UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

a

LUT UNIVERSITY
LUT SCHOOL OF ENGINEERING SCIENCE

DISERTAČNÍ PRÁCE

Jazyková aproximace výstupů fuzzy modelů

Katedra matematické analýzy a aplikací matematiky
Školitelé: **doc. RNDr. Jana Talašová, CSc.**
prof. Mikael Collan, Ph.D.
prof. Pasi Luukka, Ph.D.
Vypracoval: **Mgr. Tomáš Talášek**
Studijní program: P1104 Aplikovaná matematika
Studijní obor: 1103V004 Aplikovaná matematika
Forma studia: Prezenční
Rok odevzdání: 2019

PALACKÝ UNIVERSITY OLOMOUC
FACULTY OF SCIENCE

and

LUT UNIVERSITY
LUT SCHOOL OF ENGINEERING SCIENCE

DISSERTATION THESIS

The linguistic approximation of fuzzy models outputs

Palacký University Olomouc

LUT University

Department of Mathematical Analysis and Applications of Mathematics
Supervisors: **doc. RNDr. Jana Talašová, CSc.**
prof. Mikael Collan, Ph.D.
prof. Pasi Luukka, Ph.D.
Author: **Mgr. Tomáš Talášek**
Study program: P1104 Aplikovaná matematika
Field of study: 1103V004 Aplikovaná matematika
Form of study: Full-time
Year of submission: 2019

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Mgr. Tomáš Talášek

Název práce: Jazyková aproximace výstupů fuzzy modelů

Typ práce: Disertační práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

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

Rok obhajoby práce: 2019

Abstrakt: Jazyková aproximace je proces, kterým se přiřazují jazykové termy matematickým objektům, přičemž tyto objekty jsou často výstupy fuzzy modelů. Na jazykovou aproximaci klademe požadavek, aby jazykové termy dostatečně "reprezentovaly význam" aproximovaných objektů nebo aby alespoň reprezentovaly ty charakteristiky těchto objektů, které jsou pro daný účel klíčové. Výběrem vhodné metody pro jazykovou aproximaci se však odborná literatura téměř nezabývá. V průběhu let bylo představeno několik přístupů k jazykové aproximaci, avšak doposud nebylo provedeno žádné důkladné srovnaní těchto přístupů. Tato disertační práce si klade za cíl napravit tento nedostatek pomocí analytického frameworku, který umožní navrhovatelům (a zároveň i uživatelům) modelů s pomocí vizualizace znázornit možné výsledky jazykové aproximace při použití rozdílných vzdáleností/podobností fuzzy množin, porovnat jejich chování a identifikovat případné nedostatky/problémy zvolených vzdáleností/podoboností. K výběru vhodné vzdálenosti/podobnosti fuzzy množin (pro potřeby jazykové aproximace) tedy v této práci přistupujeme jiným způsobem - pomocí identifikace potenciálních problémů postupně vylučujeme ty vzdálenosti/podobnosti, které se v daném případě nechovají dle potřeby.
Hlavní přínos práce spočívá v návrhu frameworku, který umožňuje analyzovat to, jak volba vzdálenosti/podobnosti fuzzy čísel ovlivňuje výsledek jazykové aproximace. Framework je navržen tak, aby byl snadno použitelný, nekladl na vysoké nároky na znalosti uživatele, umožňoval přímé srovnání vlivu vzdálenosti/podobnosti na výsledek jazykové aproximace a výsledky byly snadno vizualizovatelné. Přestože se práce zaměřuje převážně na trojúhleníková a lichoběžníková fuzzy čísla (symetrická i nesymetrická), je zde představena i modifikace frameworku, která umožňuje analýzu a vizualizaci výsledků jazykové aproximace fuzzy výstupů, které lze obdržet pomocí Mamdaniho fuzzy inference. Na této modifikaci mimo jiné ukazujeme, jak snadno lze framework upravit pro analýzu dalších typů výstupů. Dále je v práci představena nová metoda pro jazykovou aproximaci, která je postavena na myšlence fuzzy 2-

tuples. Tato metoda jazykové aproximace se od jiných metod liší tím, že požaduje jen omezený počet jazykových termů (tj. uživatel modelu nemusí rozumět velkému počtu jazykových termů), ale díky 2-tuples může být i tak výstupem jazykové aproximace nekonečně mnoho jazykových termů, přičemž každý takovýto term se skládá z jednoho ze zvolených jazykových termů a také z informace o jeho “posunu”. Tím je zajištěna snadnost porozumění výsledku jazykové aproximace.
Disertační práce dále obsahuje 11 publikací, na kterých se Tomáš Talášek významně spolupodílel. Tyto publikace shrnují dosažené matematické výsledky autora a umožňují detailnější náhled na zkoumání toho, jak volba vzdálenosti/podobnosti fuzzy množin ovlivňuje výsledek jazykové aproximace.

Klíčová slova: jazyková aproximace, fuzzy číslo, vzdálenost, podobnost, jazyková škála, numerické šetření, 2-tuples

Počet stran: 209 včetně příloh a publikací

Počet příloh: 2 přílohy + 11 plných textů článků

Jazyk: anglický

BIBLIOGRAPHICAL IDENTIFICATION

Author: Mgr. Tomáš Talášek

Title: The linguistic approximation of fuzzy models outputs

Type of thesis: Dissertation thesis

Department: Department of Mathematical Analysis and Applications of Mathematics

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

The year of presentation: 2019

Abstract: Linguistic approximation is a process of assigning linguistic labels to various mathematical objects, frequently ones that are obtained as outputs of fuzzy models. Such an assignment cannot be arbitrary – the usual requirement on linguistic approximation is for the linguistic label to "represent the meaning" of the approximated object sufficiently, or at least to reflect its characteristics that are the most important for the given purpose. How to define such a sufficiency, or in other words how to recognize an appropriate method of linguistic approximation, however, remains an unresolved issue. Over the years several various approaches for linguistic approximation was introduced but almost no proper comparison of these approaches was made. This thesis strives to resolve this issue by suggesting a universal analytical framework that helps the designers (and also users) of the models to visualize the performance of linguistic approximation under different distance/similarity measure of fuzzy numbers and to use this visualization to compare their performance and identify the potential drawbacks of using selected distance/similarity measures. It therefore approaches the issue of sufficiency of the linguistic approximation from behind – mainly pointing out the problems and thus ruling out some of the not-well-functioning distance/similarity measures.
The contribution of the thesis lies in the proposal of a framework for the analysis of performance of different distance/similarity measures of fuzzy numbers such that its use is straightforward, it requires only limited knowledge from his potential user, it allows for a direct comparison of the performance of different distance/similarity measures in the given context and it provides results by means of graphical representation. Although most of the thesis focuses on the frequently used shapes of fuzzy numbers (triangular, trapezoidal; both symmetrical and asymmetrical), we also propose a modification of the framework which allows for the analysis and visualization of results for Mamdani-type fuzzy sets (outputs of Mamdani fuzzy inference). On

this modification we also show the simplicity of the generalization of the framework for different conditions and contexts (represented by different approximated objects etc.). Another contribution is the proposal of a novel linguistic approximation method based on the idea of fuzzy 2-tuples in the thesis. This method differs from other methods in a way that it requires only small number of linguistic terms (i.e. the decision-makers's vocabulary for the description of the results can remain reasonably limited, hence the requirements on understanding the meaning of the linguistic values can be kept to a reasonable minimum), but thanks to the 2-tuple concept it can results in an infinite number of linguistic labels composed of one of the known linguistic terms and a description of its "shift" which facilitates easy understandability of the results of the linguistic approximation.
Eleven publications on which the author participated are appended to the thesis. These publications summarize the mathematical results and provide closer insights into the issue of the investigation of the performance of linguistic approximation under different distance/similarity measures.

Key words: linguistic approximation, fuzzy number, distance, similarity, linguistic scale, numerical investigation, 2-tuples

Number of pages: 209 with appendices and publications

Number of appendices: 2 appendices + 11 full-text papers

Language: English

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. a prof. Pasiho Luukky, Ph.D. v rámci double degree smlouvy o doktorském studiu mezi Univerzitou Palackého v Olomouci, Česká republika a LUT University, Finsko. Prohlašuji, že všechny použité zdroje jsem řádně citoval a uvedl v seznamu literatury.

V Olomouci, 15. května 2019

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 prof. Pasi Luukka, Ph.D. under a doctoral double degree agreement between Palacký University, Olomouc, Czech Republic and LUT University of Technology, Finland. The literature used is listed in the list of references and duly cited in the text.

Olomouc, May 15, 2019

Signature: ....

Preface

The following text summarizes my research on linguistic approximation during the last seven years. I am glad that I was offered the possibility to study at two universities – LUT university in Lappeenranta and Palacký University in Olomouc.

I would like to express my sincere gratitude to my supervisors Jana Talašová, Mikael Collan and Pasi Luukka for their support and mentoring during my study - I know that it was not always easy with me. A big thanks to Jan stoklasa - my dear colleague and friend for his help and valuable constructive feedback during my study. I hope that I will be able to repay you in the future.

Finally, I would like to thank to my family and friends for their support and understanding during my university studies.

Olomouc, April 2019
Tomáš Talášek

Acknowledgements

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

- Grants PrF_2013_013, PrF_2014_028, PrF_2015_014, PrF_2016_025, FF_2015_014, FF_2016_007, FF_2017_011, IGA_2018_002, FF_2019_002 of the Internal Grant Agency of Palacký University Olomouc.
- Grant 313396 MFG40 - Manufacturing 4.0 of the Finnish Strategic Research Council.
- Grant GA 14 - 02424S of the Grant Agency of the Czech Republic.
- Education 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 LUT university.

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

Contents

List of publications

Publication I
Talášek, T. and Stoklasa, J., A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation under Different Linguistic Scales. *Journal of Multiple-Valued Logic and Soft Computing*, 29(5), 485–503, 2017.

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

Publication III
Talášek, T., Stoklasa, J., Collan, M. and Luukka, P., Ordering of fuzzy numbers through linguistic approximation based on Bonissone's two step method. 16^{th} *IEEE International Symposium on Computational Intelligence and Informatics 2015*, 285–290, 2015.

Publication IV
Talášek, T., Stoklasa, J. and Talašová, J., Linguistic approximation using fuzzy 2-tuples in investment decision making. *Proceedings of the* 33^{rd} *International Conference on Mathematical Methods in Economics 2015,* 817–822, 2015.

Publication V
Talášek, T. and Stoklasa, J., The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers. *Proceedings of the international scientific conference Knowledge for Market Use 2016,* 539–546, 2016.

Publication VI
Talášek, T., Stoklasa, J. and Talašová, J., The role of distance and similarity in Bonissone's linguistic approximation method – a numerical study. *Proceedings of the* 34^{th} *International Conference on Mathematical Methods in Economics 2016,* 845–850, 2016.

Publication VII
Stoklasa, J. and Talášek, T., Linguistic approximation of values close to the gain/loss threshold. *Proceedings of the* 35^{th} *International Conference on Mathematical Methods in Economics 2017,* 726–731, 2017.

Publication VIII

Talášek, T. and Stoklasa, J., Distance-based linguistic approximation methods: graphical analysis and numerical experiments. *Proceedings of the* 35^{th} *International Conference on Mathematical Methods in Economics 2017*, 777–782, 2017.

Publication IX

Talášek, T. and Stoklasa, J., Selection of tools for managerial decision support – the identification of methods of choice in linguistic approximation. *Proceedings of the international scientific conference Knowledge for Market Use 2017*, 606–613, 2017.

Publication X

Talášek, T. and Stoklasa, J., Ordering of fuzzy quantities with respect to a fuzzy benchmark – how the shape of the fuzzy benchmark and the choice of distance/similarity affect the ordering. *Proceedings of the* 36^{th} *International Conference on Mathematical Methods in Economics 2018*, 573–578 , 2018.

Publication XI

Talášek, T. and Stoklasa, J., Three-dimensional histogram visualization of the performance of linguistic approximation of asymmetrical triangular fuzzy numbers. *Proceedings of the international scientific conference Knowledge for Market Use 2018*, 445–451, 2018.

The thesis represents a text unifying the contributions presented in the separate publications listed above. The publications are listed chronologically starting with journal papers and followed by conference proceedings papers. These publications by the author of the thesis are reffered to by Roman numbers in the text.

The Publication **I** summarizes the analytical methodology proposed by the author for the assessment of performance of different distance/similarity measures in the linguistic approximation. Tomáš Talášek is the main and corresponding author of the paper; he proposed the analytical framework and carried out all the calculations.

On the other hand Publication **II** summarizes the general ideas that constituted the motivation for the actual introduction of the analytical methodology proposed in Publication **I**. The paper discusses possible uses of linguistic fuzzy models and their outputs in social sciences and takes the perspective of a layman user. As a co-author Tomáš Talášek provided the mathematical perspective to the paper and participated in the literature review of the use of fuzzy sets in social sciences.

The Publication **III** presents an attempt to suggest the use of linguistic approximation for a different purpose than just “retranslation”. Tomáš Talášek as the corresponding author proposes an algorithm based on Bonissone’s two-step linguistic approximation method for finding a partial ordering of fuzzy numbers. He also carried out the necessary calculations in Matlab. In Publication **VI** the idea of a analytical framework utilizing graphical outputs to be used by an unexperienced users was born and discussed on the case of Bonissone’s two-step linguistic approximation method. Tomáš Talášek was again the main author of the paper, wrote most of the text and generated the graphical outputs. The analytical framework was further structured and introduced in proper mathematical terms in Publications **I** and **V**. Publication **V** generalizes the framework to asymmetrical fuzzy numbers and introduces a numerical experiment. Tomáš Talášek was the main author and carried out the calculations and simulations. The issues potentially complicating the use of graphical summaries introduced by the asymmetry of approximated fuzzy numbers are addressed in Publication **XI** by the proposal of the three dimensional histogram visualization by Tomáš Talášek who is also the main author of the paper. The notion of a fuzzy ideal is considered **IX** and the proposed analytical framework with its graphical outputs is applied to gain insights into the performance of a specific distance of fuzzy numbers in the context of optimization with fuzzy goals. Tomáš Talášek is the main author responsible for the calculations and for the final form of the paper. A similar idea is discussed in **X** where the proposed analytical framework is used to investigate the ability of selected distance/similarity measures to define the order of fuzzy objects with respect to a fuzzy benchmark. Tomáš Talášek is again the main author who carried out the necessary calculations.

A further step in direction of a general analytical framework was proposed in Publication **VIII** where Mamdani-type outputs were considered as the approximated objects. Tomáš Talášek, as the main author, suggested the necessary parametrization of the Mamdani-type outputs and the modifications of the currently available analytical framework to reach a desired graphical outputs.

A thorough literature review of the available methods of linguistic approximation (its summary is presented in section 3) resulted in the suggestion of a novel linguistic approximation method based on the idea of fuzzy 2-tuples [12], that was proposed in Publication **IV**. This method extends the set of available outputs of linguistic approximation to an infinite one. At the same time it maintains the understandability focusing primarily on the elementary linguistic terms. A more detailed summary of the method is available in section 5. Tomáš Talášek is the main author of the paper proposed the idea of the method and participated on the introduction of the necessary mathematical notation.

The Publication **VII** applies the graphical summaries generated by the analytical framework to the specific context of linguistic approximation of quantities that can be framed as gains or losses. The issue of misslabeling a gain by a “loss” label and vice versa is discussed in details. Tomáš Talášek is a co-author of the paper carried out all the calculations and generated the graphical outputs.

1 Introduction

This thesis deals with the analysis of the effect of different distance and similarity measures of fuzzy numbers on the results of a classification model. There is an abudance of distance/similarity measures of fuzzy numbers to be used for various purposes. Even though their mathematical properties are being studies to some extent in the scientific literature (see e.g. [44, 32, 4]) the actual choice of a distance or a similarity measure for a model for a given purpose remains frequently arbitrary. Most of the theoretical papers suggest "the most appropriate" distance or similarity measure to be used. Unfortunately, the criteria for appropriateness are not specified and the actual knowledge of the differences in performance of the models under different distance/similarity measures is unavailable. In fact, the very methods for the analysis of the effect of specific choices of distance/similarity measures on the results of the models are missing. The thesis therefore aims on suggesting a general framework for the comparison of the performance of a specific model under different distance/similarity measures. The goal is to provide the makers of the models with insights regarding the choice of the distance/similarity measure. The thesis strives to provide graphical outputs that allow for easy comparability of the performance of these measures and that do not assume that the correct answers (e.g. in classification or linguistic approximation) are necessarily available. This framework should at least allow for the identification of differences in performance of the distance/similarity measures and of their similarities as well.

Due to the extensiveness of the selected goal in the further text we will focus only on the effect of the selection of the distance/similarity measures on the results of linguistic approximation of the given fuzzy numbers (representing the outputs of mathematical model). This does not result in a loss of generality, because linguistic approximation is a representative example of a classification model - the goal is to assign a linguistic term (label or class) from a predefined set of linguistic terms (labels or classes) that describes best the approximated output of a model (usually in a form of fuzzy set or fuzzy number). The restriction to linguistic approximation is driven by several motives:

- To deal with a clear example of a classification task. This way a clear parallel to general classification task can be easily established.
- To properly explain the application of the presented framework and to show what kind of insights can be obtained, it is suitable to demonstrate its usage on a conceivable example. Linguistic approximation provides such an example.
- Linguistic approximation does not pose any default requirements on the distance/similarity measure used for the the determination of the most appropriate linguistic label for the given fuzzy number. This allows us to apply (and thus investigate) any distance/similarity measure of fuzzy numbers in this framework.

- The performance of linguistic approximation can be easily parametrized by the adjustement of linguistic scale. This enables us to study distance/similarity measures in a controlled environment.

Linguistic approximation can, in essence, be applied to any mathematical object. Real numbers are not considered in this thesis since their one dimensional representation is straightforward and the graphical summaries would not have much additional information value. Instead we focus on the approximation of fuzzy numbers whose representation requires more than one parameter. This opens a way to multidimensional graphical representation of the outputs. Even though most of the results presented in this thesis are in 2D, generalization to more dimensions (i.e. to more parametres representing the approximated objects) is straightforward. With more parameters we however encounter the limits of convenient graphical representation.

Since there is an large number of different distance/similarity measures it is not possible to study all within a single thesis. Instead we propsed a general framework applicable to any distance/similarity measure and select examples of the most frequently used distance/similarity measures to be able to show the performance of the suggested framework. The selection of frequently used measures allows us also to draw conclusions on their (un)suitability for linguistic approximation in the given analysed settings. This adds another application contribution to the thesis.

To clarify the inteded contribution of this thesis and to explain its structure we set the following subgoals:

- To investigate the performance of the chosen distance/similarity measures in the linguistic approximation (different approximated objects, different linguistic scales). So far, the choice of the distance or similarity measure was left entirely with the creator of the model - no guideance for the choice exists; graphical representations/summaries of the performance of distance/similarity measures are not being used so far.

- Propose easy-to-understand and easy-to-use method for graphical comparison of the performance of linguistic approximation applying different distance/similarity measures

- Identify potential drawbacks of chosen distance/similarity measures in the context of linguistic approximation and their possible strange/unexpeted behaviour.

- To proposes a new method for linguistic approximation and show the adaptability of the developed analytical framework to new linguistic approximation methods on its example. The new method will provide not only a resulting linguistic term for the approximated fuzzy number but also a supplementary information describing its deviation from the approximating fuzzy number (in terms of meaning).

Since the number of pages is limited some results will be available in the appendices without detailed descriptions. Their understanding will be analogous to the results presented directly in the body of the thesis.

2 Preliminaries

In this section the mathematical preliminaries used in the thesis are be presented to unify the notation. First, the main concepts such as fuzzy sets, their properties and basic operations with them are be defined. The definitions are be based on the notation from L. A. Zadeh's paper [38] where the key contepts of fuzzy sets were formulated. Second, fuzzy numbers as the key concept representing the meanings of linguistic terms throught out the thesis are defined. Finally a chapter introducing the key concept of linguistic fuzzy modeling i.e. linguistic variables, linguistic scale and linguistic approximation follows. For more details on fuzzy sets please see for example [17, 8].

2.1 Basic notions

Let U be a nonempty set (*the universe of discourse*). A *fuzzy set* A on the universe U is defined by the mapping $A : U \to [0,1]$. A family of all fuzzy sets on U is denoted by $\mathcal{F}(U)$. 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 a *membership function* of the fuzzy set A.

Let A and B be fuzzy sets on the same universe U. The set $\text{Ker}(A) = \{x \in U | A(x) = 1\}$ represents the *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an *α-cut* of A for any $\alpha \in [0,1]$, $\text{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A. $\text{Hgt}(A) = \sup\{A(x) | x \in U\}$ denotes a height of fuzzy set. Fuzzy set A is called *normal* if $Hgt(A) = 1$, otherwise it is called *subnormal*.

We say that A is a *fuzzy subset* of B ($A \subseteq B$), if $A(x) \leq B(x)$ for all $x \in U$. A *union* of two fuzzy sets A and B on U is a fuzzy set $(A \cup B)$ on U defined as follows: $(A \cup B)(x) = \max\{A(x), B(x)\}$ and a *Łukasiewicz union* of two fuzzy sets A and B on U is a fuzzy set $(A \cup_L B)$ on U defined as follows: $(A \cup_L B)(x) = \min\{1, A(x) + B(x)\}$, $\forall x \in U$. A *intersection* of two fuzzy sets A and B on U is a fuzzy set $(A \cap B)$ on U defined as follows: $(A \cap B)(x) = \min\{A(x), B(x)\}$ and a *Łukasiewicz intersection* of two fuzzy sets A and B on U is a fuzzy set $(A \cap_L B)$ on U defined as follows: $(A \cap_L B)(x) = \max\{0, A(x) + B(x) - 1\}$, $\forall x \in U$.

Let $A_1, \ldots, A_n$ be fuzzy sets on $U_1, \ldots, U_n$ respectively. The *Cartesian product* of $A_1, \ldots, A_n$ is a fuzzy set $(A_1 \times \cdots \times A_n)$ on $U_1 \times \cdots \times U_n$ with membership function $(A_1 \times \cdots \times A_n)(x_1, \ldots, x_n) = \min\{A_1(x_1), \ldots, A_n(x_n)\}$, $\forall x_i \in U_i$, $i = 1, \ldots, n$. A fuzzy set R on $U_1 \times \cdots \times U_n$ is called an n-ary fuzzy relation. Let R be a fuzzy relation on $U \times V$ and S be a fuzzy relation on $V \times W$. The composition $(R \circ S)$ is a fuzzy relation on $U \times W$ a with membership function $(R \circ S)(x, z) = \sup_{y \in V} \min\{R(x, y), S(y, z)\}$, $\forall x \in U, z \in W$.

A *fuzzy number* is a fuzzy set A defined on the set of real numbers which satisfies the following conditions:

1. $\mathrm{Ker}(A) \neq \emptyset$ (A is *normal*);
2. A_α are closed intervals for all $\alpha \in (0,1]$ (this implies that A is *unimodal*)
3. $\mathrm{Supp}(A)$ is bounded.

A family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on $[a,b] \subset \mathbb{R}$, if $\mathrm{Supp}(A)$ is a subset of the interval $[a,b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1,a_4] = \mathrm{Cl}(\mathrm{Supp}(A))$ and $[a_2,a_3] = \mathrm{Ker}(A)$, where $\mathrm{Cl}(\mathrm{Supp}(A))$ denotes a closure of $\mathrm{Supp}(A)$.

Each fuzzy number A is determined by $\left\{[\underline{a}(\alpha), \overline{a}(\alpha)]\right\}_{\alpha\in[0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall\alpha \in (0,1]$, and the closure of the support of A $\mathrm{Cl}(\mathrm{Supp}(A)) = [\underline{a}(0), \overline{a}(0)]$. The *length* of the support of a fuzzy number A, $L(\mathrm{Supp}(A))$ can now be calculated as $\mathrm{L}(\mathrm{Supp}(A)) = \overline{a}(0) - \underline{a}(0)$.

The fuzzy number A such that $a_1 \neq a_4$ is called *linear* if its membership function is linear on $[a_1,a_2]$ if $a_1 \neq a_2$ and on $[a_3,a_4]$ if $a_3 \neq a_4$; for such fuzzy numbers we will use a simplified notation $A \sim (a_1,a_2,a_3,a_4)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by an ordered triplet $A \sim (a_1,a_2,a_4)$. Triangular fuzzy number $A \sim (a_1,a_2,a_4)$ is called *symmetrical triangular fuzzy number* if $a_2 - a_1 = a_4 - a_2$. If $A \in \mathcal{F}_N(U)$ is a linear fuzzy number and c is a real number, then $A + c = (a_1 + c, a_2 + c, a_3 + c, a_4 + c)$.

The *cardinality* of a fuzzy number A on $[a,b]$ is a real number $\mathrm{Card}(A)$ defined as follows: $\mathrm{Card}(A) = \int_a^b A(x)dx$. Let A be a fuzzy number on $[a,b]$ for which $a_1 \neq a_4$. The *center of gravity* of A is defined by the formula $\mathrm{COG}(A) = \int_a^b xA(x)dx/\mathrm{Card}(A)$. If $A = (a_1,a_2,a_4)$ is symmetrical triangular fuzzy number on $[a,b]$, then $\mathrm{COG}(A) = a_2$ (note that $\mathrm{Ker}(a) = \{a_2\}$).

2.2 Linguistic approximation

Mathematical models nowadays are capable of providing a wide variety of outputs ranging from numbers, intervals, functions to complex outputs represented in matrix or graphical forms etc. The complexity of outputs can reflect the complexity of the modeled system as well as the requirements of the user of the model. Not all types of outputs that are currently available are however intuitive to the users of the models. Consider standard (non-technical) education where numbers, intervals and functions are the most frequently used mathematical objects. If a user is not familiar with more complex mathematical entities he/she might not be able to interpret and use them correctly. One way to solve this issue is to resign on complex outputs and provide only such outputs that are understandable for their user. Another approach, the one adopted in this thesis is to assist the user of the model in his/her understanding of the more complex mathematical entities. As natural language is the most common means of communication it seems only reasonable to provide this assistance by "translating" the mathematical objects into common language. The use of natural language also allows to stress important aspects of the obtained solution or to dampen the less important ones. It can even allow to add a desired "spin" to the presented information [37].

Obviously such a translation can not constitute a one to one mapping. Even though there might be slight differences in meaning between the mathematical outputs and their natural language translations, the benefits of such a translation might outweight the risk stemming from slight alterations of meaning. Formally speaking the process of assigning linguistic labels (words in common language) to various mathematical objects is called *linguistic approximation*. If performed correctly this process enables the use of advanced mathematical models with possibly complex outputs even to unexperienced users (non-mathematicians). This thesis focuses on providing guidance for this process, more specifically it aims on answering the question of which methods should (or should not) be used for linguistic approximation to achieve the desired effect. Note, that not only understanding of the outputs, but also stressing some of their aspects as well as raising attention etc. might be the possible goals of this process.

The process of linguistic approximation was proposed by L.A. Zadeh in 1975 [40, 41, 39]. Its key concept, as defined by Zadeh, is the *linguistic variable*. A *linguistic variable* is a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is the set of its linguistic values (terms), X is an universe on which the meanings of the linguistic values are defined, G is a syntactic rule for generating the linguistic values of the variable $\mathcal{V}$. M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X.

The values of the linguistic variable can now represent the possible translations that we would like to obtain in the process of linguistic approximation. We assume that the meanings of the linguistic terms are understandable to the user of the mathematical model and as such provide usefull replacements for the mathematical entities provided by the model. We just need to be able to select the most appropriate linguistic value of the linguistic variable to provide to the user of the model. Note that not only are the meanings of the values of the linguistic variable assumed to be understood by the user, their mathematical meaning obtained by the function M are also available as mathematical entities. Surprisingly enough even though linguistic approximation methods started to be proposed already at the end of 1970s and the beginning of 1980s (see [40, 10, 35, 2] etc.) no single method seems to dominate this area. In fact, the problem of linguistic approximation is considered unsolved sufficiently even in 2006 [22] where it seems to reemerge in the context of computing with perceptions under the slightly more general label of "retranslation". Also Yager in 2004 [37] points out the unavailability of criteria to asses the most appropriate retranslation in computing with words.

Let *Out* be a output of mathematical model that needs to be assigned a linguistic label and $(\mathcal{V}, \mathcal{T}(\mathcal{V}), [a,b], G, M)$ be a linguistic variable such that $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$. The *linguistic approximation* of the output Out is the process of searching for a suitable linguistic term $\mathcal{T}_{Out}$ from $\mathcal{T}(\mathcal{V})$ which describes the meaning of the output Out the best. One of the most popular approaches to finding the linguistic term $\mathcal{T}_{Out}$ is using the "best-fit" approach:

$$T_{Out} = \arg \min_{i \in \{1,\ldots,s\}} d(T_i, Out), \tag{2.1}$$

where T_i is the meaning of the linguistic term $\mathcal{T}_i, i = 1, \ldots, s$. To be able to define a suitable distance we need the meanings of the linguistic terms to be represented by mathematical entities of the same type as Out. For the purposes of this thesis we assume Out and T_i to be fuzzy sets (fuzzy numbers) on the same universe. In this case d in the previous formula represents a general distance[1] of fuzzy sets (fuzzy numbers). In the family of "best-fit" linguistic approximation methods we are looking for the linguistic term the meaning of which is the closest (most similar) to the approximated mathematical entity (i.e. which mathematical object representing the meaning of some linguistic term from $\mathcal{T}(\mathcal{V})$ "fits best" the (features of) the approximated mathematical entity). Tah et al. [30] suggest the use of Euclidean distance to find the best-fit whereas other authors suggest [44, 7] a range of possible distances and similarities of fuzzy numbers for this purpose.

So far, the only requirement imposed on the linguistic variable used for the purposes of linguistic approximation is that the meaning of its linguistic terms is represented by fuzzy numbers. In this thesis, we however restrict ourself to special types of linguistic variables called *linguistic scale* and *enhanced linguistic scale*. These restrictions enable us to analyse the behaviour of linguistic approximation under different distance (or similarity) of fuzzy numbers. This does not restrict the use of the proposed methods for the analysis of the performance of different distance/similarity measures of fuzzy numbers in the linguistic approximation.

A *fuzzy scale* on $[a,b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on $[a,b]$, that form a Ruspini fuzzy partition (see [27]) of the interval $[a,b]$, i.e. for all $x \in [a,b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. Linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ is called a *linguistic scale* on $[a,b]$ if $X = [a,b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ and $T_i = M(\mathcal{T}_i)$, $i = 1, \ldots, n$ form a fuzzy scale on $[a,b]$. Fuzzy scale is called *uniform* when $\mathrm{L}(\mathrm{Supp}(T_i)) = 2 \cdot (b-a)/(n-1)$ for all $i = 2, ..., n-1, \mathrm{L}(\mathrm{Supp}(T_i)) = (b-a)/(n-1)$ for $i = 1$ and $i = n$, T_i form a Ruspini fuzzy partition of U, and $T_2, \ldots, T_{n-1}$ are symmetrical triangular fuzzy numbers.

Linguistic terms $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ of linguistic scale $\mathcal{T}(\mathcal{V})$ are called *elementary (level 1) terms* of linguistic scale. Linguistic variable that we obtain from a linguistic scale $\mathcal{T}(\mathcal{V})$ by extending its linguistic-term set by additional linguistic terms $\mathcal{T}_i$ to $\mathcal{T}_j$ where $i = 1, \ldots, n-1$, $j = 2, \ldots, n$ and $i < j$ (called *derived linguistic terms*) is called *enhanced linguistic scale*; $M(\mathcal{T}_i \text{ to } \mathcal{T}_j) = T_i \cup_L T_{i+1} \cup_L \cdots \cup_L T_j$. The enhanced linguistic scale thus contains linguistic values of different levels of uncertainty – from

[1]Alternatively a similarity of two fuzzy sets (fuzzy numbers) can be used. In this case, the $\arg\min$ function in formula (2.1) is replaced by $\arg\max$.

the possibly least uncertain elementary terms $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ to the most uncertain linguistic term $\mathcal{T}_1$ to $\mathcal{T}_n$ (uncertainty can be assessed by the cardinality of the meanings of these linguistic terms). Derived linguistic terms $\mathcal{T}_i$ to $\mathcal{T}_j$ are called *level* $j - i + 1$ *terms* and can be also denoted by $\mathcal{T}_{ij}$. Elementary linguistic terms $\mathcal{T}_i$ can be also denoted by $\mathcal{T}_{ii}$ (i.e. $\mathcal{T}_i = \mathcal{T}_{ii}$ to unify the notation).

3 Literature review

In this section we will provide a brief overview of linguistic approximation methods as they were proposed in the scientific literature since the seminal papers on linguistic fuzzy modeling by L. A. Zadeh [40, 41, 39]. Since the rest of this thesis focuses on "best-fit" approaches, i.e. approaches that determine the linguistic approximation based on the closeness (distance) of the approximated object and the fuzzy set meaning of the linguistic terms, we will try to present mainly alternative approaches to the "best-fit" in this chapter.

The first methods to appear in the scientific literature were multistage-ones. Eshragh and Mamdani suggested in [10] to first split the approximated fuzzy set into specific subsets, then to assign linguistic labels to this subsets and finaly to derive the linguistic approximation using connectives (and, or, etc.) and hedges (very, more or less, not). Even though this method was introduced four years after the introduction of linguistic variables it already proposes searching for more then one linguistic approximation and selecting the most simple (easiest to understand) one. More specifically the method approximates the original fuzzy set and it also finds the negation of the linguistic approximation of the negated fuzzy set (i.e. the method finds the linguistic approximation of the fuzzy set $\mathcal{A}$ and its negation $\mathcal{B}$; it suggests the first in the form "it is $\mathcal{A}$" and the negation of the second in the form "it is not $\mathcal{B}$" as possible linguistic approximations and chooses whichever one is easier to understand or less complex). This constitutes a first step to considering the semantic features of linguistic approximation. Similar approach was proposed by Dvořák in [9] in the context of Novak's fuzzy inference system.

