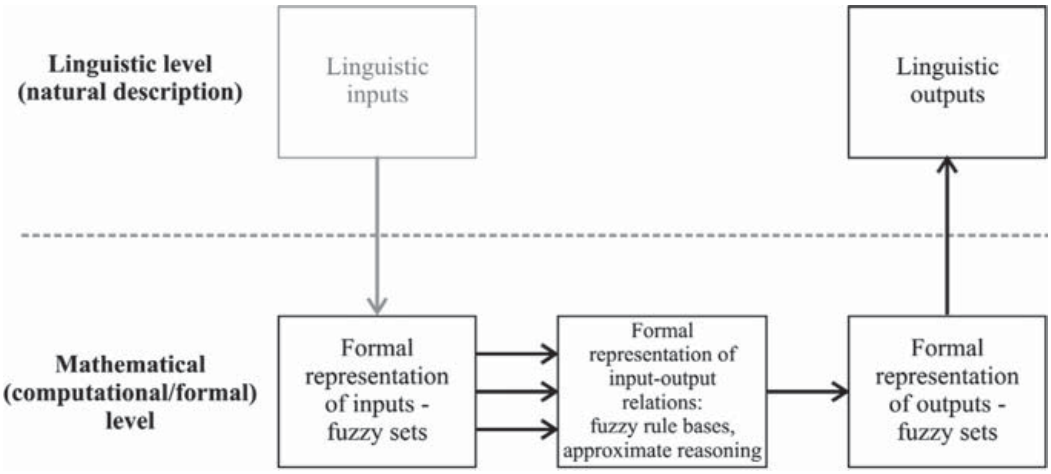


Figure 2. Scheme of the usual approach to mathematical modelling in psychology.

diagnosticians. They can train their skills against a modelled expert in the field. The main advantage of fuzzy modelling in this context compared to other mathematical tools (such as neural networks) is that when students make a mistake, they can check what they did differently from the procedure implemented in the model. As the model has an in-build linguistic level, the students can check it against the description of the process described in words, not mathematical formulas.

We can also use the fuzzy approach to assist us in everyday complex tasks which require our insight, but are repeated frequently. Using fuzzy we can build decision support tools by describing what we do in words and spare time to concentrate on more pressing matters. In psychological diagnostics, the pre-processing of data can be automatized (in a way that still reflects our habits in working with the data) to provide us with some kind of summarizing information, even to suggest possible diagnoses (using the fact that a subject can belong fully or partially to several classes).

What can fuzzy bring psychology—practical examples

Before we present some examples of the use of fuzzy methods in a humanities context, we provide a brief overview of the possible benefits of fuzzy approach to psychology. Figure 2 illustrates the use of classical mathematical methods in psychology—inputs (these may be words obtained by interview or other self-report based methods) are converted into mathematical objects (numerical inputs provided by diagnostics methods can be rescaled or used in the form they are provided) and are then processed by the selected mathematical model. The model produces results in the form of mathematical objects, which need to be interpreted appropriately. To describe the results of a mathematical model using words in a way that captures their proper meaning is not easy—this process is even more demanding if the mathematical operations performed with the inputs are complex.

If we link the inputs and the mathematical operations we perform on the inputs to their proper linguistic meanings, we get a linguistic model. This model (see Figure 3) has two

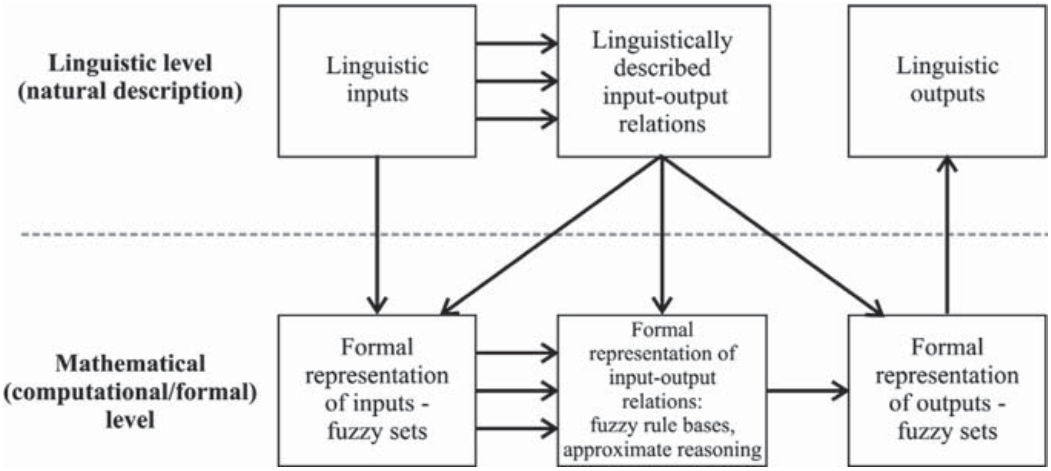


Figure 3. Scheme of the two-level linguistic approach to mathematical modelling suitable for the humanities.

levels for describing the modelled system. The first is the linguistic level, which remains comprehensible to all (even the non-expert) because it uses words to describe the variables and their relationships. The second level (computational or mathematical) reflects the linguistic level, if possible, in each step of the model. Mathematical methods therefore have to be chosen to best reflect the linguistic level (which is demanding and requires a sufficient understanding of the methods and the fuzzy approach itself). By maintaining the correspondence between the two levels of the model, interpreting the outputs of the computational level is much easier and the model remains comprehensible. Also adjustments to the model can be easily made at the linguistic level—particularly when the relationships between the variables are described using linguistic IF-THEN rules (see the example of the academic faculty evaluation system).

Academic faculty evaluation system IS HAP (example 1)

Linguistic rules—such as "If the weather is nice, then you can leave your umbrella at home" provide an easy-to-understand description of the modelled system or expert knowledge on a system. Linguistic fuzzy models can be used for knowledge storage, knowledge transfer and even to test expert knowledge. Consider that we build a linguistic model of the reasoning process of a skilled diagnostician (see Figure 7 for a simple example of such a decision process described using 25 rules, Figures 4–6 summarize the meanings of the linguistic terms used in the rules). Once it is available, we can provide it to students to see how the expert approaches the diagnostic situation. The computational level allows us to input this knowledge (albeit described in words and thus uncertain) into a computer programme against which the students can test their diagnostic conclusions and thanks to the linguistic level, they can find out which aspects of their train of thinking differs from the experts'.

Let us consider a real example of an academic faculty evaluation system called IS HAP, developed at the Faculty of Science, Palacky University in Olomouc, (see Stoklasa

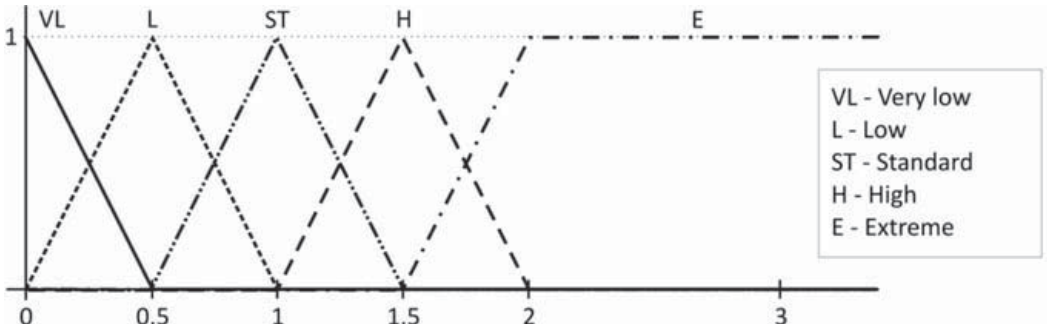


Figure 4. Linguistic scale for evaluating academic faculty in teaching used in IS HAP.

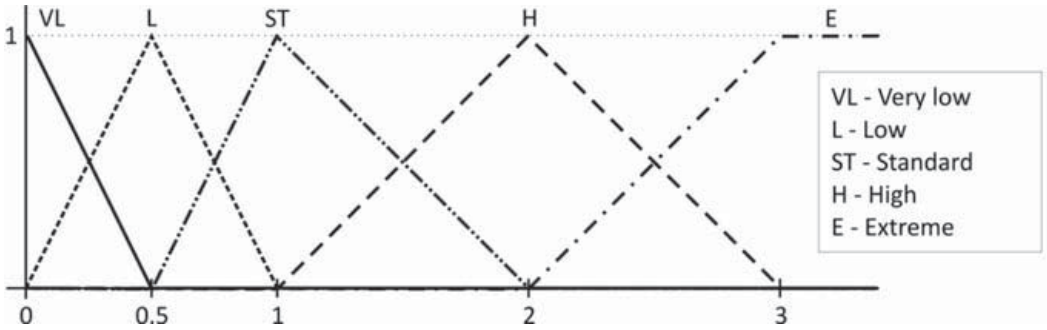


Figure 5. Linguistic scale for the evaluating academic faculty in research and development used in IS HAP—illustration of different meanings of the same linguistic terms (see Figure 4) in a different context.

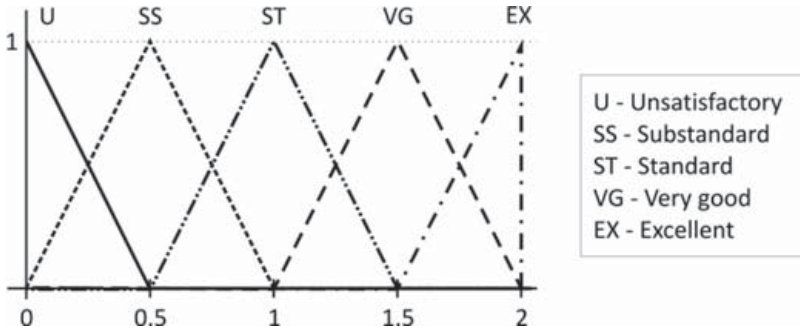


Figure 6. Linguistic scale for evaluating academic faculty used in IS HAP. The linguistic terms in this scale are used to describe outputs of the evaluation model to the users.

et al. (2011, 2013) for more details). The system is based on two inputs—evaluation of an academic faculty member in teaching (see Figure 4) and evaluation of the academic faculty member in research and development (see Figure 5). For both areas 5 linguistic values are used to describe the performance of the academic faculty member: very low, low, standard, high, extreme. The meanings of these words are modelled by the respective triangles in

Overall performance of a current staff member in teaching and RD		Research and Development performance				
		Very low	Low	Standard	High	Extreme
Teaching performance	Very low	Unsatisfactory	Unsatisfactory	Substandard	Standard	Very good
	Low	Unsatisfactory	Unsatisfactory	Substandard	Very good	Excellent
	Standard	Substandard	Substandard	Standard	Very good	Excellent
	High	Standard	Very good	Very good	Excellent	Excellent
	Extreme	Very good	Excellent	Excellent	Excellent	Excellent

Figure 7. Rule base describing the evaluation process in IS HAP—25 linguistic rules.

Figures 4 and 5. It can be seen that the meanings of the neighbouring linguistic terms overlap. This can be interpreted in the following way: as teaching performance (Figure 4) improves—moving along the horizontal axis from 0 to the right, the true linguistic description of the performance ceases to be "very low" and gradually moves to "low"; for the value of 0.5 on the horizontal axis, "low" is an entirely appropriate description and as the performance of the staff member improves, "low" ceases to be an appropriate description and "standard" becomes more appropriate up to the value of 1, where standard is entirely appropriate. This way the value of 0.9 can be interpreted as being "20% low and 80% standard"—that is "somewhere between a low and a standard performance but closer to standard".

The relationship between the evaluation in teaching and research and development is described by the rule base in Figure 7, which can be read as 25 rules thus:

> RULE 1: "if teaching performance is *low* and research and development performance is *low*, then the overall evaluation is *unsatisfactory*",
> ...
> RULE 14: "if teaching performance is *standard* and research and development performance is *high*, then the overall evaluation is *very good*",
> ...
> RULE 25: "if teaching performance is *extreme* and research and development performance is *extreme*, then the overall evaluation is *excellent*".

The meanings of the linguistic terms of the output variable are shown in Figure 6. The rule base is easy to understand and can be used not only to compute the linguistic evaluation, but also to explain to the academic faculty members what kind of behaviour will result in which particular evaluation. Although the description is highly comprehensible, the evaluation function represented by the rule base is quite a complex one (see Figure 8, Stoklasa, 2011) describes how the evaluations are computed at the mathematical level of the model). This illustrates that linguistic models are capable of describing complex relationships

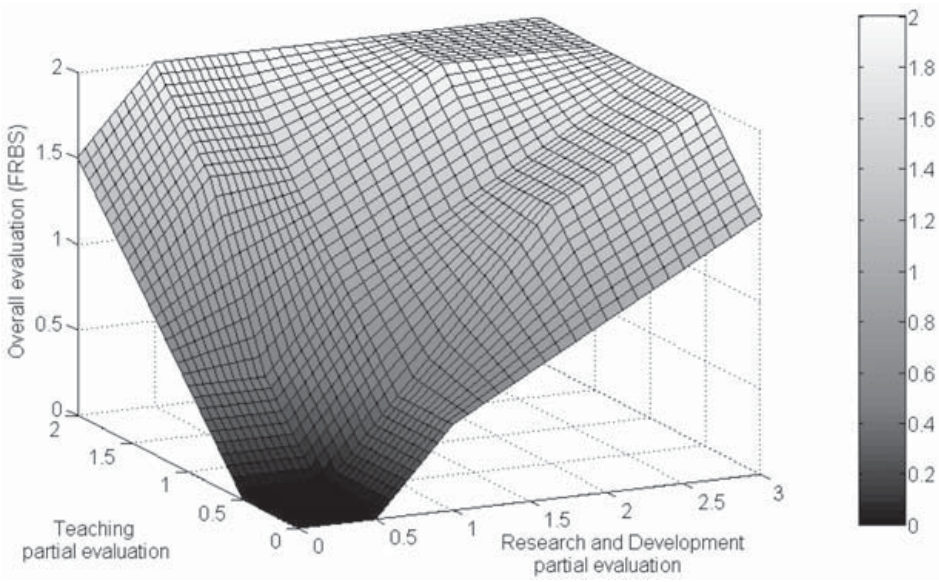


Figure 8. Plot of the evaluation function described by the fuzzy rule base from Figure 7.

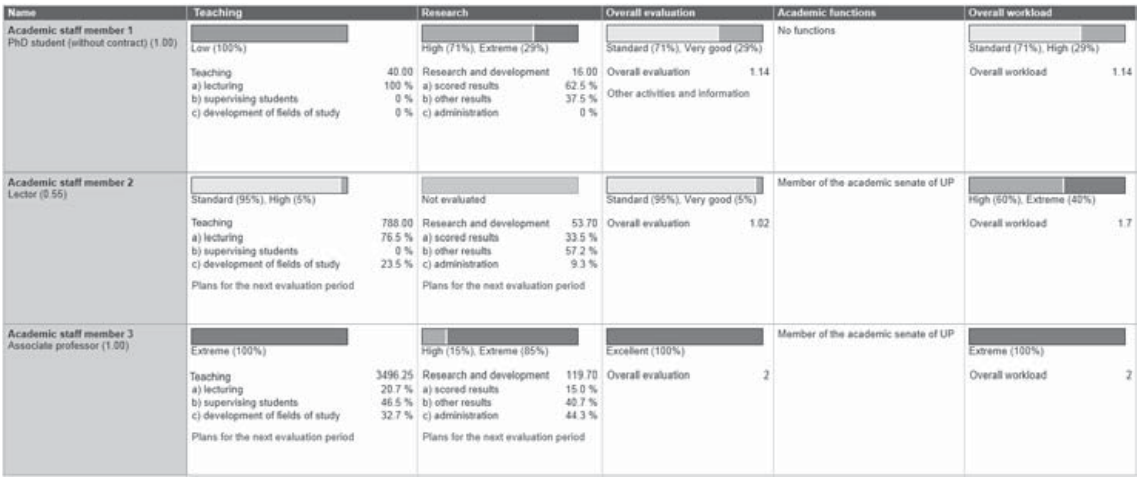


Figure 9. Example of graphical outputs (here bars in different shades of grey that sum up the evaluation information; colours are used in the actual output of IS HAP) and linguistic outputs (under the bars) from the evaluation model used in IS HAP.

in a way that is easy to understand. Also adjustments to the evaluation process can be made simply by changing the outputs (that is the "then" part of the 25 rules). The outputs can easily be transformed into colour bars (see Figure 9) by assigning a colour to each value of the output variable. If the overall evaluation is "60% standard and 40% very good", we will obtain a rectangle which will be 60% yellow and 40% light blue (that is an output that is uncertain and requires the active participation of the evaluator to be appropriately interpreted within the whole evaluation context, which is desirable).

Psychological diagnostics (example 2)

Linguistic rules can also be used to classify objects into categories. This is a typical task in psychological diagnostics for example. Again, we can obtain rules that describe under which conditions an object (a client) should be classified into which category (assigned which diagnosis). Inputs for this classification process could be complex results from several test methods, from an interview or any other source of information we might use. It may prove useful not to see the diagnoses as mutually exclusive—a client may be assigned several diagnoses. Also, we can consider situations in which we are able to find only partial evidence for assigning specific diagnoses. Figure 10 shows an example output of such a model in which we consider 6 diagnoses dg_1, ..., dg_6. These results can be interpreted such that if we have confirmed diagnosis 1, we have found partial evidence for diagnoses 2, 4 and 6 and we have found no confirmatory information for diagnoses 3 and 5.

If we also add rules that describe the conditions under which we can disprove a diagnosis, we can obtain results as depicted in Figure 11. This kind of thinking brings additional information to the diagnostics situation. We can interpret the results in the following way: diagnosis 1 can be seen as confirmed, there is contradictory information concerning diagnosis 2—it is partially confirmed and partially disproved, we have found

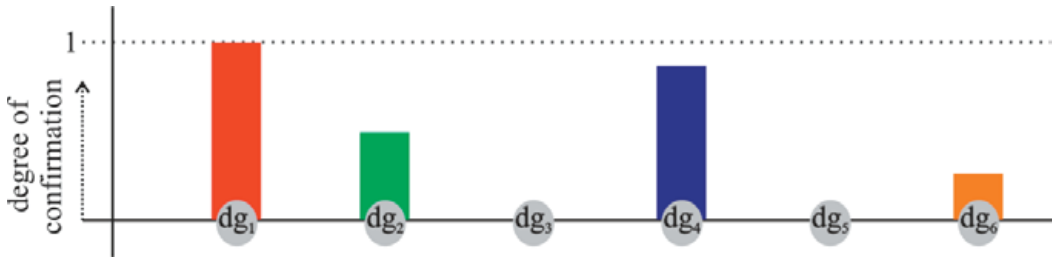


Figure 10. Example of a possible output of a fuzzy classification model—diagnostics (only confirmatory information for all diagnoses available).

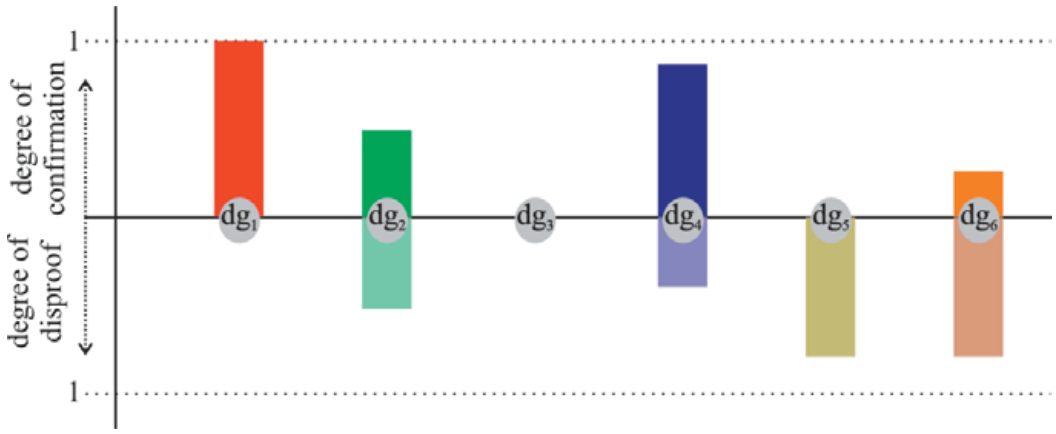


Figure 11. Example of a possible output of a fuzzy classification model—diagnostics (confirmatory information and information disproving each diagnosis available).

no information (neither confirmatory, nor disproving) for diagnosis 3, there is strong but incomplete confirmation of dg_4, but some disproving information has also been found, dg_5 can be considered as disproved, as can dg_6 (where only a small level of confirmation has been found). If we add the disproving rules, we are able to identify the ambivalent information (dg_2). We are now able to distinguish between dg_3 (complete lack of information on this diagnosis—no reason to confirm or disprove it) and dg_5 (now clearly disproved).

Conclusions

Psychology relies substantially on self-report based methods, which provide linguistic and, hence uncertain, information. Despite its uncertainty, linguistic information is sufficient to describe some systems and well suited to describe systems with human components. As such it can prove useful in that it can deal with uncertain and linguistic information in psychology, reflect the partial validity of statements and represent it formally. We have identified several prototypical issues which can signal that the use of fuzzy methodology may provide useful tools. We have discussed what the fuzzy approach can bring to the table that other mathematical tools cannot and also some possible shortcomings in the fuzzy approach.

In our two examples, we have illustrated that using the linguistic fuzzy modelling approach means we can easily understand and easily adjust models of an individual's knowledge, decision-making process and understanding of certain systems. These models operate on two levels—linguistic and formal. The formal level allows us to input the models into a computer—this way, in the case of psychological diagnostics, part of the diagnostics data can be pre-processed, based on the diagnostician's own knowledge and experience reflected in linguistic rules and the diagnostician can be provided with comprehensive output—see e.g. Figure 11. We have provided several reasons for why the fuzzy approach might be considered the tool of choice in some of the situations a psychologist may encounter. The final decision as to whether or not to try these methods now rests with the reader.

References

Arfi, B. (2010). *Linguistic fuzzy logic methods in social sciences.* Berlin Heidelberg: Springer-Verlag.

Bebčáková, I., Talašová, J., & Škobrtal, P. (2010). Interpretation of the MMPI-2 Test based on fuzzy set techniques. *Acta Universitatis Matthiae Belii ser. Mathematics 16*, 5-16.

Burisch, M. (1993). In search of theory: Some ruminations on the nature and etiology of burnout. In W. B. Schaufeli, C. Maslach, &T. Marek (Eds.), *Professional burnout: recent developments in theory and research* (pp. 75-93). Washington: Taylor & Francis.

Dubois, D., & Prade, H. (Eds.). (2000). *Fundamentals of fuzzy sets. The handbook of fuzzy sets series*. Boston, London, Dordrecht: Kluwer Academic Publishers.

Horowitz, L. M., & Malle, B. F. (1993). Fuzzy concepts in psychotherapy research. *Psychotherapy research, 3,* 131-148.

Klir, G. J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*. New Jersey: Prentice Hall.

Massaro, D. W., Weldon, M. S., & Kitzis, S. N. (1991). Integration of orthographic and semantic information in memory retrieval. *Journal of Experimental Psychology Learning, Memory and Cognition, 17,* 277-287.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological review, 85*, 172-191.

Ragin, C. C. (2000). *Fuzzy-set social sciences.* Chicago: University of Chicago Press.

Smithson, M., & Oden, C. G. (1999). Fuzzy set theory and applications in psychology. In H. J. Zimmermann (Ed.), *Practical Applications of fuzzy technologies* (pp. 557-585). Norwell: Kluwer Academic Publishers.

Smithson, M., & Verkuilen, J. (2006). *Fuzzy set theory: Applications in the social sciences.* Thousand Oaks, London, New Delhi: Sage Publications.

Smithson, M. (1986). *Fuzzy set analysis for behavioral and social science.* New York, Berlin, Heidelberg: Springer-Verlag.

Stoklasa, J., Holeček, P., & Talašová, J. (2012). A holistic approach to academic staff performance evaluation – a way to the fuzzy logic based evaluation, *Peer reviewed full papers of the 8th international conference on evaluation for practice "Evaluation as a tool for research, learning and making things better". A Conference for Experts of Education, Human Services and Policy, 18 – 20 June 2012, 2012, Pori, Finland*, 121-131.

Stoklasa, J., Jandová, V., & Talašová, J. (2013). Weak consistency in Saaty's AHP – evaluating creative work outcomes of Czech Art Colleges. *Neural network world 23*, 61-77.

Stoklasa, J., & Talašová, J. (2013). AHP based decision support tool for the evaluation of works of art - Registry of Artistic Performances., *Proceedings of the Finnish operations research society 40th anniversary workshop – FORS40, Lappeenranta 20. – 21.8.2013*, *LUT Scientific and Expertise Publications No. 13*, 44-47.

Stoklasa, J., & Talašová, J. (2011). Using linguistic fuzzy modeling for MMPI-2 data interpretation. *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011 – part II, Praha, Czech Republic*, 653-658.

Stoklasa, J., Talašová, J., & Holeček, P. (2011). Academic staff performance evaluation – variants of models, *Acta Polytechnica Hungarica* 8(3), 91-111.

Zadeh, L. A. (1975). The concept of linguistic variable and its application to approximate reasoning. *Information Sciences*, *Part 1: 8*, 199-249; *Part 2: 8*, 301-357; *Part 3: 9*, 43-80.

Zadeh, L. A. (1965). Fuzzy sets. *Inform. Control, 8*, 338-353.

Zemková, B., & Talašová, J. (2011). Fuzzy sets in HR Management, *Acta Polytechnica Hungarica 8* (3), 113- 124.

Zétényi, T. (Ed.). (1988). *Fuzzy sets in psychology.* Amsterdam, New York, Oxford, Tokyo: North-Holland.

Department of Mathematical Analysis and Applications of Mathematics,
Faculty of Science,
Palacky University in Olomouc,
tř. 17. listopadu 12
771 46 Olomouc
Czech Republic
E-mail: jan.stoklasa@upol.cz

Department of Mathematical Analysis
and Applications of Mathematics,
Faculty of Science,

Palacky University in Olomouc,
Czech Republic;

LUT Graduate School,
Lappeenranta University of Technology,
Finland
E-mail: tomas.talasek@upol.cz
and
Social Prevention Centre,
Family counselling
Olomouc,
Czech Republic
E-mail: musilova@ssp-ol.cz

PUBL. VII

Stoklasa, J., Talášek and Talašová, J., AHP and weak consistency in the evaluation of works of art - a case study of a large problem.

Accepted for publication in *International Journal of Business Innovation and Research*, Inderscience. (2014)

http://www.inderscience.com/jhome.php?jcode=ijbir

AHP and weak consistency in the evaluation of works of art – a case study of a large problem

Jan Stoklasa*

Centre of the Region Haná for Biotechnological
and Agricultural Research,
Faculty of Science,
Palacký University in Olomouc,
Olomouc, Czech Republic
Fax: +420-585-634-002
E-mail: jan.stoklasa@upol.cz
*Corresponding author

Tomáš Talášek and Jana Talašová

Department of Mathematical Analysis
and Applications of Mathematics,
Faculty of Science,
Palacký University in Olomouc,
Olomouc, Czech Republic
Fax: +420-585-634-002
E-mail: tomas.talasek@upol.cz
E-mail: jana.talasova@upol.cz

Abstract: The paper describes an evaluation methodology based on Saaty's AHP, that relaxes the classical Saaty's consistency condition and works with the concept of weak consistency. Weak consistency is seen as a minimum requirement on the consistency of experts' preferences. The relationship of weak consistency and the linguistic level of consistency description using the Saaty's scale is discussed in the paper. The benefits of using weak consistency with large problems are presented – for data input, for the flexibility of the mathematical model and to facilitate further adjustments of the evaluation methodology. In the case of pairwise comparison matrices with ordered categories, the fulfilment of weak consistency can be checked during the data input phase. This way the weak consistency of pairwise comparison matrices can be achieved even for large numbers of categories – unlike the full consistency in Saaty's sense. The paper also provides a case study of a practical application of the proposed evaluation methodology – the mathematical model for the evaluation of creative work outcomes of Czech Art Colleges. The case presented here combines evaluation based on objective criteria with peer review and suggests a possible solution to the problem of arts evaluation for funding purposes.

Keywords: analytic hierarchy process; AHP; multiple criteria evaluation; MCDM; consistency; art.

Reference to this paper should be made as follows: Stoklasa, J., Talášek, T. and Talašová, J. (xxxx) 'AHP and weak consistency in the evaluation of works of art – a case study of a large problem', *Int. J. Business Innovation and Research*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Jan Stoklasa's professional interests include multiple criteria decision support models and evaluation models using fuzzy approach and linguistic fuzzy modelling and applications of these models in human resource management, psychology, medicine and management. He also conducts research concerning the applications of fuzzy modelling in disaster management with focus on medical rescue services and in the area of university management decision support. He is a co-author of an academic faculty performance evaluation model for tertiary education institutions in the Czech Republic and of an evaluation model for creative work outcomes of Czech Art Colleges and Faculties.

Tomáš Talášek graduated in Applied Mathematics in Economy at Palacký University in Olomouc. He is currently working on his PhD thesis on systems based on fuzzy rule bases and their applications. He is also interested in fuzzy classification, software solutions for fuzzy problems and fuzzy neural networks. He is a member of the team responsible for the development of the evaluation methodology for the works of art resulting from the creative work of Czech Art Colleges and Faculties, that is currently being applied in the process of funds distribution from the state budget of the Czech Republic among public universities.

Jana Talašová is an Associate Professor of Applied Mathematics at the Faculty of Science, Palacký University in Olomouc. Her research is focused on multiple criteria evaluation and decision making, applications of the fuzzy set theory and linguistic fuzzy modelling. She has developed the theoretical basis for software tools NEFRIT and FuzzME (fuzzy multiple criteria evaluation). Her expert activities in higher education management include design of mathematical models (a model for academic faculty performance evaluation, model for evaluation of creative work outcomes at Czech Art Colleges) and analyses of the models for evaluation and funding of Czech universities.

This paper is a revised and expanded version of a paper entitled 'AHP based decision support tool for the evaluation of works of art – Registry of Artistic Performances' presented at Finnish Operations Research 40th Anniversary Workshop, Lappeenranta, Finland, 20–21 August 2013.

1 Introduction

Evaluation of works of art is a difficult task. Individual taste and preferences are an important part of the art assessment process and as such the consensual evaluation in a group of experts is not easy to obtain. However, particularly for the purposes of funds distribution on national level, tools for the assessment of performance in the area of artistic work need to be available. Multiple criteria evaluation approach able to combine

(multi) expert assessment of the quality of the output with more objective criteria proves to be justified for this purpose (see the example of Ministry of Education of the Slovak Republic, 2008). This paper aims to present an interesting evaluation methodology based on Saaty's AHP and the concept of weak consistency and a case study of such a tool developed and currently used in the Czech Republic (Stoklasa et al., 2013; Talašová and Stoklasa, 2011).

In Section 2 we introduce the necessary mathematical tools and concepts. Section 3 summarises the basic ideas of the Registry of Artistic Performances (RUV in Czech) – the evaluation criteria and categories of works of arts used in RUV. Basic principles of the methodology for evaluation of works of art in the Czech Republic are also presented here. In Section 4 the mathematical model used to determine the scores of each category of works of art is described. Section 5 presents the results of the adjustment of the mathematical model to reflect specific findings from the pilot run of RUV. The following section provides discussion of the presented results.

2 Preliminaries

Let us consider n categories we need to evaluate. The multiplicative Saaty's matrix of preference intensities S can be used to express the preferences of a group of experts among pairs of categories. The square matrix $S = \{s_{ij}\}_{i,j=1,\ldots,n}$ is required to be reciprocal, that is $s_{ij} = 1/s_{ji}$ for all $i, j = 1, 2, \ldots, n$. If experts input their intensities of preferences between categories i and j, and if we assume that category i is preferred to category j or of equal importance, the elements s_{ij} are chosen from the set $\{1, 2, \ldots, 9\}$. Saaty (see Saaty and Vargas, 2006; Saaty, 2000) suggests linguistic descriptors to be used by the experts to express their preferences (see Table 1) in his analytic hierarchy process (AHP) method.

Table 1 Saaty's scale

s_{ij}	*Linguistic meanings*
1	Category i is *equally important as* category j
3	Category i is *slightly/moderately more important than* category j
5	Category i is *strongly more important than* category j
7	Category i is *very strongly more important than* category j
9	Category i is *extremely/absolutely more important than* category j
2, 4, 6, 8	Correspond with the respective intermediate linguistic meanings

The elements s_{ij} of the matrix S are expertly defined estimations of the ratio w_i/w_j, where w_i is the evaluation of category i and w_j is the evaluation of category j. Finding the evaluations $w_1, \ldots, w_n$ of the categories based on the information provided by experts through the matrix S means finding the arguments of the minimum of expression (1).

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(s_{ij} - \frac{w_i}{w_j}\right)^2 \tag{1}$$

The evaluations w_i for all $i = 1, \ldots, n$ can also be computed using the logarithmic least square method in the form of (2).

$$w_i = \sqrt[n]{\prod_{j=1}^{n} s_{ij}} \tag{2}$$

An exact solution w_i to (1) (see e.g., Saaty, 2008) can be obtained as the i^{th} component of the eigenvector of S corresponding to its largest real eigenvalue $\lambda_{\max}$ (also known as the spectral radius of S). If $s_{ij} = w_i/w_j$ for all $i, j = 1, \ldots, n$, the matrix S is fully consistent in Saaty's sense. The Saaty's consistency condition can be also formulated in the following way:

$$s_{ik} = s_{ij} \cdot s_{jk}, \text{ for all } i, j, k = 1, 2, \ldots, n. \tag{3}$$

It is well known that the condition (3) is usually not fulfilled by expertly defined pairwise comparison matrices of larger order (particularly when linguistic descriptors from Table 1 are used in the input process). This can be expected even more frequently when working with experts outside the field of mathematics, for whom the consistency condition (3) formulated not for the linguistic labels of intensities of preferences, but for their numerical values, is not easy to interpret. When we substitute the linguistic values of Saaty's scale presented in Table 1 into (3) then the desired full consistency condition for S turns out to be very counterintuitive. Let us for example consider three categories A, B and C. According to (3) if $s_{AB} = 3$ and $s_{BC} = 3$ then s_{AC} has to be equal to the product of the two intensities of preferences, therefore for a completely consistent matrix S it holds that $s_{AC} = 9$. If we now describe the same situation using the linguistic labels proposed by Saaty, we realise that from the fact that A is *slightly more important* than B and B is *slightly more important* than C we need to infer that A is *absolutely more important* than C. This is, in our experience, not in accordance with the expectations of the experts providing input data for the model (particularly when we realise that *slightly more important* is the second lowest degree of preference that can be expressed by linguistic terms using the Saaty's linguistic scale presented in Table 1). Hence the use of linguistic labels to express experts' intensities of preferences may introduce further inconsistencies into the matrix S during the input phase.

To deal with the fact, that for larger pairwise comparison matrices the complete consistency [as formulated by condition (3)] can not be achieved, Saaty proposes an inconsistency index CI based on the spectral radius of S by (4), where n is the order of S (in our case the number of categories).

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{4}$$

The matrix S is considered consistent enough in Saaty's sense, if condition (5) holds, where CR is the so called inconsistency ratio and RI_n is the random inconsistency index of a matrix of intensities of preferences of order n (RI_n is obtained as an average of inconsistency indices for randomly generated reciprocal multiplicative matrices of intensities of preferences of the order n). Although RI_n can be computed from randomly

generated matrices, Alonso and Lamata (2006) show that it can be also estimated by the equation $RI_n = (2.7740n - 4.3513 - n)/(n-1)$.

$$CR = \frac{CI}{RI_n} < 0.1 \tag{5}$$

Other approaches to inconsistency measurement can be found for example in Lamata and Pelaez (2002), Ji and Jiang (2003) or Herrera-Viedma et al. (2004). Brunelli et al. (2013) provide a comparison of various inconsistency measures. The threshold set in (5) at 0.1 is rather arbitrary, although it has been successfully used in many real world applications.

An important deficiency of many of the inconsistency measures that can be found in the literature is, apart from the arbitrary definition of the threshold for acceptable consistency, the fact that the actual inconsistency can be assessed no sooner than the matrix S is completed. However if we require the experts to input a large matrix of preference intensities, the input process might be too time consuming to be repeated in case the sufficient consistency is not achieved.

Various methods of dealing with this issue can be found in the literature. Some involve adjustments of the final matrix so that it becomes consistent enough, proposals of methods of inputting incomplete matrix S and computing evaluations from such matrix with missing elements can also be found (see e.g., Fedrizzi and Giove, 2013). Hence we either repeat the whole input process, require additional information, change the obtained information or compute evaluations regardless of the missing information [examples of such approaches can be found in Alonso et al. (2008), Fedrizzi and Giove (2007), Harker (1987), Kwiesielewicz and van Uden (2003), Nishizawa (2005) and Xu (2004, 2005)].

If we need to obtain information from a group of experts concerning their preferences on a large set of categories, obtaining additional data for changes of the resulting matrix S or inputting the whole matrix again may not be feasible in real applications. It seems reasonable to define a minimum requirement on the consistency of preferences of the experts that has to be met. The requirement should also be such that it can be checked during the input phase. Such minimum consistency requirement was presented in Stoklasa et al. (2013) and Talašová and Stoklasa (2011) as weak consistency of the matrix S. According to Stoklasa et al. (2013) S is weakly consistent, if for all $i, j, k \in \{1, \ldots, n\}$ the implications (6) and (7) hold.

$$s_{ij} > 1 \wedge s_{jk} > 1 \quad \Longrightarrow \quad s_{ik} \geq \max\{s_{ij}, s_{jk}\} \tag{6}$$

$$(s_{ij} = 1 \wedge s_{jk} \geq 1) \vee (s_{ij} \geq 1 \wedge s_{jk} = 1) \quad \Longrightarrow \quad s_{ik} = \max\{s_{ij}, s_{jk}\} \tag{7}$$

The property $s_{ik} \geq \max\{s_{ij}, s_{jk}\}$ is referred to as max-max transitivity in for example, Herrera-Viedma et al. (2004). It is easy to prove (see e.g., Stoklasa et al., 2013) that a consistent matrix S [in the Saaty's sense according to (3)] is also weakly consistent. Weak consistency defined by (6) and (7) is a natural and reasonable requirement on the consistency of expert preferences. If we substitute the linguistic labels from Saaty's scale into (6) or (7) we get results much more consistent with experts' expectations and experience. Let us again consider, for example, that $s_{AB} = 3$ and $s_{BC} = 3$, then from (6) it follows that $s_{AC} \geqq 3$. Linguistic equivalent of the statements from the previous sentence is the following: if A is *slightly more important* than B and B

is *slightly more important* than C then A has to be at least *slightly more important* than C. Such a linguistically described minimum consistency condition does not seem counterintuitive to the experts providing inputs for the mathematical model. Although the weak consistency is in fact a relaxation of the 'full' consistency (3) required by Saaty, in Sections 4 and 5 we will provide practical evidence that the use of weak consistency with large pairwise comparison matrices has significant benefits.

Furthermore if we order the categories according to their importance before we start inputting the matrix S, the weak consistency requirement can be checked in each step of the data input phase directly by the experts, as it translates into two simple conditions – the elements of S need to be non-decreasing in each row from left to right and non-increasing from the top downwards in each column. This allows us to obtain a weakly consistent matrix at the end of the input process and no additional modifications to the matrix are necessary for this minimum consistency condition to be met.

The quasi-ordering of categories (a transitive and complete relation) can be obtained by the pairwise comparison method. The matrix of preferences and indifferences $P = \{p_{ij}\}_{i,j=1,...,n}$, where $p_{ij} = 1$ iff category i is more important than category j, $p_{ij} = 0.5$ iff categories i and j are of equal importance and $p_{ij} = 0$ iff category j is more important that category i is constructed. For all $i, j = 1, ..., n$ it holds that $p_{ii} = 0.5$ and $p_{ij} = 1 - p_{ji}$. The row sums $R_i = \sum_{j=1}^{n} p_{ij}$, $i = 1, \ldots, n$ can be used to determine the quasi-ordering of the categories according to their importance (the largest R_i corresponds with the most important category).

3 Evaluation methodology for works of art in the Czech Republic

The Czech Republic is currently trying to find a way how to put Arts and Science at the same level (also for the purposes of funds distribution). This idea is currently being pushed forward mainly by Czech Art Colleges and Faculties, national level project has been established to support the development of necessary tools and methods and an expert group on multiple criteria evaluation has been involved in the process. That is mathematics has been summoned to help with the process of evaluation of works of art by the artists themselves.

The main tool to show that Arts and Science can really 'see eye to eye' is the Registry of Artistic Performances (RUV in Czech), that has been developed in recent years and is continuously being adapted to serve this purpose as well as possible. At present the "principles and rules of financing public universities" – a basic document for funds distribution among public universities in the Czech Republic (see Ministry of Education of the Czech Republic, 2011) already uses the outputs of RUV as one of the parameters for determining the funding of universities. This suggests that the idea that RUV represents is widely accepted and that the goals it has set are seen as reasonable.

The idea of the Registry of Artistic Performances is to promote excellence in artistic performance in the Czech Republic and store information concerning the works of art created by workers (teachers) of Czech Art Colleges and Faculties and classify them into categories according to criteria that will be described further in this section. This way information on the outputs of art colleges and faculties will be stored in a systematic way – in a predefined structure that allows the assessment of quality and hence allows to provide a basis for funds distribution. As there will always be a subjective component in the assessment of works of art, the proposed methodology aims to integrate expert

evaluation (a peer review component) with some more objective criteria to assign scores to the categories.

The sum of these scores for each Art College of Faculty is used as a measure of performance of these institutions and it is used as a basis for allocating a part of the subsidy from the state budget of the Czech Republic among them. For the purposes of evaluation, the art sector in the Czech Republic is divided into seven segments – architecture, design, film, fine arts, literature, music and theatre. The works of art are evaluated according to three criteria regardless of their segment of origin. In each criterion, three different levels are distinguished (denoted by capital letters that are then used for the description of categories):

- *Relevance or significance of the piece*

  A a new piece of art or a performance of crucial significance

  B a new piece of art or a performance containing numerous important innovations

  C a new piece of art or a performance pushing forward modern trends.

  This criterion is assessed expertly in a peer review process and plays the role of a quality assessment criterion in the model (this criterion can reflect the subjectivity in the evaluation process). Each segment of art provided real-life (historical) examples for levels A, B and C and also the general linguistic specifications are customised for each segment and made available to the expert evaluators.

- *Extent of the piece*

  K a piece of art or a performance of large extent

  L a piece of art or a performance of medium extent

  M a piece of art or a performance of limited extent.

  The levels of this criterion are again specified linguistically, by examples or by measurable characteristics for each segment on such a level of accuracy that most of the ambiguity in categorising works of art according to this criterion is removed. This can be seen as reflecting the amount of work needed to produce the piece, the costs associated with it, number of people involved in the creation of the piece and so on. Respective units can be used to determine the levels of this criterion quantitatively.

- *Institutional and media reception/impact of the piece*

  X international reception/impact

  Y national reception/impact

  Z regional reception/impact.

  For this criterion lists of institutions corresponding to level X, Y and Z are provided by each segment.

Based on the levels of the three criteria 27 categories of works of art are defined (see Figure 1). Each category is characterised by a triplet of letters (one for each criterion reflecting the level of the respective criterion), e.g., AKX, BLX, or CMZ. Each of these 27 categories is assigned a score (the mathematical model used to determine scores is described in Section 4). For the purposes of the development of the mathematical model, each segment provided a list of typical works of art for each of the 27 categories. These real life examples can also be used in the peer review process by the reviewers as prototypical examples.

Figure 1 Ordering of the categories, scores assigned by the eigenvector method and by the geometrical means method (2)

Category	Relevance or significance	Extent	Institutional reception	Eigenvector method	Geom. means method	adjusted scores	Category
AKX	crucial significance	large	international	305	305	305	AKX
AKY	crucial significance	large	national	259	254	229	AKY
AKZ	crucial significance	large	regional	210	217	94	AKZ
ALX	crucial significance	medium	international	191	194	208	ALX
AMX	crucial significance	limited	international	174	171	181	AMX
ALY	crucial significance	medium	national	138	138	156	ALY
ALZ	crucial significance	medium	regional	127	124	82	ALZ
BKX	containing numerous important innovations	large	international	117	112	132	BKX
AMY	crucial significance	limited	national	97	94	110	AMY
AMZ	crucial significance	limited	regional	90	87	73	AMZ
BKY	containing numerous important innovations	large	national	79	75	66	BKY
BKZ	containing numerous important innovations	large	regional	66	66	37	BKZ
BLX	containing numerous important innovations	medium	international	62	61	60	BLX
BMX	containing numerous important innovations	limited	international	48	50	54	BMX
BLY	containing numerous important innovations	medium	national	44	46	47	BLY
BLZ	containing numerous important innovations	medium	regional	40	41	31	BLZ
BMY	containing numerous important innovations	limited	national	37	38	40	BMY
BMZ	containing numerous important innovations	limited	regional	31	30	27	BMZ
CKX	pushing forward modern trends	large	international	26	26	24	CKX
CLX	pushing forward modern trends	medium	international	24	24	21	CLX
CKY	pushing forward modern trends	large	national	19	20	19	CKY
CKZ	pushing forward modern trends	large	regional	17	18	11	CKZ
CMX	pushing forward modern trends	limited	international	16	16	17	CMX
CLY	pushing forward modern trends	medium	national	12	13	15	CLY
CLZ	pushing forward modern trends	medium	regional	10	11	10	CLZ
CMY	pushing forward modern trends	limited	national	9	9	13	CMY
CMZ	pushing forward modern trends	limited	regional	8	9	8	CMZ

Note: Two columns on the right describe adjusted scores (see Section 5).

In the first step the evaluation procedure involves the assessment of each work of art by its creator and his/her university or faculty. By self assessment the author proposes the initial classification of the respective work of art – assigning an initial triplet of letters. The proposed category has to be approved by the dean of the Faculty or the head of the Art College the creator is employed at. In the second step, this initial evaluation is assessed by the council of the respective segment of arts and either confirmation of this evaluation is issued or a second evaluation (second triplet representing the opinion of the council) is assigned. This information (either one confirmed classification of the piece of art, or two conflicting classifications) is then provided to independent reviewers in a way so that each piece of art is assessed by at least two such reviewers. The final evaluation is determined as a majority opinion of all the parties (that is as a majority based consensus on the levels of the three criteria). In indecisive cases the external (independent) evaluation is favoured. This way the peer review based quality assessment

is integrated into the evaluation process and combined with two more objective criteria in a natural way.

Each work of art submitted into RUV is intended to accrue funding for its authors' university during seven years from its submission into RUV. During this period, the institutional impact of the piece of art may change (increase). In this case the category of the piece will be reconsidered (and the resulting score of the current piece will increase accordingly).

4 Mathematical background of the evaluation – determining the scores

It is quite obvious that the three criteria used for the evaluation are not fully independent. In this case the classical approach proposed by Saaty in his AHP – that is to determine the weights of each criterion and the weights of each level within each criterion and based on these to calculate the evaluation of the categories – is not appropriate. It is better to expertly compare the 27 categories using the pair wise comparison matrix. This however constitutes a methodological problem, as Saaty's AHP was never intended for direct comparisons of so many categories (for example the consistency condition (3) is almost impossible to fulfil for such large matrices of preference intensities).

Saaty's recommendation is to split large problems (we now need to consider a 27×27 matrix) into several problems of lower dimension. To do so would, however, in this case require the experts to express their preferences concerning really vaguely defined supercategories of works of art. This is obviously a problem, as the experts are unable to provide relevant information of this kind.

As was mentioned before, the mathematical model for determining scores of the 27 categories in RUV is based on a matrix of intensities of preferences of the order 27. This matrix was required to be weakly consistent, which allows the experts to use the linguistic labels of the Saaty's scale in an intuitive way (see Section 2) and also allows for the weak consistency condition fulfilment to be checked after each pairwise comparison (and the input of the respective intensity of preference). This ensures that the resulting matrix of pairwise comparisons is weakly consistent and no more adjustments are required to determine the evaluations.

To make the input process and the consistency control as easy as possible, the first step was to determine the quasi-ordering of the 27 categories using the pair wise comparison method. The resulting ordering is presented in the first column in Figure 1. In the following text we will also see, that the initial ordering of categories that are being compared can play an important role in further adjustments of the evaluation methodology.

In the second step the experts from various segments of arts were asked to input the matrix of intensities of preferences. Having the categories ordered according to their importance resulting from the first step, the experts were instructed to provide their preferences and not to violate the weak consistency. This way a weakly consistent matrix of preference intensities was obtained (see Figure 2), with inconsistency ratio $CR = 0.1996$ [Saaty's inconsistency ratio defined by (5)]. The scores of the categories were computed using the eigenvector method and the Geometrical means method (2) and then linearly transformed so that the maximum score is 305 (the same value is assigned to a paper in a top scientific journal in the current methodology for the evaluation of

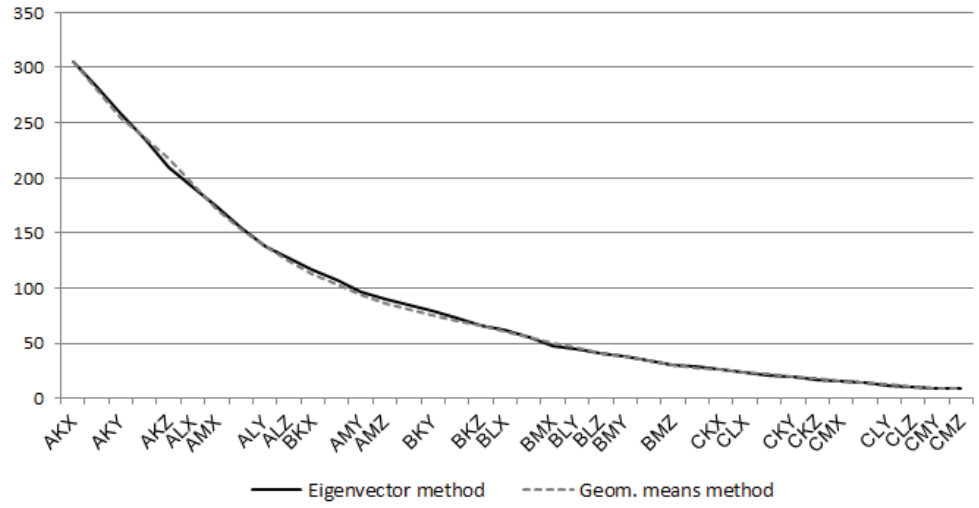
R&D outputs in the Czech Republic). The resulting scores are presented in Figure 1. The evaluation scale is exponential for both methods of computing final evaluations (see Figure 3).