One year later Wenstøp in [35] proposed a full auxiliary language to perform quantitative analysis with linguistic values. The basic representations of meaning where unimodal fuzzy sets which were represented by their coordinates in a two-dimensional space; first coordinate representing the low-high dimension (position), the second one representing their imprecision. This approach was capable of approximating multimodal fuzzy sets by splitting them into simpler ones, replacing simpler multimodal fuzzy sets by a unimodal fuzzy set from which a low-uncertain fuzzy sets were excepted to model the "valley" between the two peaks. This way a collection unimodal fuzzy sets was obtained, that could be combined using connectives into the original approximated fuzzy set. The linguistic labels were assigned based on the Euclidean distance of these simple fuzzy sets to the fuzzy sets representing the meaning of 56 available linguistic labels in the two-dimensional position-imprecision space (see Figure 3.1). The author himself however states [35, p. 106] that "...some labels are expressed rather more clumsily than in ordinary language." For example a possible output could look like this: "((not below lower medium) OR possibly unknown) EXCEPT ((medium to rather high) EXCEPT upper medium)". Given the state of computer technology at that time, the results still provide reasonable linguistic approximation of complex fuzzy sets. Modifications of this method were proposed in [28, 36].

Figure 3.1: The 56 linguistic terms and their location in the position and imprecision space. Adapted from [35, p. 105]

In the same year as Wenstøp, Bonnisone proposed in [2] another linguistic approximation method based on feature extraction and pattern recognition techniques. In its first step the method selected a predefined number of linguistic terms from a finite linguistic term set based on their semantic similarity with the approximated fuzzy object. The semantic similarity is assessed based on several key features such as cardinality, location (COG), skewness and fuzziness etc. These features are assigned weights to emphasize the difference in their importance for the given purpose. In the second step the modified Bhattacharyya distance (which takes into account the complete information represented by the membership functions) is applied to select the closest fuzzy set representation of a linguistic term to the approximated fuzzy set. This linguistic term is finally assigned as a result of the linguistic approximation. The preselection by applying weighted Euclidean distance to find the "most appropriate" candidates based on the selected features reduces the computation complexity of the whole method and allows for the application of the more advanced modified Bhattacharyya distance only on the reduced set of candidates for the linguistic approximation. The author revisited the idea of linguistic approximation again in [3].

In 1998 Kowalczyk [18] proposed the question of correct linguistic approximation of subnormal fuzzy sets and suggest to use of subnormal primitive terms. He proposes a methodology for the linguistic approximation of subnormal fuzzy sets but concludes that the linguistic approximation using subnormal meanings of linguistic terms might be difficult to interpret. Kowalczyk also stresses that the issue of selecting the most appropriate distance/similarity measure for linguistic approximation remains an open question. Even after more than twenty years since the introduction of the concept of linguistic approximation this issue remains unresolved until today. This is one of the reasons why this thesis was written.

Zwick et al. in 1987 in [44] and Degani and Bortolan in [7] point out the abundancy of available distance and similarity measures of fuzzy numbers and the lack of guidelines for their appropriate selection in linguistic approximation in various context. Even these authors do not resolve the issue of selection of appropriate distance/similarity measures in linguistic approximation. They provide some insights in the functioning of the distances/similarities, yet no general methodology for the analysis of the performance of various distance/similarity measures is suggested. Marhamati et al. [21] approach the issue of linguistic approximation from a slightly different angle in the context of computing with words. They classify linguistic approximation methods into three categories and assume that the result of linguistic approximation can be a sentence, i.e. modifiers and quantifiers are applied to the atomic terms which is well in line with the ideas of computing with perceptions (see e.g. [42, 43, 22]. Given the possibly vast number of linguistic approximations (increasing with the number of elementary/primary terms, connectives and hedges), Kowalczyk [19] suggested to use a genetic programming to speed-up the process of searching for a fitting linguistic term.

In 2004 Klir wrote a short paper called *Some Issues of Linguistic Approximation* [16] where he stated that the defuzzification of fuzzy numbers (i.e. in the process of assigning a real number value to a fuzzy number) was more extensively studied in the literature than linguistic approximation. He also contemplates about the meaning of the expression good approximation: *'There are of course various views about what the terms "good approximation" and "best approximation" are supposed to means. An epistemological position taken here is that these terms should always be viewed in information-theoretic terms. That is, a good approximation should be one in which the loss of information is small and, similarly, the best approximation (not necessarily uniuqe in this case) should be one of those in which the loss of information is minimal.'*[16, p. 5]

4 Methods for the analysis of linguistic approximation

In this section several methods for the analysis of linguistic approximation are proposed. The methods differ in several aspects:

- type of the aproximated object (triangular or more general fuzzy numbers, fuzzy sets are also considered in several cases) – this influences the number of parameters that need to be reflected in the graphical representation,
- symmetry of the approximated object (most frequently fuzzy number) – this also influences the number of parameters required for unambiguous representation of the output of the analysis; asymmetry introduces overlaps in graphical representation using fever dimensions,
- type of the linguistic scale used to provide linguistic values for the approximation (elementary and enhanced scales) – this affects the number of possible outputs of the linguistic approximation (only linguistic variables with finite lingustic terms set are considered).

Proposed methods are applicable for the analysis of any linguistic approximation method using a finite linguistic terms set. This is not a restrictive requirement since as long as there are finitely many possible linguistic approximations, we can make sure that all of them are properly understood by the user of the outputs. Given the goal of the thesis, examples of the performance of the proposed analytical methods consider distance/similarity based linguistic approximation.

4.1 Representation of the approximated objects

In this thesis we consider mainly triangular fuzzy numbers to be the objects to be approximated – both symmetrical and asymmetrical; trapezoidal fuzzy numbers as well as general Mamdani-type fuzzy set are also briefly discussed. For the purpose of visualization of the results of our analysis, we need to be able to represent the approximated objects by sufficiently low number of characteristics. The higher the number, the more complex (and less understandable) the visualization may become.

As long as symmetrical triangular fuzzy numbers are considered, each can be uniquely represented by a single point in two-dimensional space, e.g. using center of gravity (COG) and length of support of the approximated fuzzy number as coordinates (see e.g. publication **I** or [31]).

If the triangular fuzzy number is asymmetrical, the above suggested representation is no longer unique (the same point in the two-dimensional space can represent various fuzzy numbers). Nevertheless, these fuzzy numbers can be uniquely represented in three-dimensional space, e.g. using their Center of gravity, length of support and the kernel element as coordinates. A similar representation is required for symmetrical trapezoidal fuzzy numbers. Again, three dimensions are needed, e.g. center of gravity, the length of support and the length of kernel as coordinates.

In the more complex case e.g. when asymmetrical trapezoidal fuzzy numbers are taken into account, more characteristics are needed for unambiguous representation: center of gravity (COG), the length of support, the length of kernel and center of support as coordinates, etc.

Obviously a visualization using more than two dimensions is potentially problematic as it requires interactive representation (e.g. the ability of the user to rotate the plots). Therefore, in following text we will use two-dimensional plots wherever possible, providing additional information using other means where needed.

4.2 Approximating linguistic variables selected for the analyses

As mentioned before, the proposed analytical methods can be used with any approximating linguistic variable as long as its terms set is finite. For the purposes of presentation of the performance of the proposed analytical methods, the following linguistic variables on $[0, 1]$ interval are considered:

- **A uniform linguistic scale with five linguistic terms**

  In this case, we will consider a linguistic scale that contains five linguistic terms $\mathcal{T}_1, \ldots, \mathcal{T}_5$. The Meanings of these terms are represented (in respective order) by triangular fuzzy numbers $T_1 = (0, 0, 0.25), T_2 = (0, 0.25, 0.5), T_3 = (0.25, 0.5, 0.75), T_4 = (0.5, 0.75, 1), T_5 = (0.75, 1, 1)$. These fuzzy numbers form a uniform Ruspini fuzzy partition of interval $[0, 1]$ and are depicted (together with their respective linguistic terms) on Figure 4.1.

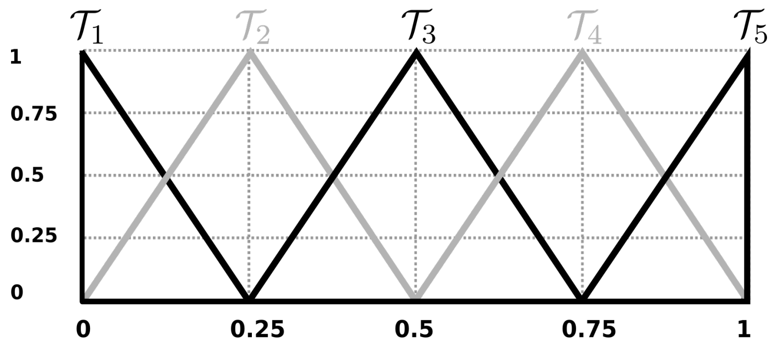


Figure 4.1: Fuzzy set meanings of the elementary linguistic terms of the 5-term uniform linguistic scale.

- **Enhanced linguistic scale derived from the uniform linguistic scale with five terms**

  This linguistic variable will contain all five elementary linguistic terms from the previous case and also derived linguistic terms "$\mathcal{T}_i$ to $\mathcal{T}_j$" denoted $\mathcal{T}_{ij}$, where $i = 1, \ldots, 4$, $j = 2, \ldots, 5$ and $i < j$ (see Section 2 for the approach how to obtain derived linguistic terms and their meaning).

For the simplicity of graphical representation each linguistic term will be assigned a specific colour further in the text. Bright red will be reserved for undecided cases where more than one lingustic term is suggested by the linguistic approximation. The generalization of the results for different universes, different numbers of elementary terms, non uniform scales, etc. is straightforward. The thesis is restricted to five elementary terms to maintain graphical representation clear enough.

4.3 Studied distance and similarity measures

It was already specified that the focus of this thesis is mainly on distance/similarity based linguistic approximation. The reason for this is that distance/similarity based linguistic approximation methods often require finite and previously known sets of linguistic terms. In the following sections we will suggest analytical methods for the assessment of linguistic approximation applied to various approximated objects. In line with the available literature [7, 44] a distance is supposed to be minimized while similarity is supposed to be maximized to obtain the best linguistic approximation.

To enhance the practical relevance of the thesis eight frequently used or investigated distance/similarity measures of fuzzy numbers have been chosen. The proposed analytical methods will be applied to all of them. This will allow us to not only clearly see the benefits (and possible limitations) of the proposed analytic methods, but also to draw conclusions concerning the usefulness of the selected distance/similarity measures in Linguistic approximation in the given setting. This, to my best knowledge, has never been done before. This way it not only clearly shows how to use the proposed analytical tools, but also provide valuable insights concerning the performance of the chosen eight distance/similarity measures in linguistic approximation.

Let A and B are trapezoidal fuzzy numbers on $[0, 1]$. Following distance/similarity measures of fuzzy numbers will be used in the further text:

- *distance measure* d_1 [25] (Formula 4.1 is a generalization of the distance used in [25] for fuzzy numbers on an interval.):

$$d_1(A, B) = \frac{\int_0^1 \mid A(x) - B(x) \mid dx}{\int_0^1 A(x)dx + \int_0^1 B(x)dx}. \tag{4.1}$$

- *distance measure* d_2 [24]:

$$d_2(A, B) = \sup_{x \in [0,1]} |A(x) - B(x)|. \tag{4.2}$$

- *modified Bhattacharyya distance* d_3 [1]:

$$d_3(A,B) = \left[1 - \int_0^1 (A^*(x) \cdot B^*(x))^{1/2} dx\right]^{1/2}, \tag{4.3}$$

where $A^*(x) = A(x)/\int_0^1 A(x)dx$ and $B^*(x) = B(x)/\int_0^1 B(x)dx$.

- *dissemblance index* d_4 [14]:

$$d_4(A,B) = \int_0^1 |\underline{a}(\alpha) - \underline{b}(\alpha)| + |\overline{a}(\alpha) - \overline{b}(\alpha)|\, d\alpha. \tag{4.4}$$

Please note that the following formulas for similarity measures of fuzzy numbers were originally defined for generalized trapezoidal fuzzy numbers[2]. Since we restrict the scope of the thesis only to linguistic approximation of triangular/trapezoidal fuzzy numbers, the formulas were adjusted for easier computation (since the height of fuzzy numbers is 1 by definition). Please check the respected references for original formulas. Let us assume that $A \sim (a_1, a_2, a_3, a_4)$ and $B \sim (b_1, b_2, b_3, b_4)$.

- *similarity measure* s_1 [5], [6]:

$$\begin{aligned} s_1(A,B) &= \left(1 - \frac{\sum_{i=1}^4 |a_i - b_i|}{4}\right) \\ &\cdot (1 - |X_A - X_B|)^{\left\lceil \frac{(a_4-a_1)+(b_4-b_1)}{2} \right\rceil} \cdot \frac{\min(Y_A, Y_B)}{\max(Y_A, Y_B)}, \end{aligned} \tag{4.5}$$

where $[X_A, Y_A]$ are the coordinates of the center of mass of fuzzy number A calculated using the following formulas:

$$Y_A = \begin{cases} \frac{\left(\frac{a_3-a_2}{a_4-a_1}+2\right)}{6}, & \text{if } a_4 \neq a_1 \\ \frac{1}{2}, & \text{if } a_4 = a_1 \end{cases}, \tag{4.6}$$

$$X_A = \frac{Y_A \cdot (a_3 + a_2) + (a_4 + a_1) \cdot (1 - Y_A)}{2}, \tag{4.7}$$

and $[X_B, Y_B]$ are coordinates of the center of mass of B defined analogically.

[2] A fuzzy set A_G on U is called *generalized trapezoidal fuzzy number* if there exists a trapezoidal fuzzy number A and $w_A \in [0,1]$ for which $A_G(x) = w_A \cdot A(x), x \in U$.

- *similarity measure* s_2 [33]:

$$s_2(A,B) = \left(1 - \frac{\sum_{i=1}^{4}|a_i - b_i|}{4}\right) \cdot \frac{\min\{Pe(A), Pe(B)\} + 1}{\max\{Pe(A), Pe(B)\} + 1}, \quad (4.8)$$

  where $Pe(A) = \sqrt{(a_1 - a_2)^2 + 1} + \sqrt{(a_3 - a_4)^2 + 1} + (a_3 - a_2) + (a_4 - a_1)$, $Pe(B)$ is defined analogically.

- *similarity measure* s_3 [11]:

$$s_3(A,B) = \left(1 - \frac{\sum_{i=1}^{4}|a_i - b_i|}{4}\right) \cdot \frac{\min\{Pe(A), Pe(B)\}}{\max\{Pe(A), Pe(B)\}} \cdot \frac{\min\{Ar(A), Ar(B)\} + 1}{\max\{Ar(A), Ar(B)\} + 1}, \quad (4.9)$$

  where $Ar(A) = \frac{1}{2}(a_3 - a_2 + a_4 - a_1)$, $Ar(B)$ is defined analogically and $Pe(A)$ and $Pe(B)$ are computed identically as in the previous measure.

- *similarity measure* s_4 [15]:

$$s_4(A,B) = \left(1 - \frac{\sum_{i=1}^{4}|a_i - b_i|}{4} \cdot d'(A,B)\right) \cdot \left(1 - \frac{|Ar(A) - Ar(B)|}{3} - \frac{\frac{|Pe(A)-Pe(B)|}{\max\{Pe(A),Pe(B)\}}}{3}\right), \quad (4.10)$$

  where $d'(A,B) = \frac{\sqrt{(X_A - X_B)^2 + (Y_A - Y_B)^2}}{\sqrt{1.25}}$, $\frac{|Pe(A)-Pe(B)|}{\max\{Pe(A),Pe(B)\}} = 0$ when $\max\{Pe(A), Pe(B)\} = 0$ and $[X_A, Y_A]$ and $[X_B, Y_B]$ are computed identically as in similarity measure s_1.

However, the coordinates of a center of mass of a rectangle defined by the following four points $[a_1, 0], [a_2, 1], [a_3, 1]$ and $[a_4, 0]$ are defined by (4.11) and (4.12). Clearly, (4.7) does not coincide with (4.11) neither does (4.6) with (4.12). Assuming that the intention was to use the coordinates of the center of mass of the fuzzy number, formulas (4.11) and (4.12) should be used.

$$X_A = \begin{cases} \frac{1}{3}\frac{a_4^2 + a_3^2 - a_2^2 - a_1^2 + a_4 a_3 - a_2 a_1}{a_4 + a_3 - a_2 - a_1}, & \text{if } a_4 < a_1 \\ a_1, & \text{if } a_4 = a_1 \end{cases}, \quad (4.11)$$

$$Y_A = \begin{cases} \frac{1}{3}\left(1 + \frac{a_3 - a_2}{a_4 + a_3 - a_2 - a_1}\right), & \text{if } a_4 < a_1 \\ \frac{1}{2}, & \text{if } a_4 = a_1 \end{cases}. \quad (4.12)$$

In this thesis we proceed to use the formulas (4.11) and (4.12) for the calculation of coordinates of the center of mass of a trapezoidal fuzzy numbers. The effect of this correction is discussed in section 4.4.2.

Distance measures d_3 and d_4 along with similarity measures s_2 and s_3 have already been extensively studied by the author (see Table 4.1). To extend the findings and to show the validity of the proposed methods for the analysis of the performance of distance/similarity measures in linguistic approximation, distance measures d_1 and d_2 and similarity measures s_1 and s_4 are also investigated in this thesis. Table 4.1 summarizes which publications focus on which distance/similarity measures and which type of approximated fuzzy numbers are taken into consideration in the publications by the author. It also indicates what underlying linguistic variable is assumed for linguistic approximation.

	$\mathbf{d_1}$	$\mathbf{d_2}$	$\mathbf{d_3}$	$\mathbf{d_4}$	$\mathbf{s_1}$	$\mathbf{s_2}$	$\mathbf{s_3}$	$\mathbf{s_4}$	Fuzzy number type	Scale used for LA
Publication I			•	•		•	•		symmetrical, triangular	elementary, enhanced
Publication II										
Publication III			•							enhanced
Publication IV				•						elementary
Publication V			•	•		•	•		asymmetrical, triangular	elementary
Publication VI			•	•		•	•		asymmetrical, triangular	enhanced
Publication VII			•	•					symmetrical, triangular	elementary
Publication VIII	•		•						Mamdani-type	enhanced
Publication IX				•					symmetrical, triangular	
Publication X			•	•		•	•		symmetrical, triangular	
Publication XI			•						asymmetrical, triangular	elementary

Table 4.1: Overview of the distance and similarity measures, types of fuzzy numbers and underlying linguistic variables studied in publications **I-XI**. Publication **II** deals with the background of the use of the concepts of fuzzy sets and linguistic modeling in social sciences, i.e. it discusses the benefits of these concepts for laymen.

As it was already mentioned, only some selected distance/similarity measures are examined in this thesis. Other, less frequently used distance measures can be found for example in [44, 7, 37, 21] etc. Similarly, less frequently used similarity measures can be found in [34, 26, 23] etc.

4.4 Analysis of linguistic approximation of symmetrical triangular fuzzy numbers

In this section we will focus on linguistic approximation of *triangular fuzzy numbers* that are *symmetrical*. This section summarizes the findings from publication **I** and extends these findings to distance measures d_1 and d_2 and similarity measures s_1 and s_4 introduced in section 4.3.

In accordance with publication **I** the performace of each distance/similarity measure for the linguistic approximation of symmetrical triangular fuzzy numbers on the interval $[0, 1]$ will be numerically investigated. Each of these fuzzy numbers $O = (o_1, o_2, o_4)$, where $o_2 - o_1 = o_4 - o_2$, represents a possible output of some mathematical model and can be represented by a 2-tuple $(\mathrm{COG}(O), \mathrm{L}(\mathrm{Supp}(O)))$;. To allow quicker calculations, this 2-tuple can be rewritten for symmetrical triangular fuzzy numbers as $(o_2, o_4 - o_1)$. Using this 2-tuple each approximated fuzzy number can be unambiguously respresented as a point in a 2D graph as is presented in Figure 4.2.

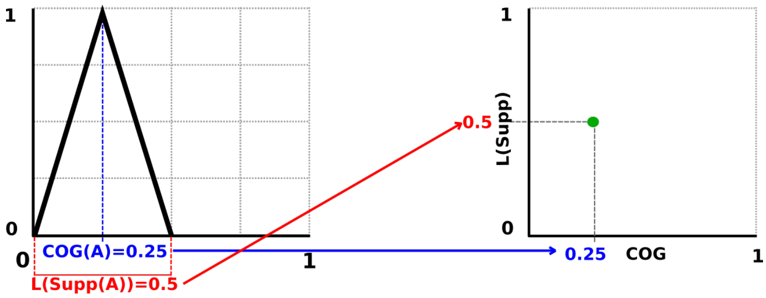


Figure 4.2: Representation of a symmetrical fuzzy number A as a point in a 2-dimensional space. The x-coordinate represents COG(A), the y-coordinate represents L(Supp(A)) and colour can be used to represent the approximating linguistic term.

To obtain a set of symmetrical triangular fuzzy numbers to be linguistically approximated, it is possible to randomly generate them as was suggested in [31]. The same procedure was later used in publications **V**, **VI** or **XI**. For a systematic analysis of the performance of distance/similarity measures it may be more appropriate to generate the fuzzy numbers on $[0, 1]$ in a "uniform way". A *grid approach*, presented in publication **I** and later used in publication **VII** can be used to generate the sample of approximated fuzzy numbers such that their representations are uniformly distributed in the $[0, 1] \times [0, 1]$ space. The grid approach is applied in this thesis. The two $[0, 1]$ intervals are therefore uniformly divided into 1 001 points each and these points represent possible centers of gravity/length of support of the approximated fuzzy numbers. Using the cartesian product we obtain 1002001 2-tuples that represent symmetrical triangular fuzzy numbers. Not all of these fuzzy numbers are defined on $[0, 1]$ interval (e.g. two tuple $(1, 1)$ represents the fuzzy number $(0.5, 1, 1.5)$). We restrict our analysis to the fuzzy numbers defined on the $[0, 1]$ interval only. Thus obtaining the set $Out_1 = \{O_1, \ldots, O_{500000}\}$ that contains 500 000 symmetrical triangular fuzzy numbers on interval $[0, 1]$ (see section 4 of publication **I** for more information).

4.4.1 Linguistic approximation of symmetrical triangular fuzzy numbers using a linguistic scale

At first, we will consider a uniform linguistic scale with five linguistic terms $\mathcal{T}_1, \ldots, \mathcal{T}_5$. The meanings of these terms are represented (in respective order) by triangular fuzzy numbers $T_1 = (0, 0, 0.25), T_2 = (0, 0.25, 0.5), T_3 = (0.25, 0.5, 0.75), T_4 = (0.5, 0.75, 1), T_5 = (0.75, 1, 1)$ that form a uniform Ruspini fuzzy partition of interval $[0, 1]$.

Each of the distance and similarity measures from section 4.3 is applied to identify the linguistic approximation of each fuzzy number from the set Out_1. Results describing the performance of the Bhattacharyya distance d_3 are depicted in Figure 4.3. Each approximating linguistic term is assigned a different colour, ares with same colour represent fuzzy numbers that are linguistically approximated by the same linguistic term. White areas consist of the representations of such symmetrical triangular fuzzy numbers, that are not defined on $[0,1]$. This graphical representation was designed to provide insights into the performance of a selected distance/similarity measure that would be easily understandable. From Figure 4.3 we can e.g. see, that result of linguistic approximation is highly COG driven; the length of support plays only minor role.

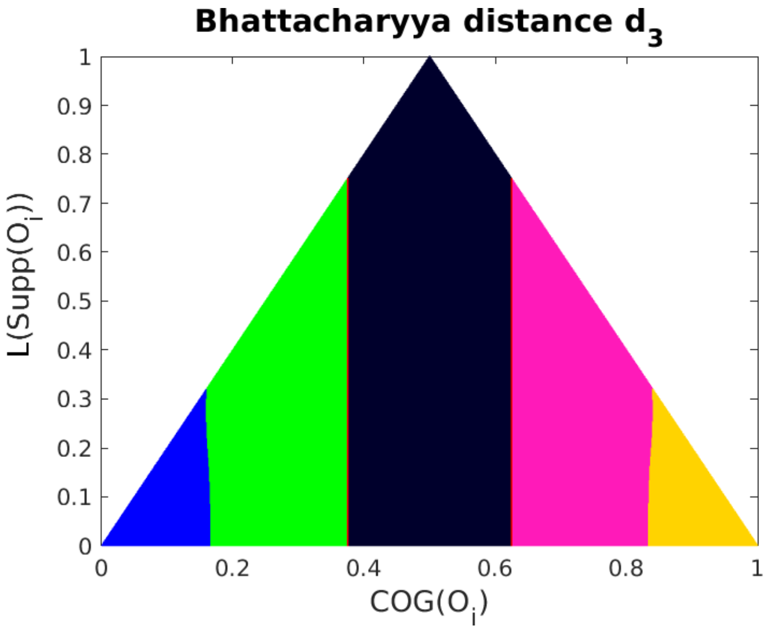


Figure 4.3: A graphical representation of the results of linguistic approximation of symmetrical triangular fuzzy numbers using the Bhattacharyya distance d_3 and a linguistic scale. Each colour represents one term of the five term linguistic scale: $\mathcal{T}_1$ (blue), $\mathcal{T}_2$ (green), $\mathcal{T}_3$ (black), $\mathcal{T}_4$ (pink) and $\mathcal{T}_5$ (yellow). Red colour (visible on the borders between the black area and its neighborouing areas) represents ambiguous cases, i.e. cases when more than one linguistic term is assigned.

Using this graphical representation more distance and similarity measures can be compared. Figure 4.4 summarizes the performance for all the investigated distance/similarity measures presented in section 4.3. Differences in their performance are clearly visible. To provide more details about the performance of linguistic approximation under selected measures we add the Table 4.2, that sumarizes the frequencies of assignment of each of the elementary linguistic terms by linguistic approximations under distance/similarity measures. This information can be used not only to verify our findings based on the graphical summary provided by Figure 4.4 (see the following list), but also to highlight some unexpected/easily overlooked behaviour of linguistic approximation under some distance/similarity measures. This becomes more important when enhanced linguistic scales are used and higher-level linguistic terms start to be assigned (see e.g. the low frequency of assignment of linguistic terms $\mathcal{T}_{12}$ and $\mathcal{T}_{45}$ by linguistic approximation using distance measure d_1 in section 4.4.2, which can be easily overlooked in Figure 4.7). Based on the direct comparison of the performance of distance/similarity measure graphically summarized in Figure 4.4 we can draw the following conclusions:

- Linguistic approximation of symmetrical triangular fuzzy numbers with the length of the support higher then approximately 0.4 are not dependent on the choice of distance/similarity measure. The results of the lingusitic approximation in these cases depend only on the center of gravity of the approximated fuzzy numbers. Therefore if the possible outputs of the model are only fuzzy numbers with higher cardinality, the choice of a distance/similarity measure (out of the once discussed in this thesis) is of no consequence. In such cases it is therefore reasonable to use measure that are e.g. easy to compute or readily available in the software we are using.

- Distance measure d_2 does not seem to be appropriate for the linguistic approximation of triangular symmetrical fuzzy numbers. Firstly, lingusitic terms $\mathcal{T}_1$ (blue) and $\mathcal{T}_5$ (yellow) are not used at all. The set of obtainable linguistic terms is thus reduced, moreover the border terms (i.e. the terms with meanings closest to the endpoints of $[0, 1]$ interval) are eliminated. This could by undesirable, because the border terms can be the most important ones (e.g. excellent evaluation; extremly dangerous...). Secondly, there are four "triangle-shaped" areas (red) that represent fuzzy numbers, for which a unambiguous linguistic approximation can not be determined (distance measure d_2 selects more than one lingustic term as a result of linguistic approximation).

- The remaining distance measures d_1, d_3 and d_4 provide similar results. The results of linguistic approximation using Bhattacharyya distance d_3 depend almost exclusively on the center of gravity of the approximated fuzzy number (we can see from Figure 4.3 that the border between linguistic terms $\mathcal{T}_1$ (blue) and $\mathcal{T}_2$ (green) is not completly vertical, nor is the border between $\mathcal{T}_4$ (pink) and $\mathcal{T}_5$ (yellow)). The results of linguistic approximation using distance measures d_1 and d_4 exhibit the same pattern, the border between the blue and the green area is more dependent on the length of the support of the approximated fuzzy numbers. The same holds for the border between the pink and the yellow areas for these two distances.

- Differences between the outputs of linguistic approximation using the four selected similarity measures are clearly visible. Linguistic approximation using the similarity measure s_1 provides results similar to Bhattacharyya distance d_3 – i.e. it focuses mostly on the center of gravity of approximated fuzzy numbers.

- Similarity measure s_2 is more focused on the shape of the approximated fuzzy number (it uses perimeters of fuzzy numbers) than s_1. Therefore fuzzy numbers with smaller length of support and center of gravity close to the borders of the interval $[0, 1]$ tend to be linguistically approximated by linguistic term $\mathcal{T}_1$ instead of $\mathcal{T}_2$ or $\mathcal{T}_5$ instead of $\mathcal{T}_4$. This is due to the fact, that narrow triangles close to 0 or 1 are more similar to T_1 ot T_5 respectively (note, that the perimeter of T_1 or T_5 is smaller than the perimeter of T_2, T_3 or T_4.

- The performance of similarity measures s_3 and s_4 is significantly different from the performance of all the other similarity and distance measures considered in this thesis. The s_3 and s_4 measures focus not only on the perimeters (as similarity measure s_2 does), but also on the areas of fuzzy numbers. This results in the amplification of the effect observed for similarity measure s_2. Note that even some fuzzy numbers that were consistently linguistically approximated by the middle term $\mathcal{T}_3$ (black) by all the previous measures (with the exception of distance measure d_2) are approximated by linguistic terms $\mathcal{T}_1$ or $\mathcal{T}_5$ under s_3 and s_4. This effect is even stronger using similarity measure s_4. In essence, $\mathcal{T}_2$ and $\mathcal{T}_4$ are never assigned as lingusitic approximation of low uncertain fuzzy numbers.

Level 1	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$	Ambiguous
$\mathbf{d_1}$	26 654	113 595	218 000	113 595	26 654	1 502
$\mathbf{d_2}$	0	93 500	187 000	93 500	0	126 000
$\mathbf{d_3}$	26 694	113 556	218 000	113 556	26 694	1 500
$\mathbf{d_4}$	28 405	111 823	218 000	111 823	28 405	1 544
$\mathbf{s_1}$	28 940	111 307	218 000	111 307	28 940	1 506
$\mathbf{s_2}$	47 360	92 885	218 000	92 885	47 360	1 510
$\mathbf{s_3}$	90 096	52 582	213 500	52 582	90 096	1 144
$\mathbf{s_4}$	100 001	45 231	208 554	45 231	100 001	982

Table 4.2: Frequencies of assignment of each of the elementary linguistic terms $\mathcal{T}_1, \dots, \mathcal{T}_5$ as linguistic approximations of the symmetrical triangular fuzzy numbers from the set Out_1 by the examined distance/similarity measures $d_1, d_2, d_3, d_4, s_1, s_2, s_3$ and s_4. The column ambiguous represents cases where more than one linguistic term was recommended, i.e. the shortest distance/maximum similarity of the approximated fuzzy number to the meanings of the linguistic terms was identical for two or more terms.