Figure 2 Saaty's matrix of preference intensities as provided by the experts from various segments of arts

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
	AKX	AKY	AKZ	ALX	AMX	ALY	ALZ	BKX	AMY	AMZ	BKY	BKZ	BLX	BMX	BLY	BLZ	BMY	BMZ	CKX	CLX	CKY	CKZ	CMX	CLY	CLZ	CMY	CMZ
1 AKX	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	9	9	9	9	9	9	9	9	9	9
2 AKY		1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	9	9	9	9	9	9
3 AKZ			1	3	3	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	9	9	9	9	9	9
4 ALX				1	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9	9	9	9	9
5 AMX					1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9	9	9
6 ALY						1	3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9	9
7 ALZ							1	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	9
8 BKX								1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	7
9 AMY									1	3	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	7
10 AMZ										1	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	7
11 BKY											1	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	7
12 BKZ												1	3	5	5	5	5	5	5	5	5	5	7	7	7	7	7
13 BLX													1	5	5	5	5	5	5	5	5	5	7	7	7	7	7
14 BMX														1	3	3	3	5	5	5	5	5	7	7	7	7	7
15 BLY															1	3	3	5	5	5	5	5	7	7	7	7	7
16 BLZ																1	3	5	5	5	5	5	7	7	7	7	7
17 BMY																	1	5	5	5	5	5	7	7	7	7	7
18 BMZ																		1	5	5	5	5	5	5	5	5	5
19 CKX																			1	3	5	5	5	5	5	5	5
20 CLX																				1	5	5	5	5	5	5	5
21 CKY																					1	3	3	5	5	5	5
22 CKZ																						1	3	5	5	5	5
23 CMX																							1	5	5	5	5
24 CLY																								1	3	3	3
25 CLZ																									1	3	3
26 CMY																										1	3
27 CMZ																											1

Note: Consensus through all segments was required.

Figure 3 Comparison of the resulting scores of categories using the eigenvector method and the geometrical means method



Note: The distances between neighbouring categories reflect the respective intensities of preferences.

Hence the difference in scores of the two most preferred categories of works of art is significantly greater than the difference of scores of any other pair of neighbouring categories. Excellence of works of art can be promoted this way. This is completely in accordance with the intended purpose of the Registry of Artistic Performances. The differences between the resulting scores computed for all the evaluated categories by both methods are negligible (see Figure 3), the scores computed by the eigenvector method were selected to be used in the pilot run of RUV.

5 Modifications of the evaluations of categories – possibilities provided by the proposed evaluation methodology

After the first year pilot testing of RUV, 3,902 works of art were assembled in the RUV database and evaluated. A detailed analysis of the inputs and the evaluation methodology itself (including the determined scores of categories) was performed. The results of the analysis confirmed, that the three chosen criteria and their levels are sufficient to capture the structure of artistic production in the Czech Republic and reflect the specifics of each segment of arts well. The analysis also revealed, that a majority (over 66%) of the works of art submitted into RUV was classified as having only regional reception or impact (that is were assigned the letter Z as the third letter of the identifying triplet) – such proportion of Z results was unexpected. As one of the goals of RUV is to promote excellence of artistic performance, distributing a substantial sum of money among works of art with only regional impact will not help to achieve this goal.

As the works of art generate funding for seven years from their submission into RUV, it can be expected, that really good works of art will improve as for the Institutional and media reception/impact criterion soon. Based on this idea and the requirement to promote excellence in arts, it was decided to substantially lower the scores of all categories with regional reception/impact. To do this in a well justified and systematic manner, the 27×27 matrix of pairwise comparisons had to be reconstructed (intensities of preferences provided again as the view on the categories and what they represent had changed). First of all the desired ordering of categories was established. This ordering reflected that for each level of the relevance and significance criterion all the categories with Z were shifted lower within the original ordering to be the last three for the respective level of relevance/significance) – see Figure 4.

The approach to determine scores of categories described in Section 4 was applied. As weak consistency of the pairwise comparison matrix was required and the categories were again ordered according to their importance, the values in the matrix were required to be non-decreasing from left to right in every row and from the bottom up in every column. This allowed the experts to concentrate in fact on the shifts of intensities of preferences between pairs of categories. Thanks to this, the input process was quite quick and effortless and resulted in a weakly consistent matrix presented in Figure 4 for which the $CR = 0.1229$. The resulting evaluations obtained from this matrix are summarised in the two columns on the right in Figure 1 (the ordering of the categories is the original one in this figure to enable easy comparison of the new and original scores of categories). The desired effect of lowering substantially the scores of categories with regional reception/impact is illustrated in Figure 5. It can be easily seen, that the black solid line representing the adjusted scores is lower than the line representing the original scores for all categories with Z. Scores of categories of the type A?X and B?X – that

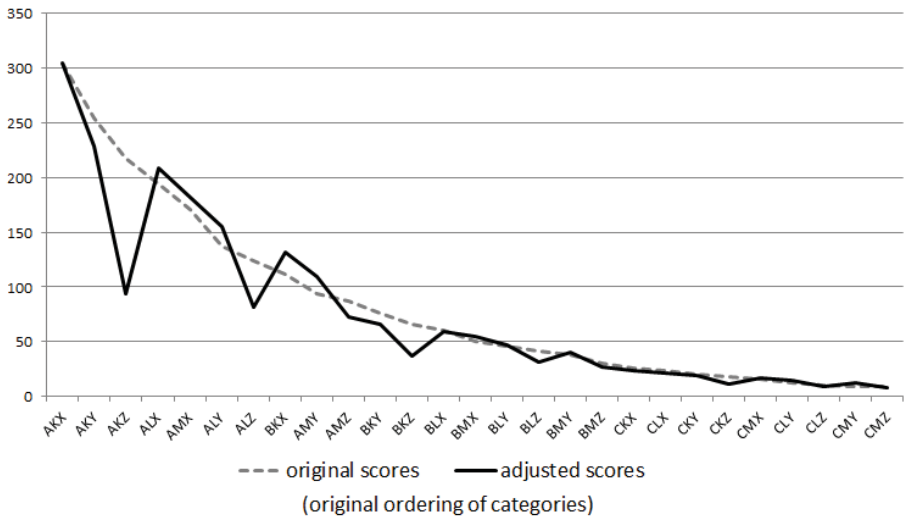
is categories with international reception and with high evaluations according to the relevance/significance criterion – have increased. This also is completely in accordance with the excellence promotion goal, as these are the categories that are most desired.

Figure 4 Complete Saaty's matrix of preference intensities for the reordered categories (see online version for colours)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
		AKX	AKY	ALX	AMX	ALY	BKX	AMY	AKZ	ALZ	AMZ	BKY	BLX	BMX	BLY	BMY	BKZ	BLZ	BMZ	CKX	CLX	CKY	CMX	CLY	CMY	CKZ	CLZ	CMZ
1	AKX	1	5	5	5	5	5	5	7	7	7	7	7	7	7	7	7	7	7	9	9	9	9	9	9	9	9	9
2	AKY	1/5	1	3	3	3	5	5	5	5	5	5	5	5	5	7	7	7	7	7	7	7	7	9	9	9	9	9
3	ALX	1/5	1/3	1	3	3	5	5	5	5	5	5	5	5	5	7	7	7	7	7	7	7	7	9	9	9	9	9
4	AMX	1/5	1/3	1/3	1	3	3	3	5	5	5	5	5	5	5	7	7	7	7	7	7	7	7	7	7	9	9	9
5	ALY	1/5	1/3	1/3	1/3	1	3	3	3	3	5	5	5	5	5	5	5	7	7	7	7	7	7	7	7	7	9	9
6	BKX	1/5	1/5	1/5	1/3	1/3	1	3	3	3	5	5	5	5	5	5	5	5	5	7	7	7	7	7	7	7	9	9
7	AMY	1/5	1/5	1/5	1/3	1/3	1/3	1	3	3	3	3	3	3	5	5	5	5	5	5	5	5	7	7	7	7	7	9
8	AKZ	1/7	1/5	1/5	1/5	1/3	1/3	1/3	1	3	3	3	3	3	3	5	5	5	5	5	5	5	5	7	7	7	7	7
9	ALZ	1/7	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1	3	3	3	3	3	3	3	5	5	5	5	5	5	5	5	7	7	7
10	AMZ	1/7	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1	3	3	3	3	3	3	5	5	5	5	5	5	5	5	5	7	7
11	BKY	1/7	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1	3	3	3	3	3	5	5	5	5	5	5	5	5	5	7	7
12	BLX	1/7	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1	3	3	3	3	3	5	5	5	5	5	5	5	5	7	7
13	BMX	1/7	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1/3	1	3	3	3	3	5	5	5	5	5	5	5	5	5	7
14	BLY	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1/3	1	3	3	3	3	5	5	5	5	5	5	5	5	7
15	BMY	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1/3	1	3	3	3	3	3	5	5	5	5	5	5	7
16	BKZ	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1	3	3	3	3	3	5	5	5	5	5	7
17	BLZ	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1	3	3	3	3	5	5	5	5	5	7
18	BMZ	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1	3	3	3	3	3	5	5	5	5
19	CKX	1/9	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1	3	3	3	3	3	5	5	5
20	CLX	1/9	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1/3	1	3	3	3	3	3	5	5
21	CKY	1/9	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1	3	3	3	3	5	5
22	CMX	1/9	1/7	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1	3	3	3	5
23	CLY	1/9	1/9	1/9	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1	3	3	5
24	CMY	1/9	1/9	1/9	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1/3	1	3	3
25	CKZ	1/9	1/9	1/9	1/9	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1	3
26	CLZ	1/9	1/9	1/9	1/9	1/9	1/9	1/7	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1/3	1/3	1
27	CMZ	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/7	1/7	1/7	1/7	1/7	1/7	1/7	1/7	1/7	1/7	1/5	1/5	1/5	1/5	1/5	1/5	1/3	1/3	1/3	1

Note: For each level of significance the categories of works of art with regional reception/impact are placed last – these categories are marked grey.

Figure 5 Comparison of the original scores of categories with the adjusted scores

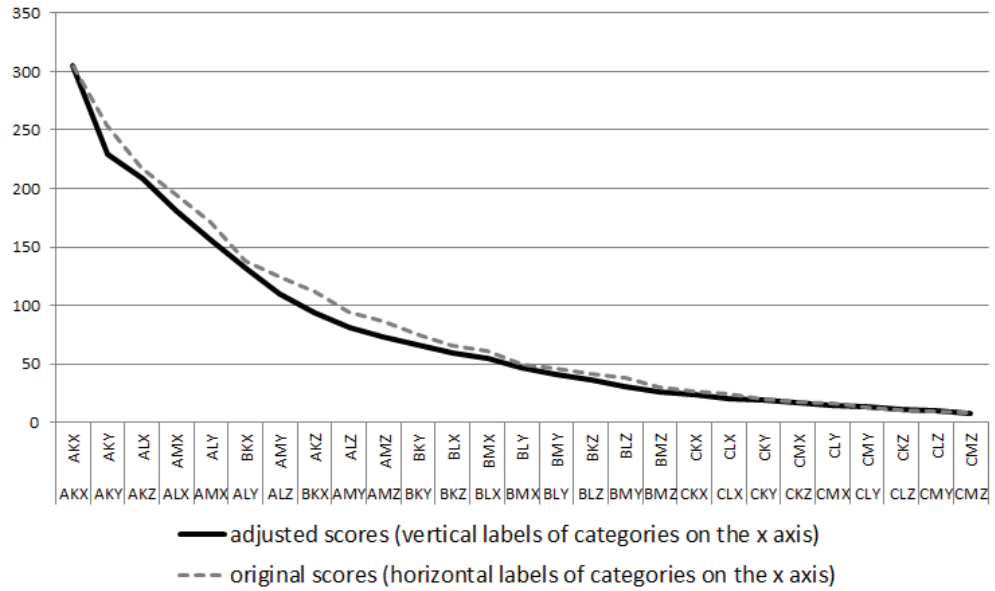


Notes: Categories on the x axis are ordered according to the initial ordering. Lower scores resulting from the proposed modification are evident.

6 Discussion

The case study of RUV on the proposed evaluation methodology and the underlying mathematical model provides a good example that Saaty's AHP approach can be successfully modified to large problems. The weakening of the Saaty's original consistency condition (3) by introducing a weaker requirement represented here by the weak consistency conditions (6) and (7) can provide the necessary space for the adaptation of the method for large problems. The weak consistency combined with initial ordering of categories provides an easy to use approach to obtain sufficiently consistent matrix of preference intensities in these cases. We can even observe that the (linguistic) concretisation of the relationships between certain levels of criteria and setting them into the context of the evaluation task can result in even higher consistency of the experts' preferences. In this case the additional information provided was "good works of art with regional reception/impact will soon get higher levels of reception/impact, there is therefore no need to evaluate these good pieces high too soon, as they have seven years to prove their reception is wider than regional". We can consider the shift from $CR = 0.1996$ for the original weakly consistent matrix to $CR = 0.1229$ for the matrix with readjusted order of categories to be a significant improvement (as it gets much closer even to the Saaty's sufficient consistency defined by $CR < 0.1$).

Figure 6 Comparison of the original scores of categories computed using the eigenvector method and the adjusted scores (where works of art with regional reception are assigned lower scores)



Notes: The vertical labels of the x axis correspond to the new adjusted scores, the horizontally typed labels correspond to the original scores. The intensities of preferences between neighbouring categories are not reflected in this figure.

We can also observe, that the AHP combined with weak consistency is quite robust to changes in the intensities of preferences. Figure 6 provides a graphical comparisons of the scores of categories computed from the original matrix of preference intensities (dashed grey line) and from the adjusted matrix (solid black line). The differences in scores of categories of the same order are not large and diminish with decreasing preference (that is with the increasing order) of the categories. Weak consistency also provides a good framework for the use of linguistic labels of Saaty's scale as was stressed in Section 2 – the linguistic description of the weak consistency condition seems to work well with the experts and their experience and expectations.

The proposed evaluation methodology is also quite flexible to changes. The adjustment of the scores based on redefinitions of given levels of criteria or the change in their interpretation is not only possible, but can be done quickly and with results that are required and expected – this is well illustrated by Figure 5.

7 Conclusions

The paper deals with the Saaty's AHP method in multiple criteria evaluation in such settings, where a large number of categories needs to be evaluated. It presents an evaluation methodology based on a modification of the Saaty's original approach. The consistency of experts' preferences expressed through the matrix of preference intensities is not assessed using the Saaty's consistency condition – instead a weaker condition is proposed. This weak consistency condition works well with large matrices of preference intensities, it also enables the consistency to be checked during the input phase and allows for easy adjustability of the evaluation methodology. The importance of an intuitive correspondence between the linguistic labels used to input the intensities of preferences (Saaty's linguistic scale) and the consistency condition is accented here.

The utility of the proposed evaluation methodology in multiple criteria evaluation setting has been discussed theoretically and illustrated on an example of its practical application – a case of the evaluation methodology of works of art within the Registry of Artistic Performances in the Czech Republic. The presented case does not only deal with the difficult task of the evaluation of works of art – it also provides a means for integrating the peer review component into the process of evaluation and combining it with measurable evaluation criteria using the evaluation methodology described in the first part of the paper. The Registry of Artistic Performances is currently being used in the Czech Republic for the distribution of a part of the subsidy from the state budget among Art Colleges and Faculties. As such it is in our opinion a good example of the usefulness of the proposed evaluation methodology.

Although presented here on an example of arts evaluation, the proposed evaluation methodology based on pairwise comparisons can be applied in many other contexts. It is well suited to be applied in situations where expert evaluation (an particularly evaluation provided by experts not much familiar with mathematics) plays a crucial role and where many classes/objects need to be evaluated. As it is able to include both peer review based evaluation and classical evaluation criteria, it can be also of applied to the evaluation of R&D outcomes and other tasks where quality assessment plays an important role.

Acknowledgements

The research presented in this paper was funded in part by the Educational policy fund – indicator F (Registry of Artistic Performances – RUV) from the state budget of the Czech Republic and partially also by the grant PrF_2013_013 Mathematical models of the Internal Grant Agency of Palacky University in Olomouc.

References

Alonso, S., Chiclana, F., Herrera, F., Herrera-Viedma, E., Alcalá-Fdez, J. and Porcel, C. (2008) 'A consistency-based procedure to estimate missing pairwise preference values', *International Journal of Intelligent Systems*, Vol. 23, No. 2, pp.155–175.

Alonso, J.A. and Lamata, M.T. (2006) 'Consistency in the analytic hierarchy process: a new approach', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 14, No. 4, pp.445–459.

Brunelli, M., Canal, L. and Fedrizzi, M. (2013) 'Inconsistency indices for pairwise comparison matrices: a numerical study', *Annals of Operations Research*, Vol. 211, No. 1, pp.493–509.

Fedrizzi, M. and Giove, S. (2013) 'Optimal sequencing in incomplete pairwise comparisons for large-dimensional problems', *International Journal of General Systems*, Vol. 42, No. 4, pp.366–375.

Fedrizzi, M. and Giove, S. (2007) 'Incomplete pairwise comparison and consistency optimization', *European Journal of Operational Research*, Vol. 183, No. 1, pp.303–313.

Harker, P.T. (1987) 'Incomplete pairwise comparisons in the analytic hierarcy process', *Mathematical Modelling*, Vol. 9, No. 11, pp.837–848.

Herrera-Viedma, E., Herrera, F., Chiclana, F. and Luque, M. (2004) 'Some issues on consistency of fuzzy preference relations', *European Journal of Operational Research*, Vol. 154, No. 1, pp.98–109.

Ji, P. and Jiang, R. (2003) 'Scale transitivity in the AHP', *The Journal of the Operational Research Society*, Vol. 54, No. 8, pp.896–905.

Kwiesielewicz, M. and van Uden, E. (2003) 'Ranking decision variants by subjective paired comparisons in cases with incomplete data', in Kumar, V. et al. (Eds.): *Computational Science and its Applications – ICCSA, Lecture Notes in Computer Science*, Vol. 2669, pp.208–215, Berlin/Heidelberg, Springer-Verlag.

Lamata, M.T. and Pelaez, J.I. (2002) 'A method for improving the consistency judgemets', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 10, No. 6, pp.677–686.

Ministry of Education of the Czech Republic (2011) *Zásady a pravidla financování veřejných vysokých škol pro rok 2012* (*Principles and Rules of Financing Public Universities for 2012*), Praha, Czech Republic.

Ministry of Education of the Slovak Republic (2008) *Smernica č. 13/2008-R zo 16. októbra 2008 o bibliografickej registrácii a kategorizácii publikačnej činnosti, umeleckej činnosti a ohlasov*, Slovak Republic.

Nishizawa, K. (2005) *Estimation of Unknown Comparisons in Incomplete AHP and it's Compensation*, Report of the Research Institute of Industrial Technology, Nihon University, No. 77.

Saaty, T.L. (2008) 'Relative measurement and its generalization in decision making, why pairwise comparisons are central in mathematics for the measurement of intangible factors – the analytic hierarchy/network process', *RACSAM*, Vol. 102, No. 2, pp.251–318.

Saaty, T.L. (2000) *The Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*, Vol. VI of the AHP Series, 478 pp., RWS Publ.

Saaty, T.L. and Vargas, L.G. (2006) *Decision Making with the Analytic Network Process: Economic, Political, Social and Technological Applications with Benefits, Opportunities, Costs and Risks*, Springer, New York.

Stoklasa, J., Jandová, V. and Talašová, J. (2013) 'Weak consistency in Saaty's AHP – evaluating creative work outcomes of Czech Art Colleges', *Neural Network World*, Vol. 23, No. 1, pp.61–77.

Talašová, J. and Stoklasa, J. (2011) 'A model for evaluating creative work outcomes at Czech Art Colleges', *Proceedings of the 29th International Conference on Mathematical Methods in Economics*, Part II, Praha, Czech Republic, Praha, pp.698–703.

Xu, Z.S. (2004) 'Goal programming models for obtaining the priority vector of incomplete fuzzy preference relation', *International Journal of Approximate Reasoning*, Vol. 36, No. 3, pp.261–270.

Xu, Z.S. (2005) 'A procedure for decision making based on incomplete fuzzy preference relation', *Fuzzy Optimization and Decision Making*, Vol. 4, No. 3, pp.175–189.

PUBL. VIII

Collan, M., Stoklasa, J. and Talašová, J., On academic faculty evaluation systems - more than just simple benchmarking. *International Journal of Process Management and Benchmarking*, 4(4), 437–455, 2014. DOI: 10.1504/IJPMB.2014.065522

http://www.inderscience.com/jhome.php?jcode=ijpmb

On academic faculty evaluation systems – more than just simple benchmarking

Mikael Collan*

School of Business,
Lappeenranta University of Technology,
Skinnarilankatu 34, FIN-53851, Lappeenranta, Finland
E-mail: mikael.collan@lut.fi
*Corresponding author

Jan Stoklasa

Centre of the Region Haná for Biotechnological
and Agricultural Research,
Faculty of Science,
Palacký University in Olomouc,
Šlechtitelů 586/11, 783 71 Olomouc, Czech Republic
and
LUT Graduate School,
Lappeenranta University of Technology,
Skinnarilankatu 34, FIN-53851, Lappeenranta, Finland
E-mail: jan.stoklasa@upol.cz

Jana Talasova

Department of Mathematical Analysis
and Applications of Mathematics,
Palacky University in Olomouc,
17. listopadu 1192/12, 771 46 Olomouc, the Czech Republic
E-mail: jana.talasova@upol.cz

Abstract: Academic faculty evaluation is a yearly recurring part of the management process at most universities and it is an issue that is getting more and more attention, as universities all over the world are required to become increasingly accountable for their performance and efficiency to their stakeholders. Designing good academic faculty evaluation systems is not a simple problem because multiple issues and a large number of criteria should be considered and aggregate in a sensible way. To highlight the diversity of existing academic evaluation systems, we present and shortly compare real world systems from four universities in three different countries. We argue that as there are no best practices or guidelines available for academic faculty evaluation systems the topic requires more research attention from both, the human resources management side and from the systems design side.

Keywords: academic faculty evaluation; multiple criteria evaluation; support system; personnel management.

Reference to this paper should be made as follows: Collan, M., Stoklasa, J. and Talasova, J. (xxxx) 'On academic faculty evaluation systems – more than just simple benchmarking', *Int. J. Process Management and Benchmarking*, Vol. X, No. Y, pp.000–000.

Biographical notes: Mikael Collan received his MSc in Social Sciences in the year 2000 and the DSc in Economics and Business Administration in 2004 from Abo Akademi University, Turku, Finland. He is currently a Professor of Strategic Finance with the Lappeenranta University of Technology Business School in Finland. His research is focused on profitability analysis, asset valuation methods and the application of fuzzy logic and real options in management decision support.

Jan Stoklasa is an Applied Mathematician and a Psychologist currently working at Palacky University in Olomouc, Czech Republic. He is finishing his PhD studies at Lappeenranta University of Technology, Finland. His research focuses on multiple criteria decision support and evaluation models using crisp methods and linguistic fuzzy modelling. The application areas of these models include human resource management, psychology, medicine and management. He also conducts research concerning the applications of fuzzy modelling in disaster management, with special focus on the medical rescue services and in the area of university management decision support.

Jana Talašová is an Associate Professor of Applied Mathematics at Palacky University in Olomouc, Czech Republic. Her research is focused on multiple criteria evaluation and decision making, applications of the fuzzy set theory and linguistic fuzzy modelling. Her current research focuses on mathematical models for multiple criteria evaluation in the area of higher education management: models for academic faculty evaluation, for the evaluation of creative work outcomes of art colleges and for the evaluation and funding of universities.

1 Introduction

Development of faculty evaluation systems at universities has become a more pressing issue, as the organisation of tertiary education is changing and the focus of universities is shifting in many countries (Bana e Costa and Oliveira, 2012). As funding of universities in many countries becomes more performance-based, see e.g., Hicks (2012), it is natural that also the academic faculty is evaluated according to their performance. The issue of quality of the activities performed by universities and by academic faculty has also become important. As measures are being taken to build quality assurance systems for universities national systems for evaluating academic faculty have been defined in some and are in the process of being defined in many European countries (Minelli et al., 2006; Elmore, 2008).

What needs to be stressed is that academic faculty evaluation is not a simple HR benchmarking problem but a rather complex issue with multiple criteria to be considered and with some 'academic' particularities. For example, faculty evaluation systems should be able to consider both, teaching and research contributions, as well as the contributions within and outside the traditional university sphere such as work in the academic or teaching community or in the area of the academia-industry collaboration. Within each of

these sub-areas the multiple different forms of academic contribution should be considered.

Academic freedom to choose one's main focus has been one of the drivers in academia for ages, this freedom of having chosen a profile is something that an evaluation system should be able to consider that is, people in academia specialise. Specialisation in research usually means less achievement in teaching, if this is tolerated a system should be able to reflect it. All in all, there are many issues on many different levels and within the different levels within the academic systems that make them a complex and an interesting focus for research. For some background on current faculty evaluation systems, see, e.g., Bana e Costa and Oliveira (2012).

In general, staff evaluation systems should be constructed only after careful analysis of the needs, the goals and the culture of the institution – this makes optimal systems for different organisations different. There are most likely best practices that can be identified for academic faculty evaluation but as was observed already above there is very little research on the issue. What remains is that there are many open questions and more identified problems than solutions.