4.4.2 Linguistic approximation of symmetrical triangular fuzzy numbers using an enhanced linguistic scale

In this section, we will expand the uniform linguistic scale with five linguistic terms $\mathcal{T}_1, \dots, \mathcal{T}_5$ from the previous section by adding derived linguistic terms $\mathcal{T}_{ij}$, where $i = 1, \dots, 4$, $j = 2, \dots, 5$ and $i < j$. The meanings of the elementary terms remain

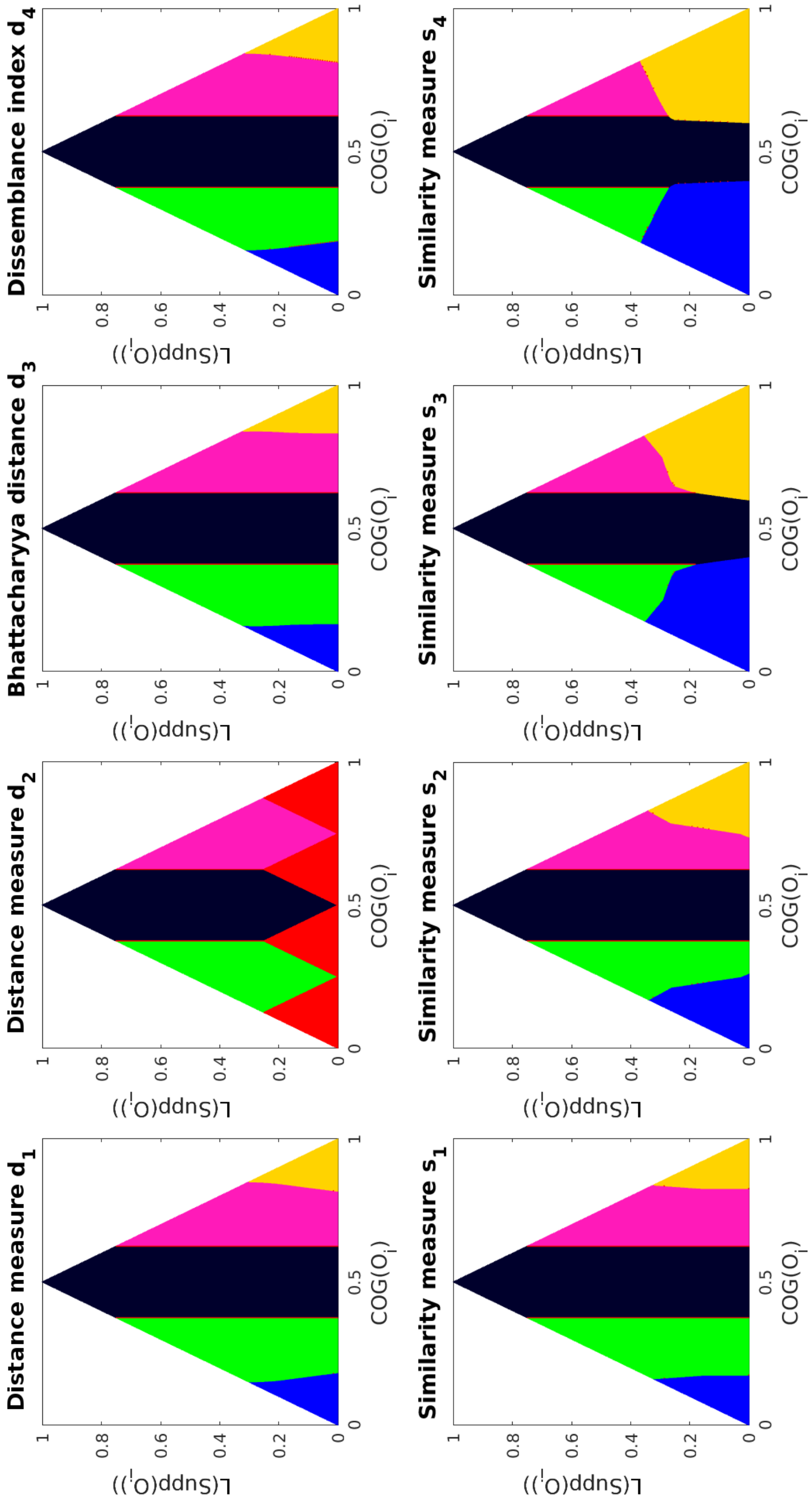


Figure 4.4: A graphical summary of the performance of the chosen distance and similarity measures in the linguistic approximation of symmetrical triangular fuzzy numbers on $[0, 1]$ using a linguistic scale. Each colour represents one term of the five term linguistic scale: $\mathcal{T}_1$ (blue); $\mathcal{T}_2$ (green), $\mathcal{T}_3$ (black), $\mathcal{T}_4$ (pink) and $\mathcal{T}_5$ (yellow). Red colour represents ambiguous cases, i.e. cases when more than one linguistic term is assigned.

the same (and are represented by triangular fuzzy numbers) and the meanings of the enhanced linguistic terms are represented by trapezoidal fuzzy numbers (obtained as Łukasiewicz unions of the meanings of the respective elementary linguistic terms).

In line with the previous section each of the distance and similarity measures considered in section 4.3 is applied here to identify the linguistic approximation of each fuzzy number from the set Out_1. An example of the results describing the performance of one chosen distane measure, the Bhattacharyya distance d_3, is depicted in Figure 4.5. Again, each approximating linguistic term is assigned a different colour; areas with the same colour represent fuzzy numbers that are linguistically approximated by the same linguistic term. Colours that have been used in the previous section still represent the same linguistic terms ($\mathcal{T}_1, \ldots, \mathcal{T}_5$), while new colours represent derived linguistic terms $\mathcal{T}_{ij}$. Differences resulting from the use of enhanced linguistic scale can be directly studied by the comparison of Figures 4.3 and 4.5. When the enhanced linguistic scale is used for linguistic approximation the information respresented by the length of support of the approximated fuzzy number now plays a much more significant role for the investigated Bhattacharyya distance d_3. This can not be observed under the linguistic scale. When the center of gravity of the approximated fuzzy number lies half way between the centers of gravity of the fuzzy numbers representing the meanings of the neighbouring elementary linguistic terms a derived linguistic term is suggested as a linguistic approximation instead of the elementary ones. The higher the length of support of the approximated fuzzy number the further its center of gravity can be from the middle point for the linguistic approximation to assigned derived term. When the length of support is higher then approximately 0.75, alementary linguistic terms are no longer assigned and the linguistic approximation suggests derived linguistic terms only.

Before we focus on the direct comparison of the performance of linguistic approximation using the selected distance/similarity measures in combination with the enhanced linguistic scale, we need to deal with one important issue. As we have stated in section 4.3, the similarity measures s_1 and s_4 require the calculation of x and y coordinates of the center of mass of the fuzzy numbers. However we have pointed out that the original formulas (4.6) and (4.7) do not provide the coordinates of the center of mass; to obtain the correct coordinates formulas (4.11) and (4.12) should be used. In Figure 4.6 we show the difference between the originally proposed formulas (4.6) and (4.7) and the correct formulas for the calculation of the center of mass of the fuzzy number (4.11) and (4.12) in the calculation of the similarity measure s_1. More specifically we show how the results of the linguistic approximation under enhanced linguistic scale differ between the two approaches to the calculation of the center of mass of a fuzzy number. The left graph represents the results using the original formulas while the right graph shows the results obtained using formulas (4.11) and (4.12) in s_1. While the original formula for the calculation of s_1 can result in the use of derived linguistic term - $\mathcal{T}_{23}$ (purple) and $\mathcal{T}_{34}$ (brown), the use of the correct formulas (4.11) and (4.12) no longer suggest any derived linguistic terms to be used as a linguistic approximation under s_1. It is evident that the results of linguistic approximation are affected by the choice of the

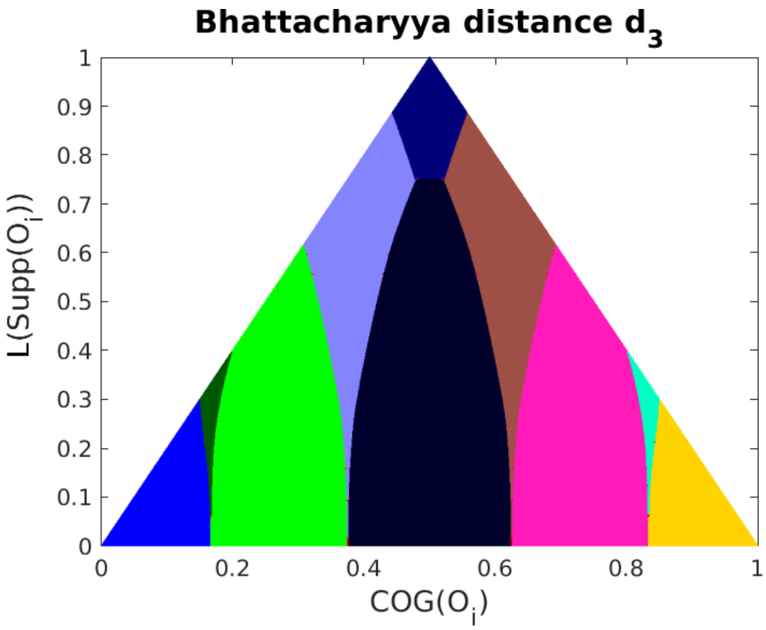


Figure 4.5: A graphical representation of the results of linguistic approximation of symmetrical triangular fuzzy numbers using the Bhattacharyya distance d_3 and an enhanced linguistic scale. Elementary linguistic terms are represented by the same colours as in figure 4.3 (i.e. blue, green, black, pink and yellow respectively). Other colours represent derived linguisted terms.

formulas for the coordinates of the center of mass of a fuzzy number. As the original idea in s_1 and s_4 is to use the x and y coordinates of the centers of mass of the fuzzy numbers we have decided to use the correct formulas (4.11) and (4.12) in the thesis.

As in the previous section, we will use graphical representation of the results of linguistic approximation using the selected distance and similarity measures. Figure 4.7 summarizes the performance for all the investigated distance/similarity measures presented in section 4.3 in combination with enhanced linguistic scale. Frequencies of assignment of each of the linguistic terms (elementary and derived) by linguistic approximation are presented in Table 4.3. Based on such a direct comparison of the performance of distance/similarity measure in the given context we can draw the following conclusions:

- Similarity measure s_1 is the only measure that suggest the same linguistic approximation regardless if standard or enhanced linguistic scale is used. Derived linguistic terms are never assigned. All the other studied distance/similarity measure suggest derived linguistic terms as linguistic aproximation for some of the symmetric triangular fuzzy numbers on $[0, 1]$ when the enhanced linguistic scale is considered.

- Distance measures d_1 and d_3 are the only measures for which a level 3 linguistic term can be selected as a result of linguistic approximation - linguistic term

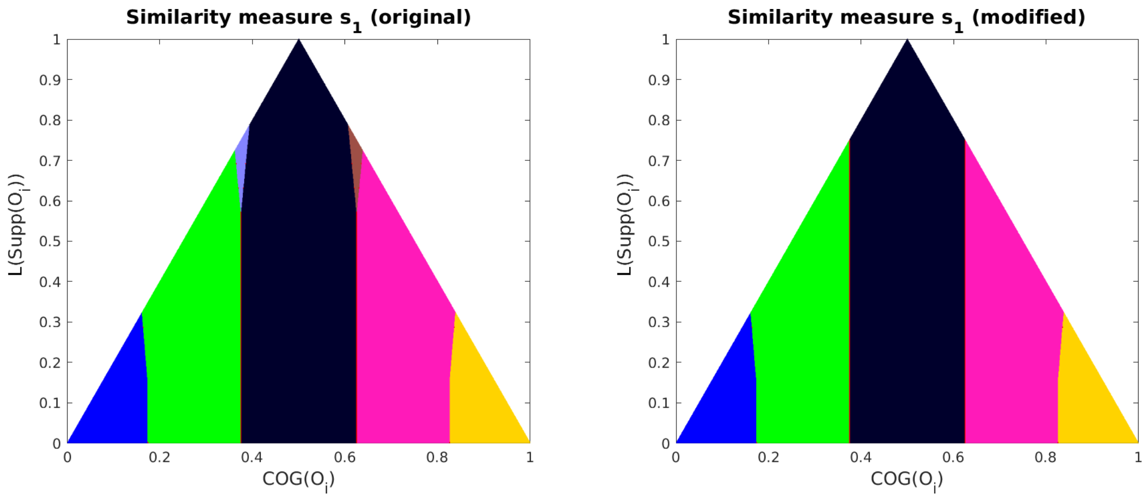


Figure 4.6: A comparison of the effect of different approaches to the calculation of the center of mass of a fuzzy number on the results of linguistic approximation using s_1 with enhanced linguistic scales. Formulas (4.6) and (4.7) are applied in the left graph, formulas (4.11) and (4.12) are applied in the right graph.

$\mathcal{T}_{24}$ (dark blue). Other distance/similarity measures can select only elementary linguistic terms or level 2 terms (except for s_1 where even level 2 terms are not used at all). This means that except for d_1 and d_3 all the measures assigned relatively low-uncertain linquistic approximations even to fuzzy numbers with high cardinality.

- Linguistic approximation using distance measure d_1 can result in level 2 linguistic terms $\mathcal{T}_{12}$ (dark green) and $\mathcal{T}_{45}$ (aqua). This could be easily overlooked, because in our numerical investigation, only 50 fuzzy numbers were approximated by either of these two level 2 terms (e.g. fuzzy numbers with the length of the support approximately equal to 0.3 and the center of gravity approximately equal to 0.15 or 0.85). In order to avoid potential overlookings of this type, I strongly suggest to accompany the graphical representation of the performance of linguistic approximation by a table representing the relative frequencies of assignment of individual linguistic terms - see Table 4.3.

- With the exception of Bhattacharyya distance d_3, all the distance/similarity measures provide linguistic approximations identical to those under the linguistic scale (section 4.4.1) for all fuzzy numbers whose length of support is below a certail threshold. The value of this threshold varies from approximatelly 0.3 for distance measure d_1, approximately 0.5 for distance measures d_2 and d_4 to approximatelly 0.6 for similarity measures s_2, s_3 and s_4. We have already stated that the results suggested by s_1 are identical regardles of the linguistic scale, i.e. the respective threshold would be 1. This implies that the choice of the approximating linguistic scale is of consequence only if the approximated fuzzy numbers are going to have the lengths of support above the respective thresholds. Note that for the Bhattacharyya distance d_3 the threshold is essentially zero which means that level 2 linguistic terms can be obtained as linguistic approximation even for very low-uncertain fuzzy numbers.

- Distance d_2 again does not seem to be appropriate for the linguistic approximation - both problems from previous section remain (linguistic terms $\mathcal{T}_1$ and $\mathcal{T}_5$ are not used at all and there are "triangle-shaped" red areas that represent fuzzy numbers that can not be unambiguously linguistically approximated). The use of the enhanced linguistic scale introduces two level 2 linguistic terms as a result of linguistic approximation for fuzzy numbers with high cardinality - $\mathcal{T}_{23}$ (purple) and $\mathcal{T}_{34}$ (brown). This is also the case of distance measure d_4.

- Similarity measures s_2, s_3 and s_4 provide similar results of linguistic approximation. Derived linguistic terms are suggested (as linguistic approximations) for fuzzy numbers with high cardinality. However, while s_2 (similarly to d_2 and d_4) may result only in $\mathcal{T}_{23}$ (purple) and $\mathcal{T}_{34}$ (brown) linguistic terms, the remaining similarity measures s_3 and s_4 may result also in $\mathcal{T}_{12}$ (dark green) and $\mathcal{T}_{45}$ (aqua); the only other measure that can provide these linguistic terms as a result of linguistic approximation is the Bhattacharyya distance d_3. Note, that areas representing fuzzy numbers that are linguistically approximated by $\mathcal{T}_{12}$ or $\mathcal{T}_{45}$ are significantly larger in the case of similarity measure s_4 than s_3.

- The performance of linguistic approximation using investigated distance/similarity measures showed similar characteristics for all the measures with the exceptions of Bhattacharyya distance d_3. Bhattacharyya distance assigns higher level (derived) linguistic terms to fuzzy numbers with much lower cardinalities than all the other measures. Also the "borders" between the areas of fuzzy numbers linguistically approximated by the same linguistic terms are curved - for other measures, the areas are rather straight.

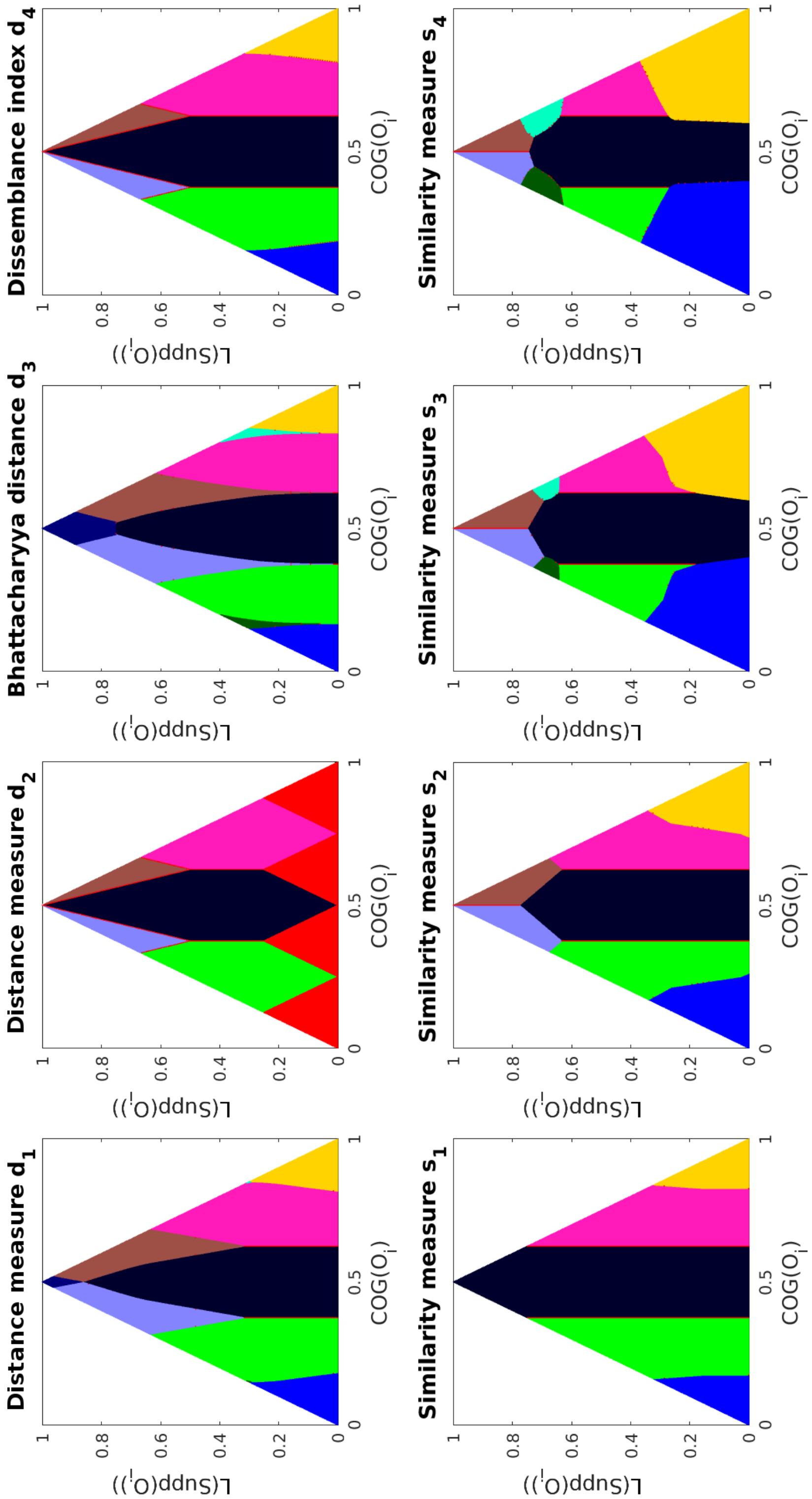


Figure 4.7: A graphical summary of the performance of the chosen distance and similarity measures in the linguistic approximation of symmetrical triangular fuzzy numbers on $[0, 1]$ using enhanced linguistic scale. Elementary linguistic terms are represented by the same colours as in figure 4.4 (i.e. blue, green, black, pink and yellow respectively). Other colours represent derived linguisted terms. Red colour represents ambiguous cases, i.e. cases when more than one linguistic term is assigned.

Level 1	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$
$\mathbf{d_1}$	26 650	101 270	156 474	101 270	26 650
$\mathbf{d_2}$	0	88 375	155 751	88 375	0
$\mathbf{d_3}$	25 707	93 062	138 922	93 062	25 707
$\mathbf{d_4}$	28 405	106 698	186 751	106 698	28 405
$\mathbf{s_1}$	28 940	111 307	218 000	111 307	28 940
$\mathbf{s_2}$	47 360	90 804	175 672	90 804	47 360
$\mathbf{s_3}$	90 096	49 765	172 272	49 765	90 096
$\mathbf{s_4}$	100 001	41 895	165 361	41 895	100 001

Level 2	$\mathcal{T}_{12}$	$\mathcal{T}_{23}$	$\mathcal{T}_{34}$	$\mathcal{T}_{45}$
$\mathbf{d_1}$	25	42 252	42 252	25
$\mathbf{d_2}$	0	20 834	20 834	0
$\mathbf{d_3}$	4 282	48 973	48 973	4 282
$\mathbf{d_4}$	0	20 834	20 834	0
$\mathbf{s_1}$	0	0	0	0
$\mathbf{s_2}$	0	23 249	23 249	0
$\mathbf{s_3}$	3 870	19 539	19 539	3 870
$\mathbf{s_4}$	9 200	15 704	15 704	9 200

Level 3	$\mathcal{T}_{13}$	$\mathcal{T}_{24}$	$\mathcal{T}_{35}$
$\mathbf{d_1}$	0	2 490	0
$\mathbf{d_2}$	0	0	0
$\mathbf{d_3}$	0	16 970	0
$\mathbf{d_4}$	0	0	0
$\mathbf{s_1}$	0	0	0
$\mathbf{s_2}$	0	0	0
$\mathbf{s_3}$	0	0	0
$\mathbf{s_4}$	0	0	0

Ambiguous
642
125 831
60
1 375
1 506
1 502
1 188
1 039

Table 4.3: Frequencies of assignment of each of the elementary linguistic terms $\mathcal{T}_1, \ldots, \mathcal{T}_5$ and also of the derived level 2 linguistic terms $\mathcal{T}_{12}$, $\mathcal{T}_{23}$, $\mathcal{T}_{34}$, $\mathcal{T}_{45}$ and level 3 linguistic terms $\mathcal{T}_{13}$, $\mathcal{T}_{24}$ and $\mathcal{T}_{35}$ as linguistic approximations of the symmetrical triangular fuzzy numbers from the set Out_1 by $d_1, d_2, d_3, d_4, s_1, s_2, s_3$ and s_4. The frequencies for use of higher level linguistic terms are not presented, because they were not selected as a output of the linguistic approximation for any of the approximated fuzzy numbers. The column ambiguous represents cases where more than one linguistic term was recommended.

4.5 Analysis of linguistic approximation of asymmetrical triangular fuzzy numbers

The framework introduced in section 4.4 will be now applied on linguistic approximation of *asymmetrical* triangular fuzzy numbers. Although the generalization to asymmetrical fuzzy numbers may seem to be straightforward (technically methods introduced in section 4.4 could be directly applied on asymmterical triangular fuzzy numbers without adjustments), there is one significant complication, that needs to be taken into account. Each asymmetrical triangular fuzzy number $O = (o_1, o_2, o_4)$ can again be represented by a 2-tuple $(\mathrm{COG}(O), \mathrm{L}(\mathrm{Supp}(O)))$, however this 2-tuple representation possibly represents more than one asymmetrical triangular fuzzy number. This introduces a complication into our framework, because one point in the 2D graphical visualization space represents possibly several different asymmetrical triangular fuzzy numbers (technically a single point can represents an infinite number of different fuzzy numbers). Figure 4.8 presents an example of two different asymmetrical triangular fuzzy numbers $A = (0.1, 0.1, 0.4)$ and $B = (0, 0.3, 0.3)$ for which $\mathrm{COG}(A) = \mathrm{COG}(B) = 0.2$ and $\mathrm{L}(\mathrm{Supp}(A)) = \mathrm{L}(\mathrm{Supp}(B)) = 0.3$, i.e. both can be represented by the same 2-tuple $(0.2, 0.3)$.

There are several possible solutions to this problem. One would be using such a representation of the triangular fuzzy numbers that would represent each of them by a unique vector of values. Such vectors would however need to have more than two components. In other words, applying this solution we loose the easy to follow 2D graphical representation. E.g. switching to a 3D graphical representation, non-transparency can become an issue.

If we insist on 2D graphical summaries, we can apply the methods from section 4.4, but account for the loss of information stemming from the 2D vector representation of asymmetrical triangular fuzzy number in some way. Direct application of the previously defined analytical framework may result in one point being coloured by several colours (representing different linguistic terms) at the same time. This is impossible to present graphically in a single plot. We can resolve this issue (in accordance with publication **V**) by using more than one 2D graph for the graphical representation of the performance of the linguistic approximation under selected measures to prevent this "overlapping" of coloured areas in the graphical presentation. Or to be more precise to clearly show which $(\mathrm{COG}(O), \mathrm{L}(\mathrm{Supp}(O)))$ representations of fuzzy numbers can result in the assignment of each linguistic term.

This can be done in two possible ways: 1) for each linguistic term we can use separate graphical representation – coloured area represents fuzzy numbers that are linguistically approximated by the given linguistic term; 2) linguistic terms are divided into groups in such a way that the 2D graphical representation of fuzzy numbers linguistically approximated by terms from one group is not overlapping (see Figure 4.9). This reduces the required number of plots.

We now need a suitable method for the generation of asymmetrical triangular fuzzy numbers on interval $[0, 1]$ to study the effect of the selection of distance/similarity measure on their linguistic approximation. Again, it is possible to

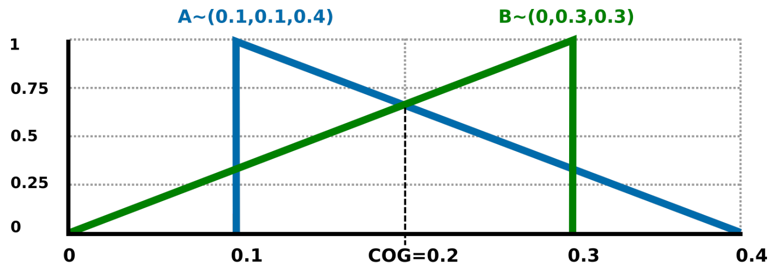


Figure 4.8: Example of two different asymmetrical triangular fuzzy numbers A (blue) and B (green) that have identical center of gravity ($\mathrm{COG}(A) = \mathrm{COG}(B) = 0.2$) and length of support ($\mathrm{L}(\mathrm{Supp}(A)) = \mathrm{L}(\mathrm{Supp}(B)) = 0.3$).

randomly generate these fuzzy numbers (i.e. to randomly generate the triplet of their significant values; this approach was chosen in publications **V** and **XI**). However, in the further text we will adjust the *grid approach* that was originally used to generate the set of symmetrical triangular fuzzy numbers Out_1 in previous sections for the generation of asymmetrical triangular fuzzy numbers. This approach allows for uniform sampling from the set of asymmetrical triangular fuzzy numbers on $[0, 1]$ and also for the adjustment of precision and hence of the computational speed. As we already mentioned, these fuzzy numbers can not be unambiguously represented using 2-tuple $(\mathrm{COG}(O), \mathrm{L}(\mathrm{Supp}(O)))$. For this reason we will use standard representation of triangular fuzzy numbers using their significant values. These values will be uniformly distributed in the $[0, 1] \times [0, 1] \times [0, 1]$ space. These three intervals will be uniformly divided into 151 points each[3]. Using the cartesian product we obtain 3 442 951 candidates on triangular fuzzy numbers. However, some of these candidates do not represent a fuzzy number (e.g. triplet $A = (1, 0, 0)$ is not a fuzzy number, because $a_1 > a_2$) and thus it makes no sense to use them in further analysis of the performance of linguistic approximation. After this restriction (i.e. restricting the set of candidates on triangular fuzzy numbers to the set of actual fuzzy numbers), we obtain the set Out_2 that contains 585 125 asymmetrical triangular fuzzy numbers on the interval $[0, 1]$.

Linguistic approximation of asymmetrical triangular fuzzy numbers using a linguistic scale

In accordance with section 4.4 where the performance of linguistic approximation of symmetrical triangular fuzzy numbers under different distance/similarity measures was studied, we will firstly consider a uniform linguistic scale with five linguistic

[3] In the case of assymetrical triangular fuzzy numbers the interval $[0, 1]$ is divided into significantly less points then in the case of symmetrical triangular fuzzy numbers. This is not an oversight - due to the fact that now each fuzzy number is represented by a triplet of its significant values, the initial cartesian product results in an even larger set of “potential candidates” on asymmetrical triangular fuzzy numbers then in the case of symmetrical triangular fuzzy numbers. Obviously, the partition of the $[0, 1]$ intervals can be adjusted.

terms $\mathcal{T}_1, \ldots, \mathcal{T}_5$. Their meanings (in accordance with previous text) are assumed to be represented by triangular fuzzy numebrs $T_1 = (0, 0, 0.25), T_2 = (0, 0.25, 0.5), T_3 = (0.25, 0.5, 0.75), T_4 = (0.5, 0.75, 1), T_5 = (0.75, 1, 1)$ that form uniform Ruspini fuzzy partition of interval $[0, 1]$. Again, each of studied distance/similarity measures introduced in section 4.3 is applied to identify the linguistic approximation of each fuzzy number from the set Out_2.

Results of the performance of linguistic approximation using Bhattacharyya distance d_3 are depicted in Figure 4.9. Colours represent the same linguistic terms as in section 4.4 and therefore the direct comparison with the performance of linguistic approximation of symmetrical triangular fuzzy numbers under Bhattacharyya distance depicted in Figure 4.3 is possible (and recommended). Please note, that assymetrical triangular fuzzy numbers are a generalization of symmetrical triangular fuzzy numbers and therefore the findings from previous sections also apply in the case of asymmetrical triangular fuzzy numbers. Therefore new findings should be perceived as a generalization/extension of prior findings.

From Figure 4.9 can see, that the area representing the linguistically approximated fuzzy numbers is larger than in the case of symmetrical triangular fuzzy numbers. This is especially evident in the case of fuzzy numbers with the length of support equal to one - if symmetrical triangular fuzzy numbers are considered, there is only a single fuzzy number with that property: $(0, 0.5, 1)$. But in the case of asymmetrical triangular fuzzy numbers, this property is fullfilled for any fuzzy number $(0, x, 1), x \in [0, 1]$, i.e. for infinitely many fuzzy numbers. Centers of gravity for the “borderline” fuzzy numbers $(0, 0, 1)$ and $(0, 1, 1)$ are $\mathrm{COG}(0, 0, 1) = 1/3$ and $\mathrm{COG}(0, 1, 1) = 2/3$. Also it is clearly notable that unlike in the case of symmetrical triangular fuzzy numbers, the results of linguistic approximation using Bhattacharyya distance do not depend as strongly on the center of gravity as in the symmetrical case.

Figures 4.10 and 4.11 summarize the performance for all the investigated distance/similarity measures presented in section 4.3. First figure depicts asymmetrical triangular fuzzy numbers linguistically approximated by linguistic terms $\mathcal{T}_1, \mathcal{T}_3$ and $\mathcal{T}_5$ while the second figure depicts fuzzy numbers linguistically approximated by linguistic terms $\mathcal{T}_2$ and $\mathcal{T}_4$. Each term is again represented by a specific colour (in accordance with the previous sections). Unlike in the case of symmetrical triangular fuzzy numbers, ambiguous cases (represented by red colour) are depicted separatelly in Figure 4.12. This is necessary, because a single point in the 2D space can represent both a fuzzy number that can not be linguistically approximated (i.e. an ambiguous case) and also a fuzzy number that can be linguistically approximated. Therefore for the sake of clarity ambiguous cases are investigated in a separate figure.

To provide even more insight into the performance of linguistic approximation in this case we can also plot the areas of possible colour overlaps. In another words we can identify areas in the $(\mathrm{COG}(O), \mathrm{L}(\mathrm{Supp}(O)))$ space, where each point represents a set of fuzzy numbers with identical 2-tuple representation, but possibly different linguistic approximation, see Figure 4.13. In the further text we will call these areas *linguistic approximation grey zones*.

Table 4.4 presents a different view on the performance of different distance/similarity measures of fuzzy numbers in the linguistic approximation – through absolute frequencies of assignment of each linguistic term. Comparing the results with the Table 4.2 should take into an account different cardinalities of the sets of approximated fuzzy numbers Out_1 and Out_2. Clearly, ambiguous cases are much less frequent under d_1 and d_3 as long as asymmetrical fuzzy numbers are considered.

Considering Figures 4.10, 4.11, 4.12 and 4.13 and Table 4.4 we can now draw for example the following conclusions:

- Distance measures d_1, d_3 and similarity measure s_1 exhibit similar properties. The result of linguistic approximation using these measures is again (as in the case of symmetrical triangular fuzzy numbers) mainly center of gravity dependent. For all these measures the $(\text{COG}(O), \text{L}(\text{Supp}(O)))$ representation is an acceptable simplification; in other words the center of gravity and the cardinality of the approximated fuzzy number are a good predictor of the result of linguistic approximation. The d_4 measure behaves in a similar manner, but the amount of ambiguous cases as well as the area of the linguistic approximation grey zones are larger. For all these four measures a situation where more than two linguistic approximations would be suggested for the same $(\text{COG}(O), \text{L}(\text{Supp}(O)))$ point is practically ruled out.

- Distance measures d_2 and d_4 are the only ones that assign linguistic terms $\mathcal{T}_1$ and $\mathcal{T}_5$ to fuzzy numbers with high length of support (over 0.5). Note, that for d_2 the term $\mathcal{T}_1$ is assigned only to fuzzy numbers of the type $(0, 0, x), x \in (0, 1]$ and $\mathcal{T}_5$ to fuzzy numbers of the type $(x, 1, 1), x \in [0, 1)$.

- Using d_2 the linguistic approximation of low-uncertain fuzzy numbers (as well as of fuzzy singletons) is virtually impossible (see Figure 4.12).

- High length of support in combination with the use of d_2, s_2 or s_3 can result in three linguistic labels being assigned to a single $(\text{COG}(O), \text{L}(\text{Supp}(O)))$ two-tuple (see the (0.5, 0.75) points in Figures 4.10 and 4.11 corresponding with these measures). Note that this feature is not present in the linguistic approximation of symmetrical triangular fuzzy numbers using these measures.

- Again, under s_3 and s_4 similarity measures, linguistic terms $\mathcal{T}_2$ and $\mathcal{T}_4$ are never assigned to low-uncertain fuzzy numbers. Also a similar problem as in the previous item is present, this time for $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ or $\mathcal{T}_3$, $\mathcal{T}_4$ and $\mathcal{T}_5$ terms. Not to mention, that two plots for each of these measures (assuming a five element scale) are not sufficient to avoid overlaps. However to maintain comparability with the analysis of the other six measures, the two-plot representation (Figure 4.10 and 4.11) is maintained for these two measures as well.