Potential problems in building and instituting include, for example, finding reliable sources of inputs for the evaluation, finding a proper balance between the focus on quality and the orientation on measurable performance and finding tools that are able to reflect well the complexity of the evaluation context while still remaining understandable to the people working with the tools. The design of faculty evaluation systems calls for interdisciplinary cooperation as the above mentioned problems reflect competencies of different disciplines among which are human resources, decision-support systems and perhaps software engineers. Stakeholders, that is, academic faculty that are being evaluated and their superiors, should not be left out of the design process either.

As decision support systems, evaluation systems can provide information for various purposes, for example:

- faculty development
- identification of problems and bottlenecks
- information on the goals set by the faculty member and how they are reached
- promotion and tenure decisions
- identification of individual's skills and talents – better composition of teams and assignment of work
- declaration of relevant/beneficial activities (an evaluation system itself can provide information concerning the organisational goals and preferred ways of achieving them) – this in fact is an information transfer function of the evaluation system
- clarification of the possibility of mutual compensation among different areas of activities (how much can one specialise in one field at the expense of another)
- outplacement.

Benefits of an evaluation system for academic faculty that are being evaluated should also be considered, they may include for example:

- summary of all the work/activities that are considered beneficial or relevant to the institution as a source of information for professional CV or for grant applications
- proof of an individual's skills and preparedness for promotion, or a tenured position
- a record of professional career, optimally of its development in time
- identification of strengths and weaknesses
- partial guarantee of the objectivity of the evaluation (and the consequences of evaluation) – a tool that helps preventing injustice, or discrimination.

There are many approaches to academic faculty evaluation, ranging from manually operated simple rule-based approaches (or scorecards) to mathematically advance multiple criteria evaluation software systems (Uzoka, 2008). In our comparison of selected academic staff evaluation models, we concentrate on some key features of the models in the case description:

- areas of activities that are being evaluated (research, teaching, other)
- collection of information for the evaluation
- the actual 'scoring' model used to produce the evaluation.

Other selected relevant features are analysed in the comparison of the cases:

- sources of information for the evaluation (self-assessment, objective data, peer-review, student evaluations, committee, outside evaluation)
- possibility to reflect/accept specialisation of academic faculty
- the form of final evaluation (report, single number).

So far the research in the area of academic faculty evaluation and evaluation models has concentrated mainly on the evaluation of pedagogical activities of academic faculty, student learning outcomes and student evaluation of teaching performance and quality. Adjunct faculty evaluation has been also specifically considered (see e.g., Langen, 2011). Evaluation in the area of research activities has also been explored for example in 'scientometrics', where studies, methods and methodologies were created. The issue of bringing together the evaluations in various areas of academic activities into evaluation systems has received a significantly lower attention.

This paper continues by briefly describing four cases of actual academic faculty evaluation systems that are in place in Finland, in the Czech Republic and in the USA to illustrate the diversity of systems in place and the differences that can be found in actual academic evaluation systems. These four cases were selected after an analysis of various academic faculty evaluation models, currently used in universities in the Czech Republic (Talasova and Stoklasa, 2010; Jan Evangelista Purkyne University, 2012; Masaryk University, 2012; Tomas Bata University in Zlín, 2012), in Finland (Board of the Turku School of Economics, 2004; Lappeenranta University of Technology, 2011) and elsewhere (University of Technology Sydney, 2009; Wayne State University, 2009; McGill University, 2012; Flinders University, 2012; Texas A&M University, Kingsville, 2011a).

After the presentation of the four cases the paper closes with a summary and a discussion.

2 Case studies of several real live academic faculty evaluation models

The following four cases are intended to illustrate the main features of the models used in the real world and to go into details only if it is needed for understanding of the evaluation itself, or the aggregation method of partial evaluations used in the system. References to the relevant resources are provided for the reader to have an opportunity of getting deeper insights into the evaluation methodologies. Each case is constructed as follows: first, a short introduction is given of the university, where the system is used; second, a presentation of the evaluation system is given with focus on evaluation of teaching, research and other tasks; third, the collection of input data for the evaluation within the university is discussed; fourth; observations and issues of particular interest within the system are discussed. This paper excludes the detailed presentation of the national context of these systems. The choice of cases presented is based on the authors' personal experience and/or deep understanding of the selected systems and is not a representative selection of academic evaluation systems as a whole; however, we believe that the cases illustrate the diversity of academic evaluation systems rather well.

2.1 Case 1: University of Turku (FIN)

2.1.1 Introduction of the university

University of Turku (UTU) is a multidisciplinary scientific university located in the city of Turku on the South-Western coast of Finland. UTU is one of the largest universities in Finland. In the beginning of 2010 The UTU and the Turku School of Economics (TSE) were merged into one university called the UTU.

2.1.2 The evaluation system

The evaluation system in UTU is based on a 'performance points system' that originates from the TSE and has been used there since 2005. The system has been taken into use in the whole UTU for the year 2011. The performance points system is a performance measure of research and research related activities. The system cannot be used as a holistic system in the evaluation of academic faculty, that is to say, that teaching or other pedagogical activity is not included in a standardised system.

2.1.3 Teaching evaluation

For teaching performance an ad-hoc system is in use: the employer representative does a heuristic case by case evaluation of teaching performance.

2.1.4 Research evaluation

The system consists of four types of activities that accrue performance points. These are divided into the following categories (A–D):

A publications

B scientific expert assignments

C international teaching and research mobility

D research funding gathered.

For each category there are a number of sub categories, for example, for the category A publications, there are six sub-categories:

A1 monographs and sections in monographs

A2 refereed journal publications

A3 conference publications

A4 theses

A5 publications in scientific outlets without a referee practice

A6 citations (SSCI+SCI).

Each one of the sub-categories is further divided into publication types, for which research points are awarded separately and depending on if the publication is international or national. An example of the division into and on the level of publication types is shown in Figure 1 similar division exists for almost all sub-categories.

Figure 1 A part of the UTU research point guidelines (w. translation)

	Domestic	International
A 2.1. High quality journal article • blind peer review • more than one reviewer • highly rated journal	6	12
Points deducted if elements of quality are missing:		
- no blind review	-1	-1
- only one reviewer	-1	-1
- not a highly rated journal	-1	-1 to -4
Points at minimum	3	6

2.1.5 Other academic tasks evaluation

Research points do not only accrue from publication activities but also activity within the scientific community is rewarded by research points. Activity in editorial boards of journals, in organising conferences and within leading positions of scientific

organisations are taken into consideration and acting as an expert in, for example, committees of the Finnish Academy of Sciences are considered meritorious. Also acting as review for journals and conferences are rewarded by points. Acting as a faculty appointed reviewer or opponent to a dissertation or as an expert with regards to selection of faculty positions and acting in other expert tasks for, e.g., the Finnish Parliament, the EU Commission, or such yield points. Also being rewarded a prize for scientific achievements will yield research points. Faculty mobility and collected research funding are also considered.

2.1.6 Collection of evaluation data

The collection of research points is done variably, unit by unit, usually by ad-hoc excel sheets maintained by a nominated person (usually the department secretary) that collects the information from the academic faculty; most often by circulating points submission requests by email. As the points' collection is not standardised and there are colourful practices within different organisations regarding the collection of points, some research merits that would generate research points may possibly end up never being reported. The excel sheets are then sent to 'central administration' where the information is aggregated. It is not unusual that the same (research related) information is collected at UTU even three times by different organisations within the university. The reported research points are approved by a research board.

2.1.7 Observations

Some observations that can be made about the system include the fact that impact factor (IF) of journals does not have a direct effect on the points given to publications in international journals and that the points that a single international journal article can fetch are at maximum twelve points while a refereed conference proceedings article in an international conference fetches four points. Also it is notable that citations by others of the researcher's work in articles in SSCI and/or SCI databases will yield three research points each.

The research points are not used directly in determining the remuneration of academic faculty but a high annual research point accumulation is considered as a clearly positive indication of research activities. The research points are also used in ranking departments (at least in the TSE) and in the evaluation of faculties (in the new UTU). The research points are calculated in the same way for all academic faculty members from junior to senior.

2.2 Case 2: Lappeenranta University of Technology (FIN)

2.2.1 Introduction of the university

Lappeenranta University of Technology (LUT) is a medium size university located in the South-East of Finland, specialising in the nexus of technology and business.

2.2.2 The evaluation system

Academic faculty performance is evaluated with a points system that awards a maximum of 255 points for yearly performance, as an average of a two-year observation period.

Points are awarded for teaching and supervision, publication and raising research funding. Also research visits abroad and pedagogical studies are rewarded. In the Finnish (national) university remuneration system academic faculty is divided into eleven levels depending on how demanding the task is, according to a nationally agreed upon 'demand level chart'. In the LUT evaluation system the academic faculty is divided into three sub categories, determined by seniority and the level of their tasks. 'Junior level' is the levels 1–4 of the national system, 'middle level' is the levels 5–7 and senior level is the levels 8–11. The 'junior' level includes, e.g., doctoral students, the middle level includes academic faculty up to junior professors with fixed term contracts. 'Senior' level includes professors with fixed term or permanent position obtained through a (faculty appointed external) expert assessment of competence.

LUT uses the same system for the evaluation for all categories but for each of the three categories the amount of awarded points for different types of research merits is different. That is, senior faculty receives a lower amount of points for the same merits than junior faculty.

2.2.3 Teaching evaluation

Evaluation points are given for teaching as an average of two years' teaching performance and the feedback from the courses is taken into consideration. This is done in a way that the number of credit points given for the courses given during two years are multiplied by three (a weight for 'normalising' the teaching score) and then multiplied by average student feedback (scale 1–5) divided by three ('half way' feedback score). That is if there are on average 12 credits per year and average feedback is 3.5, then the evaluation points received are 42. If there are more than 200 or more than 400 enrolled students on the course the points are multiplied by 1.2 or 1.5. Points are given also for supervised completed bachelor degrees (max ten points) and for supervised completed master's degrees (max 15 points).

2.2.4 Research evaluation

We take the evaluation points awarded for publication activities as a more detailed example of the system. For the 'middle level' international level refereed (journal) publications are rewarded 20*TSC points, so that:

$$TSC = R * RIM * RCD * RR * RRM * RI$$

where

TSC	the total score of the publication
R	for a refereed article (R = 1), non-refereed (R = 0)
RIM	when the publication outlet has an IF (1.25), no IF (1)
RCD	for multi-disciplinary publication within the LUT (1.1) otherwise (1); with multidisciplinary is meant the collaboration between authors from different faculties and between different departments within the technical and natural sciences
RR	for publications with Russian universities (1.15), otherwise (1)

RRM for publications having to do with Russian markets (1.15), otherwise (1)

RI for publications with other international universities or organisations (1.1), otherwise (1)

This means that refereed article with IF done by authors from two faculties at the LUT in collaboration with a researcher from a Russian and an Italian university about Russian markets will yield a multiplicator of ~2.00 while the minimum multiplicator for a refereed publication is 1.00.

National level refereed (journal) publications and international conference publications yield four points per publication (with a maximum of five conference publications counted). Scientific monographs will fetch at maximum 40 points (depending on the level and quality), book sections in refereed books account for 12 points. The system gives a push to publish in refereed journals with an IF. The maximum number of evaluation points that can be awarded for publication activities is capped at 75.

2.2.5 Other academic tasks evaluation

Organising research financing for the university results in maximum 20 points, the maximum is awarded for 200,000 € of annual funding gathered. Pedagogical studies and internationalisation (long term staff mobility) and the whole university meeting the set goals (university level goals) also contribute to the evaluation points (together max 60 points).

2.2.6 Collection of evaluation data

The instrument for collecting the information is an excel sheet that automatically calculates the points accumulation after the inputs are fed into the sheet, the sheet includes both the research and the teaching merits. The academic faculty members are responsible for reporting their own evaluation points and the corresponding reference information etc. to back it up; if they do not report they will be evaluated based on zero points accumulation. There is a very clear incentive to report all merits that accrue evaluation points.

2.2.7 Observations

The points-accumulation is directly connected to academic faculty remuneration level for the period after the evaluation. The actual salary level is set according to discussions with the faculty member's superior, but in case there are no 'special circumstances' the points accumulation is a very strong indicator of the remuneration level. The remuneration matrix is agreed in collective negotiations between the university employers and the Finnish academic employees union AKAVA. The level of personal achievement can account for a maximum of 46% of a staff member's salary.

Figure 2 The LUT faculty evaluation result matrix and the direct connection to the remuneration matrix (salary matrix in force from 1 March 2012) (see online version for colours)

	Points required for each acchievement level								
Demand level	1	2	3	4	5	6	7	8	9
5	<50	50	60	70	80	90	105	120	135
6	<55	55	70	85	100	115	130	145	165
7	<60	60	80	95	110	130	150	170	200

Teaching- and researchpersonnel

Jobdemand level	Workperformance level								
	1	2	3	4	5	6	7	8	9
1	1 747,01 €	1 816,89 €	1 923,46 €	2 028,28 €	2 133,10 €	2 239,67 €	2 344,49 €	2 451,06 €	2 555,88 €
2	1 921,63 €	1 998,50 €	2 115,71 €	2 231,01 €	2 346,31 €	2 463,53 €	2 578,83 €	2 696,05 €	2 811,34 €
3	2 114,13 €	2 198,70 €	2 327,66 €	2 454,50 €	2 581,35 €	2 710,31 €	2 837,16 €	2 966,12 €	3 092,97 €
4	2 403,36 €	2 499,49 €	2 646,10 €	2 790,30 €	2 934,50 €	3 081,11 €	3 225,31 €	3 371,91 €	3 516,12 €
5	2 787,19 €	2 898,68 €	3 068,70 €	3 235,93 €	3 403,16 €	3 573,18 €	3 740,41 €	3 910,43 €	4 077,66 €
6	3 254,17 €	3 384,34 €	3 582,84 €	3 778,09 €	3 973,34 €	4 171,85 €	4 367,10 €	4 565,60 €	4 760,85 €
7	3 754,51 €	3 904,69 €	4 133,72 €	4 358,99 €	4 584,26 €	4 813,28 €	5 038,55 €	5 267,58 €	5 492,85 €
8	4 543,23 €	4 724,96 €	5 002,10 €	5 274,69 €	5 547,28 €	5 824,42 €	6 097,01 €	6 374,15 €	6 646,75 €
9	5 119,79 €	5 324,58 €	5 636,89 €	5 944,08 €	6 251,26 €	6 563,57 €	6 870,76 €	7 183,07 €	7 490,25 €
10	5 796,43 €	6 028,29 €	6 381,87 €	6 729,66 €	7 077,44 €	7 431,02 €	7 778,81 €	8 132,39 €	8 480,18 €
11	6 703,08 €	6 971,20 €	7 380,09 €	7 782,28 €	8 184,46 €	8 593,35 €	8 995,53 €	9 404,42 €	9 806,61 €

To the best of our knowledge, LUT is the only university in Finland that has a system that directly and systematically connects the academic faculty evaluation result to the remuneration matrix, see Figure 2. The system is well documented and information about it is available to the employees in the intranet of the university. There is also a yearly cycle to develop and enhance the system continuously.

2.3 Case 3: Palacky University in Olomouc, Faculty of Science (CZE)

2.3.1 Introduction of the university

Palacky University (PU) in Olomouc is one of the oldest universities in Central Europe and with almost 23,000 undergraduate students on eight faculties is one of the largest in the Czech Republic.

2.3.2 The evaluation system

Faculty of Science at the PU has created and uses an information system for the evaluation of academic faculty, the system is called ‘IS HAP’. The evaluation by the system includes almost every aspect of academic faculty activity; performance of each member of the academic faculty is evaluated in pedagogical, research and development (R&D), as well as other areas of activities. The system uses only easy to verify and objective data and is designed to be easy to work with for the evaluator and the academic faculty being evaluated. As such the system is designed to provide only an information support for the evaluation process – the context of the evaluation and ‘soft data’ relevant for this purpose need to be reflected during the evaluation interview of the faculty member with its superior. The evaluation system is designed to reflect the performance of a given academic faculty member as well as possible. This is achieved by not just calculating a simple average of partial evaluations in separate areas of activity, but by using intelligent (soft) aggregation. This type of aggregation (by a linguistic fuzzy rule-base) is transparent and comprehensible even to a layman as it is described verbally and provides verbal outputs.

The IS HAP system, after several years of its development, provides a sophisticated mathematical background of the evaluation mechanism, yet still well understood by the evaluators, an intuitive online interface for gathering input data and clear way of presentation of evaluation outputs. For more details concerning the development of the system see (Stoklasa et al., 2011).

2.3.3 Teaching evaluation

Three areas of activities are taken into consideration for pedagogical performance evaluation: lecturing, student supervision and work associated with the development of the fields of study. Each particular activity is assigned a score mainly based on the time used for the task.

2.3.4 Research evaluation

The evaluation of research and development activities is based on the national Czech guidelines for R&D evaluation (Government office of the Czech Republic, 2013), but also other activities, like project management, editorial board memberships and the like are included. The most important role in the evaluation of R&D outcomes in the Czech national system is played by journals with non-zero IF and issued patents. The Czech national system uses formula (1) for the evaluation of scientific papers published in journals with non-zero IF. This formula uses the rank of the journal in a decreasing sequence of all journals in the current field ordered according to the IF. It is meant to minimise the differences between the evaluations of various scientific fields regarding the IF of journals. Each paper is assigned a certain score (J_{imp}) depending on the journal it was published in. High evaluation is assigned to papers published in the journals with the highest IF in their field, whereas papers published in journals with low IF (relative to the current field) are assigned a lower evaluation (1).

$$J_{imp} = 10 + 295 \cdot factor, \tag{1}$$

where

$$factor = \frac{1 - N}{1 + \frac{N}{0.057}}, \tag{2}$$

and

$$N = \frac{P - 1}{P_{\max} - 1} \tag{3}$$

and

N is the normalised rank of the journal in the respective field according to IF

P is the rank of the journal in a decreasing sequence ordered according to the IF (according to *Journal Citation Report*)

P_{max} is the total number of journals in the current field according to *Journal Citation Report*.

Such evaluation results in a score from 10 to 305 for each paper. Two high IF journals 'nature' and 'science' open to all fields are treated separately, each paper published in these journals is awarded 500 evaluation points (the same amount of points is awarded for a European, American or Japanese patent). The score assigned by this method to a paper is then divided among the authors of the paper based on their relative contribution to the paper.

2.3.5 Other academic tasks evaluation

The system takes into account the secretarial and managerial activities performed by each member of the academic faculty (understood as activities that drain time away from and thus reduce the performance in teaching and research).

2.3.6 Collection of evaluation data

The IS HAP system is currently being used in the form of a web-based application, which is accessible through the internet. Each academic faculty member fills in an online form summarising his/her activities in the previous 12 months. All items in the form are divided into categories and subcategories. A brief help for understanding and inputting any of the items is also available. The current version of the software is able to communicate with the main information system of the PU in Olomouc and to draw information directly from this system. This reduces the time necessary to complete the form. All the filled in forms of a particular department are accessible to the head of the department and all the forms within a faculty are accessible to the dean. There is an apparent need of simplifying the input phase for the academic faculty, as evaluation is seen as an obligation – as something that keeps people from doing their work (which is an attitude quite common in some countries of Central/Eastern Europe).

2.3.7 Observations

Even though the mathematical apparatus used to calculate the final evaluation is relatively complicated, the results obtained are presented in way comprehensible even to a layman – that is by using linguistic terms and graphical presentation (see Figure 3). The output of the evaluation (obtained by a fuzzy rule-based system) provides a rough piece of information which still gives the evaluator a sufficient idea concerning the overall performance of the faculty member. The main advantage of using a fuzzy rule-based aggregation is that it allows to set-up the shape of the aggregation function used in the evaluation of academic faculty members completely in line with the evaluator's requirements; for example, giving the evaluator the possibility to appreciate excellence achieved in one specific area more than in other areas. The IS HAP system provides an easy to understand overall view of the academic faculty performance, to enable the identification of possible problems and discrepancies. A more thorough analysis of all the evaluation data (partial evaluations and even single items from the forms) is readily available through the system and allows a deeper understanding of the possible reasons for a given evaluation in detail.

Figure 3 A sample output of IS HAP – overview of evaluations of all academic faculty members (see online version for colours)

From the human resource management perspective, the most important part of the evaluation are perhaps the filled-in forms and the partial evaluations in the areas of interest, the overall evaluations are easy to understand and also useful for quick orientation in large numbers of evaluation outputs. The evaluation results are given in a verbal form on all aggregation levels. Both the pedagogical and R&D areas are evaluated by a standard scoring system. There are different standard scores set up for academic faculty members of different seniority. The standard scores are set up to reflect the characteristics of the faculty on which the evaluation is performed. The evaluation representing a partial evaluation of a member of the academic faculty in a certain evaluated area of activities (pedagogical activities, R&D) is determined as a multiple of the respective standard, or expectation, for the faculty member's position. For better clarity and easier interpretation, these numbers are not provided and instead are transformed into verbal evaluations using linguistic scales; evaluations are also provided in a graphical form, see Talasova and Stoklasa (2010). The IS HAP system can also be used as a HR management tool on all levels of management.

The evaluation by the IS HAP system is not directly connected with academic faculty remuneration – it is intended primarily for human resource management purposes. But the outputs it provides can be easily compared with the remuneration of academic faculty and discrepancies can be identified and eliminated.

2.4 Case 4: A&M University, Kingsville (USA)

2.4.1 Introduction of the university

The A&M University in Kingsville, Southern Texas is a university unit with over 6,000 students from more than 43 countries and over 300 academic faculty members. It is a part of the Texas A&M University System (over 50,000 student in total; over 1,200 academic faculty members) and uses the same academic faculty system described by the *Faculty Handbook* (see Texas A&M University, Kingsville, 2011a).

2.4.2 The evaluation system

All academic faculty members are being evaluated by the same evaluation model (regardless of if they are tenured or not), the evaluation is annual. The person responsible for the evaluation is the head of the department; the evaluation is confirmed by the dean. Four areas of activities are being considered. Before the beginning of the evaluation period, the academic faculty member discusses his/her plans for the period and set weights for each of these four categories (the sum of weights has to be equal to one). This is done with respect to the faculty member's last year evaluation. The relevant categories with the respective weight ranges are: *teaching performance* weights range: (0.25–0.65); *research and scholarly activities* weights range: (0.15–0.55); *professional growth and activities* weights range: (0.05–0.45) and *non-teaching activities supportive of university programmes* weights range: (0.15–0.55).

Each area is in the end expertly evaluated by the academic faculty member's superior on the following seven point scale where each level is characterised linguistically (see Texas A&M University, Kingsville, 2011b):

- EXEMPLARY (seven points) – (awarded rarely) – connected with the highest degree of productivity and effectiveness.
- EXCEPTIONAL (six points) – performance high above the average level (that is the level of expectations for GOOD).
- OUTSTANDING (five points) – better performance than GOOD (faculty member surpasses the expectations for the evaluation GOOD).
- GOOD (four points) – this is the average expected performance of academic faculty. The performance is considered productive and effective. This level of evaluation should be attained by all faculty members in all evaluated areas. This evaluation is still seen in favourable light.
- ACCEPTABLE (three points) – performance, that is considered to meet the requirements for academic faculty on the A&M University.
- DEFICIENT (two points) – performance barely satisfies expectations, there is apparent room for improvement – this evaluation should be accompanied with a plan for improvement.
- UNACCEPTABLE (one point) – the performance in the given area is not productive or effective; it does not meet the requirements. Again plan for improvements will be an integral part of the evaluation report.

It can be easily seen that the nature of evaluation based on such scale is subjective. The linguistic level however provides easy to understand interpretation of the evaluation in each area.

2.4.3 Teaching evaluation

The evaluation in this area consists of two parts, each of which has an assigned weight (agreed upon by the faculty member and his/her superior): *student ratings of instruction* with weight range of (0.25–0.5) and *other evidence of teaching performance* (weight range (0.5–0.75) where the faculty are expected to prove appropriate behaviour and

performance in teaching activities). The faculty members are expected to provide evidence of preparation for teaching (instructional materials, syllabi, outlines etc.). It is also possible to provide other evidence of effective teaching, such as teaching portfolio, reflective self-review, workshops or other training conducted or provided for others, peer reviews, colleague reviews, trained observers, feedback from current students and many other materials.

The student ratings of instruction have to meet strict criteria to be reflected in the evaluation. Adjustments to compensate for known biases are also made. The role of student ratings of instruction it to identify problematic aspects of instruction (with an average score of two or lower out of five) and only the presence of such problems is reflected in the loss of evaluation points.

2.4.4 Research evaluation

The evaluation of research activities reflects the 'scholarship of discovery', the 'scholarship of education', 'scholarship of teaching' and the 'scholarship of integration'. The academic faculty member has to fill in an unstructured form, provide an overview of his/her outputs in the R&D area, the summary of creative and artistic endeavours, contract research, participation in curricular innovation and similar relevant activities; evidence of the existence of all the stated outputs must be provided. Based on these materials the faculty member's performance in this area is assessed by his/her superior and evaluated using the seven point scale.

2.4.5 Other academic tasks evaluation

Evaluation in the areas of *professional growth and activities* and *non-teaching activities supportive of university programmes* is again based on the material provided by the academic faculty member and evaluated by the superior using the seven point scale.

2.4.6 Collection of evaluation data

Apart from the student ratings of instruction (which have a separate methodology, guidelines, forms and assessment methods), all the data that the faculty member deems relevant for his/her evaluation in the respective area are gathered through unstructured forms filled in by the faculty, accompanied by the necessary documents proving the existence or quality of the reported outputs.