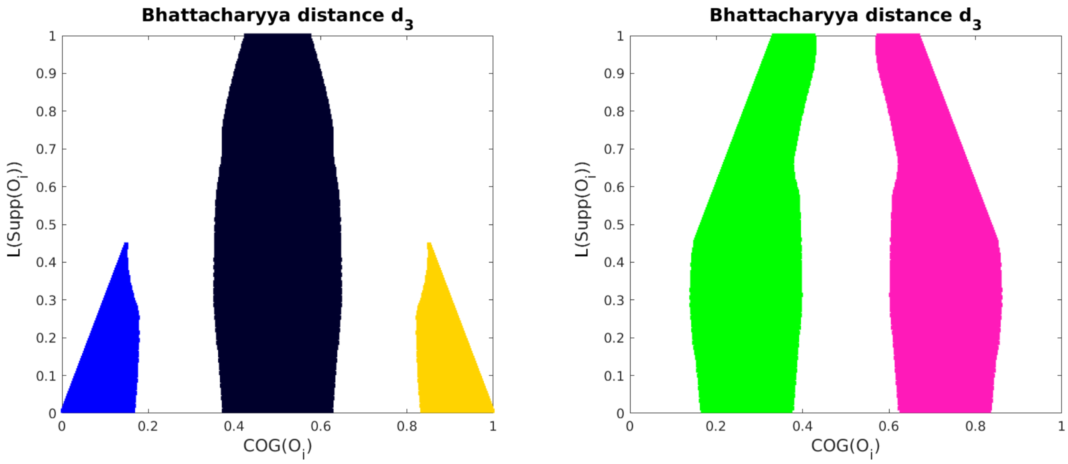


Figure 4.9: A graphical representation of the results of linguistic approximation of asymmetrical triangular fuzzy numbers using the Bhattacharyya distance d_3 and a linguistic scale. Colours on the left subfigure represents "odd" linguistic terms $\mathcal{T}_1$ (blue), $\mathcal{T}_3$ (black) and $\mathcal{T}_5$ (yellow) and colours on the right subfigure represents "even" linguistic terms $\mathcal{T}_2$ (green) and $\mathcal{T}_4$ (pink).

Level 1	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$	**Ambiguous**
$\mathbf{d_1}$	11 258	140 143	282 305	140 143	11 258	18
$\mathbf{d_2}$	150	152 638	238 687	152 638	150	40 862
$\mathbf{d_3}$	11 350	135 316	291 781	135 316	11 350	12
$\mathbf{d_4}$	14 541	143 774	267 047	143 774	14 541	1 448
$\mathbf{s_1}$	14 497	136 332	283 193	136 332	14 497	274
$\mathbf{s_2}$	23 364	142 638	251 333	142 638	23 364	1 788
$\mathbf{s_3}$	44 287	122 560	249 749	122 560	44 287	1 682
$\mathbf{s_4}$	52 302	97 712	284 713	97 712	52 302	384

Table 4.4: Frequencies of assignment of each of the elementary linguistic terms $\mathcal{T}_1, \ldots, \mathcal{T}_5$ to the asymmetrical triangular fuzzy numbers from set Out_2 by linguistic approximation using each of the examined distance/similarity measures $d_1, d_2, d_3, d_4, s_1, s_2, s_3$ and s_4. Frequencies of cases where more than one linguistic terms were recommended are also presented as an ambiguous cases.

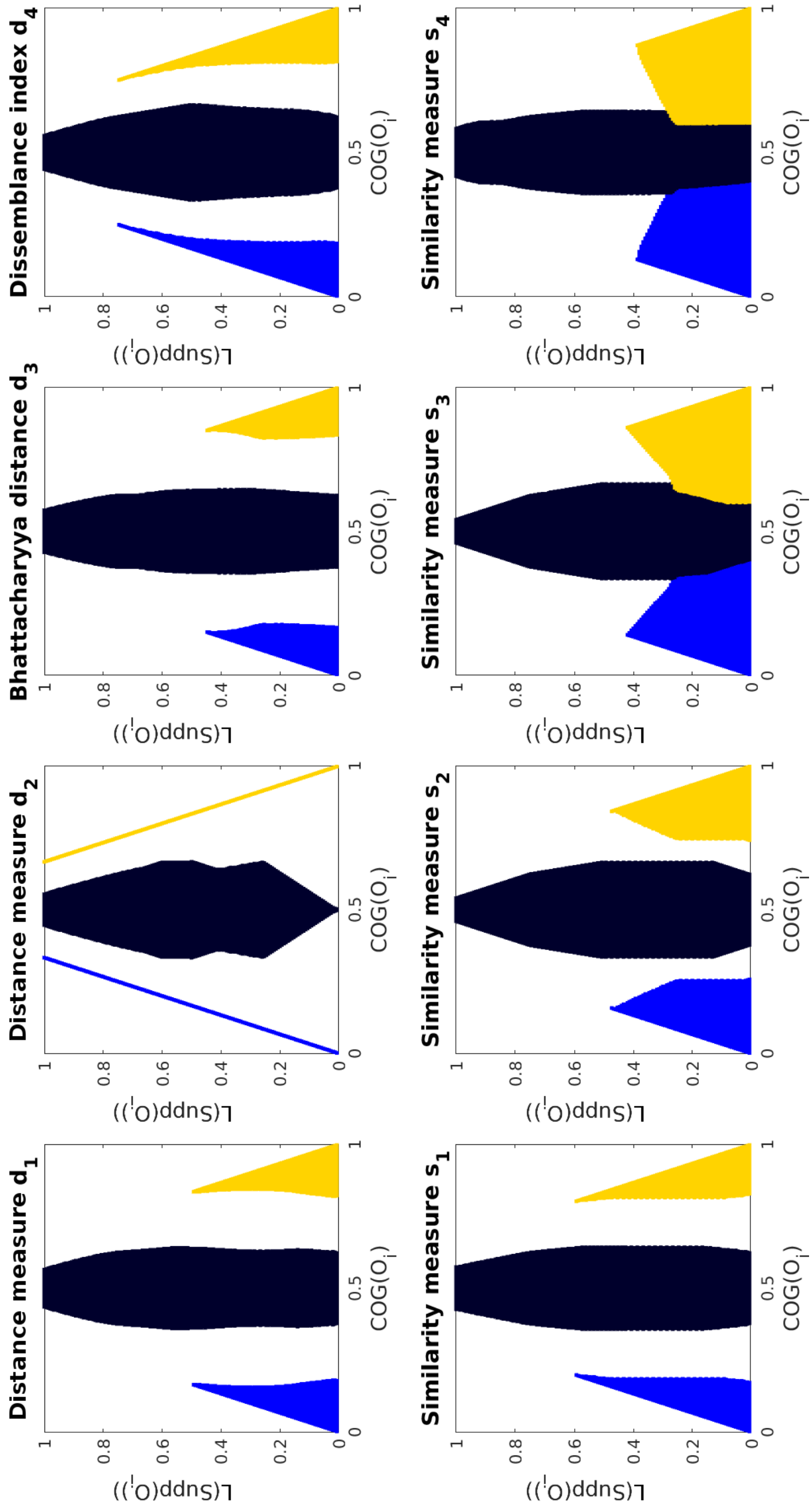


Figure 4.10: A graphical summary of the performance of the chosen distance and similarity measures in the linguistic approximation of asymmetrical triangular fuzzy number on $[0, 1]$ using 5-term linguistic scale. Colour represents three selected terms of the linguistic scale: $\mathcal{T}_1$ (blue), $\mathcal{T}_3$ (black) and $\mathcal{T}_5$ (yellow). The remaining linguistic terms $\mathcal{T}_2$ and $\mathcal{T}_4$ are depicted in Figure 4.11.

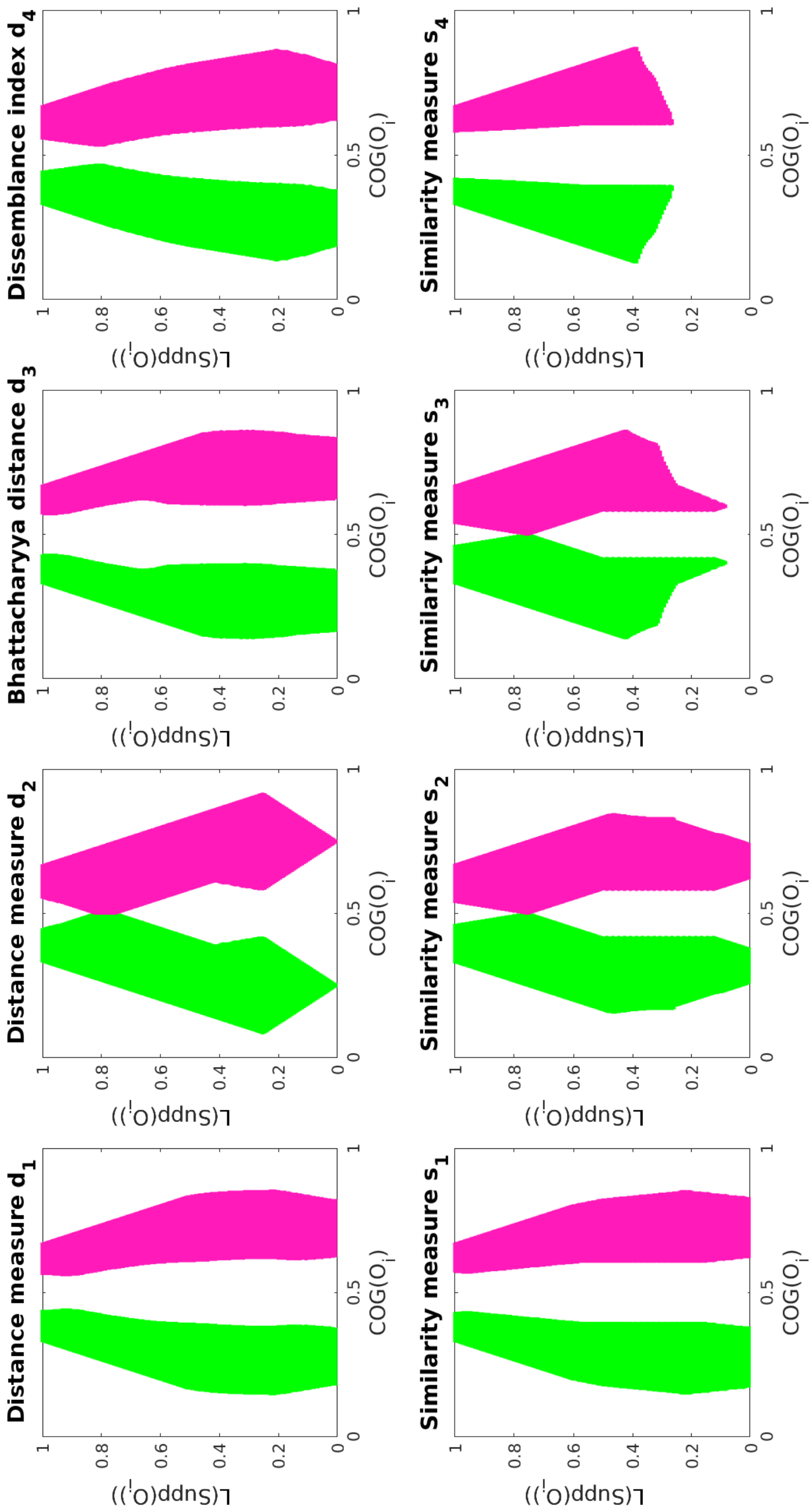


Figure 4.11: A graphical summary of the performance of the chosen distance and similarity measures in the linguistic approximation of asymmetrical triangular fuzzy number on $[0, 1]$ using 5-term linguistic scale. Colour represents two selected terms of the linguistic scale: $\mathcal{T}_2$ (green) and $\mathcal{T}_4$ (pink). The remaining linguistic terms $\mathcal{T}_1$, $\mathcal{T}_3$ and $\mathcal{T}_5$ are depicted in Figure 4.10.

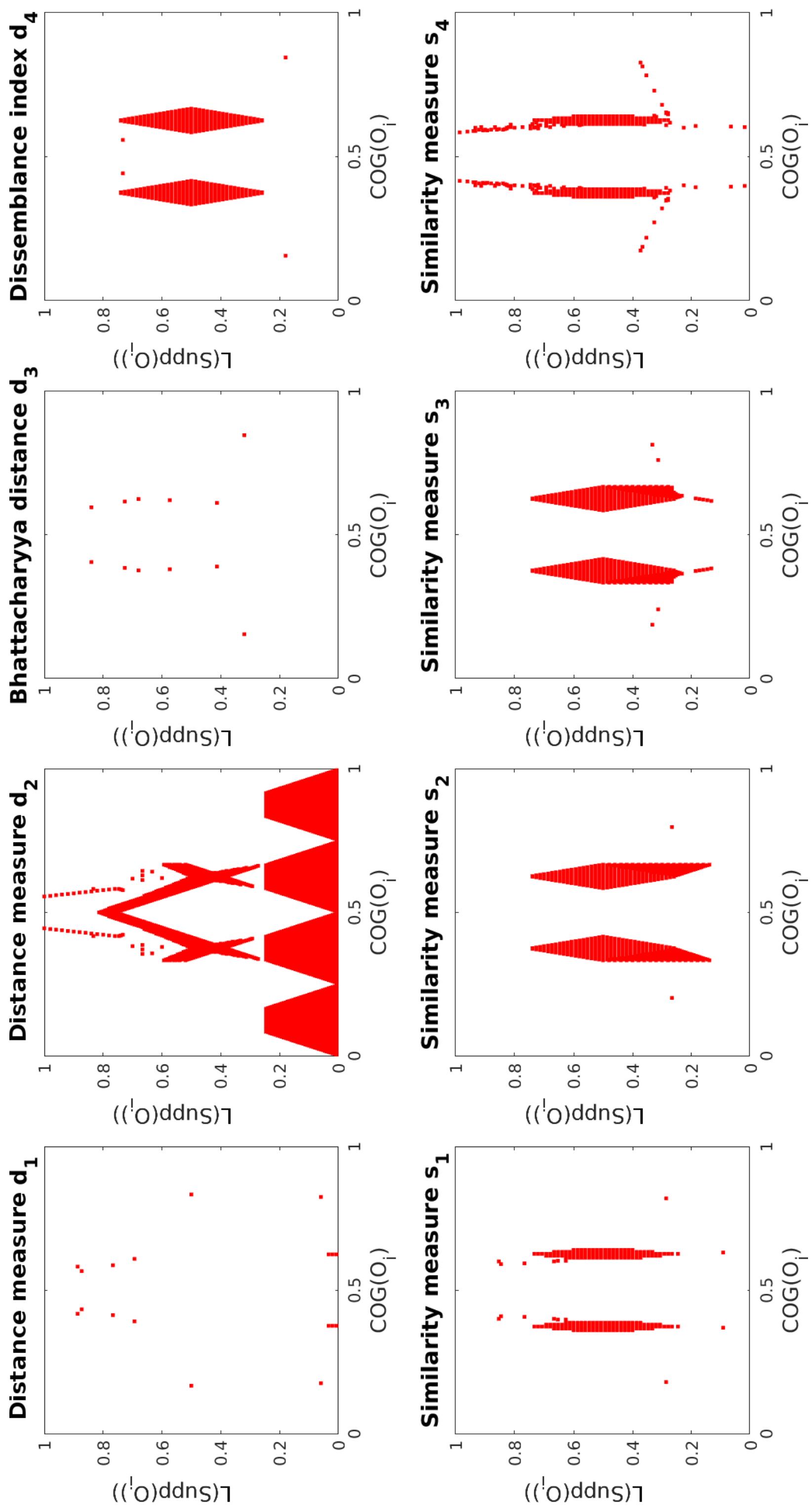


Figure 4.12: A graphical summary of the ambiguous cases (red areas) in linguistic approximation of asymmetrical triangular fuzzy numbers on $[0, 1]$ using 5-terms linguistic scale. All eight selected distance/similarity measures are considered.

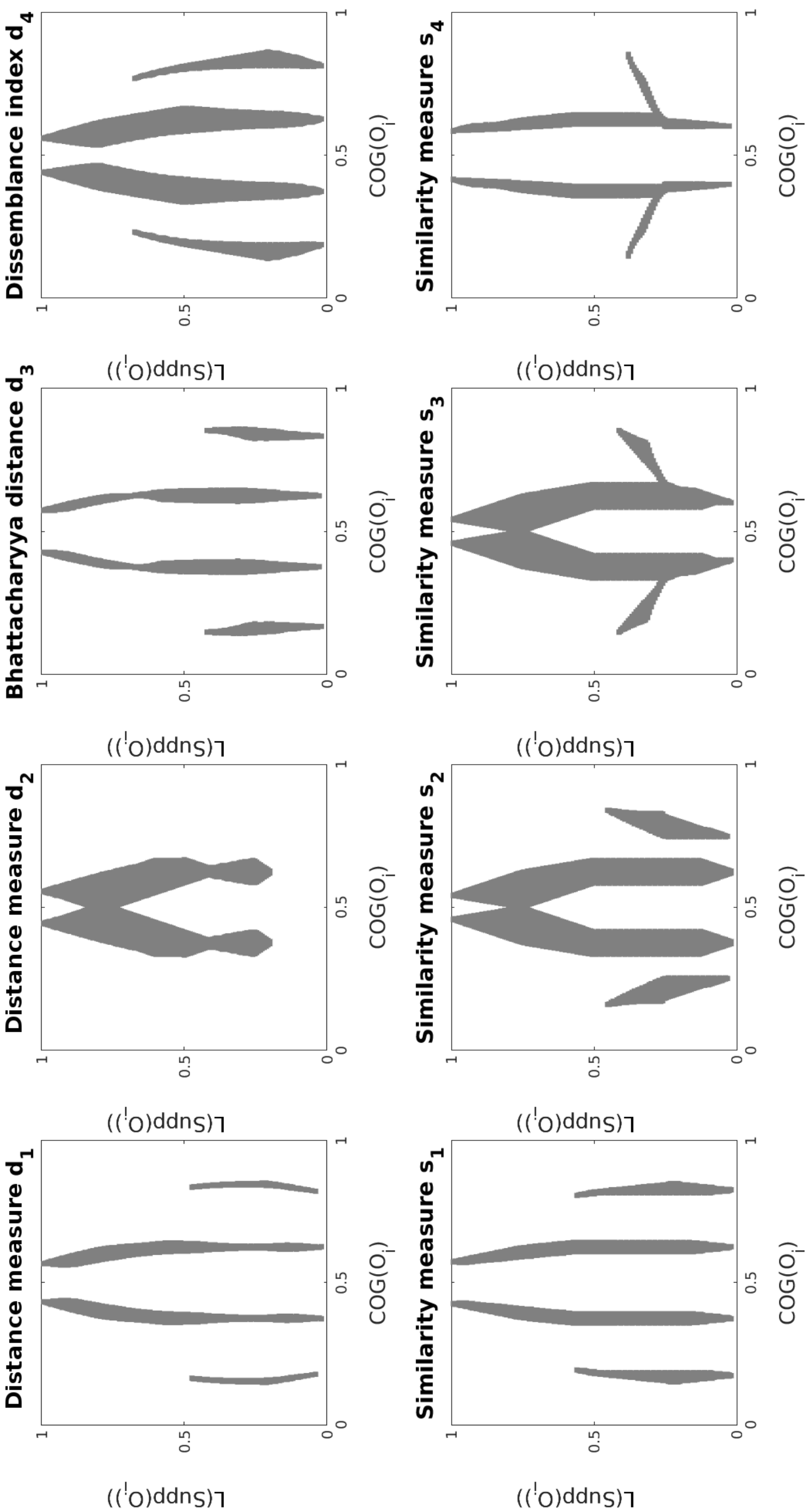


Figure 4.13: A graphical summary of the linguistic approximation grey zones of the chosen distance and similarity measures in the linguistic approximation of asymmetrical triangular fuzzy number on $[0, 1]$ using 5-terms linguistic scale.

As we already mentioned, the use of the $(\mathrm{COG}(O), \mathrm{L}(\mathrm{Supp}(O)))$ two-tuple representation (i.e. the requirement for two-dimensional graphical representation) in combination with asymmetrical fuzzy numbers as linguistically approximated objects introduces the grey zones. Each point in the grey zone represents a variety of asymmetrical fuzzy numbers approximated by different linguistic terms. The information on the relative frequency of the use of these terms is, so far, not available in our analysis method. It is however possible to obtain these relative frequencies and even to visualize them graphically using three-dimensional histograms as was proposed in Publication **XI**.

This framework was designed in a way to follow and extend the findings from previous analysis. Each fuzzy number from the set Out_2 is again represented by the two-tuple $(\mathrm{COG}(O), \mathrm{L}(\mathrm{Supp}(O)))$ and its linguistic approximation (using selected distance/similarity measure) is calculated. If the previous analysis method was applied, all the necessary data (which is possibly time consuming to obtain) is already available. The intervals $[0, 1]$ representing the universe for the center of gravity and the universe for the length of support are uniformly divided into n parts each. This introduces a uniform partition of the $[0, 1] \times [0, 1]$ universe into n times n two-dimensional areas. Each area represents a subset $Out_2^{i,j}, i = 1, \ldots, n, j = 1, \ldots, n$ of Out_2 that contains only assymetrical triangular fuzzy numbers from Out_2 with the respective $(\mathrm{COG}(O)$ and $\mathrm{L}(\mathrm{Supp}(O))$. Cardinalities of $Out_2^{i,j}, i = 1, \ldots, n, j = 1, \ldots, n$ define the three-dimensional histogram in Figure 4.15 (for $n = 20$ and $n = 10$) representing the distribution of approximated asymmetrical triangular fuzzy numbers obtained using the grid approach described previously[4].

At this point fuzzy numbers belonging to the same bin can still be linguistically approxited by different linguistic terms. To get clear insights into the actual relative frequencies of assignment of each of the linguistic terms in each bin the graphical representation presented in Figure 4.14 is suggested. Under this representation each linguistic term is assigned a three-dimensional histogram representing its usage (the relative frequencies of its use) in each bin for a selected distance/similarity measure. If a two-dimensional representation is prefered/required, a top-down view of the three-dimensional histograms can be used. In some cases this may be more suitable for paper (non-interactive) presentation of the results (see Figure 4.16).

Similar conclusions as those obtained for the previous graphical summaries can be derived for the three dimensional histograms. Since histograms are suggested here as an additional piece of information we leave a thorough analysis of these outputs to the interested readers and refer them to the Appendix A, where the results for all seven remaining distance/similarity measures can be found.

Analogously to the section 4.4.2, the findings from this section can be extended by using an enhanced linguistic scale instead of the linguistic scale. Again, fuzzy numbers from the set Out_2 can be used in the analysis and the results can be obtainded and visualized in analogy to the visualization proposed in this section. An example of such analysis for Bhattacharyya distance d_3 is depicted in Figure

[4]Slight asymmetry of the three-dimensional histograms in Figure 4.15 is caused by rounding.

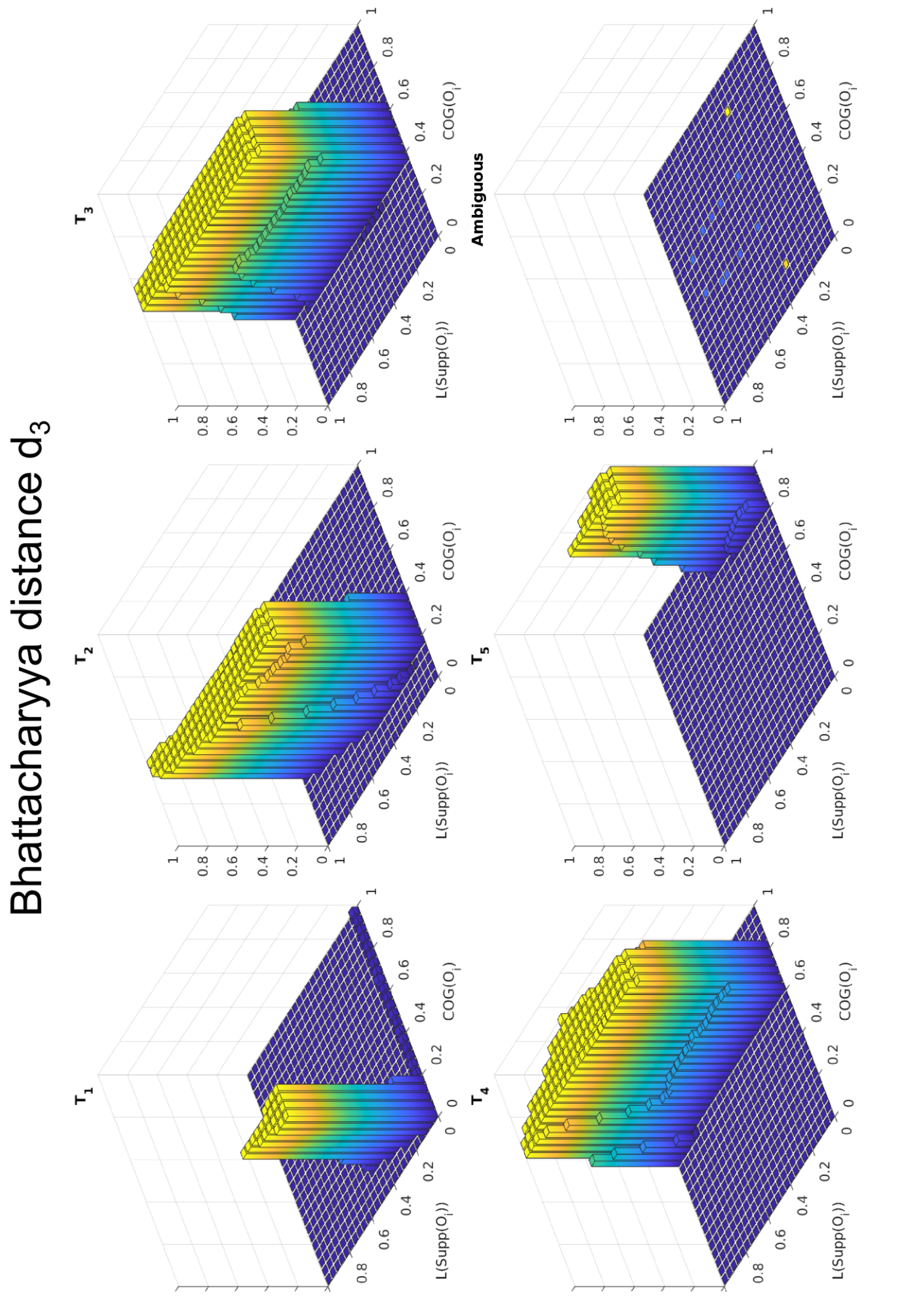


Figure 4.14: Three-dimensional histogram representation of the performance of Bhattacharyya distance d_3 in the linguistic approximation of asymmetrical triangular fuzzy numbers on $[0, 1]$ using a 5-term linguistic scale. Each subfigure summarizes the relative frequencies of suggesting the given linguistic term for the fuzzy numbers belonging to the respective bin (feature-wise). Fuzzy numbers that can be equally well approximated by more linguistic terms at the same time are depicted in the subfigure labelled ambiguous.

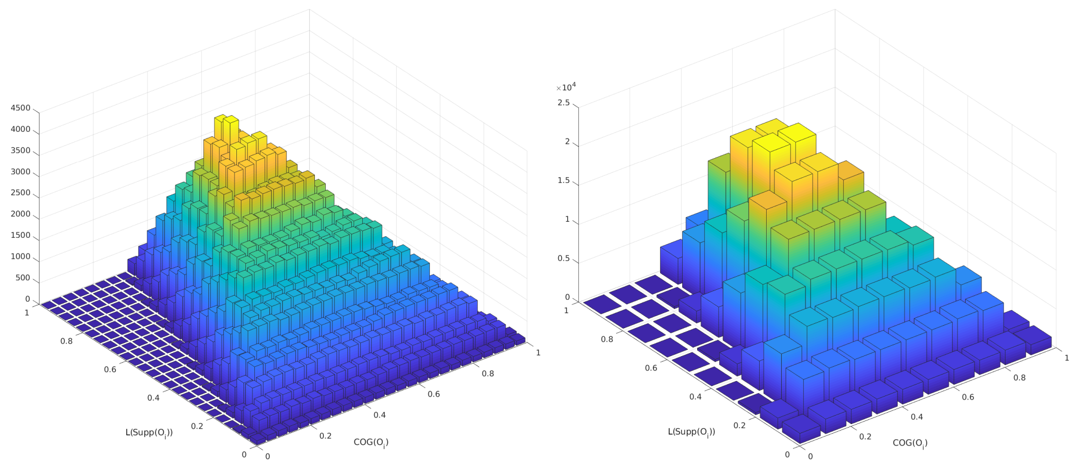


Figure 4.15: Three-dimensional histogram representation of the absolute frequencies of the asymmetrical triangular fuzzy numbers from the set Out_2 in each bin. 20 times 20 bin representation (left) and 10 times 10 bin representation (right).

4.17. This graphical representation however requires eight subfigures to properly describe the possible results of linguistic approximation using one distance/similarity measure. Due to the limited space within the thesis the graphical summaries of the results for the remaining seven distance/similarity measures are available in the Appendix B.

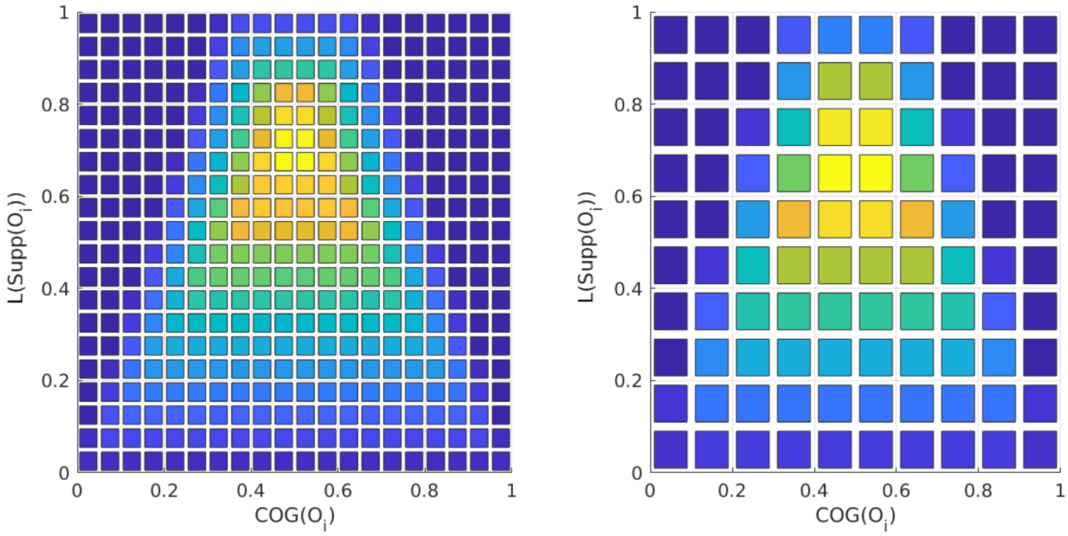


Figure 4.16: Top-down view of the three-dimensional histograms depicted in Figure 4.15. 20 times 20 bin representation (left) and 10 times 10 bin representation (right). This way three dimensional histograms are converted into “heat maps”.

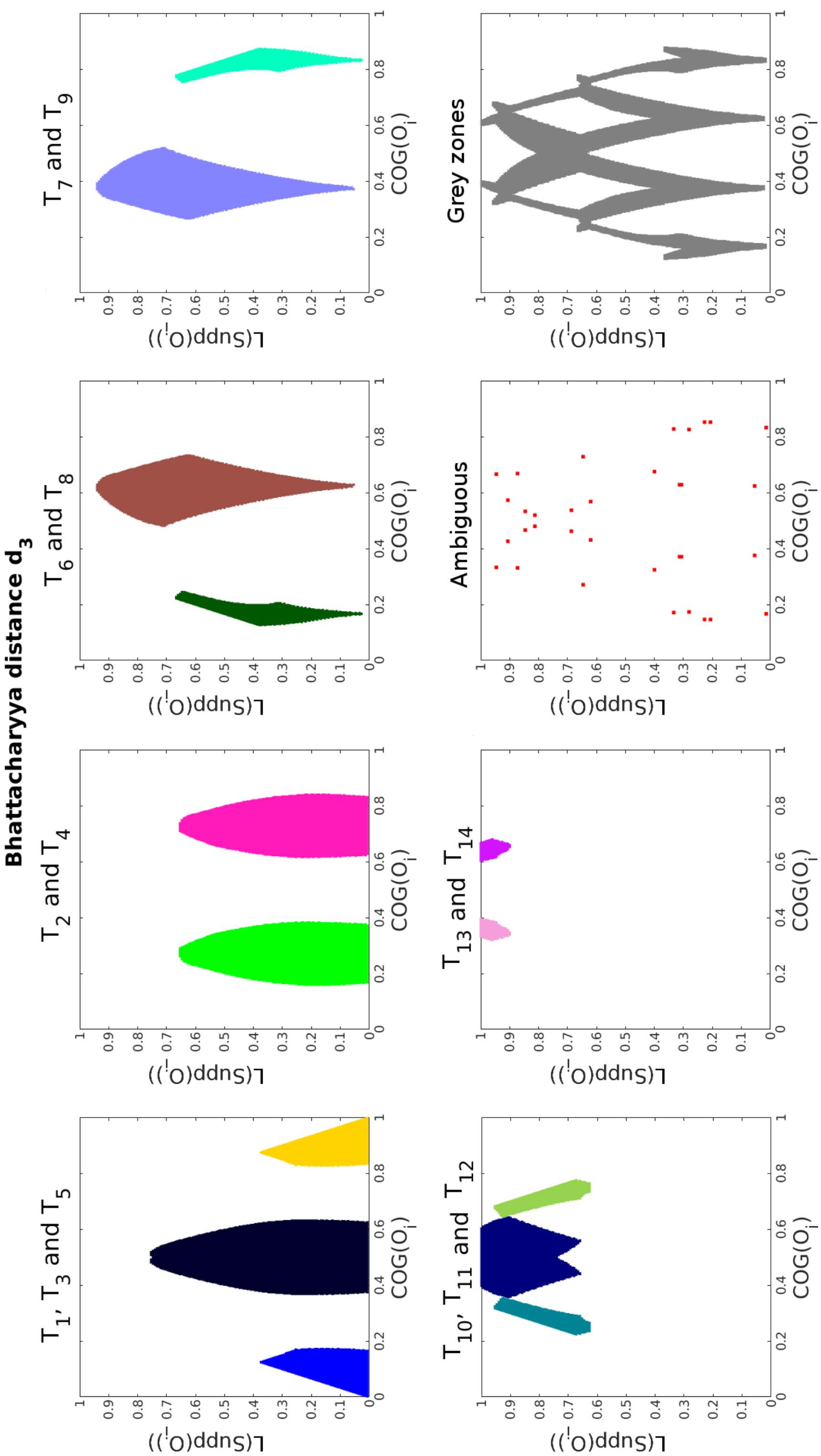


Figure 4.17: A graphical summary of the results of linguistic approximation of asymmetrical triangular fuzzy numbers using the Bhattacharyya distance d_3 and a 5-term enhanced linguistic scale. Each colour represents one term of the enhanced linguistic scale, the assignment of colours to the linguistic terms is indicated above each subplot (linguistic terms that are never assigned to any element of Out_2 are not considered in the summary). Red colour is reserved for ambiguous cases, i.e. cases when more than one linguistic term is assigned and grey colour represents the grey zones.

4.6 A note on the linguistic approximation of more general objects: Mamdani-type fuzzy sets

In previous sections we were studying the performance of linguistic approximation of triangular fuzzy numbers under selected distance/similarity measures and discussed their possible extension to trapezoidal fuzzy numbers. Several methods for analysis of the performace of these measures were proposed. However, in this section we will propose methods for the analysis of performance of linguistic approximation of more general Mamdani-type fuzzy sets[5]. These fuzzy sets can be obtained as the outputs of Mamdani fuzzy inference [20] and the performance of the linguistic approximation of such outputs was already studied in Publication **VIII**.

Before we start with the analysis of the performance, we will shortly introduce the process of the Mamdani fuzzy inference. Let $(\mathcal{U}_j, \mathcal{T}(\mathcal{U}_j), X_j, G_j, M_j)$, $j = 1, \ldots, m$ be m linguistic scales representing the inputs of the fuzzy inference system and let $(\mathcal{V}, \mathcal{T}(\mathcal{V}), Y, G, M)$, $j = 1, \ldots, m$ be a linguistic scale representing the output of the fuzzy inference system. Let

$$\text{If } \mathcal{U}_1 \text{ is } \mathcal{A}_{i1} \text{ and } \ldots \text{ and } \mathcal{U}_m \text{ is } \mathcal{A}_{im}, \text{ then } \mathcal{V} \text{ is } \mathcal{B}_i,$$

be a set of n rules describing the relationship between the input and the output linguistic variables, where $\mathcal{A}_{ij} \in \mathcal{T}(\mathcal{U}_j)$ and $\mathcal{B}_i \in \mathcal{T}(\mathcal{V})$, $M_j(\mathcal{A}_{ij}) = A_{ij}$ and $M(\mathcal{B}_i) = B_i$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Then for the input $(A'_1, A'_2, \ldots, A'_m)$ consisting of m fuzzy sets A'_j defined on X_j, $j = 1, \ldots, m$ the output of the fuzzy inference Out_M (Mamdani-type fuzzy set on Y) is computed using:

$$Out_M(A'_1, A'_2, \ldots, A'_m) = (A'_1 \times \cdots \times A'_m) \circ \bigcup_{i=1}^{n} (A_{i1} \times \cdots \times A_{im} \times B_i). \tag{4.13}$$

An example of such output is depicted in figure 4.18. Please note that the output is not a fuzzy number on Y. In general the height of Out_M need not be 1 or the fuzzy set need not be unimodal (its α-cuts may not be closed intervals for all $\alpha \in (0, 1]$). As long as Out_M is a general fuzzy set and not a fuzzy number, dissemblance index d_4 and of all four similarity measures $s_1, \ldots s_4$ investigated in this thesis from the analysis can not be used, because these measures are not applicable on general fuzzy sets.

In the further text we will restrict ourselves to cases when the output linguistic variable $\mathcal{V}$ contains five elementary linguistic terms $\mathcal{B}_1, \ldots, \mathcal{B}_5$ and their meanings are represented by triangular fuzzy numbers $B_1 = (0, 0, 0.25), B_2 = (0, 0.25, 0.5), B_3 = (0.25, 0.5, 0.75), B_4 = (0.5, 0.75, 1)$ and $B_5 = (0.75, 1, 1)$ respectively, all defined on $[0, 1]$. The same linguistic scale was used in the analysis of the performance of the linguistic approximation of triangular fuzzy numbers in sections 4.4 and 4.5. For example let us consider Mamdani-type outputs obtained as a union of three

[5] In the case of another frequently used fuzzy inference proposed by Sugeno[29] the issue of linguistic approximation is much less complicated, since the usual output of Sugeno fuzzy inference is a real number.

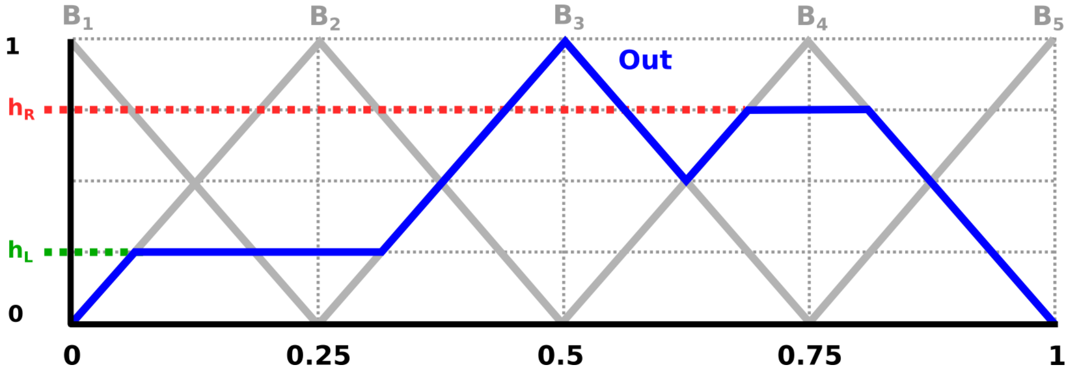


Figure 4.18: Example of Mamdani-type fuzzy set.

neigbouring fuzzy sets as depicted in Figure 4.18. Formally such outputs can be represented as

$$Out_M(x) = \max\left\{\min\{h_L, B_2(x)\}, \min\{h_M, B_3(x)\}, \min\{h_R, B_4(x)\}\right\}, \qquad (4.14)$$

where $h_L, h_M, h_R \in [0,1]$. Assuming, that $h_M = 1$, each such fuzzy set can be uniquely represented using two-tuple (h_L, h_R). Note that the two-tuple representation using (h_L, h_R) corresponds well with the need to approximate outputs of Mamdani fuzzy inference, because $h_L(h_M, h_R)$ can be interpreted as the maximum firing strength of a rule resulting in $\mathcal{B}_2(\mathcal{B}_3, \mathcal{B}_4)$ respectively. Such a representation thus uses representation directly obtainable from the fuzzy inference. Therefore, similarly as in the case of symmetrical triangular fuzzy numbers, each approximated fuzzy sets can be unambiguously represented as a point in the two-dimensional space. Colors of these points can again represent the result of the linguistic approximation.

In line with the previous analysis, it is now necessary to generate a set all types of fuzzy sets that can be represented by 4.14 to be linguistically approximated. Again, it is possible to randomly generate this set, but in the further text we will employ a more systematic approach which is based on the grid approach, as in the Publication **VIII**. The two $[0,1]$ intervals are uniformly divided into 1 001 points each and these points represent values h_L and h_R. Using the cartesian product we obtain set Out_3 that contains 1002001 two-tuples that represent the approximated fuzzy sets. Unlike in the previous sections, no further restriction need to be introduced because each of these two-tuples represents a fuzzy set obtainable by 4.14.

Each of the distance measures d_1, d_2 and d_3 is applied to identify the linguistic approximation of each fuzzy set from the set Out_3 using an enhanced linguistic scale derived from the elementary linguistic terms $\mathcal{B}_1, \ldots, \mathcal{B}_5$. Based on the direct comparison of results depicted in Figure 4.19 and Table 4.5 we can draw for example the following conclusions:

- Linguistic approximation in all three cases can results in any of the four linguistic terms $\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}, \mathcal{B}_{24}$ or it is indecisive among two or more of them

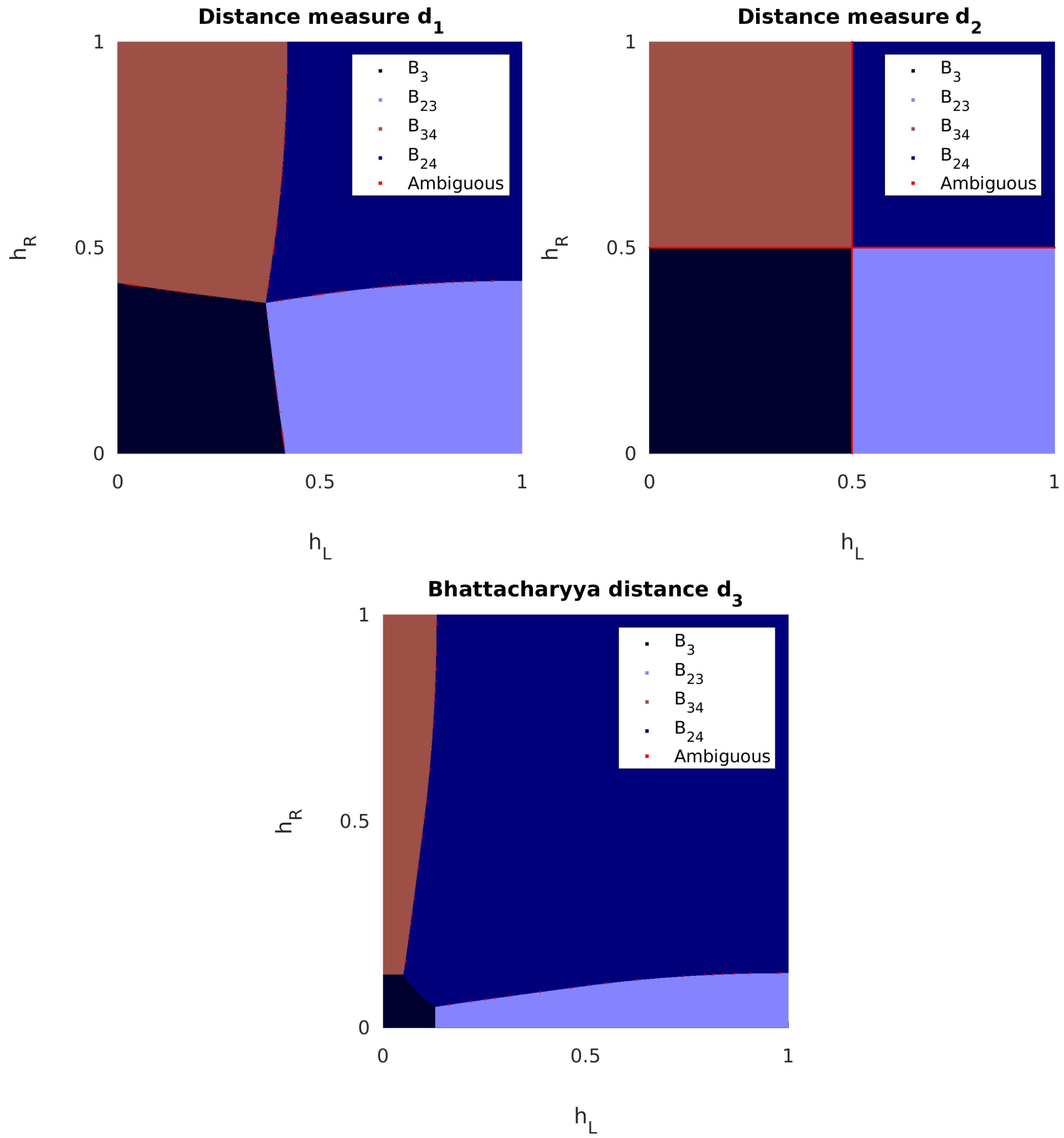


Figure 4.19: The results of the linguistic approximation of 1 002 001 fuzzy sets from Out_3 using distance measures d_1, d_2 and d_3. Enhanced linguistic scale with five elementary terms is used. Approximated Mamdani-type fuzzy sets are represented by points with coordinates (h_L, h_R). The colour of each point represents the linguistic term assigned as the most appropriate linguistic approximation for the given fuzzy set, colours represent the same linguistic terms (elementary and derived) as in the previous sections of the thesis. Ambiguous cases are depicted using red colour.

(such situations are denoted as ambiguous). Only one of the possible results of linguistic approximation is an elementary linguistic term $\mathcal{B}_3$. This is understandable, since a union of three neighboring fuzzy sets (modified meanings of $\mathcal{B}_2$ and $\mathcal{B}_4$, and the meaning of $\mathcal{B}_3$) is being approximated.

- From the Figure 4.19 it is evident, that the results of linguistic approxima-

	$\mathcal{B}_3$	$\mathcal{B}_{23}$	$\mathcal{B}_{34}$	$\mathcal{B}_{24}$	**Ambiguous**
d_1	151 383	246 741	246 741	357 042	94
d_2	250 000	250 000	250 000	250 000	2 001
d_3	13 128	89 698	89 698	809 437	40

Table 4.5: Frequencies of assignment of linguistic terms $\{\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}, \mathcal{B}_{24}\}$ and Ambiguous cases obtained for 1 002 001 fuzzy sets from Out_3 in the linguistic approximation using d_1, d_2 and d_3. Unlisted linguistic terms (e.g. $\mathcal{B}_1, \mathcal{B}_{13}, \dots$) were not assigned to any approximated Mamdani-type fuzzy set. Frequencies of ambiguous cases are strongly dependent on the chosen partition of $[0, 1]$.

tion depend significantly on the choice of the distance measure. For example distance measure d_2 tends to overuse the elementary linguistic term $\mathcal{B}_3$ and it is the only one for which the knowledge of two thresholds (one for h_L one for h_R) contains the full information concerning the result of the linguistic approximation.

- The result of linguistic approximation in the case of distance measure d_2 can be established solely on the values of h_L and h_R - this is evident from the "square-shaped" structure. If h_L or H_R is equal to 0.5, the linguistic approximation results in an ambiguous case. If h_L and h_R are both lower then 0.5, the fuzzy set is approximated by elementary term $\mathcal{B}_3$. On the other hand, if both h_L and h_R are greater than 0.5, the linguistic approximation results in the term $\mathcal{B}_{24}$. The remaining cases result in $\mathcal{B}_{23}$ ($\mathcal{B}_{34}$), if $h_L > 0.5$ and $h_R < 0.5$ ($h_L < 0.5$ and $h_R > 0.5$). Moreover the Table 4.5 shows that each linguistic term was assigned to the exactly same number of fuzzy sets. Also note, that for $h_L = h_R = 0.5$ we obtain a fuzzy set that can be linguistically approximated equally well by all four linguistic terms $\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}, \mathcal{B}_{24}$. This means that if the values h_L and h_R of the approximated fuzzy set are both close to 0.5, even a small change of h_L or h_R could result in the assignment of a different linguistic label.

- Use of the Bhattacharyya distance d_3 results in the most cases (87.78%) in the most uncertain linguistic term $\mathcal{B}_{24}$ (the only level 3 linguistic term from the set of resulting terms). On the contrary, the elementary linguistic term $\mathcal{B}_3$ is assigned rather sparsely (1.31%). Moreover, if the value of h_L or h_R is higher than ≈0.13, the term $\mathcal{B}_3$ is never assigned. Similarly, if h_L and h_R are both higher than ≈0.135, the only possible resulting linguistic term is $\mathcal{B}_{24}$. This is in compliance with the findings from the previous section where we have shown that the Bhattacharyya distance tends to suggest linguistic terms that are supersets of the approximated object meaning-wise.

- Distance measure d_1 exhibits features similar to both d_2 and d_3. In line with d_3 it suggests the level 3 term $\mathcal{B}_{24}$ most frequently (35.63%), but its requirements on the approximating fuzzy number being a superset of the approximated object is weaker (the linguistic term $\mathcal{B}_3$ is suggested in 15.11% of the cases compare to the 1.31% frequency under d_3). In this sense it seems to be a compromise between d_2 and d_3. Again as in the case of d_2 there exist a fuzzy set that can be linguistically approximated by all four obtainable linguistic terms. Its representation is $((\sqrt{3}-1)/2, (\sqrt{3}-1)/2)$.

- Note, that based on the Table 4.5 the ambiguous cases seem to be rather exceptional - 94 and 40 in the cases of d_1 and d_3 respectively. Only in the case of distance measure d_2, the linguistic approximation resuls in 2 001 ambiguous cases. This is caused by the fact, that if h_L or h_R are equal to 0.5 the linguistic approximation will automatically result in an ambiguous case (direct result of the particular choice of grid). However, using all three distance measures there is a possibility of obtaining infinitely many ambiguous cases on the boarders of differently coloured areas (again depending on the choice of grid).

To conclude with, the distance measure d_2 is the fastes to compute (it has the simplest computation formula). In this case, the formula is no even needed, since the knowledge of the thresholds $h_L = 0.5$ and $h_R = 0.5$ is sufficient to be able to determine the result of linguistic approximation. Its overuse of the elementary linguistic term $\mathcal{B}_3$ even for fuzzy sets for which h_L and h_R are close to 0.5 can however be considered an undesirable property. Fuzzy sets with high uncertainty are in these cases assigned a linguistic approximation with a low-uncertain meaning. If the desired linguistic approximation should be such that its meaning is a superset of the approximated fuzzy set, Bhattacharyya distance d_3 seems to be the method of choice. If the superset property is not that important, the distance measure d_1 represents a reasonable compromise between d_3 and the simple d_2.

We have thus presented the possible application of the analytic framework on a selected type of Mamdani-type fuzzy set. In general, we can assume n elements in the underlying linguistic scale. In this case a general Mamdani-type output can be expressed as $\max\{\min\{h_1, B_1(x)\}, \ldots, \min\{h_n, B_n(x)\}\}$ and thus represented by an n-tuple $(h_1, \ldots, h_n)$ which would require an n-dimensional graphical representation for the analysis.

5 Linguistic approximation of fuzzy numbers using fuzzy 2-tuples

In the previous section we have presented several methods for the analysis of the outputs of linguistic approximation using different distance/similarity measures. All of these methods assumed a finite set of linguistic labels to be assigned as a result of linguistic approximation (either standard or enhanced linguistic scales were assumed). This section presents a new method for the linguistic approximation of fuzzy numbers that was proposed in publication **IV**. This new method is based on the idea of fuzzy 2-tuples that was introduced by Herrera and Martinez [12, 13].

This method stresses the ability of its user to interpret the results easily. It therefore works only with elementary linguistic terms i.e. with a chosen linguistic scale. The user of this method is thus not required to understand the grammar used to derive new linguistic terms from the elementary ones, he/she also does not need to be able to define their meaning. The restriction on the elementary linguistic term set usually means, that the ability of the linguistic approximation to capture the meaning of the approximated objects that are far from the meanings of the elementary terms is compromised to some extent. We therefore propose a compensation of this disadvantage that in fact offers infinitely many possible results of the linguistic approximation. All of the results use the elementary linguistic terms as the center-piece of information. A second piece of information is provided that describes the possible shift of meaning of the elementary terms (including the direction of this shift).

Given that the approximating linguistic scale frequently uses only a very limited number of linguistic terms, it is easy to check (and ensure) that the user of the method understands them well. Also note, that the elementary linguistic terms of a linguistic scale can be ordered. The results of the linguistic approximation are in this new method provided as 2-tuples $(\mathcal{T}, \beta)$, where $\mathcal{T}$ is the resulting linguistic term and β is a real number representing the shift of the meaning of the linguistic term $\mathcal{T}$, see [12, 13]. For example instead of forcing either of the neighboring linguistic terms *Average* and *Good* as the result on linguistic approximation, we can also consider *Average* $+ \Delta_1$ or *Good* $- \Delta_2$, where Δ_1 and Δ_2 are non-negative real numbers. The known meanings of the linguistic terms can thus be shifted closer to each other. The original meanings of the linguistic terms remain unchanged, their ordering as well and the operation of shifting the meaning to the left or to the right on a real number universe is simple enough to understand intuitively. Moreover, the fuzzy 2-tuple representation can be transformed into a fully linguistic one, e.g. *slightly better than Average*. In the next subsection, the linguistic approximation method using fuzzy 2-tuples will be summarized.

5.1 Proposed method for linguistic approximation

In the following text we will assume that the outputs of the mathematical model that we are going to linguistically approximate are fuzzy numbers defined on the interval $[a, b]$, $a < b$, $a, b \in \mathbb{R}$. For the purpose of linguistic approximation we define linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), [a, b], G, M)$, where $\mathcal{T}_e(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_n\} \subseteq \mathcal{T}(\mathcal{V})$ constitutes a set of n elementary linguistic terms ($n > 1$), their meanings $\mathcal{M}(\mathcal{T}_i) = T_i, i = 1, \ldots, n$ are represented by triangular fuzzy numbers that form a uniform linguistic scale on $[a, b]$:

$$\begin{aligned} M(\mathcal{T}_1) &= (a, a, a + \Delta), \\ M(\mathcal{T}_i) &= (a + (i-2) \cdot \Delta, a + (i-1) \cdot \Delta, a + i \cdot \Delta), \; i = 2, \ldots, n-1, \\ M(\mathcal{T}_n) &= (b - \Delta, b, b), \end{aligned} \tag{5.1}$$

where $\Delta = (b-a)/(n-1)$. Note, that when the linguistic variable is defined on $[0,1]$ and has five elementary linguistic terms, we obtain the linguistic scale that is depicted in Figure 4.1 and is used in previous section.

A fuzzy set F which is a fuzzy superset of any fuzzy set on $[a,b]$ is defined as

$$F(x) = \begin{cases} 1 & \text{if } x \in [a,b], \\ 0 & \text{if } x \notin [a,b], \end{cases} \tag{5.2}$$

which can also be denoted as $F \sim (a,a,b,b)$.

As was already stated because the set of elementary linguistic terms is limited in some cases we might not be able to select the linguistic term, that would fit the the approximated fuzzy number well enough. In Publication **VII** we have shown that this can become a serious issue e.g. when the approximating linguistic terms can be categorized into gains and loses. In this particular case low granularity of the approximating scale (particularly when only elementary linguistic terms are considered) can result in a gain being approximated by a linguistic term with a loss "meaning" and vice versa. To resolve this we will extend the set of elementary linguistic terms $\mathcal{T}_e(\mathcal{V})$ into $\mathcal{T}(\mathcal{V})$ by allowing for the shifting of the linguistic terms to the right or to the left within their universe of discourse:

$$\mathcal{T}(\mathcal{V}) = \mathcal{T}_e(\mathcal{V}) \cup \{\mathcal{T}_i^\beta;\ \beta \in [-0.5\cdot\Delta, 0.5\cdot\Delta),\ i = 1,\ldots,n\},$$

where $\mathcal{T}_i^\beta$ denotes "$\mathcal{T}_i$ *shifted by* β" and can be denoted by 2-tuple $(\mathcal{T}_i, \beta)$. $\mathcal{T}_i$ represents the elementary linguistic term from the set $\mathcal{T}_e(\mathcal{V})$ and β represents a shift of the meaning of the linguistic term $\mathcal{T}_i$. If $0 < \beta < 0.5\cdot\Delta$ then the meaning of the linguistic term is shifted to the right. Analogously if $-0.5\cdot\Delta \leq \beta < 0$ then the meaning of the linguistic term $\mathcal{T}_i$ is shiftet to the left. If $\beta = 0$ the meaning of the linguistic term $\mathcal{T}_i$ does not change and $\mathcal{T}_i^\beta$ coincides with $\mathcal{T}_i$. The shift of the meaning of linguistic term $\mathcal{T}_i$ by $\beta > 0$ is depicted in Figure 5.1.

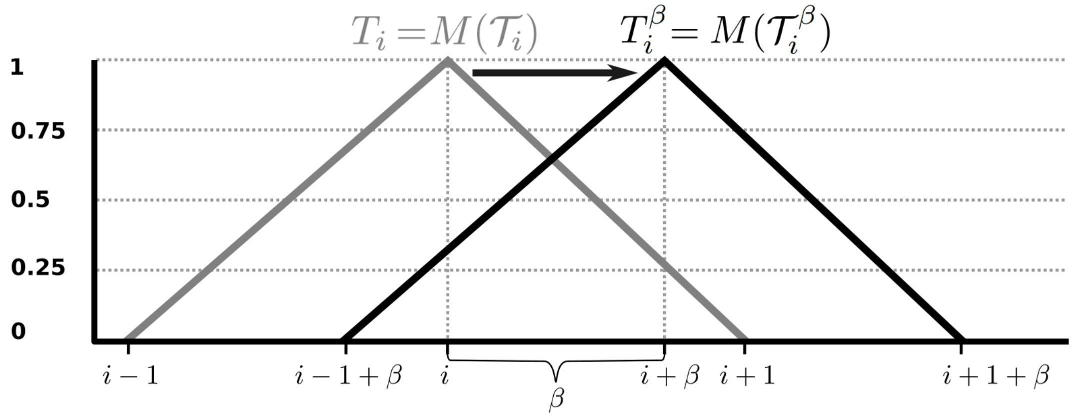


Figure 5.1: The meaning of the linguistic term $\mathcal{T}_i$ (grey) and the meaning of the linguistic term $\mathcal{T}_i$ shifted by β (black); $\mathcal{T}_i, \mathcal{T}_i^\beta \in \mathcal{T}(\mathcal{V})$

Once we have introduced the shifted linguistic terms, thus defining infinitely many elements of $\mathcal{T}(\mathcal{V})$ it is necessary to define semantic rule that describes how to assign a fuzzy-number meaning, $T_i^\beta = M(\mathcal{T}_i^\beta)$ to every element of the set of linguistic labels $\mathcal{T}(\mathcal{V})$.

Introducing the shift may result in some of the meanings of the linguistic terms "sliding out" of the $[a,b]$. For computational purposes we therefore extend the universe $[a,b]$ to $[a-\Delta, b+\Delta]$ and modify the meaning of $\mathcal{T}_1$ and $\mathcal{T}_n$ from T_1 to T_1' and from T_n to T_n' in the following way:

$$\begin{array}{lcl} T_1' & \sim & (a-0.5\cdot\Delta, a-0.5\cdot\Delta, a, a+\Delta), \\ T_n' & \sim & (b-\Delta, b, b+0.5\cdot\Delta, b+0.5\cdot\Delta). \end{array} \tag{5.3}$$

The meaning of the remaining elementary linguistic terms remains unaltered (hence $T_i = T_i'$, $i = 2, \dots, n-1$). Please note, that this process is just a technicality and thus not results in the change of the meaning of the elementary linguistic terms within the $[a,b]$ interval. Finally, the fuzzy numbers representing the meaning of the shifted linguistic terms are computed using:

$$M(\mathcal{T}_i^\beta) = (T_i' + \beta) \cap F, \ i = 1, \dots, n, \ \beta \in [-0.5\cdot\Delta, 0.5\cdot\Delta). \tag{5.4}$$

However, it is not reasonable to shift the meaning of linguistic term $\mathcal{T}_1$ ($\mathcal{T}_n$) to the left (right), since such an operation results in a fuzzy number the restriction of which on the interval $[a,b]$ is no longer normal. For this reason $\beta \in [0, 0.5\cdot\Delta)$ and $\beta \in [-0.5\cdot\Delta, 0]$ will be considered in the case of $\mathcal{T}_1^\beta$ and $\mathcal{T}_n^\beta$ respectively. Having defined the shifts and their meaning this way, the ordering of the the shifted terms $\mathcal{T}_i^{\beta_1}$ and $\mathcal{T}_j^{\beta_2}$ depends only on the ordering of $\mathcal{T}_1$ and $\mathcal{T}_j$, for $i \neq j$, $i, j = 1, \dots, n$.

The result $\mathcal{T}_{i^*}^{\beta^*}$ of linguistic approximation of fuzzy number Out on $[a,b]$ is computed using:

$$M(\mathcal{T}_{i^*}^{\beta^*}) = \begin{cases} \underset{\beta\in[0,0.5\cdot\Delta)}{\arg\min}\, d(Out, T_1^\beta) & \text{if } i = 1, \\ \underset{\beta\in[-0.5\cdot\Delta,0.5\cdot\Delta)}{\arg\min}\, d(Out, T_i^\beta) & \text{if } i = 2, \dots, n-1, \\ \underset{\beta\in[-0.5\cdot\Delta,0]}{\arg\min}\, d(Out, T_n^\beta) & \text{otherwise,} \end{cases} \tag{5.5}$$

where d is the selected distance measure of fuzzy numbers (similarity measure can also be used, but the arg min must be substituted by arg max in the previous formula). The resulting linguistic term $\mathcal{T}_{i^*}^{\beta^*}$ can be denoted using the fuzzy 2-tuple as $(\mathcal{T}_{i^*}, \beta^*)$. The linguistic approximation proposed in this chapter will therefore be refered to as *fuzzy 2-tuple linguistic approximation* in the further text.

Moreover, if the user of the model prefers a fully linguistic description of the evaluation, we can use for example Table 5.1 that suggests a linguistic intrerpretation of values of β. For example if the result of the proposed method is a 2-tuple (*Average*, -0.3) the resulting fully linguistic description would be *worse than Average*. We assume here that the user of such a linguistic approximation understands this as worst than average but definitely better than the previous value of the linguistic scale. For novice users this can be a direct part of the linguistic approximation output.

Another benefit of the use of fuzzy 2-tuples for linguistic approximation is that the results of such an approximation can be ordered. In other words $(\mathcal{T}_i, \beta)$ is preferred over $(\mathcal{T}_j, \gamma)$ if $i > j$ or if $i = j$ and $\beta > \gamma$, $i, j = 1, \ldots, n$, assuming a benefit-type scale.

Negative β value	Linguistic description	Positive β value	Linguistic description
$[-0.05 \cdot \Delta, 0)$	About	$(0, 0.05 \cdot \Delta]$	About
$[-0.2 \cdot \Delta, -0.05 \cdot \Delta)$	Slightly worse than	$(0.05 \cdot \Delta, 0.2 \cdot \Delta]$	Slightly better than
$[-0.35 \cdot \Delta, -0.2 \cdot \Delta)$	Worse than	$(0.2 \cdot \Delta, 0.35 \cdot \Delta]$	Better than
$[-0.5 \cdot \Delta, -0.35 \cdot \Delta)$	Noticeably worse than	$(0.35 \cdot \Delta, 0.5 \cdot \Delta)$	Noticeably better than

Table 5.1: An example of a posssible partition of the feasible β values with the respective linguistic descriptions.

Several examples of the linguistic approximation of triangular fuzzy numbers $A \sim (0.4, 1.6, 2.8)$, $B \sim (1.6, 3.2, 3.2)$ and $C \sim (2.4, 1, 1)$ on $[0, 4]$ interval using fuzzy 2-tuples and similarity measure s_4 are depicted in Figure 5.2. The symmetrical triangular fuzzy number A is linguistically approximated by the linguistic term $\mathcal{T}_3^{-0.4}$ represented by the fuzzy number $T_3^{-0.4}$. Note, that the kernels of both the approximated and the approximating fuzzy numbers coincide - this is an expected behaviour, because both fuzzy numbers are symmetrical and triangular. Results of the linguistic approximation of asymmetrical triangular fuzzy numbers are demonstrated in the remaining cases. Fuzzy numbers B and C have the same shape, but the second one is “shifted” to the right. In the case of fuzzy number B the resulting approximation is linguistic term $\mathcal{T}_4$ with $\beta = -0.33$, i.e. with meaning shifted to the left. Even though the kernel of $T_4^{-0.33}$ lies more to the left than the kernel of T_4 (which is already closer to zero than the kernel of B), its center of gravity is closer to the center of gravity of B than the center of gravity of T_4 is. Since s_4 is based on the area, perimeter, center of gravity and significant points of fuzzy numbers this is an expected behaviour of the linguistic approximation (the area and perimeter of T_4 and $T_4^{-0.33}$ are identical). However, in the case of fuzzy number C, the result of the linguistic approximation is the linguistic term $\mathcal{T}_5$ with $\beta = -0.3$. We can see that the meaning of the approximating linguistic term $M(\mathcal{T}_5^{0.3})$ is a trapezoidal fuzzy number. Moreover, this fuzzy number is “pulled out” to compensate for the fact that the cardinality of the fuzzy number C is larger then the cardinality of T_5.