2.4.7 Observations

Although all the activities of the academic faculty have to be well-documented, the evaluation process seems to offer a large area for subjectivity. Each area is evaluated on a seven point scale with linguistically defined levels, where the evaluation is determined expertly by the superior based on the materials provided by the faculty member. The final evaluation is computed as a weighted average of the expertly set evaluations for each area. The resulting number is however not used (which is a good thing, as its interpretation is not easy) instead a written report justifying and explaining the final evaluation is compiled by the superior. If there is room for improvement, this has to be stressed in the report and possible ways of solving the identified problems need to be suggested. This approach to evaluation means a great deal of work both for the faculty

that is being evaluated (providing materials for the evaluation) and for the evaluators (going through the provided materials).

3 Summary and discussion

The four presented cases of academic faculty evaluation systems all have their own structure and method in evaluation. The formal part of the UTU system is a research activities focused scorecard that lists and rewards for a wide number of research related academic tasks, also other than publications, but teaching achievement is not evaluated in a structured way. The LUT system is also a scorecard-based system that includes teaching and publication activities; for the part of research the system mostly concentrates on publication merits and fund raising. Table 1 compares the systems by listing selected issues and presenting how the issues are considered in each system.

Table 1 Selected characteristics of the evaluation systems

	UTU	*LUT*	*PU*	*A&M*
Sources of evaluation data	Self-reporting (checked for correctness)	Self-reporting, student feedback, funding database	Online self-reporting, R&D database, teaching IS, evaluation context added by the superior	Self-reporting w. documentation, student feedback
Research evaluation	Yes	Yes	Yes	Yes
Pedagogical evaluation	No	Yes	Yes	Yes
Averaging over more than one year	No	Yes	No	No
Publication rewarded	Yes	Yes	Yes	Yes/indirectly
Fund raising rewarded	Yes	Yes	Yes	Yes/indirectly
Other academic tasks rewarded	Yes	No	Yes	Yes
Reflect/accept specialisation of academic faculty	No	No	Can be reflected by the aggregation rule-base, very flexible	Reflected by the weights of evaluated areas
Aggregation method used to produce final evaluation	Simple addition	Rules and addition w. averaging over years	Fuzzy linguistic approach (rule-base)	Weighted average-based; seven point scale for each sub-area. Weights set in the preceding.

Table 1 Selected characteristics of the evaluation systems (continued)

	UTU	*LUT*	*PU*	*A&M*
Fuzzy/linguistic approach	No	No	Yes	Partial linguistic
Seniority consideration	No	Yes	Yes	No
Software supported	No	No	Yes	No
Direct connection to salary	No	Yes	No	No
Role of the academic faculty member's direct superior	None – collected evaluation goes to a panel	Yearly review-based on the evaluation, interview, makes final evaluation	Interprets outputs of the evaluation system, makes an interviews, makes the final evaluation and implications	Actively assesses the evaluation and actually makes the evaluation based on data. Sets weights for following period.
Form of final evaluation (report, single number)	No actual evaluation – final score available	Score, discussion	Linguistic description of performance, colour bars, evaluation interview	Number score, narrative report by the superior

The Palacky University (PU) system, the IS HAP, is a web-based software system that considers both research and pedagogical performance and provides output also in a linguistic form that is enhanced with colour coding. The A&M University system (A&M) depends on using a linguistic seven point scale to assess a number of aspects of academic achievement and relies quite heavily on supervisorial assessment. Of the four systems the PU system is the most advanced by its usability design (a working software), the UTU system takes only research related performance systematically into consideration, the LUT system includes a direct connection between the performance and the salary level and the A&M system uses a set of previously determined weights to account for personal focus of work.

The only one of the systems that uses averages over more than a year is the LUT system; this is interesting as it is often not in the hands of the academic faculty member when, for example an article is published and thus they have limited capability in affecting their research evaluation for one single year. Publication is rewarded in all systems (although in the A&M system indirectly), this is no surprise, what is less obvious is that all four systems also reward for raising research funds.

Other academic tasks such as participation in editorial boards or other academic positions of trust are rewarded with performance points in the UTU and PU systems, such tasks become indirectly under scrutiny in the A&M system. Seniority of the academic faculty, when assessing performance is taken into consideration in the LUT and the PU systems. PU system is software supported and gives linguistic outputs in the academic faculty evaluation (approximate reasoning, fuzzy outputs) and also automatically calculates an overall evaluation for the academic faculty members, according to an

intelligent aggregation method. The UTU and the PU systems are not primarily intended for the purpose, but can be used as indicators in the determination of academic faculty salary, while the output from the LUT system is more than just indicative; it offers a clear relationship with the performance and an exact salary level from the Finnish national academic faculty remuneration system. The A&M system requires that a narrative of the reasons behind the evaluation decision is created to support the evaluation. The characteristics presented in Table 1 show that there are similarities between the systems, but equally there are many differences.

Each of these systems represents a real world case of academic faculty performance evaluation and as such offers insights into how university management views academic faculty evaluation. It has been the goal of this paper to present the four cases and to shortly discuss them; future research in this vein will include collecting data from more academic evaluation systems and the analysis of these, to be able to draw conclusions about the different types of systems in place and in the hopes of learning about what in them can be characterised as being 'best practice'.

References

Bana e Costa, A. and Oliveira, M. (2012) 'A multicriteria decision analysis model for faculty evaluation', *Omega*, Vol. 40, No. 4, pp.424–436.

Board of the Turku School of Economics (2004) *Decision Regarding the Research Points Evaluation System*, Turku School of Economics, Turku.

Elmore, H. (2008) 'Toward objectivity in faculty evaluation', *Academe*, No. 3, pp.38–40.

Flinders University (2012) *Performance Management* [online] http://www.flinders.edu.au/ppmanual/staff/performance-management/performance-management_home.cfm (accessed 11 December 2013).

Government Office of the Czech Republic (2013) *Methodology for Evaluation of Research Organisations and Outcomes of Completed Programmes (Valid for 2013–2015)* (in Czech) [online] http://www.vyzkum.cz/storage/att/373C18E8F5E1311F5B8AF2BD17FAB115/M2013_v95.pdf (accessed 11 December 2013).

Hicks, D. (2012) 'Performance-based university research funding systems', *Research Policy*, Vol. 41, No. 2, pp.251–261.

Jan Evangelista Purkyne University (2012) *Academic Staff Evaluation Criteria for Personal Extra Pay Distribution at the Faculty of Environment* (in Czech) [online] http://fzp.ujep.cz/dokumenty/kritosoh.pdf (accessed 11 December 2013).

Langen, J.M. (2011) 'Evaluation of adjunct faculty in higher education institutions', *Assessment & Evaluation in Higher Education*, Vol. 36, No. 2, pp.185–196.

Lappeenranta University of Technology (2011) *Sisäinen ohje (Internal Instructions Regarding the Faculty Evaluation)*, dated 24 March 2011, Lappeenranta, LUT.

Masaryk University (2012) *Determination of Criteria for Pedagogical and Other Activities Evaluation*, Masaryk University, Faculty of Law (in Czech) [online] http://is.muni.cz/do/law/ud/predp/archiv_predpisy/4862802/Pokyn_dek._c._7-2009_urceni_kriterii_pro_hodnoceni_ped._a_j.cin..pdf (accessed 14 February 2012).

McGill University (2012) *Academic Performance Evaluation* [online] http://www.mcgill.ca/medicine-academic/performance/ (accessed 11 December 2013).

Minelli, E., Rebora, G. et al. (2006) 'The impact of research and teaching evaluation in universities: comparing an Italian and a Dutch case', *Quality in Higher Education*, Vol. 12, No. 2, pp.109–124.

Stoklasa, J., Talasova, J. et al. (2011) 'Academic staff performance evaluation – variants of models', *Acta Polytechnica Hungarica*, Vol. 8, No. 3, pp.91–111.

Talasova, J. and Stoklasa, J. (2010) 'Assessing academic staff performance using multiple criteria evaluation models', *2nd International Conference on Applied Operational Research*, Turku, Finland, Uniprint.

Texas A&M University, Kingsville (2011a) *Faculty Handbook of A&M University* [online] http://www.tamuk.edu/ senate/_pdf_files/handbookaugust2011.pdf (accessed 23 November 2012).

Texas A&M University, Kingsville (2011b) *Texas A&M University-Kingsville: Summary of Annual Evaluation of Faculty* [online] http://www.tamuk.edu/agnrhs/forms/Form_Summary_of_Annual_Evaluation_of_faculty__2_.doc (accessed 11 December 2013).

Tomas Bata University in Zlín (2012) *Pedagogical and Creative Activities Evaluation* (in Czech) [online] http://www.utb.cz/file/38664_1_1 (accessed 11 December 2013).

University of Technology Sydney (2009) *Performance Management* [online] http://www.hru.uts.edu.au/performance /reviewing/rating.html (accessed 14 February 2012).

Uzoka, F-M. (2008) 'A fuzzy-enhanced multicriteria decision analysis model for evaluating university academics' research output', *Information Knowledge Systems Management*, Vol. 7, No. 3, pp.273–299.

Wayne State University (2009) *Guidelines for Evaluation of Academic Staff* [online] http://www.aaupaft.org/pdf/AcStaffguidelines_2009-10.pdf (accessed 14 February 2012).

PUBL. IX

Talašová, J. and Stoklasa, J., Fuzzy approach to academic staff performance evaluation. *Proceedings of the 28^{th} International Conference on Mathematical Methods in Economics 2010*,, 621–626, 2010.

Fuzzy approach to academic staff performance evaluation

Jana Talašová[1], Jan Stoklasa[2]

Abstract. This paper describes a fuzzy model of academic staff evaluation. The performance of each member of academic staff is evaluated in both pedagogical, and research and development areas of activities. Input data are acquired from a form filled in by the staff where particular activities are assigned a score according to their importance and time requirements. Both areas of evaluation are assigned standard total scores – different for senior assistant professors, associate professors, and professors. A partial evaluation of a member of academic staff in a specific area is determined as a multiple of the respective standard for his or her position. A linguistic fuzzy expert system is then used to aggregate both partial evaluations – for pedagogical, and research and development areas of activities.
The proposed model also takes into account the load of managerial activities for each member of academic staff (understood here as activities draining his or her time and thus reducing the performance in both areas of evaluation mentioned above). Another fuzzy expert system is used to adjust the evaluation according to the managerial activity load of a particular academic staff member. The overall work load of academic staff members is thus described in words.

Keywords: evaluation, academic staff, aggregation, fuzzy model.

JEL Classification: C44, I20
AMS Classification: 90B50

1 Introduction

The intention to create a mathematical model for the purpose of comprehensive academic staff performance assessment in two main areas of interest – Pedagogical Activities (PA), and Research and Development (R&D) – was formulated at Palacky University, Faculty of Science as early as 2006. In connection with the model preparation we studied general problems of the quality assessment in high education institutions (see [1] for the Czech Republic and [2] for EU), fundamentals of human resources management (see [3]), and at the same time we were looking for the optimal mathematical tools (see [4, 5]). Various academic staff evaluation models currently used in USA (see e.g. [6]), Canada ([7]), and Australia ([8,9]) were subjected to a detailed analysis. Later, even the models recently designed at various Czech universities (see [10, 11, 12]) were analysed. The analysis concentrated on both practical and mathematical aspects of these evaluation models and resulted in designing several academic staff evaluation models (see e.g. [13, 14]). The proposed models differed both in the manner how members of academic staff are evaluated in separate areas of their activity and in the aggregation method for these partial evaluations (weighted average, OWA and WOWA operators were used; for the theory of the aggregation operators see [4]). Considering the faculty management requirements on properties of the evaluation function, a model using linguistic fuzzy modelling approach has been eventually selected.

General requirements on the model were as follows: It should (1) include, if possible, every aspect of academic staff activity; (2) use only easily proven and objective data; and (3) be easy to work with. Other requirements were for the final evaluation: (4) to maximally reflect staff benefit to the Faculty; and (5) not to be a simple average of partial evaluations in separate areas of activity, but appreciate excellent performance in both evaluated areas (PA, R&D).

The main objective of the model is to globally assess the performance and overall work load of each academic staff member in regular time intervals (annually). To achieve this, detailed information in unified form concerning particular activities and outcomes of a particular academic staff member will be gathered. Aggregated overall evaluation information will also be available (at different levels of aggregation). As far as the aggregated evaluation is concerned, the desired output of the model was neither to arrange members of academic staff in order of their performance, nor to obtain a single number interpretable only with difficulty. A basic piece of

[1] Palacký University, Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, Address: tř.17. listopadu 1192/12, 771 46 Olomouc, Czech Republic, e-mail: talasova@inf.upol.cz.
[2] Palacký University, Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, Address: tř.17. listopadu 1192/12, 771 46 Olomouc, Czech Republic, e-mail: jan.stoklasa01@upol.cz.

information concerning the performance of the academic staff was considered sufficient. Such assignment implied the use of linguistic fuzzy modelling – linguistic variables, rule bases, and approximate reasoning (i.e. the use of fuzzy expert systems – for more details see [5]).

2 Evaluation model

2.1 Basic structure

The performance of each member of academic staff is evaluated in both pedagogical, and research and development areas of activities. Input data are acquired from a form filled in by the staff where particular activities are assigned scores according to their importance and time requirements. Three areas are taken into consideration for pedagogical performance evaluation: (a) lecturing, (b) supervision of students, and (c) work associated with the development of fields of study. The research and development activity evaluation is based on the methodology valid for the evaluation of R&D results in the Czech Republic (see [15]) but other important activities (grant project management, editorial board memberships etc.) are also included. Both pedagogical and R&D areas are assigned a standard score – different for senior assistant professors, associate professors, and professors. The number representing partial evaluation of a staff member in a certain area is determined as a multiple of the respective standard for his or her position. To achieve a better clarity and easier interpretation these numbers are transformed into verbal evaluation using linguistic-scale values.

A linguistic fuzzy expert system is therefore used to aggregate both partial evaluations – for pedagogical and R&D areas of activities. The main advantage of this type of aggregation is that it allows setting up the shape of the aggregation function completely in accordance with evaluator's requirements (e.g. to appreciate excellence achieved in one of the areas). This type of aggregation is transparent and comprehensible even to a layman as it is described in linguistic terms. The overall aggregated evaluation is also available as a linguistic expression.

Our model also takes into account the load of academic office and management activities for each member of academic staff (understood here as activities draining his or her time capacity and hindering maximum performance in each area). The overall aggregated evaluation for pedagogical and R&D activities is being aggregated with such a rate, with which this "activity load" criterion is met, by means of another fuzzy expert system. This results in a language description of staff members' overall workload. Academic office and management activity load (AM) can also be taken into account afterwards.

2.2 Input data

We have designed comprehensible forms for pedagogical activity evaluation, research and development related activity evaluation, and overall academic office and management activity load assessment. These forms are to be filled in by each academic staff member using either numerical values (number of lectures and seminars, number of certain types of publications, held academic office position) or, where necessary, by a particular activity, outcome, or office specification (e.g. precise citation of the paper, conference paper, or specification of the office that is not specified in the forms).

Data gathering itself is a significant asset of the proposed evaluation system. It is important to have both detailed information and aggregated evaluations (at different levels of aggregation) available in the evaluation process.

2.3 Evaluated activities, used scales

All the activities in previously mentioned areas of interest (PA, R&D, AM) are assigned scores with respect to their respective significance. It is necessary to emphasize that the scoring scales used to describe the performance in each of the three areas of interest differ. For pedagogical activities, the scores reflect mostly time demands, partly also professional demands. Obviously, achieving twice the standard score in this area of interest (defined by the assessor) should be considered extreme performance. For research and development related activities we constructed the scale based on the current R&D outcome evaluation methodology valid for the Czech Republic (which is known to favour excellence – paper scores rise up quickly along with the prestige of the journal the paper is published in; see [15]). The scores of other activities in this area of interest were determined by comparing these activities with the previously mentioned outcomes. It is quite possible for excellent researchers (according to the described quality of the scale) to exceed the standard score severalfold. Scores in the area of academic office and management activities roughly reflect the percentage of working time the staff member has to dedicate to his or her office. The differences among these three scales constitute a significant setback for their aggregation.

2.4 Evaluation in particular areas of interest - PA, R&D

There are standard scores defined for different staff categories (assistant professor, associate professor, professor) in the area of PA and R&D. For pedagogical activities, the same standard score is defined for all staff categories (associate professors and professors may compensate the lower number of hours of frontal teaching, for example, by supervising Ph.D. students), while for R&D related activities the standard score is set somewhat higher for associate professors and even higher for professors.

The total scores of a particular staff member in the area of pedagogical activities on one hand and research and development activities on the other hand are computed by adding up the scores of performed activities and achieved outcomes. These summarized scores are linearly transformed into standardized partial evaluations (by means of predefined standard scores) in the respective areas using the following formula:

$$h_{i,j} = \frac{b_{i,j}}{b^{st}_{i,j}}, \quad i = 1, 2, \; j = 1, 2, 3,$$

where $b_{i,j}$ is the current staff member score in the j-th category (1 – assistant professor, 2 – associate professor, 3 – professor) for the i-th area of interest (1 – PA, 2 – R&D), $b^{st}_{i,j}$ is the predefined standard score for the i-th area of interest and in the j-th category, $h_{i,j}$ is the standardized partial evaluation of a staff member in the j-th category for the i-th area of interest. The calculated standardized partial evaluations describe such a multiple of the standard score that corresponds with the staff member performance.

In order to interpret the standardized partial evaluation for PA and R&D linguistically, we define linguistic scales on the corresponding continuous evaluation scales $[0,a_i]$, i=1,2. A **linguistic scale** is a special case of **linguistic variable** (see [5]), where the meanings of linguistic terms are fuzzy numbers that form a fuzzy partition of the original interval. Fig. 1 and 2 show how the mathematical meanings of the linguistic scales *PA – staff member performance* and *R&D – staff member performance* (with elements *very low*, *low*, *standard*, *high*, and *extreme*) can be modelled. Using the values of the linguistic scale *R&D – staff member performance* we can interpret, for example, the standardized partial evaluation 1.75 in the area of R&D as a performance that is 25% standard and 75% high. The following figures show how to reduce the difference in the character of the original evaluation scales using linguistic fuzzy scales with differently defined meanings of their linguistic terms.

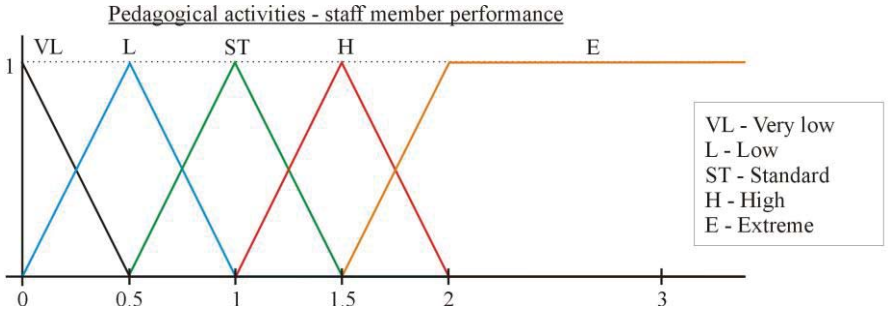


Figure 1 Pedagogical activities – staff member performance

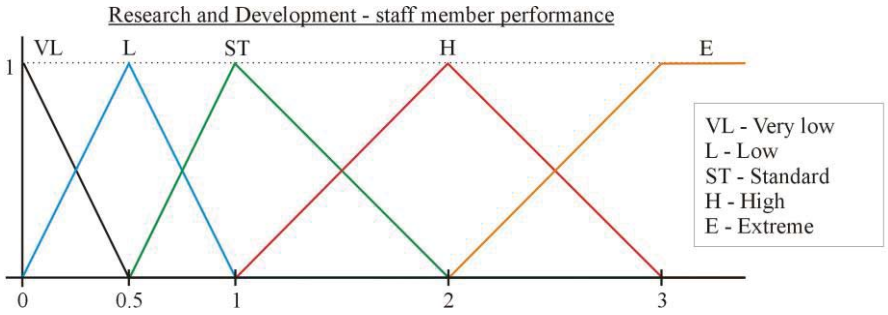


Figure 2 Research and Development – staff member performance

2.5 Evaluation of the overall PA and R&D performance

Through defining the linguistic scales we get optimal conditions for the aggregation function of the evaluation criteria *pedagogical activities* and *research and development* to be described linguistically – using the **fuzzy rule base**. The rule base is designed to reflect the fact that high or extremely high performance in any of the areas of

interest (PA, R&D) ensures good overall evaluation; on the other hand, performance that is worse than standard in one area and low in the other area of interest results in a significantly negative evaluation of the current worker performance. The fuzzy rules are e.g. as follows:

If *PA performance* is *standard* and *R&D performance* is *standard*, then *overall PA + R&D performance* is *standard* .

If *PA performance* is *standard* and *R&D performance* is *high*, then *overall PA + R&D performance* is *very good*.

If *PA performance* is *high* and *R&D performance* is *standard*, then *overall PA + R&D performance* is *very good*.

If *PA performance* is *high* and *R&D performance* is *high*, then *overall PA + R&D performance* is *excellent*.

Based on the proper **fuzzy inference algorithm**, knowing the numerical values of both input variables (PA, R&D – standardized scores), we can derive the numerical value of the output variable *Overall PA and R&D performance of a current staff member*. This overall evaluation value can be linguistically interpreted using a fuzzy scale with linguistic terms *unsatisfactory*, *substandard*, *standard*, *very good*, and *excellent* (see Fig. 3). Meanings of these linguistic terms (represented by linear fuzzy numbers) were defined with respect to expertly described overall evaluations of couples of values from input fuzzy numbers kernels. For example, if a staff member's performance (its standardized value) in PA is 1.25 (which means that it is 50% standard and 50% high) and his or her performance (its standardized value) in R&D is 1.75 (meaning 25% standard and 75% high), then the overall evaluation using the rule base is 2.25 (that is 75% very good and 25% excellent).

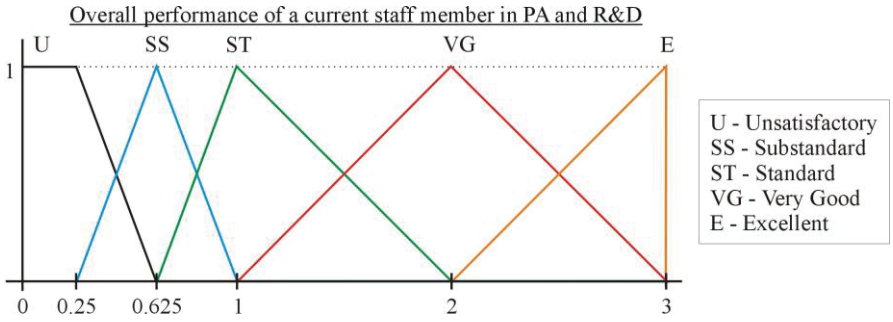


Figure 3 Overall PA and R&D performance of a particular staff member

2.6 Academic office and management activities – determination of overall staff member work load

The input data-forms contain, apart from sections concerning teaching, R&D, and related activities, also a part concerning academic offices (dean, vice dean, academic senate chairman, etc.), membership in institutions and committees where a particular staff member stands as university representative (Council of Higher Education Institutions, R&D Council, etc.), and managerial activities that can be assigned neither to pedagogical nor to research and development activities (such as the head of department). These activities are not truly "evaluated" by the model. We just see them as activities that prevent the staff member from achieving his or her full potential performance in PA or R&D. Each academic office is assigned a score that reflects the percentage of standard working time that this specific office consumes. A linguistic scale (see Fig. 4) can be used to describe the academic office and management activity load of a particular staff member linguistically. Its linguistic terms are *zero*, *light*, *medium*, *heavy*. A fuzzy rule base was developed for determining the academic staff member overall performance. *Overall PA and R&D performance of a particular staff member* and *Academic office and management activity load* are the inputs to this rule base and *Academic staff member overall performance* is the output variable. The fuzzy rule base can contain rules like these:

If *PA + R&D performance* is *standard* and *AM activity load* is *zero*, then *overall performance* is *standard* .

If *PA + R&D performance* is *standard* and *AM activity load* is *light*, then *overall performance* is *high*.

If *PA + R&D performance* is *standard* and *AM activity load* is *medium*, then *overall performance* is *extreme*.

The numerical value of *Academic staff member overall performance* can be interpreted using a linguistic scale with terms *very low*, *low*, *standard*, *high*, and *extreme* (see Fig. 5).