5.2 Example of the analysis of the performance of the fuzzy 2-tuple linguistic approximation under similarity measure s_4

As long as the value of β is not reported to the decision maker, the fuzzy 2-tuple linguistic approximation provides only one of the n-elementary linguistic term as an output. Under this simplification its results can be compared directly with the

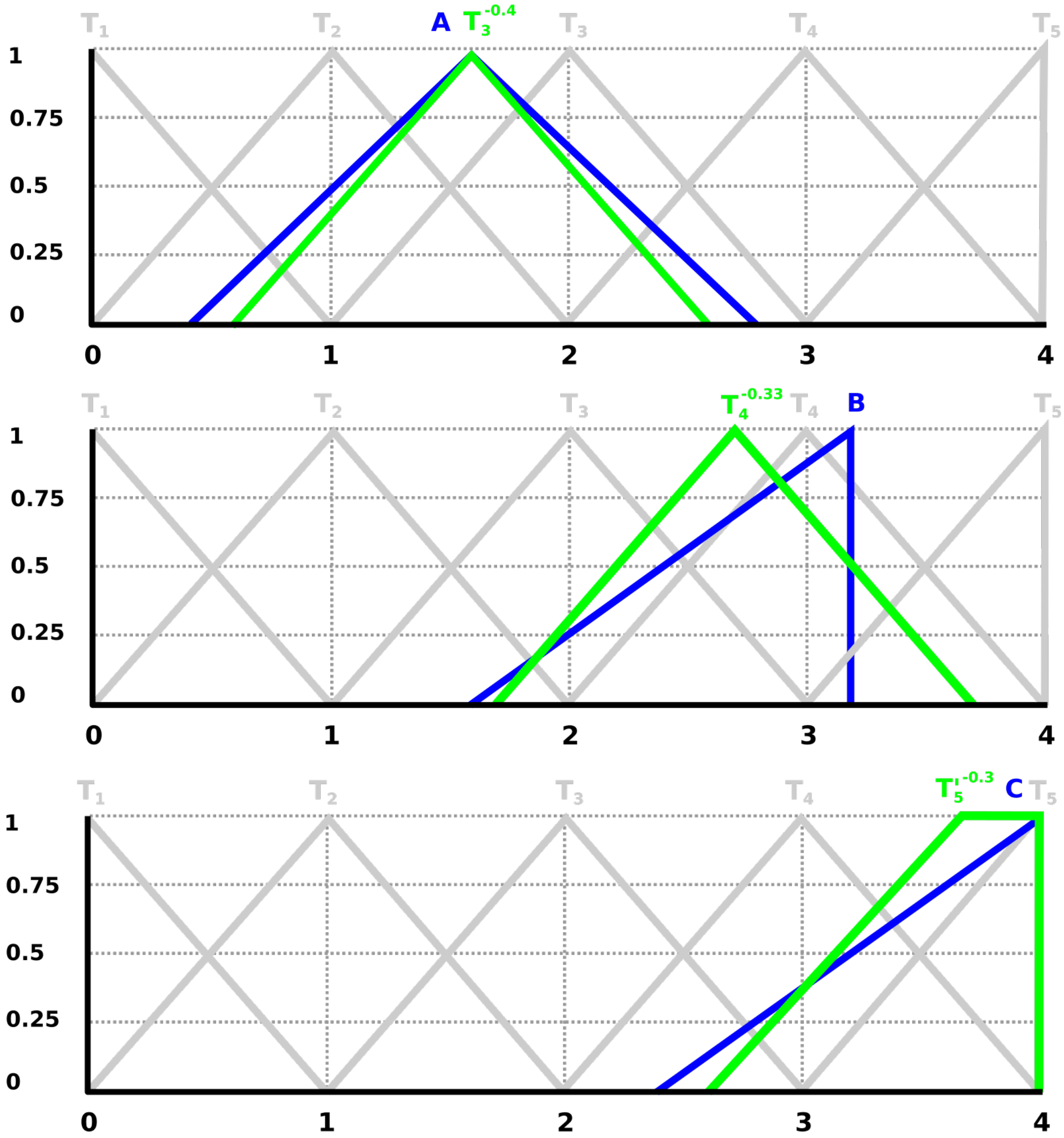


Figure 5.2: Examples of the results of fuzzy 2-tuple linguistic approximation employing distance measure s_4. Each subfigure presents the outcome of linguistic approximation of fuzzy number $A \sim (0.4, 1.6, 2.8)$ (top), $B \sim (1.6, 3.2, 3.2)$ (middle) and $C \sim (2.4, 4, 4)$ (bottom). Each example uses linguistic variable containing five uniformly distributed elementary linguistic terms the meanings of which are fuzzy numbers on [0, 4] interval.

results of the methods studied in the previous chapters. We will therefore start the analysis of the performance of fuzzy 2-tuple lingustic approximation with its comparison on with the results obtained in section 4.4.1. For simplicity, let us assume only a single distance/similarity measure, in this case the similarity measure s_4 was chosen.

Again, we will consider the uniform linguistic scale with five linguistic terms

$\mathcal{T}_1, \ldots, \mathcal{T}_5$ with meanings represented by fuzzy numbers $T_1 = (0, 0, 0.25), T_2 = (0, 0.25, 0.5), T_3 = (0.25, 0.5, 0.75), T_4 = (0.5, 0.75, 1), T_5 = (0.75, 1, 1)$ as a basis for the linguistic approximation. We will use the set Out_1 of 1 002 001 symmetrical triangular fuzzy numbers that was generated using the grid approach in section 4.4. Each fuzzy number from the set Out_1 is linguistically approximated using the fuzzy 2-tuple method and the results are visualized in the same way as in the section 4.4 (the value of β is not reflected in the visualization so far). The results are depicted in Figure 5.3 (left) together with the results of linguistic approximation using the best-fit approach (right).

Even though the similarity measure s_4 was used in both cases the visualization clearly shows some differences. Mainly, some symmetrical triangular fuzzy numbers with the length of support higher than 0.4 are lingustically approximated by the "outer" linguistic terms $\mathcal{T}_1$ and $\mathcal{T}_5$ in the fuzzy 2-tuple linguistic approximation. Also the "border" between linguistic terms $\mathcal{T}_1$ and $\mathcal{T}_2$ and also $\mathcal{T}_4$ and $\mathcal{T}_5$ is more curved in the case of the fuzzy 2-tuple linguistic approximation. And finally, some fuzzy numbers with the length of the support lower than 0.25 are linguistically approximated by linguistic terms $\mathcal{T}_2$ and $\mathcal{T}_4$, unlike in the case of linguistic approximation using the best-fit approach.

We can see that even this regarding the value of β the fuzzy 2-tuple approach provides a different perspective (i.e. different results) than the linguistic approximation approaches discussed previously. The method for the analysis of the performance of linguistic approximation under selected distance/similarity measures proposed in previous chapters can be adjusted to include the "shift" represented by the value β as well. This adjustment however results in the transition from two-dimensional to three-dimensional visualization, where the added dimension reflects the value of β. An example of such a visualization is depicted in Figure 5.4. From the visualization we can conclude several observation (obviously the ability to rotate the 3D representation is required to get full information):

- Fuzzy numbers approximated by the linguistic term $\mathcal{T}_1$ (blue) with the length of support lower than ≈ 0.25 are linguistically approximated by 2-tuple $(\mathcal{T}_1, 0)$. For the remaining fuzzy numbers approximated by $\mathcal{T}_1$ the value of β is higher than 0 and depends almost exclusively on the length of support of the approximated fuzzy number. Also note, that fuzzy numbers approximated by the linguistic term $\mathcal{T}_5$ (yellow) show an analogous behaviour, only the values of β are negative for fuzzy numbers with higher length of support.

- In the case of fuzzy numbers linguistically approximated by the remaining linguistic terms $\mathcal{T}_2, \mathcal{T}_3$ and $\mathcal{T}_4$, the value of β depends on the value of the center of gravity.

- Fuzzy numbers with center of gravity slightly lower than 0.375 (higher than 0.625) and the length of support close to zero are linguistically approximated by the linguistic term $\mathcal{T}_2$ ($\mathcal{T}_4$) with β close to 0.5 (-0.5) when the fuzzy 2-tuple linguistic approximation is used. Note that $\mathcal{T}_2(\mathcal{T}_4)$ are never assigned as results of the best-fit linguistic approximation using s_4 to fuzzy numbers with low length of support. The fuzzy 2-tuple linguistic approximation on the one hand does not "skip" from $\mathcal{T}_1$ to $\mathcal{T}_3$ ($\mathcal{T}_3$ to $\mathcal{T}_5$), but on the other hand the assignment of $\mathcal{T}_2$ ($\mathcal{T}_4$) to fuzzy numbers with low length of support is still rare and slightly counter-intuitive (see Figure 5.4).

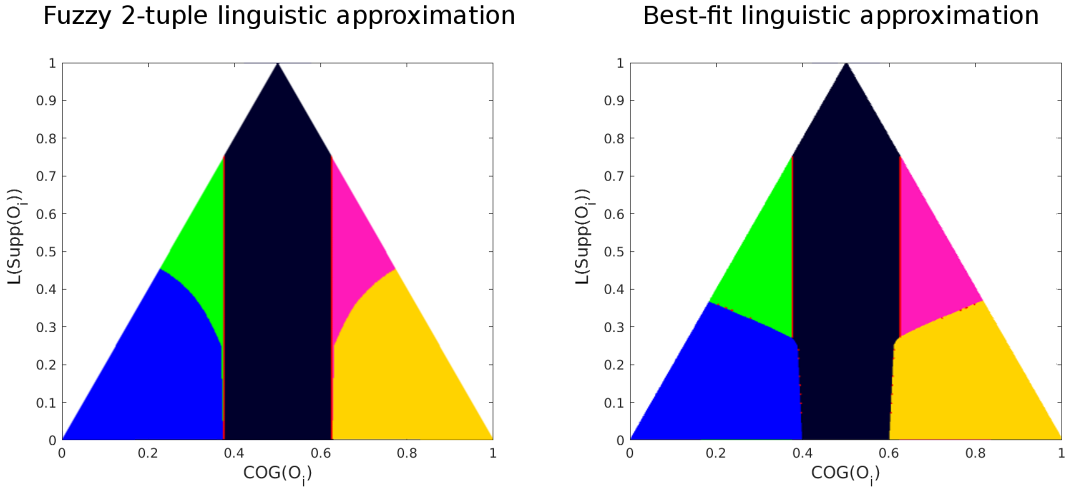


Figure 5.3: A graphical summary of the performance of the similarity measure s_4 in the fuzzy 2-tuple linguistic approximation (left) and the linguistic approximation using the best-fit approach of symmetrical triangular fuzzy numbers on $[0, 1]$ using a five-term linguistic scale. Each colour represents one term of the five-term linguistic scale: $\mathcal{T}_1$ (blue); $\mathcal{T}_2$ (green), $\mathcal{T}_3$ (black), $\mathcal{T}_4$ (pink) and $\mathcal{T}_5$ (yellow). Red color represents ambiguous cases, i.e. cases when more than one linguistic term is assigned.

Although the proposed fuzzy 2-tuple linguistic approximation method was presented here using a uniform partition by triangular fuzzy numbers on $[a, b]$, it is possible to generalize the method to non-uniform partitioning of $[a, b]$ as well and also on partitioning by non-triangular fuzzy numbers. Moreover, an enhanced linguistic scale can be used instead of the standard linguistic scale.

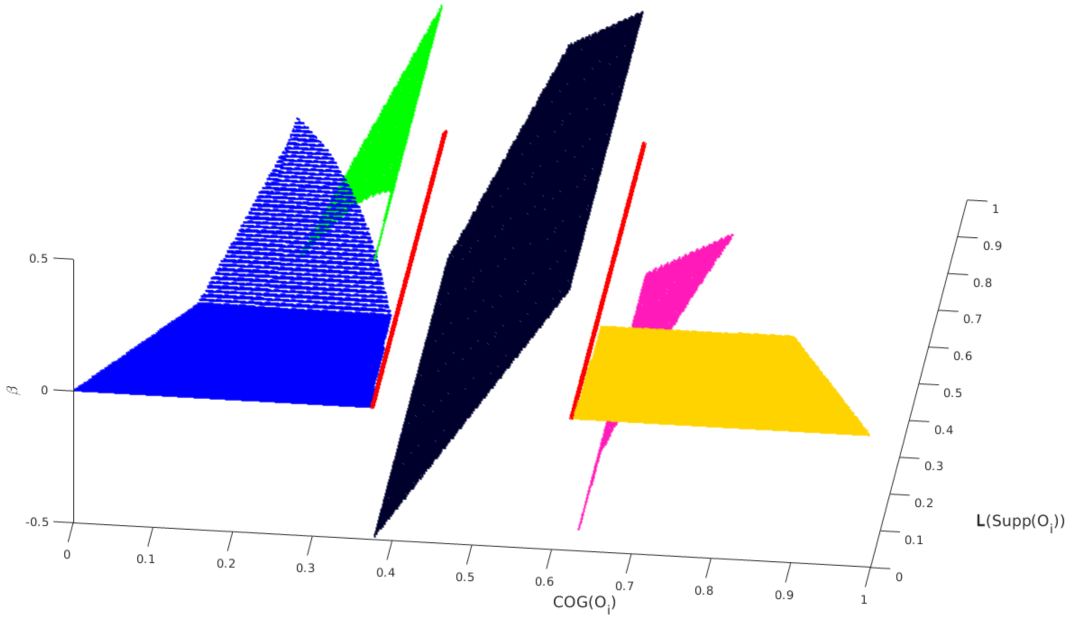


Figure 5.4: A three-dimensional visualization of the performance of the similarity measure s_4 in the fuzzy 2-tuple linguistic approximation of symmetrical triangular fuzzy numbers on $[0, 1]$ using a five-term linguistic scale. The x and y axes represent center of gravity and length of support of approximated fuzzy numbers. Each colour represents one term of the five-term linguistic scale: $\mathcal{T}_1$ (blue); $\mathcal{T}_2$ (green), $\mathcal{T}_3$ (black), $\mathcal{T}_4$ (pink) and $\mathcal{T}_5$ (yellow). Red colour represents ambiguous cases, i.e. cases when more than one linguistic term is assigned. The z axis (vertical) represents the values of β.

6 Conclusion and summary of contributions

At the beginning of the thesis we have set out the goal of suggesting a general analysis framework for the assessment of performance of different distance/similarity measures in linguistic approximation. While linguistic approximation (and methods thereof) has been in existence since 1970s the issue of appropriate selection of the distance/similarity measure has remained far from the center of scientific inquiry. This does not mean that the problem was ignored or not being pointed out in the literature. The solutions to it, however, were partial at best until now. This thesis and the publications of the author appended to it proposed an analytical framework that allows for the following:

- Any distance/similarity measure of fuzzy numbers can be analysed.
- Any linguistic variable with a finite number of linguistic terms can be considered.
- The results of the analysis are visualized in such a way, that facilitates understandability by non-profesionals mathematicians.

- Any type of fuzzy sets that can be represented by a reasonable low number of features can be considered to represent the approximated fuzzy object.
- The visual representations allow for direct comparison of the performance of different distance/similarity measures. This way strange or undesirable behaviour of some of the distance/similarity measures can be identified. Non-problematic measures can then be selected to be used in linguistic approximation. Note, that the proposed method does not require the knowledge of a "correct" linguistic approximation for the approximated objects.

To our knowledge, the proposed analytical framework is the only one currently available with the above mentioned properties. Even though the thesis restricts itself to linguistic approximation, a more general application context can be also considered. We should point out that the linguistic approximation is essentially a classification task (the set of all possible objects to be approximated is classified into classes denoted by the respective linguistic labels) – the proposed framework can thus be adapted for the analysis of classification methods as well, the graphical representation of the results is just limited by the number of features necessary to characterize the approximated objects. Nevertheless even if we restrict the results just to the linguistic approximation domain, the research gap constituted by the non-existence of a "road map" of distance/similarity measures for this context is now at least bridged.

To clearly show the use of the proposed methodology (and thus to meet the subgoals set in the introduction) we have analysed the performance of eight frequently used distance/similarity measures of fuzzy numbers in combination with standard and enhanced linguistic scales. We have considered different types of objects to be linguistically approximated ranging from symmetrical triangular fuzzy numbers through asymmetrical ones to a general family of Mamdani-type fuzzy sets. We generate graphical summaries of the results of the analysis and provide additional informations in terms of frequencies, relative frequencies, three dimensional histograms etc. to enable sufficient insights into the working of the distance/similarity measures. Wherever appropriate we point out the shortcomings and/or failures of the distance/similarity measures based on these analytical outputs.

To show the generalizability of the analytical framework we propose a brand new linguistic approximation method based on fuzzy 2-tuples. This method provides linguistic approximation in terms of a linguistic label accompanied by a number representing a shift of its meaning to the left or to the right. Even though such a result of linguistic approximation is not similar to any of the linguistic approximation methods discussed in the thesis (multistage methods, best-fit approaches) we show how to generate graphical outputs analogous to those proposed for the standard best-fit methods.

Moreover the new linguistic approximation method by fuzzy 2-tuples extends the finite set of results of linguistic approximation to an infinite one. On the other hand it uses only elementary linguistic terms for which we assume complete knowledge and understanding by the user of the model. The shift of meaning is suggested to be represented either by a number or by linguistic term approximating the magnitude of the shift and its direction. The method therefore allows for a much finer representation of the approximated objects as long as the type of approximated object is not too different from the default meanings of the elementary linguistic terms. Another aspect important for practical applications of this linguistic approximation approach is that the fuzzy 2-tuple representations can be ordered.

The main application area for the results and methods presented in this thesis is the area of linguistic fuzzy modeling, computing with words and perceptions and expert systems (i.e. systems working with or representing the knowledge, experience or skill of a human being). In these areas fuzzy sets can be considered frequent outputs of the models and the linguistic labels summarizing their meaning are required by the very nature and purpose of the models. The ability to provide appropriate linguistic approximation is also vital for the design of user-system interfaces in mathematical modeling in general – not only to make the outputs of the models clear to their users, but also to stress the important aspects of these outputs that could remain unnoticed otherwise, i.e. to create a needed "spin" for the outputs.

It is my sincere hope that this thesis and the proposed methodology will allow for a wider spread of linguistic approximation which in term means a wider use of linguistic fuzzy models in real-life applications. In fact the ability of providing reasonable, intuitive and understandable linguistic approximation can open the results of sophisticated mathematical models to a wider audience of users.

A Three dimensional histogram representations of the performance of $d_1, d_2, d_4, s_1, \ldots s_4$ in linguistic approximation

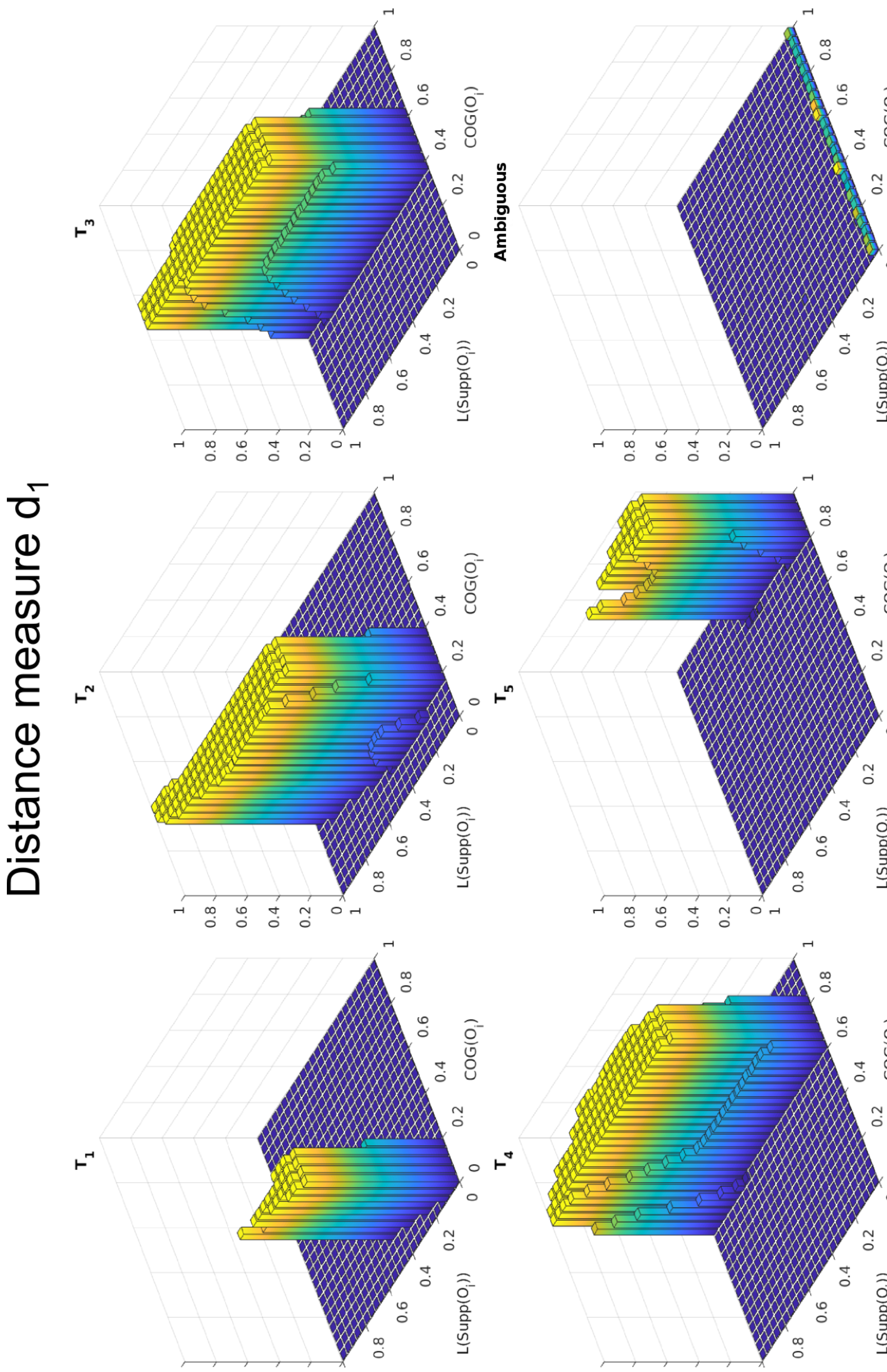


Figure A.1: Three-dimensional histogram representation of the performance of distance measure d_1 in the linguistic approximation of asymmetrical triangular fuzzy numbers on $[0, 1]$ using a 5-term linguistic scale. Each subfigure summarizes the relative frequencies of suggesting the given linguistic term for the fuzzy numbers belonging to the respective bin (feature-wise). Fuzzy numbers that can be equally well approximated by more linguistic terms at the same time are depicted in the subfigure labelled ambiguous.

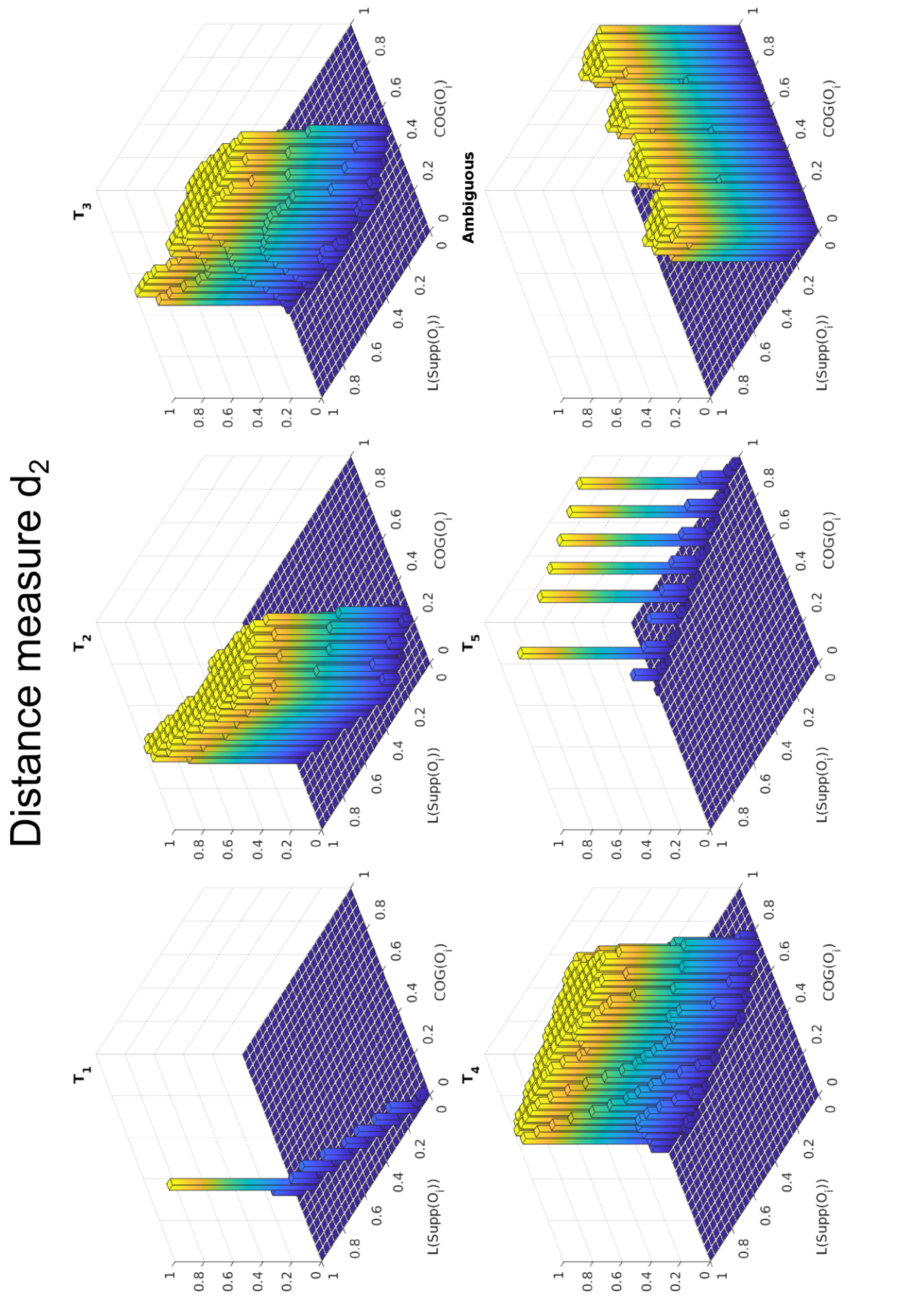


Figure A.2: Three-dimensional histogram representation of the performance of distance measure d_2 in the linguistic approximation of asymmetrical triangular fuzzy numbers on $[0, 1]$ using a 5-term linguistic scale. Each subfigure summarizes the relative frequencies of suggesting the given linguistic term for the fuzzy numbers belonging to the respective bin (feature-wise). Fuzzy numbers that can be equally well approximated by more linguistic terms at the same time are depicted in the subfigure labelled ambiguous.

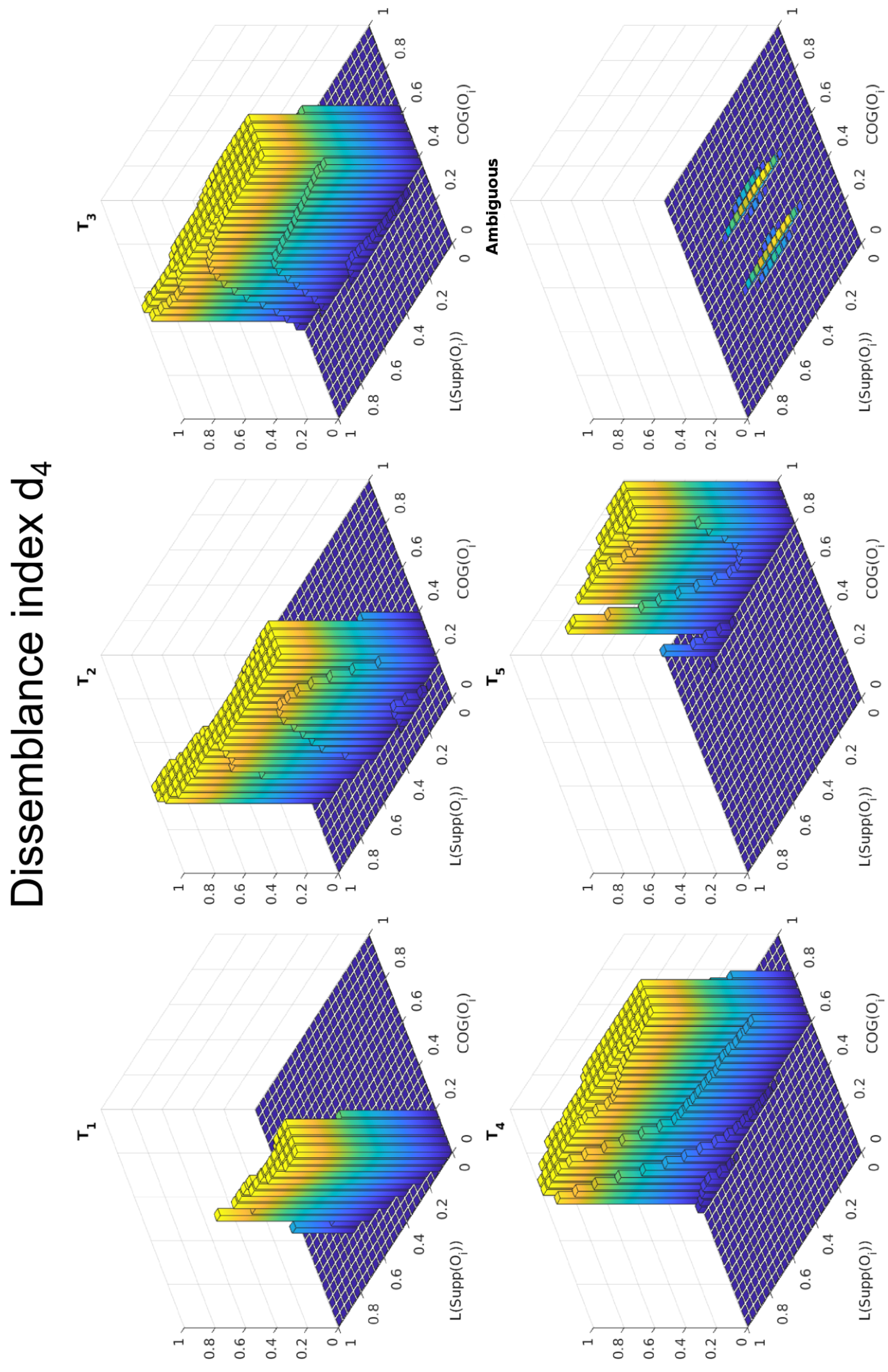


Figure A.3: Three-dimensional histogram representation of the performance of dissemblance index d_4 in the linguistic approximation of asymmetrical triangular fuzzy numbers on $[0, 1]$ using a 5-term linguistic scale. Each subfigure summarizes the relative frequencies of suggesting the given linguistic term for the fuzzy numbers belonging to the respective bin (feature-wise). Fuzzy numbers that can be equally well approximated by more linguistic terms at the same time are depicted in the subfigure labelled ambiguous.

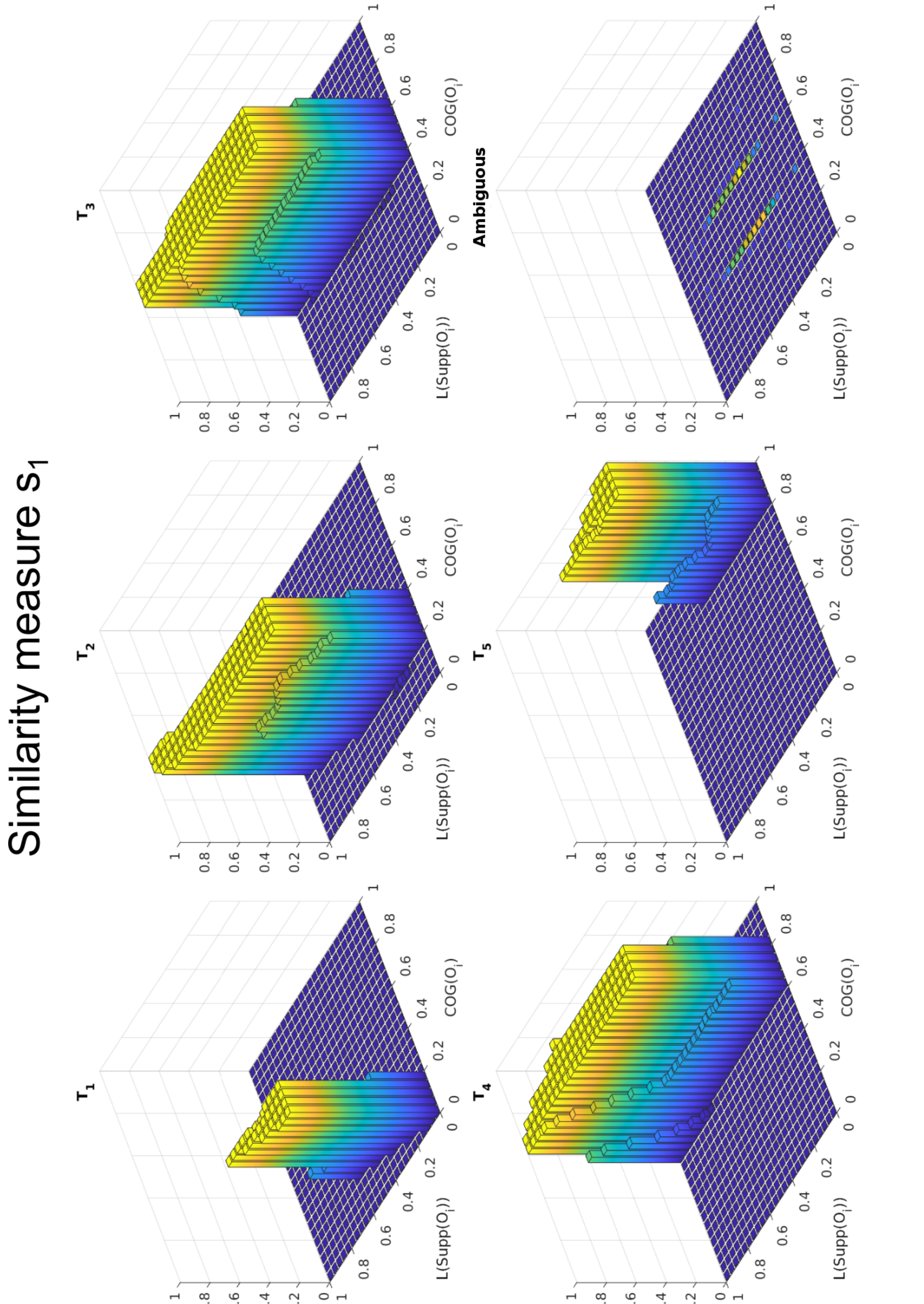


Figure A.4: Three-dimensional histogram representation of the performance of similarity measure s_1 in the linguistic approximation of asymmetrical triangular fuzzy numbers on $[0, 1]$ using a 5-term linguistic scale. Each subfigure summarizes the relative frequencies of suggesting the given linguistic term for the fuzzy numbers belonging to the respective bin (feature-wise). Fuzzy numbers that can be equally well approximated by more linguistic terms at the same time are depicted in the subfigure labelled ambiguous.