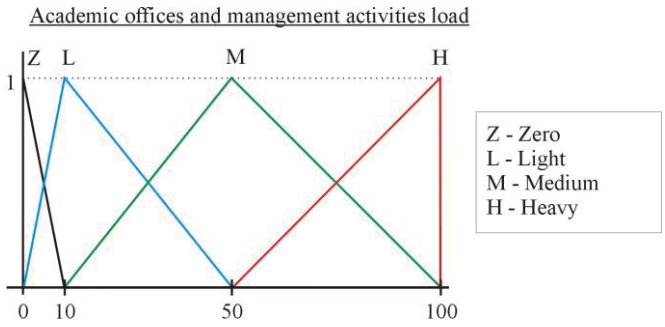


Figure 4 Academic office and management activity load

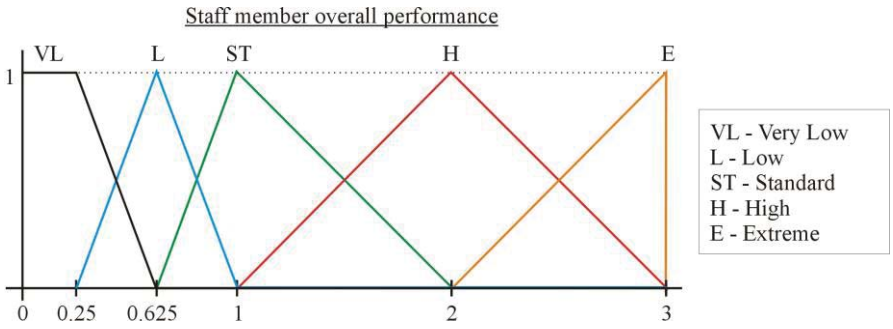


Figure 5 Academic staff member overall performance

3 Conclusion

In the paper we describe the academic staff performance evaluation model that differs from other models namely in that point that it uses linguistic fuzzy modelling. The proposed solution pursues two main goals: (1) to describe every aspect of academic staff member performance that is beneficial to the university using only objective and easily verifiable information; (2) to provide aggregated information concerning the staff member performance and work load in a vivid and easy to understand form. Aggregated information may be inaccurate but its informational value is sufficient to pinpoint an existing problem. A deeper performance analysis, utilizing all the information available in the system, is also possible.

The purpose of the introduced performance evaluation system is mainly management of human resources at universities. The fact that academic staff members are somewhat specific people (with great amount of freedom in choosing content of their activities, as compared with other workers) is also taken into account. In contrast to other performance evaluation models used in the Czech Republic, our model is not primarily intended for funds distribution purposes. However, it would be easy to compare the proposed aggregated evaluation of an academic staff member with his or her remuneration (with respect to work position) in order to detect potential discrepancies.

The developed performance evaluation system is beneficial to academic staff members as well – it serves as a record of their activities for their own needs, it provides feedback about their performance (and how the employer sees this performance) and, last but not least, ensures objectivity of evaluation on employer's part. Aggregated information available in an easy to understand form is an important management tool for the executives, namely the heads of departments. Long-term use of the model offers the opportunity to observe dynamics of staff member performance over time, which can be seen as another valuable asset of our model.

Acknowledgements

The presented research was supported by the student grant *PrF_2010_00 - Mathematical and computer science models and structures* obtained from Internal Grant Agency of the Palacky University in Olomouc.

References

[1] Chvátalová, A., Kohoutek, J., Šebková, H. (eds.): *Zajišťování kvality v českém vysokém školství*. Aleš Čeněk, Plzeň 2008, ISBN 978-80-7380-154-0.

[2] *European Association for Quality Assurance in Higher Education* [online] <http://www.enqa.eu/>

[3] Hroník, F. *Hodnocení pracovníků*. Grada Publishing, Praha, 2006.

[4] Torra, V., and Narukawa, Y.: *Modeling Decisions*. Springer, Heidelberg, 2007. ISBN 978-3-540-68789-4

[5] Talašová, J.: *Fuzzy metody vícekriteriálního hodnocení a rozhodování*. Univerzita Palackého, Olomouc, 2003, ISBN 80-244-0614-4.

[6] *2009-10 Guidelines for Evaluation of Academic Staff* [online]. Wayne State University [cited 30. 5. 2010]. <http://www.aaupaft.org/pdf/AcStaffguidelines_2009-10.pdf>

[7] *Academic Performance Evaluation* [online]. c2010, last revision May 15, 2010, McGill University [cited 30. 5. 2010]. <http://www.mcgill.ca/medicine-academic/performance/>

[8] *Performance Management* [online]. c2009, University of Technology Sydney [cited 30. 5. 2010]. <http://www.hru.uts.edu.au/performance /reviewing/rating.html>

[9] *Performance Management* [online]. Flinders University [cited 30. 5. 2010]. <http://www.flinders.edu.au/ppmanual/review.html>

[10] *Určení kritérií pro hodnocení pedagogické a jiné činnosti (vytíženosti)* [online]. Brno, Masarykova univerzita, Právnická fakulta [cited 30. 5. 2010]. <http://www.law.muni.cz/dokumenty/7601>

[11] *Kritéria hodnocení akademických pracovníků FŽP pro udělování osobních příplatků* [online]. Ústí nad Labem, Univerzita Jana Evangelisty Purkyně, Fakulta životního prostředí [cited 30. 5. 2010]. <http://fzp.ujep.cz/dokumenty/kritosoh.pdf>

[12] *Hodnocení pedagogických a tvůrčích aktivit* [online]. Zlín, Univerzita Tomáše Bati ve Zlíně, Fakulta aplikované informatiky [cited 30. 5. 2010]. <http://web.fai.utb.cz/cs/docs/SD_09_09.pdf>

[13] Talašová, J., Pavlačka, O.: *Návrh modelu hodnocení akademických pracovníků na Přírodovědecké fakultě Univerzity Palackého v Olomouci*. Research report. Faculty of Science, Palacký University, Olomouc 2006.

[14] Talašová, J., Stoklasa, J., Pavlačka, O., Holeček, P.: *Nový návrh modelu hodnocení akademických pracovníků na PřF UP.* Research report. Faculty of Science, Palacký University, Olomouc 2009.

[15] *Metodika hodnocení výsledků výzkumu a vývoje* [online]. Výzkum a vývoj v České republice [cited 30.5. 2010],<http://www.vyzkum.cz/storage/att/CDDC542199F1640B59A7D1E841B7151C/Metodika%202009_schv%c3%a1leno.pdf>

PUBL. X

Stoklasa, J. and Talašová, J., Using linguistic fuzzy modeling for MMPI-2 data interpretation. *Proceedings of the* 29^{th} *International Conference on Mathematical Methods in Economics 2011 - part II*, 653–658, 2011.

Using linguistic fuzzy modeling for MMPI-2 data interpretation

Jan Stoklasa[1], Jana Talašová[2]

Abstract. Psychological diagnostics is a crucial psychological activity. It involves systematic acquisition of a large amount of information, data classification, interpretation and final derivation of conclusions. It is desirable to develop systems able to speed up the process and reduce the risk of errors.
This paper considers possibilities of linguistic fuzzy modeling for psychological data analysis and evaluation; perspectives of knowledge transfer are discussed. We describe the process of conversion-symptoms identification based on data provided by MMPI-2 (Minnesota Multiphasic Personality Inventory). Linguistic fuzzy rules are introduced to represent the expert knowledge of the process in three stages – protocol validity, data appropriateness, and "conversion V" obviousness.
Finally, a fuzzy-rule-base aggregation of the three evaluations of a MMPI-2 profile is introduced. Sugeno's fuzzy inference algorithm is used. A fuzzy classification of conversion-symptom presence into three categories (present, possibly present and not present) is performed in this step. The model is implemented in Excel.

Keywords: Linguistic fuzzy modeling, MMPI-2, psychological diagnostics, fuzzy classification.

JEL Classification: C44
AMS Classification: 91E10

1 Introduction

Psychological diagnostics is usually the first step of any psychological intervention. Thorough analysis of all the data obtained by various diagnostic methods (test methods and clinical methods) is needed to gain a valid understanding of client's current state and situation. Unfortunately the amount of data can easily exceed the analytical capacities of a diagnostician. If we take into account that the client's are available in many different forms – linguistic descriptions, pictures, numbers or intervals (results of some diagnostic methods), scales, and even subjective impressions – the aggregation and interpretation of such data becomes a nontrivial task. A diagnostician also needs to be aware of the context and usually employs his expert knowledge and experience in this process.

Once we see the process from this perspective, various problems arise. As the amount of data to be processed and interpreted grows, so does the room for mistakes and misinterpretations. The time consumption of this process is also a point to be considered. Any tool of error elimination that would reduce the time of data processesing and interpretation would be most welcome. In this paper we introduce a linguistic fuzzy model for a particular psychodiagnostic method and present one particular diagnosis that meets these requirements.

In this paper we are presenting a tool for psychologists for conversion-symptoms identification based on the MMPI-2 results. The international classification of diseases – 10th revision (see [11]) denotes dissociative (conversion) disorders as the F44 category. In our application, we narrow the scope and consider only the subcategories F44.4 – F44.7. This group of disorders is called dissociative motor and sensory disorders and can be roughly characterized by neurological symptoms such as numbness or paralysis with no underlying neurological causes.

Minnesota Multiphasic Personality Inventory second revision (MMPI-2) and its previous version are the most widely used psychological inventories for psychopathology assessment worldwide (see [2]) and in the Czech Republic as well (see [4,8]). The MMPI was developed by Hathaway and McKinley [3] in 1940, later Netík adapted the second revision into Czech in 2002 (see [4]). This method was chosen for our research because it provides various means of validity assessment, is widely used, and much research has been done since 1940

[1] Palacky University in Olomouc/Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, , 77146 Olomouc, Czech Republic, jan.stoklasa@upol.cz;
Palacky University in Olomouc/Faculty of Arts, Dept. of Psychology, Křížkovského 10, 771 80, Olomouc, Czech Republic, dzoni@seznam.cz.
[2] Palacký University/Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, jana.talasova@upol.cz.

concerning its validity and reliability. When a client answers the 567 items of MMPI-2, the method then provides 567 answers to these items, 10 clinical scales scores, more than 7 validity scale scores and over 70 content and supplementary scales scores. In this paper we propose a linguistic fuzzy model to help the diagnostician recognize the presence of conversion symptoms. The process uses data provided by MMPI-2 and diagnostic criteria suggested by Greene in [2]. It also draws on diagnostician's expert knowledge and experience (these are implemented in the fuzzy rules). Requirements imposed on the model by psychologists were: a) to suggest interpretation of the data in the framework of conversion-symptoms identification, b) comprehensible environment and results; and c) results justification (to provide more than just the highest level of data aggregation). We also wanted to consider possible research and knowledge transfer possibilities.

2 Used mathematical apparatus

In this paper we use the concept of fuzzy sets and linguistic fuzzy modeling introduced by Zadeh in [12, 13]. Let U be a nonempty set (the universe). A fuzzy set A on U is defined by the mapping $A:U\to[0,1]$. For each $x\in U$ the value $A(x)$ is called a membership degree of the element x in the fuzzy set A and $A(\cdot)$ is called a membership function of the fuzzy set A. R will denote the set of all real numbers.

The height of a fuzzy set A is a real number $\mathrm{hgt}(A)=\sup_{x\in U}\{A(x)\}$. A union of fuzzy sets A and B on U is a fuzzy set $A\cup B$ on U with a membership function for all $x\in U$ given by $(A\cup B)(x)=\max\{A(x),B(x)\}$. An intersection of fuzzy sets A and B on U is a fuzzy set $A\cap B$ on U with a membership function for all $x\in U$ given by $(A\cap B)(x)=\min\{A(x),B(x)\}$. Let A be a fuzzy set on U and B be a fuzzy set on V. Then the Cartesian product of A and B is the fuzzy set $A\times B$ on $U\times V$ with the membership function defined for all $(x,y)\in U\times V$ by $(A\times B)(x,y)=\min\{A(x),B(y)\}$.

A fuzzy number C is a fuzzy set on R satisfying three conditions: 1) the kernel of C, $\mathrm{Ker}(C)=\{x\in\mathsf{R}\mid C(x)=1\}$, is a nonempty set; 2) all the α-cuts of the fuzzy set C, $C_\alpha=\{x\in\mathsf{R}\mid C(x)\ge\alpha\}$, are closed intervals for all $\alpha\in(0,1]$; and 3) the support of the fuzzy set C, $\mathrm{Supp}(C)=\{x\in\mathsf{R}\mid C(x)>0\}$, is bounded.
If the $\mathrm{Supp}(C)\subseteq[a,b]$, we call C a fuzzy number on the interval $[a,b]$. The family of all fuzzy numbers on the interval $[a,b]$ is denoted by $F_N([a,b])$.

Let $A_1, A_2, ..., A_n\in F_N([a,b])$, then we say that $A_1, A_2, ..., A_n$ form a fuzzy scale on $[a,b]$ if these fuzzy numbers form a Ruspini fuzzy partition (see [5]) on $[a,b]$ (i.e. $\sum_{i=1}^{n}A_i(x)=1$, for all $x\in[a,b]$) and are numbered in accordance with their ordering.

A linguistic variable is a quintuple $(X,T(X),U,M,G)$ where X is the name of the linguistic variable, $T(X)$ is the set of its linguistic values, U is the universe, on which the mathematical meanings of the linguistic terms are defined, G is a syntactical rule (grammar) for generating linguistic terms from $T(X)$, and M is a semantic rule (meaning), that assigns to every linguistic term $A\in T(X)$ its meaning $M(A)$ as a fuzzy set on U. A linguistic variable $(X,T(X),[c,d],M,G)$, $T(X)=\{A_1,A_2,...,A_n\}$ is called a linguistic scale if $A_1=M(\mathrm{A}_1),A_2=M(A_2),...,A_n=M(A_n)$ are fuzzy numbers forming a fuzzy scale on $[c,d]$.

Let $(X_j,T(X_j),[c_j,d_j],M_j,G_j)$, $j=1,...,m$, and $(Y,T(Y),[c,d],M,G)$, be linguistic variables (usually linguistic scales). Let $A_{ij}\in T(X_j)$ and $M_j(A_{ij})=A_{ij}\in F_N([c_j,d_j])$, for all $i=1,...,n$, $j=1,...,m$. Let $B_i\in T(Y)$ and $M(B_i)=B_i\in F_N([c,d])$, $i=1,...,n$. Then the following scheme is called a linguistically defined function (base of fuzzy rules).

$$\begin{aligned}&\text{If } X_1 \text{ is } A_{11} \text{ and ... and } X_m \text{ is } A_{1m}, \text{ then } Y \text{ is } B_1.\\ &\text{If } X_1 \text{ is } A_{21} \text{ and ... and } X_m \text{ is } A_{2m}, \text{ then } Y \text{ is } B_2.\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ &\text{If } X_1 \text{ is } A_{n1} \text{ and ... and } X_m \text{ is } A_{nm}, \text{ then } Y \text{ is } B_n.\end{aligned}\qquad(1)$$

Using the approach of Sugeno & Yasukawa [7], we consider the rule base (1) and an m-tuple of crisp input values (a_1, a_2, ... , a_m). By entering these observed values into the linguistically defined fuzzy function, we get

the output $b^{SY} = \left(\sum_{i=1}^{n} h_i \cdot b_i\right) / \left(\sum_{i=1}^{n} h_i\right)$, where $h_i = \min\{A_{i1}(a_1), A_{i2}(a_2), ..., A_{im}(a_m)\}$ and b_i is the center of gravity of B_i, defined by the formula $b_i = \int_{y \in V} B_i(y) \cdot y \, dy / \int_{y \in V} B_i(y) \, dy$, i=1,...,n. The advantage of this approach is that the linguistic terms B_i, i=1,...,n, can be used for the linguistic description of the output. If b^{SY} lies in the intersection of supports of two neighboring fuzzy numbers $B_{i_k}, B_{i_{k+1}}$ then the output b^{SY} can be characterized as being $B_{i_k}(b^{SY})$ percent of B_{i_k} and $B_{i_{k+1}}(b^{SY})$ percent of $B_{i_{k+1}}$.

The same can be also done by using Takagi-Sugeno fuzzy controller (presented in [9]), where the consequent parts of the rules are modeled by constant functions. The fuzzy controller of Takagi & Sugeno [9] considers a rule base in the following form:

If x_1 is A_{11} and ... and x_m is A_{1m}, then $y = g_1(x_1, ..., x_m)$.
If x_1 is A_{21} and ... and x_m is A_{2m}, then $y = g_2(x_1, ..., x_m)$. (2)
.....
If x_1 is A_{n1} and ... and x_m is A_{nm}, then $y = g_n(x_1, ..., x_m)$.

Here $x_1, x_2, ..., x_m$ are the input variables, $A_{i1}, A_{i2}, ...A_{in}$ are fuzzy numbers on intervals $[c_j, d_j]$ for all j=1, ..., m, and $y = g_i(x_1, ..., x_m)$ describes the control function for the i-th rule, i=1, ..., n. Let us consider again an m-tuple of crisp input values $a_1, a_2, ..., a_m$, $a_j \in [c_j, d_j]$ for all j=1,2,...,m. The output of Takagi-Sugeno fuzzy controller is computed as $b^{TS} = \sum_{i=1}^{n} h_i \cdot g_i(a_1, a_2, ..., a_m) / \sum_{i=1}^{n} h_i$, where h_i, i=1,2,...,n, are defined in the same way as in the Sugeno-Yasukawa algorithm. If $y = g_i(x_1, ..., x_m) = b_i$, $b_i \in \mathbb{R}$ for all i=1,2,...,n, we speak about the Sugeno fuzzy controller; its input-output function is in the form $b^S = \sum_{i=1}^{n} (h_i \cdot b_i) / \sum_{i=1}^{n} h_i$. If we take b_i as representatives of $B_i = M(B_i)$, i=1,2,...,n, that were used in the fuzzy rule base (1), we get the same using the Sugeno algorithm as before by the Sugeno-Yasukawa algorithm (Sugeno & Yasukawa represent B_i's by their centers of gravity, in the following text we will use elements of kernels of triangular fuzzy numbers B_i).

3 Methods

Greene in [2] and Netík in [4] suggest diagnostic criteria for detecting the presence of conversion symptoms. We combine these with the expert knowledge of one skilled diagnostician to construct a four phase linguistic fuzzy model that meets all the requirements given in the introduction section. The expert, drawing on his experience with MMPI-2 and conversion patients, softened the criteria formulated in Greene [2], thus creating a linguistic description of appropriate scale values which are represented by fuzzy numbers in the model. We have identified four phases of MMPI-2 data assessment.

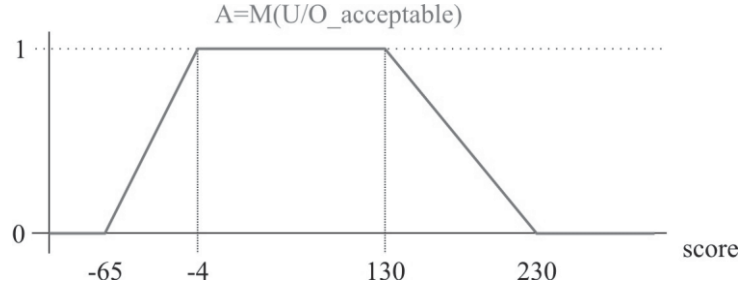


Figure 1 A fuzzy number representing the meaning of "acceptable scores" of the U/O reporting validity scale .

Validity assessment is based on 7 validity scales (?, TRIN, VRIN, U/O reporting, L, F, Fb – see [2] for details). For each of these scales we define the meaning of the linguistic term "acceptable scores" by a fuzzy number (the universe of this fuzzy number is given by all the possible values of the respective scale) – Figure 1 provides an example. Validity rate of a particular MMPI-2 protocol is then determined through (3), where $?', TRIN', ..., Fb'$ are fuzzy singletons representing the respective scale scores. Validity rate is a real number from $[0,1]$ and the following holds: validity rate $= 1 -$ invalidity rate .

$$\mathrm{hgt}\left\{\left(?' \times TRIN' \times ... \times Fb'\right) \cap \left(M(?\_acceptable) \times M(TRIN\_acceptable \times ... \times M(Fb\_acceptable))\right)\right\} \quad (3)$$

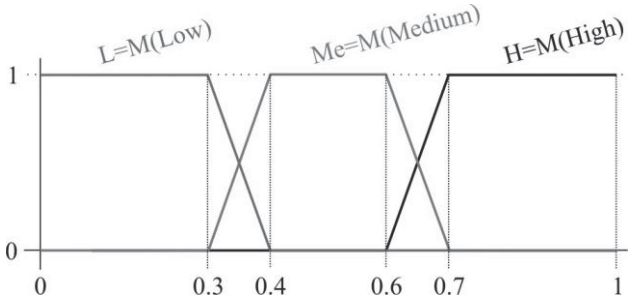


Figure 2 Fuzzy scale for validity rate interpretation.

In order to describe the resulting validity rate linguistically, we define a linguistic scale (see Figure 2). This way we are able to interpret for example the validity rate 0.34 as being 60% low and 40% medium.

The next step of the diagnostic process is to assess the MMPI-2 protocol "at first sight". We set up a one-rule "filter" that distinguishes between MMPI-2 protocols that can indicate converse symptoms and those that are not supporting such diagnosis at all. This discrimination is based on relationships among the 10 clinical scales scores. We call this step **data appropriateness determination**. Again, acceptable relationships are described linguistically and a fuzzy number meaning is assigned to each of them. The resulting appropriateness rate is from $[0,1]$ and can be interpreted linguistically using the fuzzy scale from Figure 3.

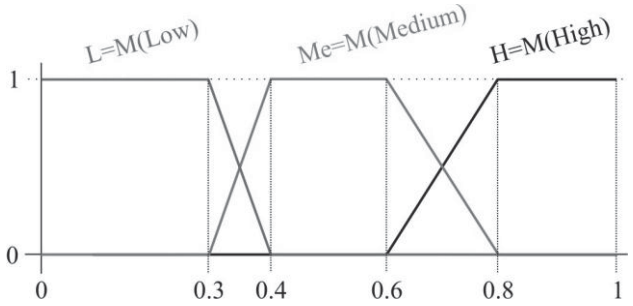


Figure 3 Fuzzy scale for appropriateness and converse V obviousness rate interpretation.

The most specific identification of converse symptoms (see [2]) is the so called "Converse V", which is such a configuration of three clinical scales scores (Hypochondrias (*Hs*), Depression (*D*) and Hysteria (*Hy*)), where both *Hs* and *Hy* scores are above the *D* score, all are "clinically significant" and none of the scores Hs and Hy is "too large". In other words the plot of these scales scores should resemble the shape of the letter V. We have transformed this description into the following:

1. (*Hs* – *D*) is *significant* and (*Hy* – *D*) is *significant*.
2. (*Hs* – *D*) is *very_significant* or (*Hy* – *D*) is *very_significant*
3. *Hs_Hy_ratio* is *acceptable*, where

$$Hs\_Hy\_ratio = \begin{cases} \dfrac{\max(Hs-D, Hy-D)}{\min(Hs-D, Hy-D)}, & \text{if } \min(Hs-D, Hy-D) \neq 0, \\ 100 & \text{else.} \end{cases}$$

Figure 4 shows obvious and indistinct "converse V" shape described by these three conditions. Again an **obviousness rate** from $[0,1]$ is obtained and can be linguistically interpreted using the fuzzy scale in Figure 3.

Validity rate	Appropriateness rate	Converse V obviousness rate	Conversion symptoms presence
High	High	High	Present
High	High	Medium	Present
High	Medium	High	Present
...	...	...	...
Medium	Medium	High	Possibly present
...	...	...	...
Anything	Anything	Low	Not present

Table 1 A part of the rule base for conversion symptoms presence determination.

From the validity, appropriateness and converse V obviousness rates we can now determine, whether conversion symptoms are present or not. We have 11 fuzzy rules available (see Table 1, where the meaning of the linguistic term "Anything" is described by the fuzzy set $AN = M(Anything)$; $AN(x) = 1$ for all $x \in U$; U is the universe on which the meanings of linguistic terms of the respective linguistic variable are defined on). With the three real values of the validity, appropriateness and converse V obviousness rates as inputs, we use a modified Sugeno & Yasukawa approach to fuzzy control to derive outputs. We see this as a fuzzy classification problem. We have 3 classes (Not_present – nr. 0, Possibly_present – nr. 1, Present – nr. 2). Our modification is that the consequent parts of the rules are represented not by centers of gravity, but by the number of the class. Numbers of the classes form a cardinal scale. This allows us to perform fuzzy classification (we accept partial membership to two neighboring classes) and obtain a number from $[0, 2]$ that can again be interpreted linguistically.

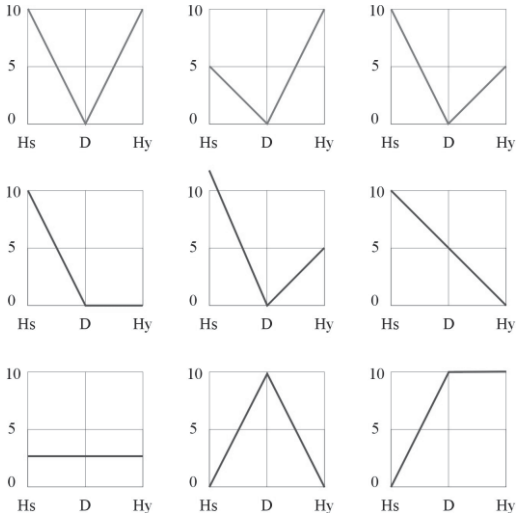


Figure 4 Plots of possible Hs, D and Hy scores configurations. The top row depicts examples of obvious converse V shapes (obviousness rate = 1), the middle and bottom rows depict examples of indistinct converse V shape (obviousness rate = 0).