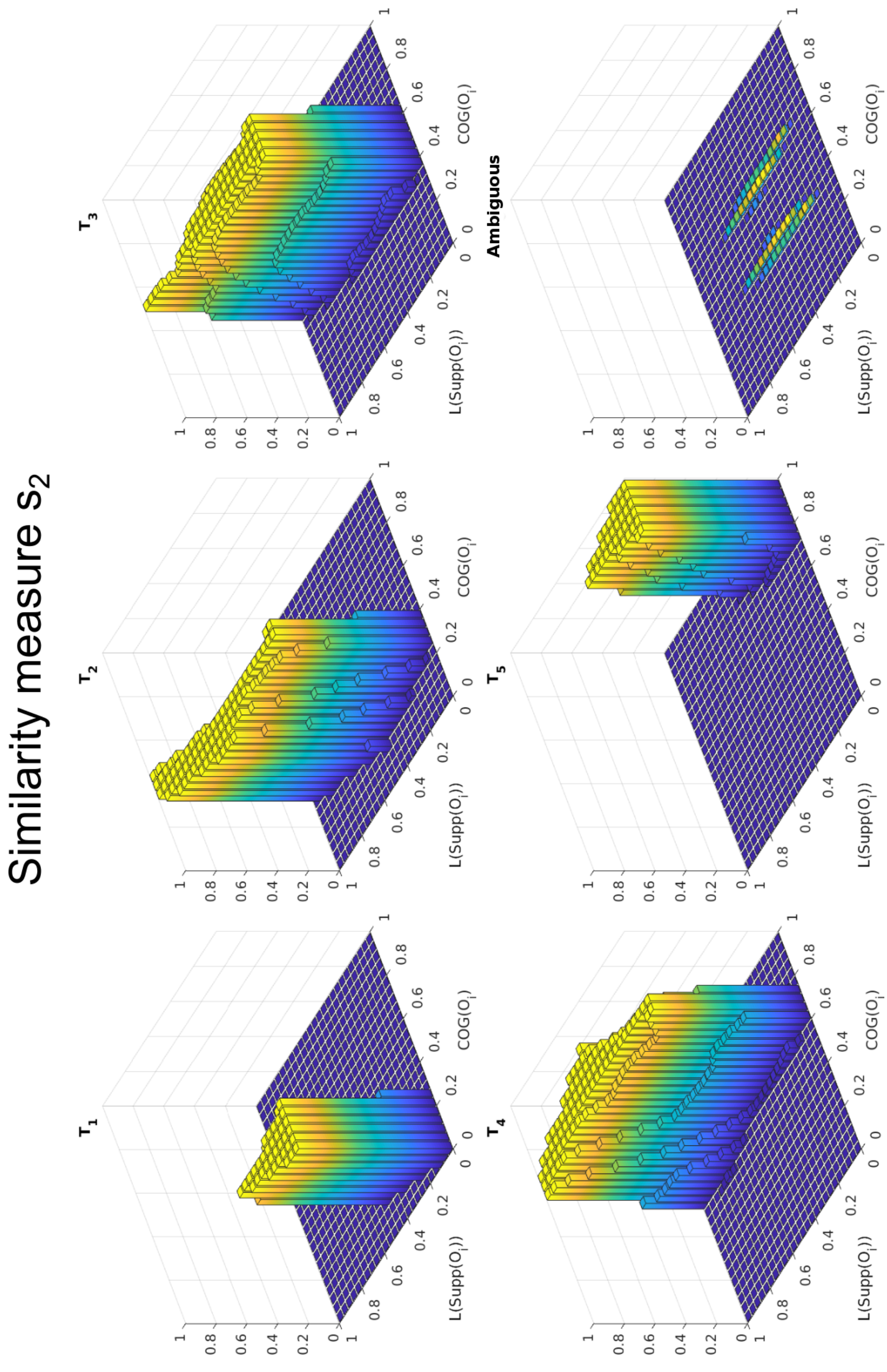


Figure A.5: Three-dimensional histogram representation of the performance of similarity measure s_2 in the linguistic approximation of asymmetrical triangular fuzzy numbers on $[0, 1]$ using a 5-term linguistic scale. Each subfigure summarizes the relative frequencies of suggesting the given linguistic term for the fuzzy numbers belonging to the respective bin (feature-wise). Fuzzy numbers that can be equally well approximated by more linguistic terms at the same time are depicted in the subfigure labelled ambiguous.

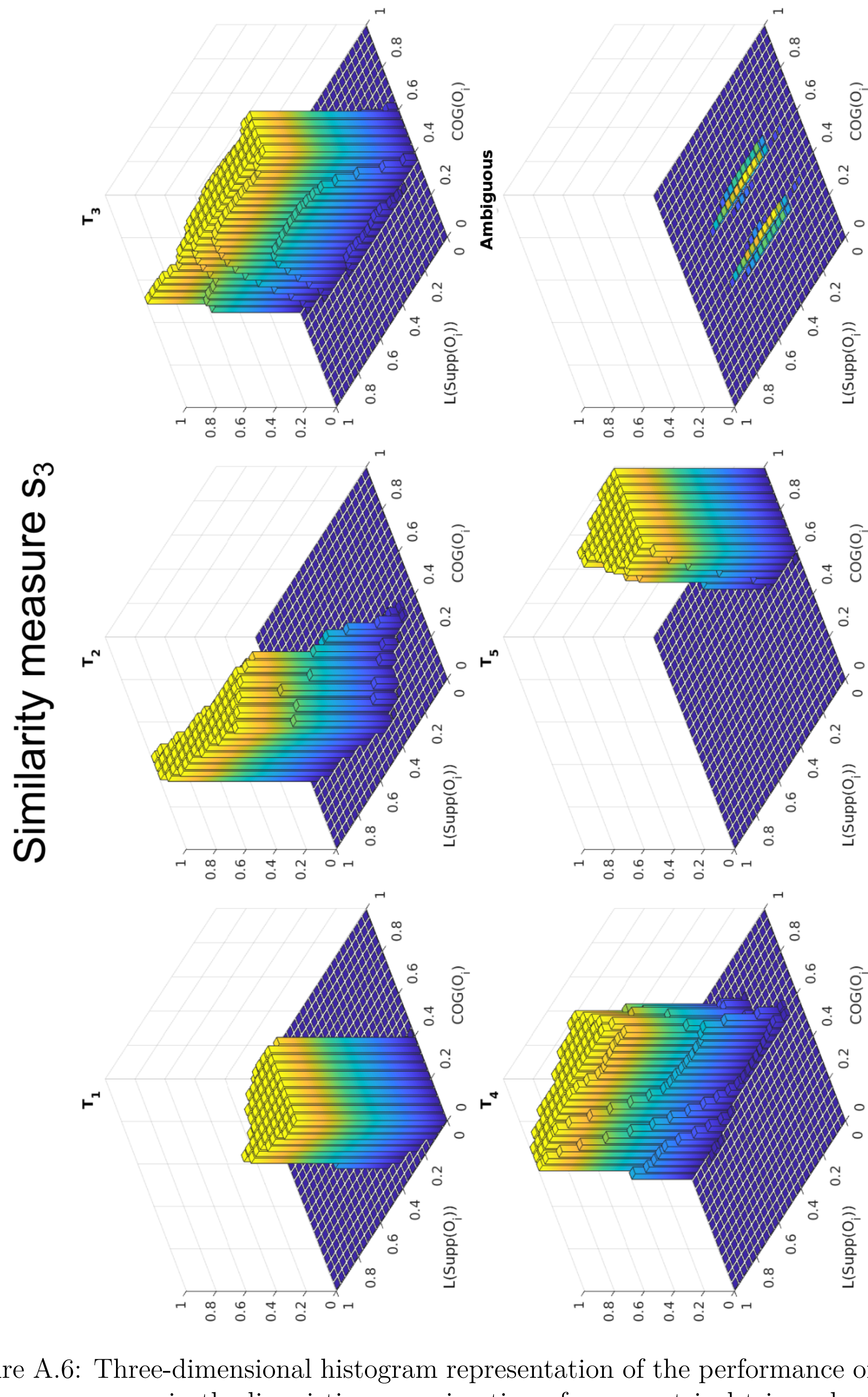


Figure A.6: Three-dimensional histogram representation of the performance of similarity measure s_3 in the linguistic approximation of asymmetrical triangular fuzzy numbers on $[0, 1]$ using a 5-term linguistic scale. Each subfigure summarizes the relative frequencies of suggesting the given linguistic term for the fuzzy numbers belonging to the respective bin (feature-wise). Fuzzy numbers that can be equally well approximated by more linguistic terms at the same time are depicted in the subfigure labelled ambiguous.

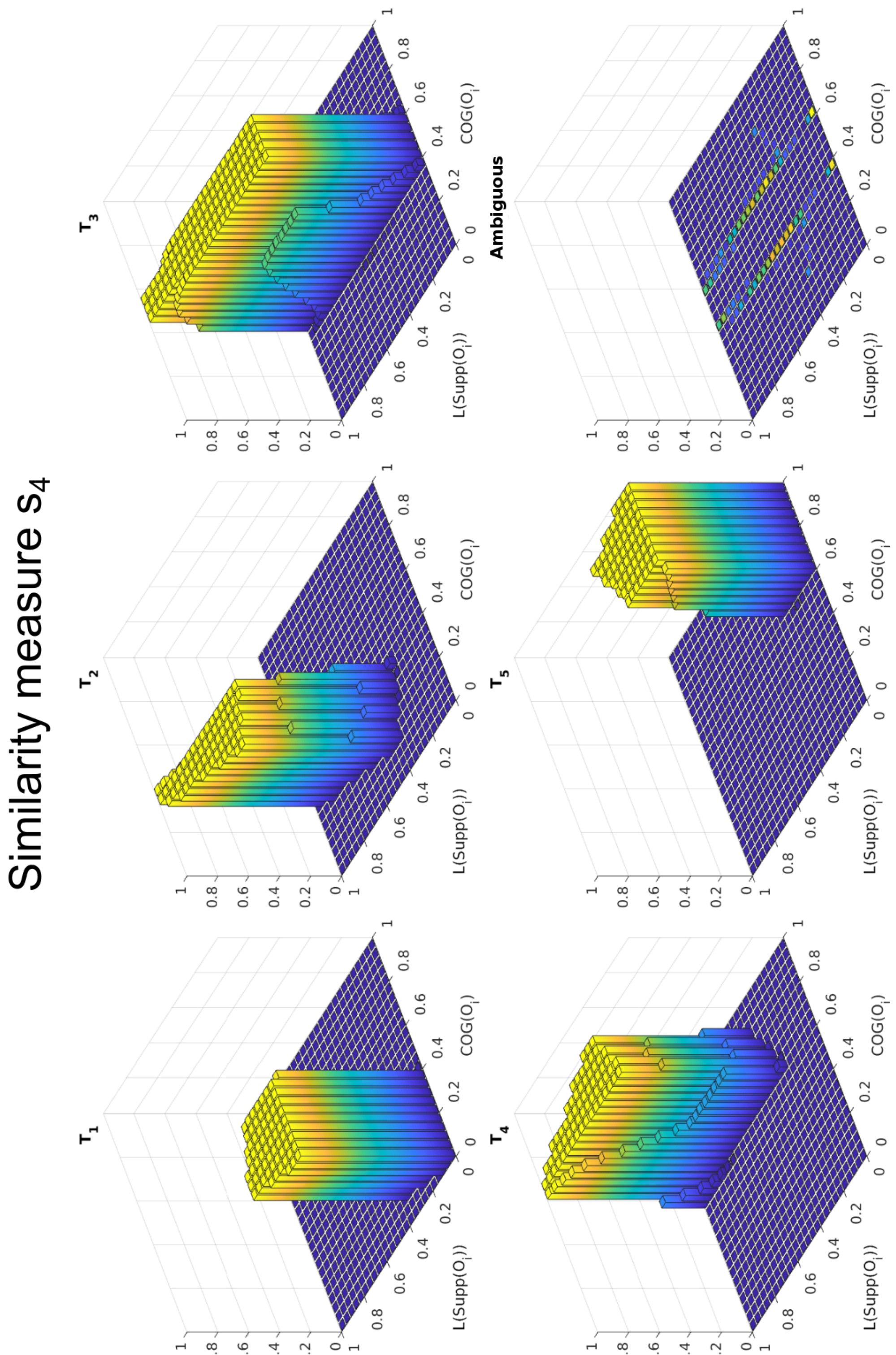


Figure A.7: Three-dimensional histogram representation of the performance of similarity measure s_4 in the linguistic approximation of asymmetrical triangular fuzzy numbers on $[0, 1]$ using a 5-term linguistic scale. Each subfigure summarizes the relative frequencies of suggesting the given linguistic term for the fuzzy numbers belonging to the respective bin (feature-wise). Fuzzy numbers that can be equally well approximated by more linguistic terms at the same time are depicted in the subfigure labelled ambiguous.

B Graphical summaries of the performance of d_1, d_2, d_4, $s_1, \ldots s_4$ in linguistic approximation of asymmetrical triangular fuzzy numbers using an enhanced linguistic scale

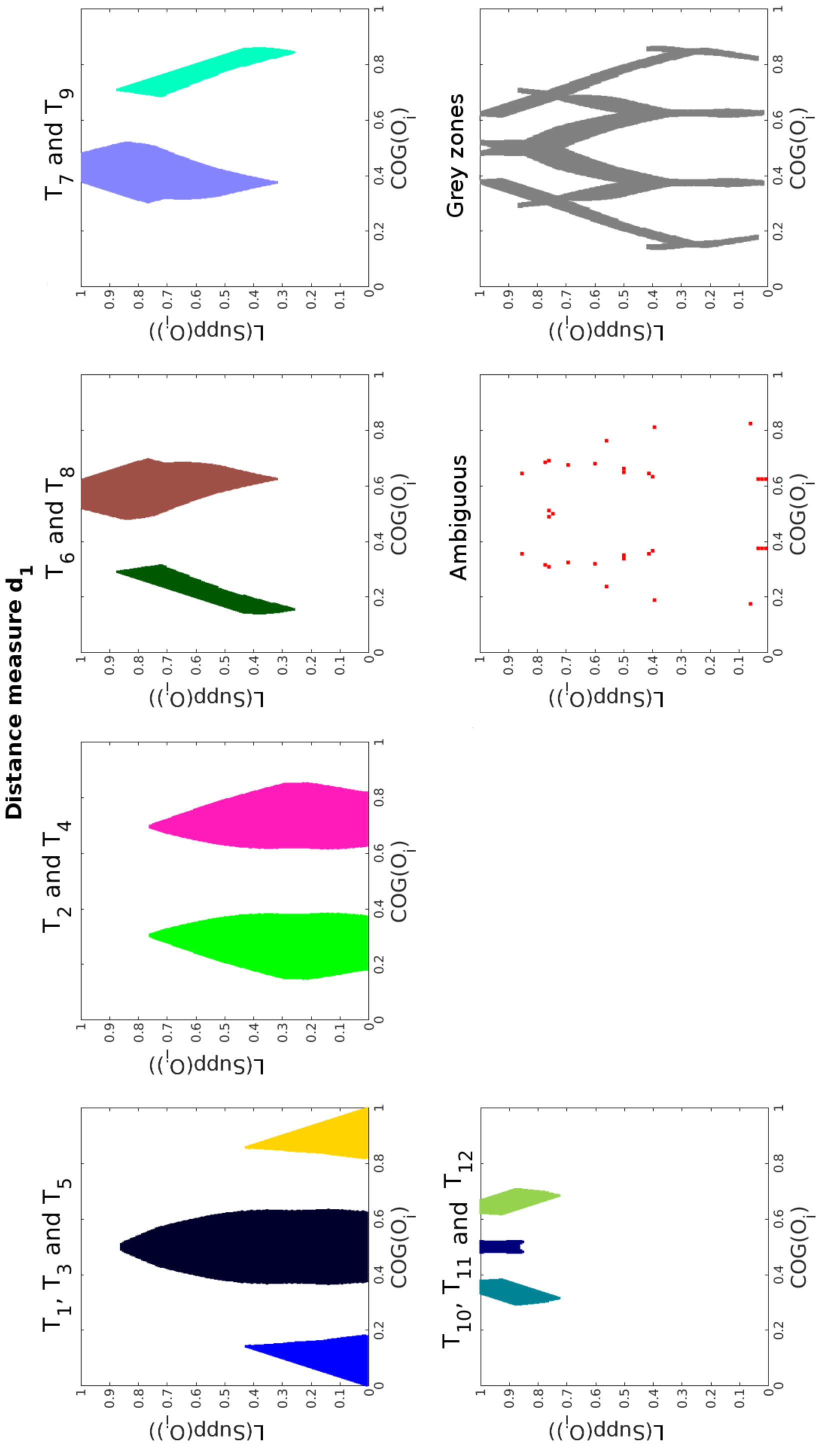


Figure B.1: A graphical summary of the results of linguistic approximation of asymmetrical triangular fuzzy numbers using the distance measure d_1 and a 5-term enhanced linguistic scale. Each colour represents one term of the enhanced linguistic scale, the assignment of colours to the linguistic terms is indicated above each subplot (linguistic terms that are never assigned to any element of Out_2 are not considered in the summary). Red colour is reserved for ambiguous cases, i.e. cases when more than one linguistic term is assigned and grey colour represents the grey zones.

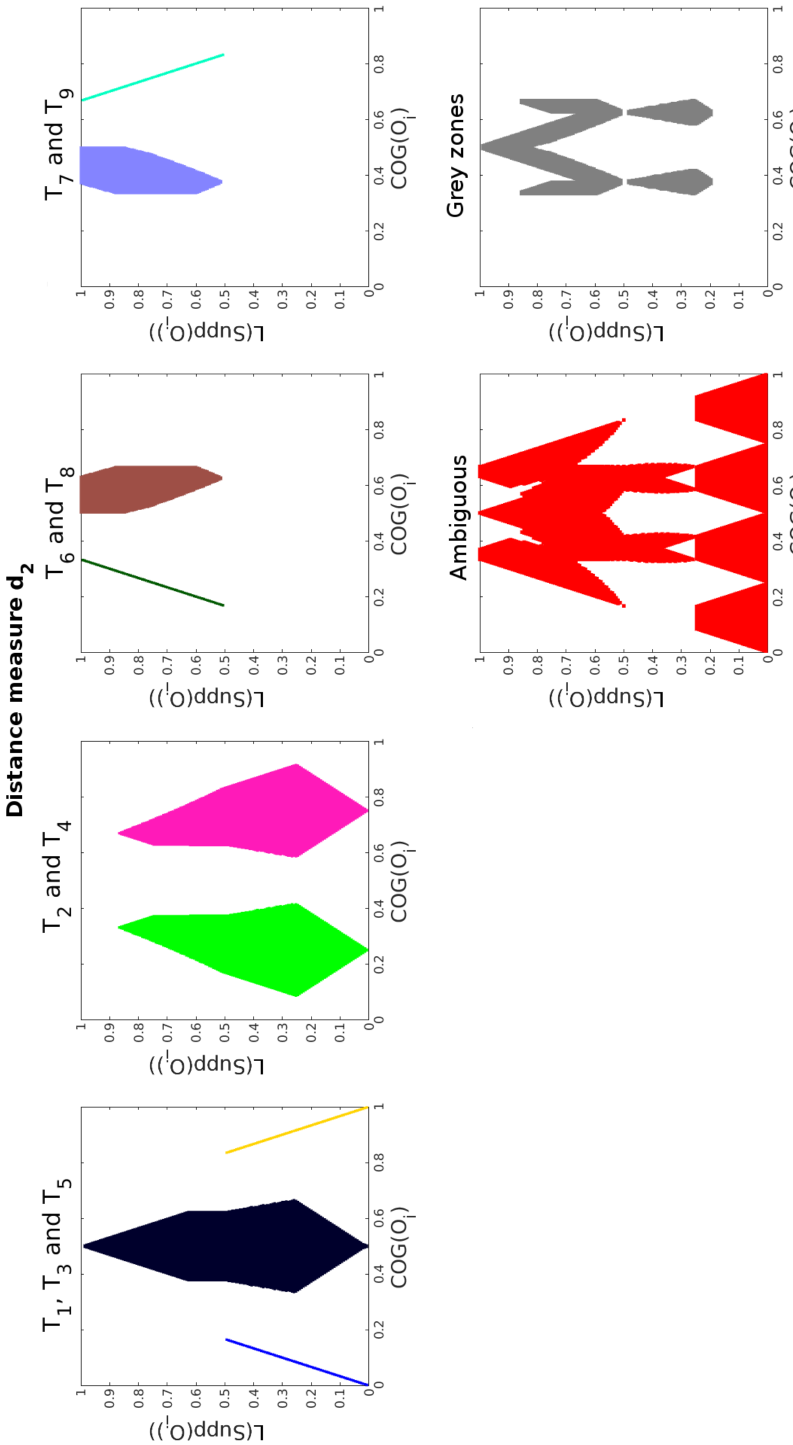


Figure B.2: A graphical summary of the results of linguistic approximation of asymmetrical triangular fuzzy numbers using the distance measure d_2 and a 5-term enhanced linguistic scale. Each colour represents one term of the enhanced linguistic scale, the assignment of colours to the linguistic terms is indicated above each subplot (linguistic terms that are never assigned to any element of Out_2 are not considered in the summary). Red colour is reserved for ambiguous cases, i.e. cases when more than one linguistic term is assigned and grey colour represents the grey zones.

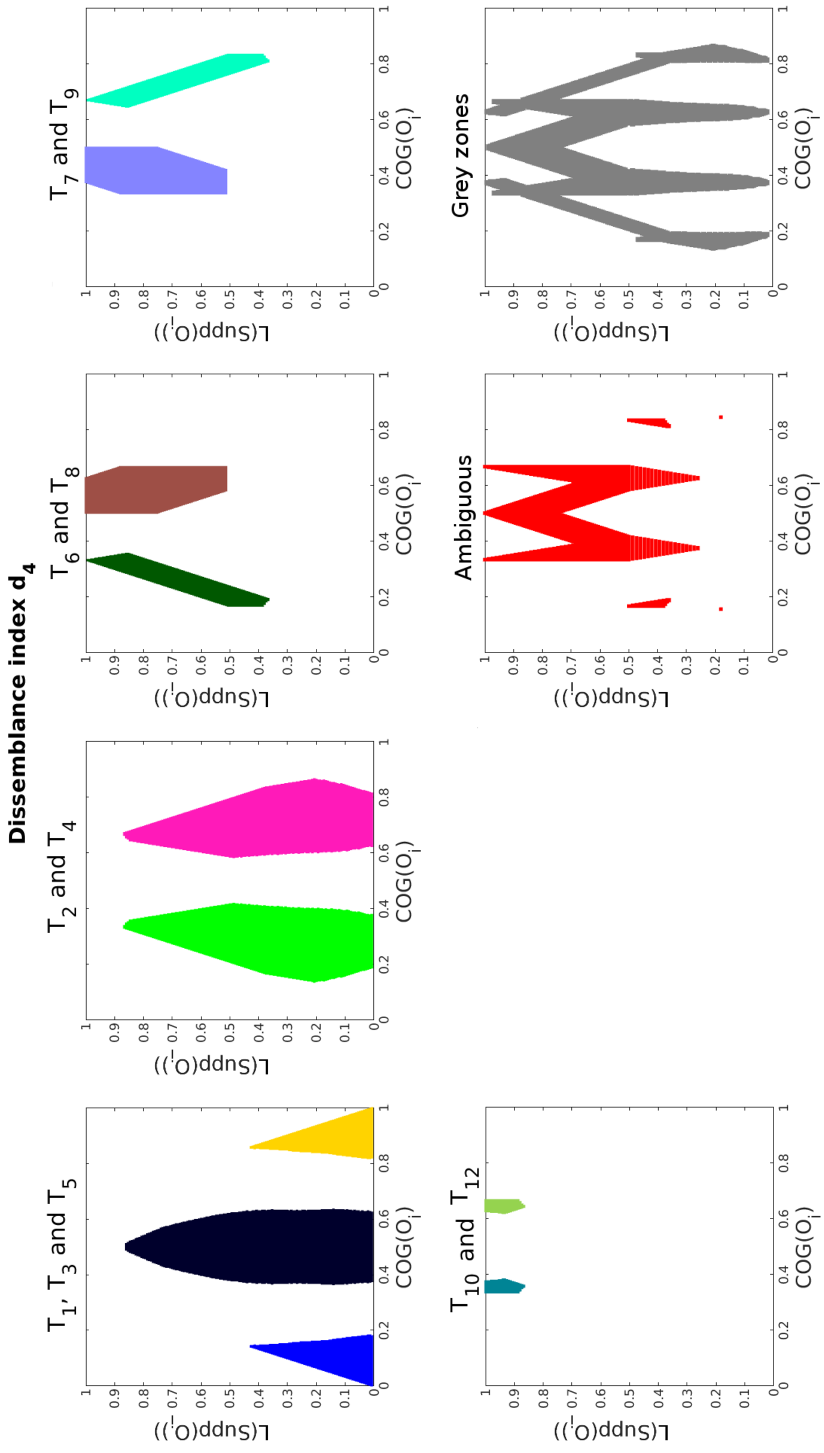


Figure B.3: A graphical summary of the results of linguistic approximation of asymmetrical triangular fuzzy numbers using the dissemblance index d_4 and a 5-term enhanced linguistic scale. Each colour represents one term of the enhanced linguistic scale, the assignment of colours to the linguistic terms is indicated above each subplot (linguistic terms that are never assigned to any element of Out_2 are not considered in the summary). Red colour is reserved for ambiguous cases, i.e. cases when more than one linguistic term is assigned and grey colour represents the grey zones.

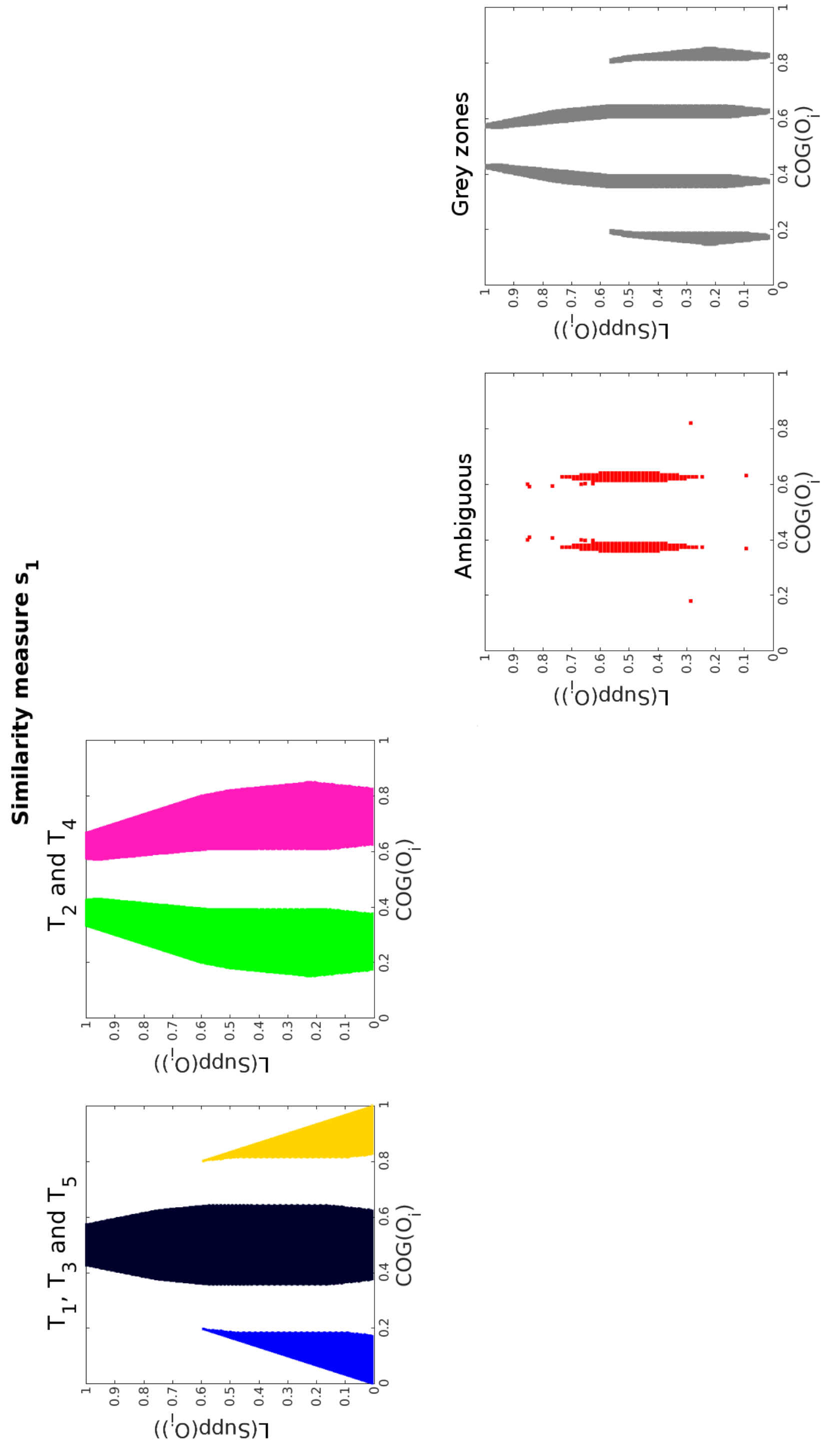


Figure B.4: A graphical summary of the results of linguistic approximation of asymmetrical triangular fuzzy numbers using the similarity measure s_1 and a 5-term enhanced linguistic scale. Each colour represents one term of the enhanced linguistic scale, the assignment of colours to the linguistic terms is indicated above each subplot (linguistic terms that are never assigned to any element of Out_2 are not considered in the summary). Red colour is reserved for ambiguous cases, i.e. cases when more than one linguistic term is assigned and grey colour represents the grey zones.

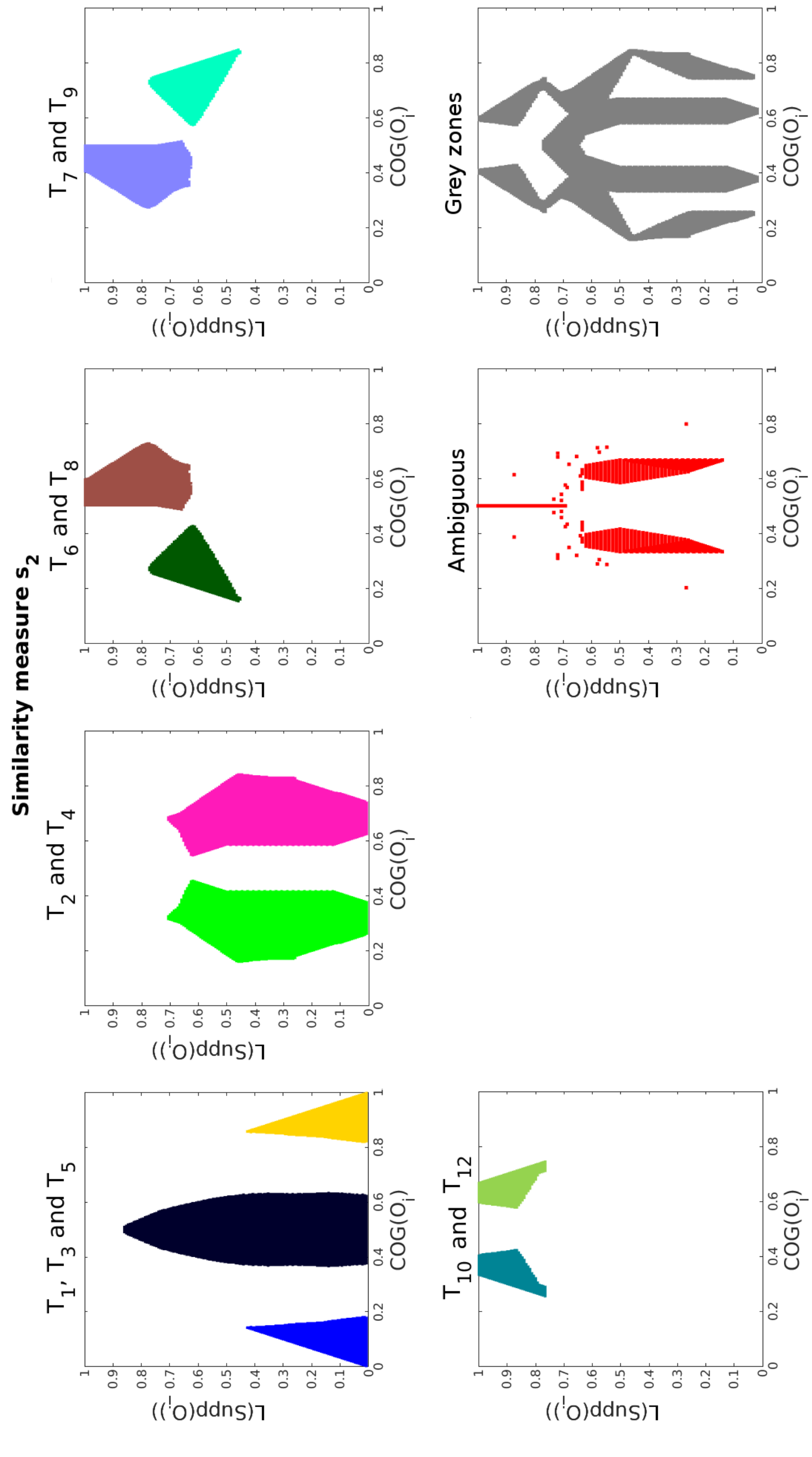


Figure B.5: A graphical summary of the results of linguistic approximation of asymmetrical triangular fuzzy numbers using the similarity measure s_2 and a 5-term enhanced linguistic scale. Each colour represents one term of the enhanced linguistic scale, the assignment of colours to the linguistic terms is indicated above each subplot (linguistic terms that are never assigned to any element of Out_2 are not considered in the summary). Red colour is reserved for ambiguous cases, i.e. cases when more than one linguistic term is assigned and grey colour represents the grey zones.