4 Results

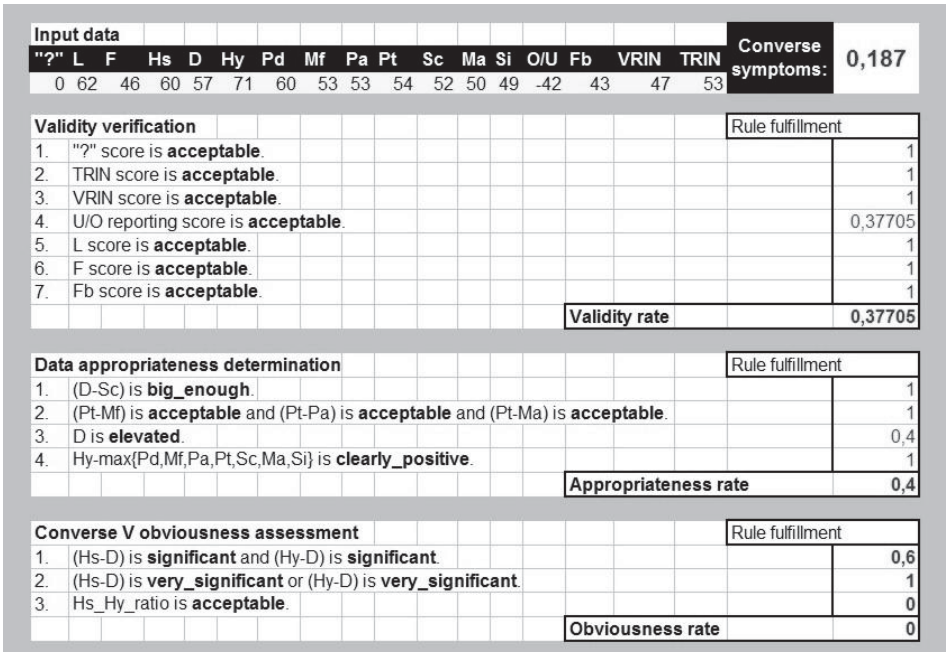
We have presented a linguistic fuzzy model for the purposes of psychological diagnostics. The above-described linguistic fuzzy model has been implemented in MS Excel (see Figure 5). It uses 17 MMPI-2 scale scores (10 clinical scales and 7 validity scales) as inputs and provides 1 overall output – it determines, whether the conversion symptoms are i) present, ii) possibly present, or iii) not present. Results on lower levels of information aggregation are also available: protocol validity rate, data appropriateness rate and converse V obviousness rate. Finally, to fully support the justification of the diagnosis, important segments of all the antecedent parts of linguistic rules and their fulfillment rates are also provided. The model reflects the experience and knowledge of one particular expert diagnostician as well as the diagnostic and interpretational guidelines contained in [2,4]. At present we are testing the model on 250 MMPI-2 protocols.

5 Discussion

We have managed to successfully capture expert knowledge and present it in a form that can be understood by psychologists not familiar with linguistic fuzzy modeling. Although our approach introduces to the process some level of uncertainty, which is inherent in the linguistic description of expertly defined rules, the results of the testing seem promising. There are still some small discrepancies between the results of our model and those obtained by strictly following the criteria e.g. in [2] or [4]. These may have many causes, including the need for general revision of some diagnostic recommendations, new norms and verification of validity of these recommendations for Czech population, and even specificity of our sample of 250 MMPI-2 protocols. At present we are fine-tuning the model to minimize these discrepancies.

Once the fine-tuning phase is completed, the presented model may prove useful in various areas. The practical diagnostic application is obvious. However, inclusion of expert knowledge into a formalized process and presenting the results in an intelligible way is the first step in knowledge transfer based on linguistic fuzzy mod-

eling. Teaching and professional training programs may benefit from such a tool as well. We may also consider its use in research – formalized and software implemented expert knowledge (or experience) can be tested and verified on large samples. Our work suggests that interdisciplinary applications of linguistic fuzzy modeling are not only possible, but even desirable.



Input data																		Converse symptoms:	0,187
"?"	L	F	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	Si	O/U	Fb	VRIN	TRIN			
0	62	46	60	57	71	60	53	53	54	52	50	49	-42	43	47	53			

Validity verification		Rule fulfillment
1.	"?" score is **acceptable**.	1
2.	TRIN score is **acceptable**.	1
3.	VRIN score is **acceptable**.	1
4.	U/O reporting score is **acceptable**.	0,37705
5.	L score is **acceptable**.	1
6.	F score is **acceptable**.	1
7.	Fb score is **acceptable**.	1
	Validity rate	**0,37705**

Data appropriateness determination		Rule fulfillment
1.	(D-Sc) is **big_enough**.	1
2.	(Pt-Mf) is **acceptable** and (Pt-Pa) is **acceptable** and (Pt-Ma) is **acceptable**.	1
3.	D is **elevated**.	0,4
4.	Hy-max{Pd,Mf,Pa,Pt,Sc,Ma,Si} is **clearly_positive**.	1
	Appropriateness rate	**0,4**

Converse V obviousness assessment		Rule fulfillment
1.	(Hs-D) is **significant** and (Hy-D) is **significant**.	**0,6**
2.	(Hs-D) is **very_significant** or (Hy-D) is **very_significant**.	**1**
3.	Hs_Hy_ratio is **acceptable**.	**0**
	Obviousness rate	**0**

Figure 5 MS Excel implementation results.

Acknowledgements

The presented research was supported by the grant *PrF_2011_022 Mathematical models and structures* obtained from the Internal Grant Agency of the Palacky University in Olomouc.

References

[1] Dubois, D., Prade, H. (Eds.).: *Fundamentals of fuzzy sets*. Kluwer Academic Publishers, Dordrecht, 2000.
[2] Greene, R. L.: *The MMPI-2: An interpretive manual*. Allyn and Bacon, Boston, 2000.
[3] Hathaway, S. R., McKinley, J. C.: A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of psychology*, 10, 249-254, 1940.
[4] Netík, K.: *The Minnesota Multiphasic Personality Inventory - 2: první české vydání*. Testcentrum, Praha, 2002.
[5] Ruspini, E.: A new approach to clustering. *Inform. Control*, 15, 1969, pp. 22-32.
[6] Sugeno, M.: An introductory survey on fuzzy control. *Information Sciences*, 36, 1985, pp. 59-83.
[7] Sugeno, M. & Yasukawa, T.: A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on fuzzy systems*, 1 (1), 1993, pp. 7-31.
[8] Svoboda, M., Řehan, V. et al. *Aplikovaná psychodiagnostika v České republice*. Psychologický ústav FF MU v Brně, Brno, 2004. [Applied psychodiagnostics in the Czech Republic]
[9] Takagi, T., & Sugeno, M.: Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on systems, man and cybernetics.*, 1 (15), 1985, pp. 116-132.
[10] Talašová, J.: Fuzzy metody vícekriteriálního hodnocení a rozhodování. Univerzita Palackého v Olomouci, Olomouc, 2003. [Fuzzy methods of multicriteria evaluation and decision making]
[11] World health organisation.: *Mezinárodní klasifikace nemocí - 10. revize: Duševní poruchy a poruchy chování* (3rd edition). Psychiatrické centrum Praha, Praha, 2006. [International Classifications of Diseases – 10th revision: Mental and behavioural disorders (3rd edition)]
[12] Zadeh, L. A.: Fuzzy Sets. *Inform. Control*, 8, 1965, pp. 338-353.
[13] Zadeh, L. A.: The concept of linguistic variable and its application to approximate reasoning. *Information sciences*, Part 1: 8, 1975, pp. 199-249, Part 2: 8 1975, pp. 301-357, Part 3: 9 1975, pp. 43-80.

PUBL. XI

Talašová, J. and Stoklasa, J., A model for evaluating creative work outcomes at Czech Art Colleges. *Proceedings of the* 29^{th} *International Conference on Mathematical Methods in Economics 2011 - part II*, 653–658, 2011.

A model for evaluating creative work outcomes at Czech Art Colleges

Jana Talašová[1], Jan Stoklasa[2]

Abstract. The Register of Artistic Performances is currently being developed in CZ that will contain information on works of art originating from creative activities of art colleges and faculties. Outcomes in various fields of artistic production will be divided into 27 categories, based on their significance, size, and international reception (each criterion classifies into three classes), and each category will be assigned a score. The total score will provide a basis for allocating a part of the state-budget subsidy among art colleges.
The paper discusses the model used to determine scores for each category. The approach is based on Saaty's method, which expertly compares significances of all 27 categories. Creating Saaty's matrix of preference intensities for abstract categories, while maintaining acceptable consistency for such a large matrix, is a difficult task. In the paper we describe a procedure for obtaining required information from a team of persons responsible for different fields of artistic production. A search for solution to this problem has led to new interpretations of Saaty's matrix elements and its consistency condition.

Keywords: Multiple criteria evaluation, Saaty's method, work of art.

JEL Classification: C44
AMS Classification: 91B74

1 Register of Artistic Performances, Classification of Works of Art

The Register of Artistic Performances (RAP) is currently being developed in the Czech Republic that should contain information on works of art originating from creative activities of art colleges and faculties (see [6]). The RAP is conceived as an analogy to the register of R&D outcomes where information on outcomes of research institutions (including universities) has been collected for some years already. In both the registers the outcomes are stored under several categories. These categories are assigned scores. The sum of scores of all the outcomes of a given university is considered an indicator of its performance in the area of creative activity. These numeric values can then be used in decisions regarding one part of the total money to be allocated among universities from the state budget.

The structure of the evaluated categories used in the Czech model was inspired, to some extent, by the artistic categories in the Slovak Republic (see [7]). However, the mathematical model used to determine scores for each category in Slovakia is quite different.

For the purposes of registration of works of art originating from creative activities of the Czech art colleges and faculties, the whole area of artistic production is divided into seven fields: fine arts, design, architecture, theatre, film, literature, and music.

Each piece of art, regardless of the field, is categorized according to the following three criteria:

- Relevance or significance of the piece;
- Extent of the piece;
- Institutional and media reception/impact of the piece.

In each criterion, three different levels are distinguished (denoted by capital letters for easier handling):

- The criterion *Relevance or significance of the piece*:
  - A – a new piece of art or a performance of crucial significance;
  - B – a new piece of art or a performance containing numerous important innovations;
  - C – a new piece of art or a performance pushing forward modern trends.

[1] Palacký University, Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, jana.talasova@upol.cz.

[2] Palacký University, Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, jan.stoklasa@upol.cz.

- The criterion *Extent of the piece*:
  - K - a piece of art or a performance of large extent;
  - L - a piece of art or a performance of medium extent;
  - M - a piece of art or a performance of limited extent.
- The criterion *Institutional and media reception/impact of the piece*:
  - X – international reception/impact,
  - Y – national reception/impact,
  - Z – regional reception/impact.

The resulting category for a piece of art is given by a combination of three capital letters – e.g. AKX, BKY, or CLZ. There are 27 categories altogether. The decision concerning the relevance or significance of the piece (choice of A, B or C) rests upon expert assessment; the experts have at their disposal general definitions of each category and examples of works of art in each category – for all three levels of each criterion and for all 7 fields of artistic production – to assist them in the decision process. As for the extent of the piece (levels K, L, M), all the classes are specified for all the fields of art. As for the institutional and media reception/impact, lists of institutions corresponding to categories X, Y, Z are available for all fields.

Le us notice, there are interactions among the three mentioned criteria. The first one (expertly defined *Relevance or significance of the piece of art*) and the third one (*Institutional and media reception/impact of the piece*) partly overlap. That means, we are not allowed to set separately the weights of criteria and the scores of levels for each of them, and then calculate the scores of categories by means of the weighted average operation. It is necessary to set directly the scores of the categories that are described by the triples of criteria levels.

2 Determining scores for particular categories of artistic production

Saaty's method (see [2, 3, 4]) served as a basis for determination of scores for all 27 categories of artistic production. However obvious it was that this mathematical tool is the most appropriate for such a task, certain challenges concerning its use were also clearly apparent: (1) a difficulty for a team of experts to express preferences with respect to abstract categories; (2) a difficulty to reach acceptable consistency of Saaty's matrix under such a large number of categories; (3) a consensus within the group of experts (professional guarantors of particular fields of art). The proposed solution to these problems will be described in the following paragraphs.

Admittedly, expressing one's opinion on intensities of preferences with respect to abstract categories is difficult. Experts, professional guarantors of artistic fields, were first asked to provide specific (historical) examples of works of art in all categories in their field. (This step was also important to ensure, or to suggest modifications to ensure, that corresponding categories be really comparable in terms of evaluation across fields.) Next, professional guarantors of each field of art set their preferences concerning pairs of categories, while considering the representatives (examples) of each category to aid them in their decisions.

Although it was possible for each of these experts to express their preferences separately, and only then to derive the collective preferences (from the individual ones), we used a different approach. The collective preferences were set directly at a team meeting of experts. The reason was that art-college experts are not used to work with mathematical models and individual inputting of required data could prove difficult for them. Achieving consensus was also intentionally preferred over averaging different opinions.

Great effort was made to find the best way of converting expert preferences concerning the 27 categories of artistic production (represented in each field of art by specific examples) into a mathematical model in order to determine their scores. To facilitate the process of inputting required data by the experts and to achieve the necessary consistency of this input, the following two-step procedure was performed:

In the first step, we have determined the order of importance of the categories by the pairwise comparison method (see [2, 5]). This method employs a matrix of preferences and indifferences $P = \left\{ p_{i,j} \right\}_{i,j=1,\ldots,27}$. For its elements it holds that:

$p_{i,j} = 1$, if the ith category is more important than the jth category;

$p_{i,j} = 0{,}5$, if the ith category is equally important as the jth category;

$p_{i,j} = 0$, if the jth category is more important than the ith category.

It is sufficient for the experts to fill in the upper right triangle of the matrix, that is, the elements $p_{i,j}, i < j$, as $p_{i,i} = 0{,}5$ and $p_{j,i} = 1 - p_{i,j}$. The row sums $D_i = \sum_{j=1}^{27} p_{i,j}, i = 1,\ldots,27$, determine the order of importance of the

categories (their quasi-ordering, transitive and complete relation, that can be described as a linear ordering of classes of indifferent elements). We need to verify consistency of the preferences in the sense of transitivity, that is, whether it holds that $p_{i,k} \geq \max\left\{p_{i,.j}, p_{j,k}\right\}$ for all $i, j, k \in \{1,...,27\}$. If the matrix is not consistent, we make a minimum amount of changes necessary for it to become so. These changes are then consulted with the team of experts and if they are approved of, we can proceed. All the changes actually made while solving our problem are summarized in Tab 1.

		1 AKX	2 AKY	3 AKZ	4 ALX	5 ALY	6 ALZ	7 AMX	8 AMY	9 AMZ	10 BKX	11 BKY	12 BKZ	13 BLX	14 BLY	15 BLZ	16 BMX	17 BMY	18 BMZ	19 CKX	20 CKY	21 CKZ	22 CLX	23 CLY	24 CLZ	25 CMX	26 CMY	27 CMZ	Prefe-rence	Preference order	
1	AKX	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	26,5	1	AKX
2	AKY		0,5	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	25	2	AKY
3	AKZ			0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	25	2	AKZ
4	ALX				0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	23,5	4	ALX
5	ALY					0,5	1,0	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	21,5	6	ALY
6	ALZ						0,5	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	20,5	7	ALZ
7	AMX							0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	22,5	5	AMX
8	AMY								0,5	1,0	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	18,5	9	AMY
9	AMZ									0,5	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	17,5	10	AMZ
10	BKX										0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	19,5	8	BKX
11	BKY											0,5	1,0	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	16	11	BKY
12	BKZ												0,5	0,0	1,0	1,0	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	14	13	BKZ
13	BLX													0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	16	11	BLX
14	BLY														0,5	1,0	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	12,5	15	BLY
15	BLZ															0,5	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	11,5	16	BLZ
16	BMX																0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	14	13	BMX
17	BMY																	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	10,5	17	BMY
18	BMZ																		0,5	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	9	18	BMZ
19	CKX																			0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	9	18	CKX
20	CKY																				0,5	1,0	0,0	1,0	1,0	1,0	1,0	1,0	6,5	21	CKY
21	CKZ																					0,5	0,0	1,0	1,0	1,0	1,0	1,0	5,5	22	CKZ
22	CLX																						0,5	1,0	1,0	1,0	1,0	1,0	7,5	20	CLX
23	CLY																							0,5	1,0	0,0	1,0	1,0	3,5	24	CLY
24	CLZ																								0,5	0,0	1,0	1,0	2,5	25	CLZ
25	CMX																									0,5	1,0	1,0	4,5	23	CMX
26	CMY																										0,5	1,0	1,5	26	CMY
27	CMZ																											0,5	0,5	27	CMZ

changes made to achieve consistency of the matrix resulting form the final order

change made from 0,5 to 1 or from 1 to 0,5 ("**small change**")

change made from 0 to 1 ("**big change**")

Table 1 The pairwise comparison matrix with highlighted changes.

In the second step, Saaty's matrix of preference intensities $S = \left\{s_{i,j}\right\}_{i,j=1,...,27}$ was constructed for categories numbered in ascending order according to their significance determined in the previous step. Again, it was sufficient to fill in the upper right triangle of the matrix. The elements $s_{i,j}, i < j,$ were set as follows:

$s_{i,j} = 1\ldots$ the i[th] and j[th] categories are **equally** important;

$s_{i,j} = 3\ldots$ the i[th] category is **slightly more important** than the j[th] category;

$s_{i,j} = 5\ldots$ the i[th] category is **strongly more important** than the j[th] category;

$s_{i,j} = 7\ldots$ the i[th] category is **very strongly more important** than the j[th] category;

$s_{i,j} = 9\ldots$ the i[th] category is **extremely more important** than the j[th] category.

It holds that $s_{i,i} = 1$ and $s_{j,i} = 1/s_{i,j}$, for the intensity of preference $s_{i,j}$ expresses the ratio of preferences between both categories.

The traditional requirement for consistency in Saaty's method, that is $s_{i,k} = s_{i,j} \cdot s_{j,k}$ for all $i, j, k \in \{1,...,27\}$, is basically unachievable. For example, consider only four arbitrary objects that are linearly ordered according to their importance. If each of them is just slightly more important than the following one, then in the case of full consistency the first one would have to be 27 times more important than the fourth. But the maximum value available for expressing intensity of preference is nine (as is shown by psychological research [3], this is the highest number of levels of importance that human is able to distinguish). We have weakened the original requirement on consistency, which was too strong, and for the purposes of our work we have requested $s_{i,k} \geq \max\left\{s_{i,.j}, s_{j,k}\right\}$ for all $i, j, k \in \{1,...,27\}$. When the categories are numbered as to their importance, this requirement is easy to verify. In addition to the fact that the matrix S has to be reciprocal (i.e. $s_{i,i} = 1$ and $s_{i,j} = 1/s_{j,i}$ for $i, j \in \{1,...,27\}$) in view of the above-mentioned condition, consistency means that the elements of S are nondecreasing from left to right and from bottom up. If the matrix, as set by the experts, is not consistent, we propose the minimum amount of changes necessary for it to become so – the team of professional

guarantors either approve of these changes or make their own to achieve consistency. Tab. 2 illustrates the changes actually made in our application in order to remove inconsistencies from the original matrix S. (Tab. 2 contains also changes induced by re-dividing the pairs of indifferent categories having originated from the pair-wise comparison method.)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
		AKX	AKY	AKZ	ALX	AMX	ALY	ALZ	BKX	AMY	AMZ	BKY	BKZ	BLX	BMX	BLY	BLZ	BMY	BMZ	CKX	CLX	CKY	CKZ	CMX	CLY	CLZ	CMY	CMZ
1	AKX	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	9	9	9	9	9	9	9	9	9	9
2	AKY		1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	9	9	9	9	9	9
3	AKZ			1	3	3	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	9	9	9	9	9	9
4	ALX				1	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9	9	9	9	9
5	AMX					1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	9	9	9	9
6	ALY						1	3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9	9
7	ALZ							1	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9
8	BKX								1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7
9	AMY									1	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7
10	AMZ										1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7
11	BKY											1	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7
12	BKZ												1	3	5	5	5	5	5	5	5	5	5	5	5	7	7	7
13	BLX													1	5	5	5	5	5	5	5	5	5	5	5	7	7	7
14	BMX														1	3	3	3	5	5	5	5	5	5	5	7	7	7
15	BLY															1	3	3	5	5	5	5	5	5	5	7	7	7
16	BLZ																1	3	5	5	5	5	5	5	5	7	7	7
17	BMY																	1	5	5	5	5	5	5	5	5	7	7
18	BMZ																		1	5	5	5	5	5	5	5	5	5
19	CKX																			1	3	5	5	5	5	5	5	5
20	CLX																				1	5	5	5	5	5	5	5
21	CKY																					1	3	3	5	5	5	5
22	CKZ																						1	3	5	5	5	5
23	CMX																							1	5	5	5	5
24	CLY																								1	3	3	3
25	CLZ																									1	3	3
26	CMY																										1	3
27	CMZ																											1

3,5,7,9	změny vyplývající z rozdělení kategorií
3,5	změny respektující párové porovnávání sousedních kategorií
5	změny pro udržení konzistence vyvolané zadáním "červených hodnot"

Table 2 Saaty's matrix of preference intensities with highlighted changes.

Under the assumption that S is close enough to an ideally consistent matrix (i.e. matrix that fulfills $s_{i,k} = s_{i,j} \cdot s_{j,k}$ for all $i, j, k \in \{1,...,27\}$), the scores of 27 categories, representing their relative importance, are calculated by Saaty's method as components of the eigenvector corresponding to the largest eigenvalue.

The resulting scores of artistic categories can also be obtained from S in a different way. The columns of S can be interpreted as repeated measurements of the relative importances of the 27 categories. These measurements are performed by the team of experts who compare all the categories with the first one, than the second one, and so on until the 27th one. From the point of view of mathematical statistics, these are compositional data, i.e. data bearing only relative information (see [1]). Information contained in this data can be expressed by estimating its mean value. A proper estimator of the mean value of this kind of data is a vector whose components are geometric means of the corresponding components of vectors representing single measurements. The relative scores of all 27 categories can be also obtained by computing geometric means of the rows of Saaty's matrix (this calculation method is known as the logarithmic least squares method, see [2]). This weaker consistency of S ($s_{i,k} \geq \max\{s_{i,.j}, s_{j,k}\}$ for all $i, j, k \in \{1,...,27\}$) is then a natural requirement that allows for an easy check on consistency of the expertly entered data. The facts that S has to be reciprocal and, with the categories ordered according to their importance, that the values of a well entered matrix S must be nondecreasing from left to right and from bottom up can serve as a good guiding principle for teams of experts in defining the preference intensities of pairs of categories.

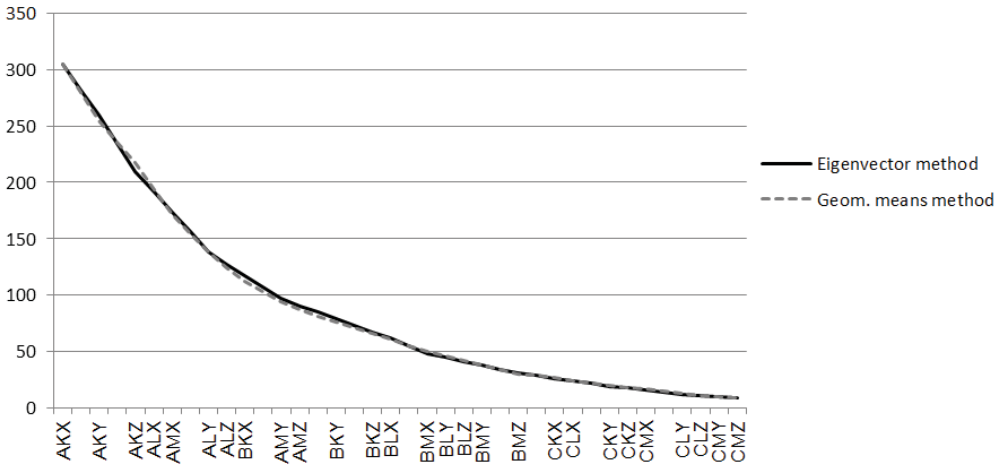


Figure 1 Graphical comparison of the eigenvector method with the geometric means method.

Category	Relevance or significance	Extent	Institutional reception	Eigenvector method	Geom. means method
AKX	Crucial significance and originality	Large	International	**305**	**305**
AKY	Crucial significance and originality	Large	National	**259**	**254**
AKZ	Crucial significance and originality	Large	Regional	**210**	**217**
ALX	Crucial significance and originality	Medium	International	**191**	**194**
AMX	Crucial significance and originality	Small	International	**174**	**171**
ALY	Crucial significance and originality	Medium	National	**138**	**138**
ALZ	Crucial significance and originality	Medium	Regional	**127**	**124**
BKX	Bearing many important inovations	Large	International	**117**	**112**
AMY	Crucial significance and originality	Small	National	**97**	**94**
AMZ	Crucial significance and originality	Small	Regional	**90**	**87**
BKY	Bearing many important inovations	Large	National	**79**	**75**
BKZ	Bearing many important inovations	Large	Regional	**66**	**66**
BLX	Bearing many important inovations	Medium	International	**62**	**61**
BMX	Bearing many important inovations	Small	International	**48**	**50**
BLY	Bearing many important inovations	Medium	National	**44**	**46**
BLZ	Bearing many important inovations	Medium	Regional	**40**	**41**
BMY	Bearing many important inovations	Small	National	**37**	**38**
BMZ	Bearing many important inovations	Small	Regional	**31**	**30**
CKX	Developing current trends	Large	International	**26**	**26**
CLX	Developing current trends	Medium	International	**24**	**24**
CKY	Developing current trends	Large	National	**19**	**20**
CKZ	Developing current trends	Large	Regional	**17**	**18**
CMX	Developing current trends	Small	International	**16**	**16**
CLY	Developing current trends	Medium	National	**12**	**13**
CLZ	Developing current trends	Medium	Regional	**10**	**11**
CMY	Developing current trends	Small	National	**9**	**9**
CMYŹ	Developing current trends	Small	Regional	**8**	**9**

Table 3 Scores obtained by the Saaty matrix eigenvector method and those determined as geometric means of rows of S.

Tab. 3 compares the scores determined by the Saaty matrix eigenvector method with those determined as geometric means of the rows. The scores are normalized so that the maximum is 305 (analogy to R&D outcomes evaluation). It is easy to see that the differences between these two methods are not significant, see Fig. 1. The Saaty matrix eigenvector method will be used in testing the model on the first real dataset, gathered by Czech art colleges and faculties for the years 2008 to 2010.

3 Conclusion

The Register of Artistic Performances and the methodology of evaluating artistic production originating from creative activities of art colleges and faculties are currently being pilot-tested in the Czech Republic. At present, our effort is focused on refining the triplets of class specification for all three criteria and for all the fields of art, and particularly on developing a most objective mechanism of expert classification of artistic production into 27 categories.

The mathematical model for score determination was developed in an effort to achieve the best possible conversion of preferences of the expert team into scores for different categories of artistic production. With Saaty's method serving as an appropriate basis, the solution to this problem required its implementation in a special procedure.

Acknowledgements

This research is conducted with the support of the Centralized Developmental Project C41 entitled *Evaluating Creative Work Outcomes Pilot Project* and financed by the Czech Ministry of Education.

References

[1] Aitchison, J. *The Statistical Analysis of Compositional Data*. *Monographs on Statistics and Applied Probability.* Chapman & Hall Ltd., London , 1986.
[2] Ramík, J.: *Analytický hierarchický process (AHP) a jeho využití v malém a středním podnikání.* Slezská univerzita v Opavě, Karviná, 2000.
[3] Saaty, T.L.: *The Brain: Unraveling the Mystery of How it Works, The Neural Network Process* . 481 pp., RWS Publ., 1999.
[4] Saaty, T.L.: *The Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process.* Vol. VI of the AHP Series, 478 pp., RWS Publ., 2000.
[5] Talašová, J.. *Fuzzy metody vícekriteriálního hodnocení a rozhodování.* Vydavatelství Univerzity Palackého, Olomouc. 2003.
[6] Zelinský, M.(ed.): *Registr uměleckých výkonů*. Akademie múzických umění v Praze, 2010.
[7] Smernica č. 13/2008-R zo 16. októbra 2008 o bibliografickej registrácii a kategorizácii publikačnej činnosti, umeleckej činnosti a ohlasov. Ministerstvo školstva Slovenskej Republiky.

PUBL. XII

Stoklasa, J., Krejčí, J. and Talašová, J., Fuzzified AHP in evaluation of R&D outputs - a case from Palacky University in Olomouc, *Proceedings of the* 31^{st} *International Conference Mathematical Methods in Economics 2013*, 856–861, 2013.

Fuzzified AHP in evaluation of R&D outputs - a case from Palacky University in Olomouc

Jan Stoklasa[1], Jana Krejčí[2], Jana Talašová[3]

Abstract. In this paper we present a method developed at the Faculty of Science, Palacky University in Olomouc for the assessment of scientific quality of books (for the purposes of funds distribution among departments). The method is based on the fuzzified AHP used to determine fuzzy weights of predefined categories of publishers prior to the evaluation process. The weights reflect the reputation of the publishers from the respective category. The following evaluation process then combines expert assessment (peer review) with the less subjective criterion of publishers reputation. In the first stage of the two-stage evaluation process, each book is evaluated by a fuzzy number which corresponds with the reputation of the publisher and is interpreted as an interval of possible evaluations - scores (the support) with a most typical score (default evaluation - the only element in the kernel of the fuzzy number). In the second stage, the default evaluation can be altered during the peer review process within the proposed interval of scores and the reasons for changes (if made) are reported. The whole development process of the method will be briefly summarized to stress the possibility of using AHP (and its fuzzified version) for visualisation and interpretation of expert preferences expressed in terms of intuitively set intervals of scores.

Keywords: AHP, fuzzy, R&D, evaluation, MCDM, management.

JEL classification: C44
AMS classification: 90C15

1 Research and development evaluation in the Czech Republic

The evaluation of R&D outputs is a complex task. In the Czech Republic, the evaluation of R&D outputs for the purposes of funds distribution among research institutions (including universities) is regulated by the Methodology for evaluation of R&D outputs [8], that is under constant development as the issues of quality assessment are becoming more and more important (see also [1]). Although some outputs are assessed according to criteria that can be assumed to correlate with scientific quality (e.g. the evaluation of papers in journals with nonzero impact factor), others are assigned a fixed score (this is the case of patents, books etc.).

On the task of the evaluation of books we will show our proposal how to combine expert assessment (peer review) with other criteria that are easier to assess - in this case the reputation of the publisher of the book. The first stage of the evaluation is based on a classification of publishers into 4 categories according to their reputation. The classification was provided by a board of experts prior to the development of the presented mathematical model. Based on pairwise comparison the categories were assigned intervals of scores (represented by triangular fuzzy numbers) with "default" evaluation for typical books published by the publishers from the current category. In the second stage, the peer review process is used to adjust the evaluation of the book according to its scientific quality within predefined limits for each category of publishers. The underlying mathematical apparatus used to reflect the preferences of the board of evaluators and to derive default evaluations for books from a given category of publishers is described

[1]Palacky University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, jan.stoklasa@upol.cz
[2]Palacky University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, jana.krejci01@upol.cz
[3]Palacky University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, jana.talasova@upol.cz

in Sections 2 and 3. The mathematical model is based on Saaty's AHP method (see [9, 10]) and its fuzzification that was introduced in [4]. Section 4 summarizes the computation of scores for categories of publishers at Palacky University in Olomouc (UP), discussess the results and presents the consecutive two stage evaluation procedure for books, that is currently being used at the Faculty of Science of UP [5].

2 Pairwise comparison matrices

Let us consider a set of objects $K_1, K_2, \ldots, K_n$ for which we need to find evaluations $h_1, h_2, \ldots, h_n$. If these evaluations are known, we can construct the multiplicative matrix of relative preferences $H = \{h_{ij}\}_{i,j=1}^n$ such that $h_{ij} = \frac{h_i}{h_j}$. The elements of such a matrix describe the relative preference of K_i over K_j. The evaluations $h_1, h_2, \ldots, h_n$ are however usually not known. To determine the values of $h_1, h_2, \ldots, h_n$ we can construct the Saaty's matrix $S = \{s_{ij}\}_{i,j=1}^n$, where the elements s_{ij} (provided by experts) describe the estimated ratio of the evaluation of K_i to the evaluation of K_j. As such, s_{ij} are expert estimations of the actual elements h_{ij} of the matrix H, hence we require the matrix S to be reciprocal, i.e. $s_{ij} = \frac{1}{s_{ji}}$ for all $i, j = 1, 2, \ldots, n$. It is easy to see that we need to set only $\frac{n(n-1)}{2}$ elements of the matrix to provide all the information necessary to complete it. For the purpose of expressing expert's intensities of preferences between pairs of objects, Saaty [9, 10] proposes to use the scale shown in Table 1. The decision maker can also use the intermediate values $2, 4, 6$ and 8 with the respective intermediate linguistic meanings (e.g. "between moderately and strongly" for 4).

numerical value of s_{ij}	linguistic description
1	decision maker is indifferent between K_i and K_j
3	K_i is moderately preferred to K_j
5	K_i is strongly preferred to K_j
7	K_i is very strongly preferred to K_j
9	K_i is absolutely preferred to K_j

Table 1 Saaty's scale.

Saaty defines the full consistency of the matrix S by $s_{ik} = s_{ij} \cdot s_{jk}$, for all $i, j, k = 1, 2, \ldots, n$. Full consistency is, however, unachievable for larger matrices. Saaty therefore introduces an inconsistency index $CI = (\lambda_{\max} - n)/(n-1)$, where n is the order of S (i.e. the number of objects that are being compared). A Saaty's matrix S is considered to be consistent enough, when it's inconsistency ratio $CR = CI/RI < 0.1$, where RI is the so called random inconsistency index representing the inconsistency of a randomly generated reciprocal pairwise comparison matrix of the order n. Other approaches to the assessment of consistency of Saaty's matrices can be found in the literature (see [2] for a comparison). Stoklasa et al. suggest the concept of weak consistency in [11, 12] as a minimum requirement on the consistency of the expertly defined Saaty's matrix S. Saaty's matrix of preference intensities is weakly consistent, if and only if for all $i, j, k \in \{1, 2, \ldots, n\}$ the following holds:

$$s_{ij} > 1 \wedge s_{jk} > 1 \quad \Longrightarrow \quad s_{ik} \geq \max\{s_{ij}, s_{jk}\}; \tag{1}$$

$$(s_{ij} = 1 \wedge s_{jk} \geq 1) \vee (s_{ij} \geq 1 \wedge s_{jk} = 1) \quad \Longrightarrow \quad s_{ik} = \max\{s_{ij}, s_{jk}\}. \tag{2}$$

This weak consistency can be checked during the input of the preference intensities by experts and represents a weakening of Saaty's notion of full consistency (for more details see [11, 12]). As follows from the Perron-Frobenius theorem, Saaty' matrix always has a maximum real eigenvalue (spectral radius - see [7]). A fully consistent Saaty's matrix has a single nonzero eigenvalue $\lambda_{\max}$, which is equal to the order of the matrix. The eigenvector corresponding to this maximum eigenvalue represents the evaluations $h_1, h_2, \ldots, h_n$ (see [9, 10]).

Alternatively to find $h_1, h_2, \ldots, h_n$ based on the expertly defined matrix S, we need to find the arguments of the minimum of the following expression:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left(s_{ij} - \frac{h_i}{h_j} \right)^2. \tag{3}$$

The minimisation problem (3) can be rewritten into:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\ln(s_{ij})-\ln(h_i)-\ln(h_j)\right)^2, \tag{4}$$

thus obtaining the logarithmic least square problem, for which the solution can be found in the form of (5), see [7] for more details.

$$h_i = \sqrt[n]{\prod_{j=1}^{n} s_{ij}} \tag{5}$$

3 Fuzzy pairwise comparison matrices

A fuzzy set A on a nonempty universal set U is defined by a mapping $A : U \to [0, 1]$. For each $x \in U$ the value $A(x)$ is called a membership degree of the element x in the fuzzy set A; $A(.)$ is called a membership function of the fuzzy set A. A fuzzy number B is a fuzzy set on $\mathbb{R}$ satisfying three conditions: 1) the kernel of B, $\mathrm{Ker}(B) = \{x \in \mathbb{R}|B(x) = 1\}$ is a nonempty set, 2) the α-cuts of B, $B_\alpha = \{x \in \mathbb{R}|B(x) \geq \alpha\}$ are closed intervals for all $\alpha \in (0, 1]$, and 3) the support of the fuzzy set B, $\mathrm{Supp}(B) = \{x \in \mathbb{R}|B(x) > 0\}$ is bounded. A triangular fuzzy number B has a membership function in the form:

$$B(x) = \begin{cases} 0, & x < b_1 \\ \frac{x-b_1}{b_2-b_1}, & b_1 \leq x \leq b_2 \\ 1, & x = b_2 \\ \frac{b_3-x}{b_3-b_2}, & b_2 \leq x \leq b_3 \\ 0, & x > b_3, \end{cases} \tag{6}$$

and can be represented by the triplet of its significant values $B = (b_1, b_2, b_3)$.

fuzzy value of s_{ij}	linguistic description
$(\frac{1}{3}, 1, 3)$	decision maker is indifferent between K_i and K_j
$(1, 3, 5)$	K_i is moderately preferred to K_j
$(3, 5, 7)$	K_i is strongly preferred to K_j
$(5, 7, 9)$	K_i is very strongly preferred to K_j
$(7, 9, 9)$	K_i is absolutely preferred to K_j

Table 2 Properly fuzzified Saaty's scale [4].

Saaty's method allows the experts to express their preferences using linguistic terms Table 1. It is surely more appropriate to represent the uncertainty of the linguistic terms not by real numbers (that is by elements from the set $\{1, 3, 5, 7, 9\}$), but by triangular fuzzy numbers. The support of each of the fuzzy numbers used is an interval defined by the single real values from the kernels of the neighboring fuzzy numbers (see Table 2). The meaning of "is indifferent between" is defined in a way such that the reciprocity of the Saaty's matrix is preserved. The fuzzification of a Saaty's scale with intermediate values (numerical and linguistic) is analogical.

There are also many ways how to assess the consistency of fuzzy Saaty's matrices (see e.g. [3, 6]). For the purposes of this application, we can approach consistency in the fuzzy case in the same way as in the crisp case - using just the real values from the kernels of the respective fuzzy number (middle significant values). Again, it is reasonable to require the matrix S to be weakly consistent - that is to require (1) and (2) to hold for the middle significant values of the fuzzy numbers.

The evaluations of the objects computed from this fuzzy Saaty's matrix will be fuzzy numbers $\tilde{h}_1, \tilde{h}_2, \ldots, \tilde{h}_n$. For simplicity and easier interpretability we can approximate them by triangular fuzzy

numbers, that is $\tilde{h}_i = (h_{i1}, h_{i2}, h_{i3})$ for all $i = 1, 2, \ldots, n$. The significant values of these triangular fuzzy numbers can then be computed using the formulas (7), (8) and (9) proposed by Krejčí in [4].

$$h_{i1} = \min \left\{ \sqrt[n]{\prod_{j=1}^{n} s_{ij}^*} \Big/ \sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} s_{kj}^*} \; ; \; s_{kj}^* \in \langle s_{kj1}, s_{kj3} \rangle, k = 1, \ldots, n, \; j = 1, \ldots, n, \; s_{kj}^* = \frac{1}{s_{jk}^*} \right\} \quad (7)$$

$$h_{i2} = \sqrt[n]{\prod_{j=1}^{n} s_{ij2}} \Big/ \sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} s_{kj2}} \quad (8)$$

$$h_{i3} = \max \left\{ \sqrt[n]{\prod_{j=1}^{n} s_{ij}^*} \Big/ \sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} s_{kj}^*} \; ; \; s_{kj}^* \in \langle s_{kj1}, s_{kj3} \rangle, k = 1, \ldots, n, \; j = 1, \ldots, n, \; s_{kj}^* = \frac{1}{s_{jk}^*} \right\} \quad (9)$$

This method of computing fuzzy evaluations of objects provides as a result fuzzy numbers, which are less uncertain that the outputs of other approaches available in the literature. The condition $s_{kj}^* = \frac{1}{s_{jk}^*}$ in (7) and (9) ensures, that no unnecessary uncertainty is added by the computation.

4 Evaluation of books at Palacky University

In this section we will show how the fuzzified AHP can be used in the evaluation of R&D outcomes (books) and what benefits there might be in using fuzzy evaluation method for these purposes. A practical example from UP will be presented. We will stress how the outputs of the fuzzified AHP can then be used in a peer-review process to reflect the quality of the current R&D output.

4.1 Definition of the problem

This paper strives to suggest an evaluation methodology for R&D outcomes that would substitute the current assignment of fixed number of points to each book (see [8]) regardless of its quality. We propose a more objective approach to the evaluation. One that assesses the quality of the publisher (represented by its reputation) and combines it with the assessment of the quality of the book itself through peer review process. The former is achieved by categorizing publishers into 4 categories according to their reputation and by expressing preferences between the categories. As the publication process involves a detailed review of the book, this criterion is taken as a basis for the evaluation. Each book is assigned an initial evaluation according to its publisher's reputation. The fact that the scientific quality may vary for books from the same publisher category can subsequently be reflected in the peer review process.

Four categories of publishers are defined according to their reputation (*Category 1* being of highest reputation, ..., *Category 4* having the lowest reputation, but still fulfilling all the requirements for scientific publishers). For each category, a comprehensive description and a few "typical" members are provided for the purpose of mutual comparison of categories. In the case of UP, Faculty of Science, the board of experts responsible for the evaluation of R&D outcomes agreed on initial intervals of scores, that books from each category of publishers might be assigned. These intervals are presented as (10). Although the intervals of scores (10) are a result of consensus of experts, it is not obvious what preference structure underlies them. As such these were not considered satisfactory for the purposes of evaluation, as the evaluation using such intervals is in fact a "black box" impossible to be understood, so the correctness of such approach is difficult to assess.

$$\begin{aligned} &\text{Category 1:} && 50 - 75 \text{ points}, \\ &\text{Category 2:} && 30 - 40 \text{ points}, \\ &\text{Category 3:} && 15 - 20 \text{ points}, \\ &\text{Category 4:} && 5 - 10 \text{ points}. \end{aligned} \quad (10)$$

4.2 Preference structure representation

The consensus expressed by (10) is however a useful input for analysis. To understand this initial idea better we will still require at least confirmation of the results (if not additional information) from group

of experts at some stages. First we can use Saaty's AHP to reconstruct the preference structure and then use it as a starting point for the development of mathematical model. If we take the centers of the intervals as the most typical scores (evaluations) of each category of publishers, calculate the values of the elements of the matrix of relative preferences and round them up to be able to describe them linguistically, we obtain (11). The matrices describe the experts' preferences on the set of categories of publishers thus expressing the desirability of publication in each category of publishers from the point of view of the Faculty.

$$\text{rounded } S = \begin{pmatrix} 1 & 2 & 4 & 8 \\ \frac{1}{2} & 1 & 2 & 5 \\ \frac{1}{4} & \frac{1}{2} & 1 & 2 \\ \frac{1}{8} & \frac{1}{5} & \frac{1}{2} & 1 \end{pmatrix} \rightarrow \text{fuzzified } \tilde{S} = \begin{pmatrix} 1 & (1,2,3) & (3,4,5) & (7,8,9) \\ \left(\frac{1}{3},\frac{1}{2},1\right) & 1 & (1,2,3) & (4,5,6) \\ \left(\frac{1}{5},\frac{1}{4},\frac{1}{3}\right) & \left(\frac{1}{3},\frac{1}{2},1\right) & 1 & (1,2,3) \\ \left(\frac{1}{9},\frac{1}{8},\frac{1}{7}\right) & \left(\frac{1}{6},\frac{1}{5},\frac{1}{4}\right) & \left(\frac{1}{3},\frac{1}{2},1\right) & 1 \end{pmatrix}. \quad (11)$$

Let us note here that in the matrices S and $\tilde{S}$ the columns (and rows) are numbered in accordance with the ordering of the categories, hence ordered according to their preferences already. As we can see from (11) the relative importances of neighboring categories of publishers are the same for all categories and are described by the lowest value associated with preference. These values can be interpreted as "one book in the higher category can be compensated by two books from the neighboring lower category". The experts were provided this interpretation and subsequently were asked to input the pairwise comparison of the neighboring categories using the linguistic terms from Table 2 (we assisted them in completing the matrix to remain weakly consistent and to achieve CR as low as possible). This way we have obtained (12) which shows that the neighboring categories were in the eyes of the experts "further from each other", than was described by (10). The interpretation that "one book from higher category can be compensated by 3 books from the neighboring lower category" was closer to the intention of the expert evaluators (as was confirmed during the discussion with the evaluators concerning the results).

$$\begin{pmatrix} 1 & (1,3,5) & ? & ? \\ \left(\frac{1}{5},\frac{1}{3},1\right) & 1 & (1,3,5) & ? \\ ? & \left(\frac{1}{5},\frac{1}{3},1\right) & 1 & (1,3,5) \\ ? & ? & \left(\frac{1}{5},\frac{1}{3},1\right) & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & (1,3,5) & (3,5,7) & (7,9,9) \\ \left(\frac{1}{5},\frac{1}{3},1\right) & 1 & (1,3,5) & (3,5,7) \\ \left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right) & \left(\frac{1}{5},\frac{1}{3},1\right) & 1 & (1,3,5) \\ \left(\frac{1}{9},\frac{1}{9},\frac{1}{7}\right) & \left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right) & \left(\frac{1}{5},\frac{1}{3},1\right) & 1 \end{pmatrix} \quad (12)$$

4.3 Resulting evaluation according to the reputation of publisher - first stage evaluation

Both matrices in (11) are weakly consistent and the Saaty's inconsistency ratio $CR = 0.0023$. We can easily see that by changing s_{24} from 5 to 4, we would obtain an absolutely consistent matrix in Saaty's sense. The consistency ratio for (12), that is in the most important aspects similar to (11), but differs by the use of linguistic labels in the input phase, is 0.0286. If we compute the evaluations using the fuzzified AHP (see Section 3), normalize the evaluations and multiply them by 100 to avoid decimals, we get:

$$\begin{array}{lllll} & \tilde{h}_1 = (41,53,61) & & \tilde{h}_1 = (41,53,62) & \tilde{h}_1 = (38,58,69) \\ \text{from (11)}: & \tilde{h}_2 = (19,28,40) & \text{from (11)}' & \tilde{h}_2 = (18,27,39) & \text{from (12)}: \quad \tilde{h}_2 = (14,25,45) \\ & \tilde{h}_3 = (9,13,20) & \text{abs. consistent:} & \tilde{h}_3 = (9,13,21) & \tilde{h}_3 = (6,11,23) \\ & \tilde{h}_4 = (5,6,9)\,. & & \tilde{h}_4 = (5,7,10) & \tilde{h}_4 = (4,5,10)\,. \end{array}$$

Clearly the evaluations of categories do not differ very much between the slightly inconsistent and absolutely consistent version of (11). As the evaluators agreed that (12) captures their preferences best, we will restrict further interpretation to (12). The initial evaluation of each book (according to the category of its publisher) is now available as a triangular fuzzy number - $\tilde{h}_1, \tilde{h}_2, \tilde{h}_3$ or $\tilde{h}_4$.

4.4 Peer review integration - second stage of the evaluation process

For the purposes of the peer-review process, the support of each of the evaluations of categories of publishers is interpreted as an interval, from which scores for books published by a publisher from the respective category can be chosen. The value corresponding to the element in the kernel of the fuzzy number is then seen as a "default" evaluation - that is evaluation of a book typical as for scientific quality for the publishers from the given category. The peer review process following the first stage is intended

to expertly assess the quality of each book. If during the peer review the book is assessed as better or worse than "typical" for the current category of publishers, the score can be adjusted accordingly within the predefined interval. This way quality of the R&D outcome itself can be reflected.

5 Conclusions

We have presented a method for R&D outputs evaluation, that combines expert assessment of the reputation of the publisher (by expressing preferences between categories of publishers) with the peer review process for quality assessment of the actual R&D output. The fuzzified AHP is used to determine the evaluations based on the reputation of publishers. The quality of the outcome is subsequently assessed by a group of experts. The model is being used at the Faculty of Science of UP. As the described method integrates the evaluation of the medium with the evaluation of the outcome itself, it may also serve as inspiration for the further development of the R&D evaluation methodology in the Czech Republic.

Acknowledgements

This research was supported by the grant PrF_2013_013 Mathematical models of the Internal Grant Agency of Palacky University in Olomouc.

References

[1] Arnold, E.: International Audit of Research, Development & Innovation in the Czech Republic, Final report, Synthesis report, 2011, [online], available from: http://audit-vav.reformy-msmt.cz/soubory-ke-stazeni/zaverecna-zprava-z-auditu-vaval/, [cited. 2013-05-03].

[2] Brunelli, M., Canal, L., Fedrizzi, M.: Inconsistency indices for pairwise comparison matrices: a numerical study. *Annals of Operations Research*, 2011.

[3] Fedrizzi, M., Marques Pereina, R. A.: Positive fuzzy matrices, dominant eigenvalues and an extension of Saaty's analytical hierarchy process, *Proceedings of IFSA World Congress, Sao Paulo*, Vol. II, Brazil, 1995, 245-247.

[4] Krejčí, J.: *Fuzzy rozšíření Saatyho AHP*. [Fuzzy extension of Saatys AHP] (in Czech), unpublished masters thesis, Palacky University in Olomouc, 2012.

[5] Krejčí, J., Jandová, V., Stoklasa, J., & Talašová, J.: *Bodové hodnocení knih.* [evaluation of monographs] (in Czech), research report, Palacky University in Olomouc, Olomouc, 2012.

[6] Ohnishi, S., Yamanoi, T., Imai, H.: A weights representation for fuzzy constraint-based AHP, *Proceeding of the Conference IPMU 2008*, 261-267.

[7] Ramík, J.: *Vícekriteriální rozhodování - analytický hierarchický proces (AHP)*. Slezská univerzita v Opavě, Karviná, 1999.

[8] RVVI, Metodika hodnocení výsledků výzkumných organizací a hodnocení ukončených programů (platná pro léta 2010, 2011 a rok 2012) [online], available from: http://www.vyzkum.cz/FrontClanek.aspx?idsekce=650022, [cited 2013-05-01].

[9] Saaty, T.L.: Relative Measurement and Its Generalization in Decision Making, Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors - The Analytic Hierarchy/Network Process, *RACSAM*, Vol. 102, No. 2, 2008, 251 - 318.

[10] Saaty, T.L.: *The Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*. Vol. VI of the AHP Series, RWS Publ., 2000.

[11] Stoklasa, J., Jandová, V. & Talašová, J.: Weak consistency in Saatys AHP - evaluating creative work outcomes of Czech Art Colleges, *Neural network world*, **23** (2013), 61 - 77.

[12] Talašová, J. & Stoklasa, J.: A model for evaluating creative work outcomes at Czech Art Colleges, *Proceedings of the* 29^{th} *International Conference on Mathematical Methods in Economics 2011 - Part II*, Praha: VŠE Praha, 2011, 698 - 703.