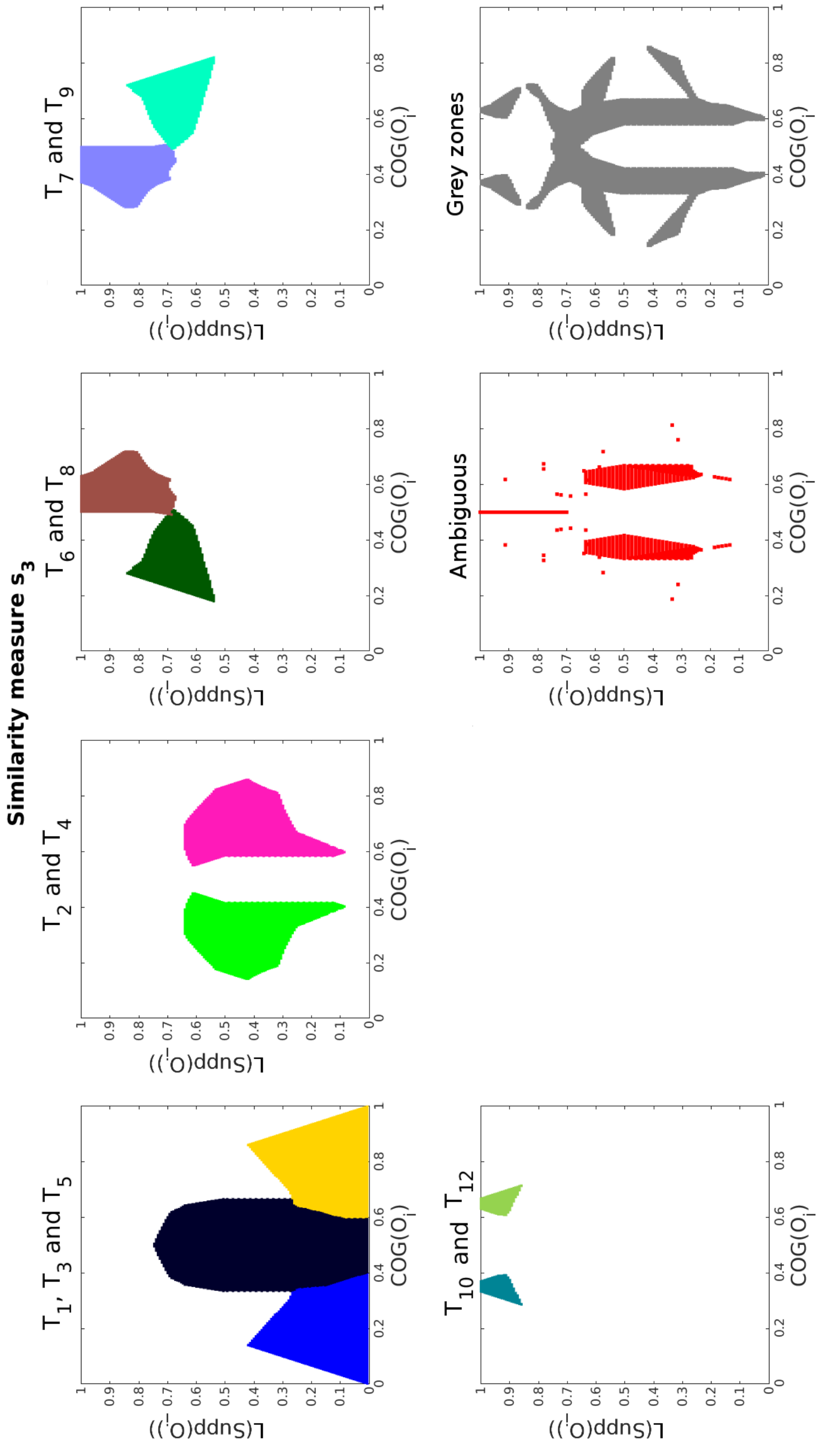


Figure B.6: A graphical summary of the results of linguistic approximation of asymmetrical triangular fuzzy numbers using the similarity measure s_3 and a 5-term enhanced linguistic scale. Each colour represents one term of the enhanced linguistic scale, the assignment of colours to the linguistic terms is indicated above each subplot (linguistic terms that are never assigned to any element of Out_2 are not considered in the summary). Red colour is reserved for ambiguous cases, i.e. cases when more than one linguistic term is assigned and grey colour represents the grey zones.

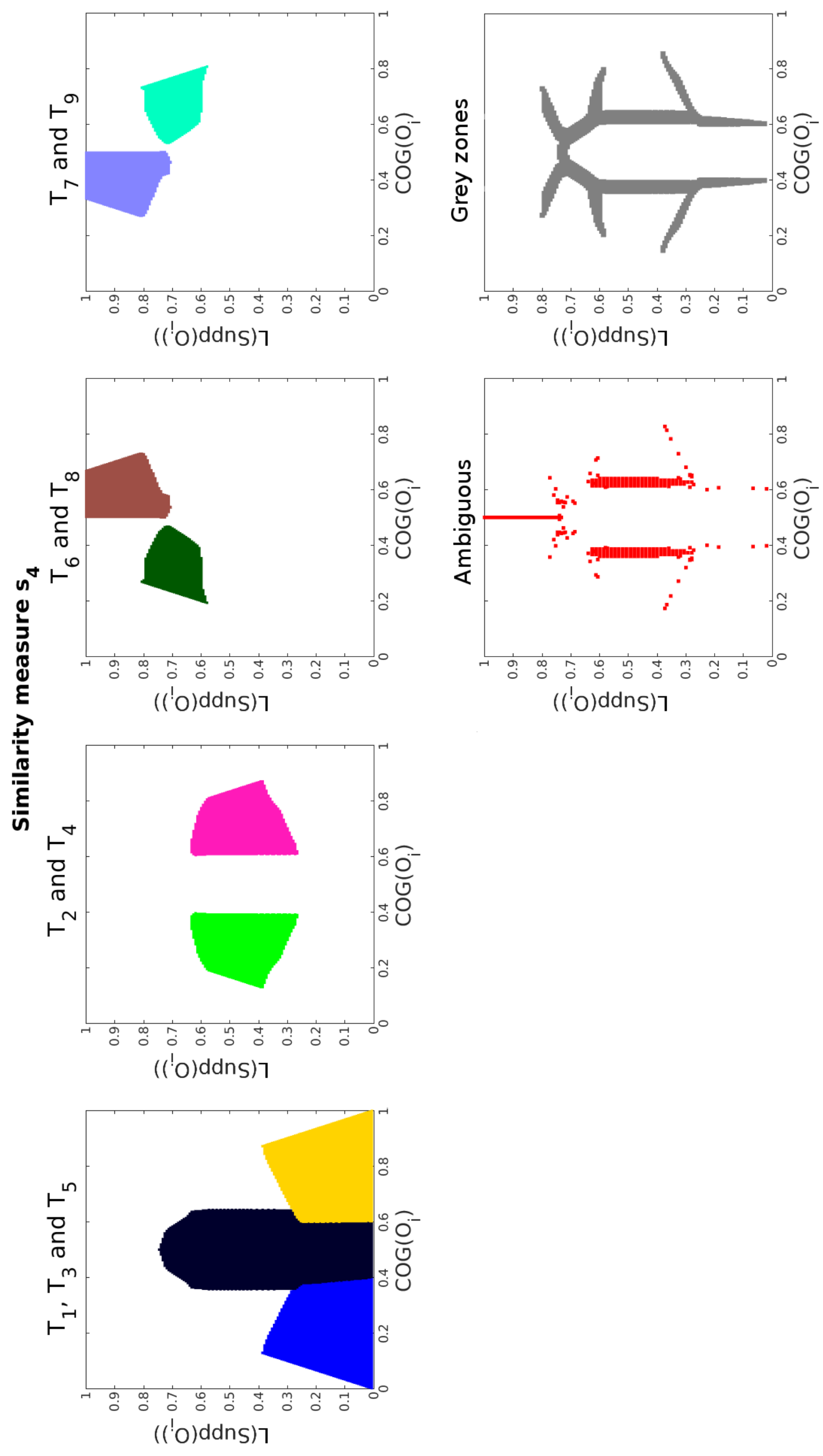


Figure B.7: A graphical summary of the results of linguistic approximation of asymmetrical triangular fuzzy numbers using the similarity measure s_4 and a 5-term enhanced linguistic scale. Each colour represents one term of the enhanced linguistic scale, the assignment of colours to the linguistic terms is indicated above each subplot (linguistic terms that are never assigned to any element of Out_2 are not considered in the summary). Red colour is reserved for ambiguous cases, i.e. cases when more than one linguistic term is assigned and grey colour represents the grey zones.

References

[1] P. P. Bonissone. A pattern recognition approach to the problem of linguistic approximation in system analysis. In *Proceedings of the IEEE International Conference on Cybernetics and Society*, pages 793–798, 1979.

[2] P. P. Bonissone. A fuzzy sets based linguistic approach: theory and applications. In *Proceedings of the 12th conference on Winter simulation*, number 1980, pages 99–111, 1980.

[3] P. P. Bonissone and K. S. Decker. Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-Off Precision and Complexity. In *Proceedings of the first Conference on Uncertainty in Artificial Intelligence (UAI1985)*, pages 217–247. 1986.

[4] 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, jul 2013.

[5] S. J. Chen and S. M. Chen. A new method to measure the similarity between fuzzy numbers. In *10th IEEE International Conference on Fuzzy Systems. (Cat. No.01CH37297)*, volume 2, pages 1123–1126, Melbourne, Australia, 2001. IEEE.

[6] S. J. Chen and S. M. Chen. Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *IEEE Transactions on Fuzzy Systems*, 11(1):45–56, feb 2003.

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

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

[9] A. Dvořák. On linguistic approximation in the frame of fuzzy logic deduction. *Soft Computing*, 3(2):111–116, 1999.

[10] F. Eshragh and E. H. Mamdani. A general approach to linguistic approximation. *International Journal of Man-Machine Studies*, 11(4):501–519, 1979.

[11] S.R. Hejazi, A. Doostparast, and S.M. Hosseini. An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers. *Expert Systems with Applications*, 38(8):9179–9185, 2011.

[12] 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.

[13] F. Herrera and L. Martinez. An approach for combining linguistic and numerical information based on the 2-tuple fuzzy linguistic representation model in decision-making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based systems*, 8(5):539–562, 2000.

[14] A. Kaufmann and M. M. Gupta. *Introduction to Fuzzy Arithmetic: Theory and Applications*. Electrical-Computer Science and Engineering Series. Van Nostrand Reinhold, New York, 1985.

[15] H. A. Khorshidi and S. Nikfalazar. An improved similarity measure for generalized fuzzy numbers and its application to fuzzy risk analysis. *Applied Soft Computing*, 52:478–486, 2017.

[16] G. J. Klir. Some issues of linguistic approximation. In *2004 2nd International IEEE Conference on 'Intelligent Systems'. Proceedings*, volume 1, page 5, 2004.

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

[18] R. Kowalczyk. On linguistic approximation of subnormal fuzzy sets. In *1998 Conference of the North American Fuzzy Information Processing Society - NAFIPS*, pages 329–333, 1998.

[19] R. Kowalczyk. On linguistic approximation with genetic programming. In José Mira, Angel Pasqual del Pobil, and Moonis Ali, editors, *Methodology and Tools in Knowledge-Based Systems*, pages 200–209, Berlin, Heidelberg, 1998. Springer Berlin Heidelberg.

[20] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1):1–13, 1975.

[21] N. Marhamati, P. Patel, E. S. Khorasani, and S. Rahimi. Towards retranslation of fuzzy values in computing with words. In *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, pages 922–928, Edmonton, AB, Canada, jun 2013. IEEE.

[22] O. Martin and G. J. Klir. On the problem of retranslation in computing with perceptions. *International Journal of General Systems*, 35(6):655–674, 2006.

[23] C.A. Murthy, S.K. Pal, and D. Dutta Majumder. Correlation between two fuzzy membership functions. *Fuzzy Sets and Systems*, 17(1):23–38, sep 1985.

[24] M. Nowakowska. Methodological problems of measurement of fuzzy concepts in the social sciences. *Behavioral Science*, 22(2):107–115, 1977.

[25] C. P. Pappis and N. I. Karacapilidis. A comparative assessment of measures of similarity of fuzzy values. *Fuzzy Sets and Systems*, 56(2):171–174, 1993.

[26] K. Patra and S. K. Mondal. Fuzzy risk analysis using area and height based similarity measure on generalized trapezoidal fuzzy numbers and its application. *Applied Soft Computing*, 28:276–284, 2015.

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

[28] B. Schott and T. Whalen. Linguistic approximation of nonconvex membership functions using "...Except..." or "...Or...". In *PeachFuzz 2000. 19th International Conference of the North American Fuzzy Information Processing Society - NAFIPS (Cat. No.00TH8500)*, pages 388–391, 2000.

[29] M. Sugeno. *Industrial Applications of Fuzzy Control*. Elsevier Science Inc., New York, NY, USA, 1985.

[30] J. H. M. Tah, A. Thorpe, and R. McCaffer. Contractor project risks contingency allocation using linguistic approximation. *Computing Systems in Engineering*, 4(2-3):281–293, 1993.

[31] T. Talášek and J. Stoklasa. Linguistic approximation under different distances/similarity measures for fuzzy numbers. In Mikael Collan and Pasi Luukka, editors, *Proceedings of the NSAIS'16 Workshop on Adaptive and Intelligent Systems 2016*, pages 49–52, Lappeenranta, 2016. LUT Scientific and expertise publications.

[32] X. Wang and E. E. Kerre. Reasonable properties for the ordering of fuzzy quantities (I). *Fuzzy Sets and Systems*, 118(3):375–385, mar 2001.

[33] S. H. Wei and S. M. Chen. A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *Expert Systems with Applications*, 36(1):589–598, jan 2009.

[34] J. Wen, X. Fan, D. Duanmu, and D. Yong. A modified similarity measure of generalized fuzzy numbers. *Procedia Engineering*, 15:2773–2777, 2011.

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

[36] T. Whalen and B. Schott. Empirical comparison of techniques for linguistic approximation. In *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569)*, volume 1, pages 93–97, Vancouver, BC, Canada, 2001. IEEE.

[37] 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, 2004.

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

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

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

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

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

[43] 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*, 46(1):105–119, 1999.

[44] R. Zwick, E. Carlstein, and D. V. Budescu. Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning*, 1(2):221–242, apr 1987.

Publication I

Talášek, T. and Stoklasa, J.

A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation under Different Linguistic Scales.

Reprinted with the permission from
Journal of Multiple-Valued Logic and Soft Computing,
Vol. 29, pp. 485–503 2017,

J. of Mult.-Valued Logic & Soft Computing, Vol. 29, pp. 485–503
Reprints available directly from the publisher
Photocopying permitted by license only

A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation Under Different Linguistic Scales

TOMÁŠ TALÁŠEK[1] AND JAN STOKLASA[2]

[1]*Dept. of Math. Analysis and Applications of Mathematics, Faculty of Science, Palacký University, Olomouc, CZ*
E-mail: tomas.talasek@upol.cz
[2]*School of Business and Management, Lappeenranta University of Technology, FI*

Accepted: August 26, 2016.

The paper investigates the behavior of linguistic approximation under two distance measures of fuzzy numbers and two similarity measures of fuzzy numbers in the context of different linguistic scales. An analytic framework for the comparison of different distance/similarity measures in the linguistic approximation based on the minimization of distance (maximization of similarity) is introduced and numerical investigation of four chosen measures is performed. The focus of this paper is narrowed to symmetrical triangular fuzzy numbers as approximated objects and several different linguistic scales are considered. The presented results provide evidence of the existence of differences in the performance of the selected measures. Preference of more uncertain approximations, reduction of uncertainty and the emergence of ambiguity regions are among the identified effects of some of the measures. Conclusions concerning the suitability of specific distance/similarity measures in different contexts are drawn and possible drawbacks of their use are identified and discussed.

Keywords: Linguistic approximation, fuzzy number, distance, similarity, Bhattacharyya distance, dissemblance index, best-fit, linguistic scale.

1 INTRODUCTION

The main focus of this paper is the process of assigning linguistic labels to the outputs of mathematical models, i.e. linguistic approximation. Yager

[14] uses also the term "retranslation" in the context of representing the outputs of mathematical models by numbers, intervals, linguistic labels or other entities, that are able to carry the core of the information, to relay the meaning, to summarize the important aspects of the output and/or to provide appropriate spin where necessary. From this point of view, linguistic approximation can be seen as one possible tool for retranslation. There are also several approaches to linguistic approximation as such. The diversity of practical applications of retranslation and the corresponding abundancy of methods and mathematical tools suggested for this purpose make the field of retranslation a rather demanding object for analysis. Behavioural investigation of the methods that are intended to carry the meaning to the users of the outputs of the models is still scarce. This paper strives to provide insights into the mechanisms of linguistic approximation techniques based on the "best-fit" approach (see [14] for an overview of methods). That is we investigate methods for the assignment of linguistic labels to the outputs of mathematical models (fuzzy numbers are considered). More specifically, we assume that the meanings of the linguistic labels are represented by fuzzy numbers (an underlying linguistic variable is assumed) and that the appropriate linguistic label is assigned based on the distance (similarity) of the approximated fuzzy number to the meanings of the values of the linguistic variable. Methods based on subsethood and multi-step methods [2, 13] are left out of the scope of this paper, as is the explicit incorporation of spin in the process [14]. This still leaves the choice of the distance/similarity measures an open issue. This paper continues in the investigation of behavior of linguistic approximation methods based on distance/similarity measures [8–10] and extends its scope to different linguistic scale structures. Since the number of available distance/similarity measures is huge (see e.g. [7, 18]), we select two representatives of distances and two representatives of similarities to perform the analysis. From this point of view, we offer a detailed investigation of the selected distance/similarity measures in linguistic approximation, but also suggest an easy-to-implement methodology for the systematic investigation of the applicability of specific distance/similarity measures in the "best-fit" linguistic approximation setting.

The structure of the article is following: mathematical preliminaries and basic notations of the theory of fuzzy sets are defined in Section 2. In the next section the process of linguistic approximation using the so called best-fit approach is defined together with four distance/similarity measures of fuzzy numbers. Analytic framework for the numerical investigation is introduced in the Section 4 and the results of this investigation are discussed in the last section.

2 PRELIMINARIES

Let U be a nonempty set (the universe of discourse). 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 $A(.)$ is called a *membership function* of the fuzzy set A. $\mathrm{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an α*-cut* of A for any $\alpha \in [0, 1]$, $\mathrm{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A. Let A and B be fuzzy sets on the same universe U. We say that A is a *fuzzy subset* of B ($A \subseteq B$), if $A(x) \leq B(x)$ for all $x \in U$.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\mathrm{Ker}(A) \neq \emptyset$ (A is *normal*); (2) A_α are closed intervals for all $\alpha \in (0, 1]$ (this implies A is *unimodal*); (3) $\mathrm{Supp}(A)$ is bounded. A family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on [a,b], if $\mathrm{Supp}(A)$ is a subset of an interval $[a, b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \mathrm{Cl}(\mathrm{Supp}(A))$ and $[a_2, a_3] = \mathrm{Ker}(A)$, where $\mathrm{Cl}(\mathrm{Supp}(A))$ denotes a closure of $\mathrm{Supp}(A)$. Each fuzzy number A is determined by $A = \left\{[\underline{a}(\alpha), \overline{a}(\alpha)]\right\}_{\alpha \in [0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall \alpha \in (0, 1]$, and the closure of the support of A $\mathrm{Cl}(\mathrm{Supp}(A)) = [\underline{a}(0), \overline{a}(0)]$. A *cardinality* of fuzzy number A on $[a, b]$ is a real number $\mathrm{Card}(A)$ defined as follows: $\mathrm{Card}(A) = \int_a^b A(x)dx$. A *union* of two fuzzy sets A and B on U (based on Łukasiewicz disjunction) is a fuzzy set $(A \cup_L B)$ on U defined as follows: $(A \cup_L B)(x) = \min\{1, A(x) + B(x)\}, \ \forall x \in U$.

The fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$ and $a_1 \neq a_4$; for such fuzzy numbers we will use a simplified notation $A = (a_1, a_2, a_3, a_4)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by ordered triplet $A = (a_1, a_2, a_4)$. Triangular fuzzy number $A = (a_1, a_2, a_4)$ is called *symmetric triangular fuzzy number* if $a_2 - a_1 = a_4 - a_2$. More details on fuzzy numbers and computations with them can be found for example in [3].

In real-life applications we often need to represent fuzzy numbers by real numbers. This process is called *defuzzification*. The most common method is to substitute fuzzy number by its center of gravity (COG). Let A be a fuzzy number on $[a, b]$ for which $a_1 \neq a_4$. The *center of gravity* of A is defined by the formula $\mathrm{COG}(A) = \int_a^b xA(x)dx/\mathrm{Card}(A)$. If $A = (a_1, a_2, a_4)$ is symmetric triangular fuzzy number on $[a, b]$, then $\mathrm{COG}(A) = \mathrm{Ker}(A) = a_2$.

A *fuzzy scale* on $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on [a,b], that form a Ruspini fuzzy partition (see [6]) of the interval $[a, b]$, i.e. for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering.

A *linguistic variable* [15–17] is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values (terms), X is an universe on which the meanings of the linguistic values are defined, G is a syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X. Linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ is called a *linguistic scale* on $[a, b]$ if $X = [a, b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ and $T_i = M(\mathcal{T}_i)$, $i = 1, \ldots, n$ form a fuzzy scale on $[a, b]$. Linguistic terms $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ of linguistic scale $\mathcal{T}(\mathcal{V})$ are called *elementary (level 1) terms* of linguistic scale. Linguistic scale can be extended using additional linguistic terms $\mathcal{T}_i$ to $\mathcal{T}_j$ where $i = 1, \ldots, n-1$, $j = 2, \ldots, n$ and $i < j$ are called *derived linguistic terms*, $M(\mathcal{T}_i \text{ to } \mathcal{T}_j) = T_i \cup_L T_{i+1} \cup_L \cdots \cup_L T_j$. The extended linguistic scale thus contains linguistic values of different levels of uncertainty – from the possibly least uncertain elementary terms $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ to the most uncertain linguistic term $\mathcal{T}_1$ to $\mathcal{T}_n$ (uncertainty can be assessed by the cardinality of the meanings of these linguistic terms). Derived linguistic terms $\mathcal{T}_i$ to $\mathcal{T}_j$ are called *level* $j - i + 1$ *terms* and can be also denoted by $\mathcal{T}_{ij}$. Elementary linguistic terms $\mathcal{T}_i$ can be also denoted by $\mathcal{T}_{ii}$ (i.e. $\mathcal{T}_i = \mathcal{T}_{ii}$ to unify the notation).

3 LINGUISTIC APPROXIMATION OF FUZZY NUMBERS

We have already stated our focus on the "best-fit" approach to linguistic approximation, based on assigning such linguistic label to the given fuzzy set, that is the closest (or most similar) in terms of its meaning represented by a fuzzy number. That is we assume an underlying linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), [a, b], G, M)$, such that $T_i = M(\mathcal{T}_i)$, $i = 1, \ldots, s$ are fuzzy numbers on $[a, b]$, and $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$ is the set of its linguistic values (possible linguistic labels). The task of a linguistic approximation is now one of finding an appropriate linguistic label from $\mathcal{T}(\mathcal{V})$ to a given fuzzy number O on $[a, b]$ (considered to be an output of a mathematical model). The linguistic approximation $\mathcal{T}_O \in \mathcal{T}(\mathcal{V})$ of the fuzzy number O is in the "best-fit" linguistic approximation framework found computing

$$T_O = \arg \min_{i \in \{1,\ldots,s\}} d(T_i, O), \tag{1}$$

where $d(A, B)$ is a distance or similarity measure* of two fuzzy numbers. To retain comparability with the results recently published in the literature

* In the case of similarity measure the *arg min* function in formula (1) is naturally replaced by *arg max*.

[8–10], we choose the following distances and similarity measures of fuzzy numbers A and B for the analysis (note, that these serve as examples of the distance and similarity measures and that the findings of this analysis can also serve as a case study of the types of behavior or differences in performance that can be identified using the analysis methodology suggested in this paper):

- *modified Bhattacharyya distance* [1]:

$$d_1(A, B) = \left[1 - \int_U (A^*(x) \cdot B^*(x))^{1/2} dx\right]^{1/2}, \tag{2}$$

  where $A^*(x) = A(x)/\text{Card}(A)$ and $B^*(x) = B(x)/\text{Card}(B)$,
- *dissemblance index* [5]:

$$d_2(A, B) = \int_0^1 |\underline{a}(\alpha) - \underline{b}(\alpha)| + |\overline{a}(\alpha) - \overline{b}(\alpha)| \, d\alpha. \tag{3}$$

- *similarity measure* (introduced by Wei and Chen [12]):

$$s_1(A, B) = \left(1 - \sum_{i=1}^{4} \frac{|a_i - b_i|}{4}\right) \cdot \frac{\min\{Pe(A), Pe(B)\} + 1}{\max\{Pe(A), Pe(B)\} + 1}, \tag{4}$$

  where $Pe(A) = \sqrt{(a_1 - a_2)^2 + 1} + \sqrt{(a_3 - a_4)^2 + 1} + (a_3 - a_2) + (a_4 - a_1)$, $Pe(B)$ is defined analogically,
- *similarity measure* (introduced by Hejazi and Doostparast [4]):

$$s_2(A, B) = \left(1 - \sum_{i=1}^{4} \frac{|a_i - b_i|}{4}\right) \cdot \frac{\min\{Pe(A), Pe(B)\}}{\max\{Pe(A), Pe(B)\}} \cdot \frac{\min\{Ar(A), Ar(B)\} + 1}{\max\{Ar(A), Ar(B)\} + 1}, \tag{5}$$

  where $Ar(A) = \frac{1}{2}(a_3 - a_2 + a_4 - a_1)$, $Ar(B)$ is defined analogically and $Pe(A)$ and $Pe(B)$ are computed identically as in the previous method.

Although these chosen measures present but a fracture of the measures available in literature, their detailed analysis can still be meaningful and present not only insights into the appropriateness of these measures in the best-fit linguistic approximation, but also a good case study of the use of the analytic framework proposed in this paper. These methods also represent an interesting selection of different approaches to represent difference between fuzzy numbers in terms of distance or similarity - d_1 focusing among other aspects on the subsethood, d_2 being based mainly on the differences between the α-cuts of the fuzzy numbers and both similarities s_1 and s_2 representing different focus on various aspects of the shape of fuzzy numbers (area, perimeter, significant values). The analysis presented in this paper can, however, be easily extended to many other distance and similarity measures of fuzzy numbers and their performance in the "best-fit" linguistic approximation.

4 NUMERICAL INVESTIGATION

In this paper, we restrict our investigation of behavior of linguistic approximation based on different distance and similarity measures applied to symmetrical triangular fuzzy numbers defined on the interval [0, 1]. These fuzzy numbers $O = (o_1, o_2, o_3)$ can be represented by 2-tuples $(o_2, o_3 - o_1)$, where the first element represents the center of gravity of O and the second element represents the length of the support of O. The simulation approach employed in [8, 10] or [9] (see Figure 1 for an example of the outputs of [8], where random generation of symmetrical triangular fuzzy numbers was employed to assess the effect of the choice of similarity/distance measure under a 5-element elementary term set of the linguistic scale) is replaced in this paper by a grid approach - i.e. by a systematic investigation of a representative sample of symmetrical triangular fuzzy numbers on [0, 1]. To obtain a set of 500 000 symmetrical triangular fuzzy numbers for the investigation, 1 001 points uniformly distributed across the interval [0, 1] are selected to represent the centers of gravity of the investigated fuzzy numbers (o_2) and the same approach is applied to generate the spreads of the fuzzy numbers ($o_3 - o_1$). This way a set of 1 002 001 symmetrical triangular fuzzy numbers $O^G = \{O_k \in \mathcal{F}_N([-0.5, 1.5]) | k = 1, \dots, 1002001\}$ is generated. Out of these we select for our investigation a subset of symmetrical triangular fuzzy numbers $O^{IN} = \left\{O_j \in O^G | O_j \in \mathcal{F}_N([0, 1])\right\}$. The set $O^{IN} = \{O_1, \dots, O_{500000}\}$ represents a uniform grid on the COG-Card(Supp) space of symmetrical triangular fuzzy numbers on [0, 1]. This allows for the comparison of frequencies and also for the use of a *crisp benchmark* (denoted "crisp" in the tables) in the analysis of the results.

This crisp benchmark assumes a partitioning of the interval [0,1] into j subintervals $I_1, \dots, I_j$, where $I_1 \cup \dots \cup I_j = [0, 1]$, and j is the number of elementary terms of the linguistic scale used for the linguistic approximation. The intervals are defined in the following way: $I_m = [x_m, y_m]$, where $x_1 = 0$, $y_j = 1$, x_m is the solution of $T_{m-1}(x_m) = T_m(x_m)$ for $m = 2, \dots, j$ and y_m is the solution of $T_m(y_m) = T_{m+1}(y_m)$ for $m = 1, \dots, j - 1$. These intervals

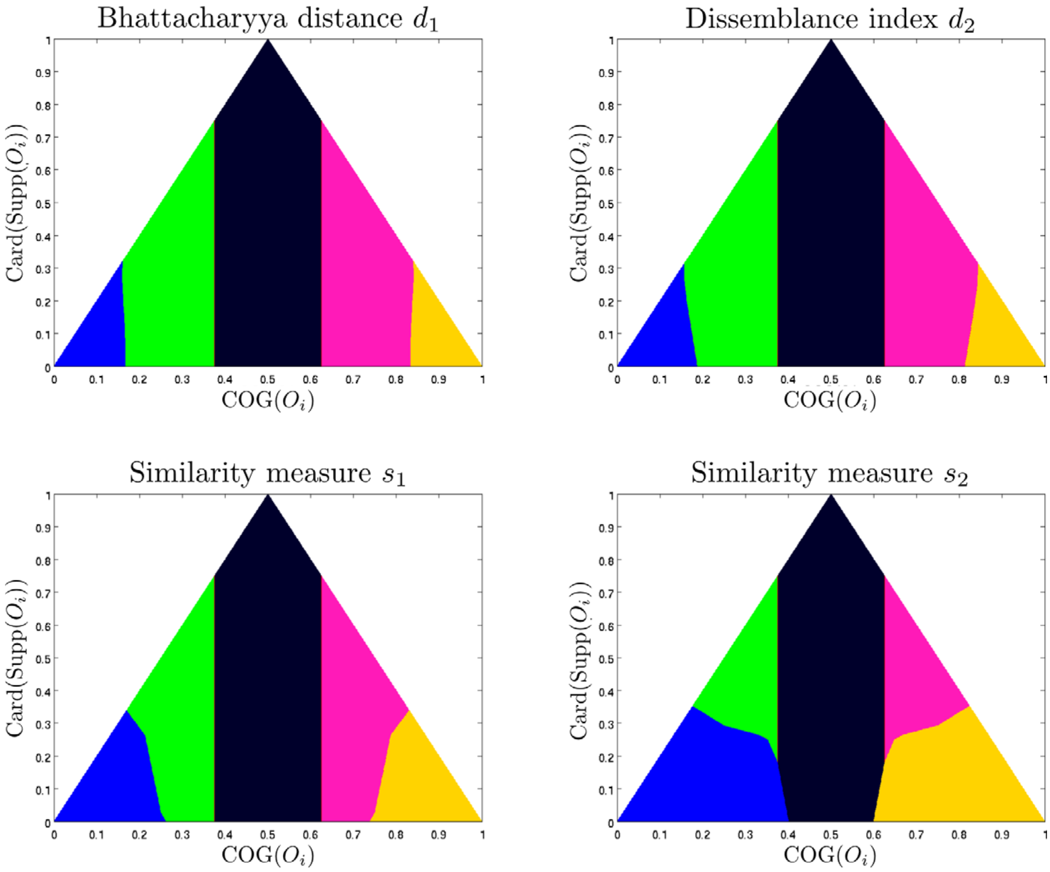


FIGURE 1
Results of the numerical experiment presented in [8]. Symmetrical triangular fuzzy numbers on [0, 1] were randomly generated and linguistically approximated using the "best-fit" approach and all of the above mentioned similarity/distance measures d_1 (top left plot), d_2 (top right plot), s_1 (bottom left plot) and s_2 (bottom right plot). Each colour correspond with one linguistic value of a linguistic scale (uniform, 5 elementary terms) used for the linguistic approximation.

then represent the meanings of the elementary linguistic terms $\mathcal{T}_1, \ldots, \mathcal{T}_j$. The linguistic approximation is then performed based on the determination to which interval the value of the COG of the approximated fuzzy set belongs. This in essence means that we are looking for the maximum membership degree of the COG of the approximated fuzzy set to the meanings of the linguistic labels and assigning the label where this membership degree is maximal.

Since triangular fuzzy numbers are used, each of them can be uniquely represented by its COG and the cardinality of its support. This way each generated fuzzy number in O^{IN} is represented by a single point in the triangular area in Figures 2 - 5; in Figure 6 only a subset of the set O^{IN} is presented, i.e. the points (in red) correspond with those fuzzy numbers for which an unambiguous linguistic approximation can not be found in the "best-fit" context. To stress the effect of using extended linguistic scales for linguistic approximation, Figure 7 presents a decomposition of the plot for d_1

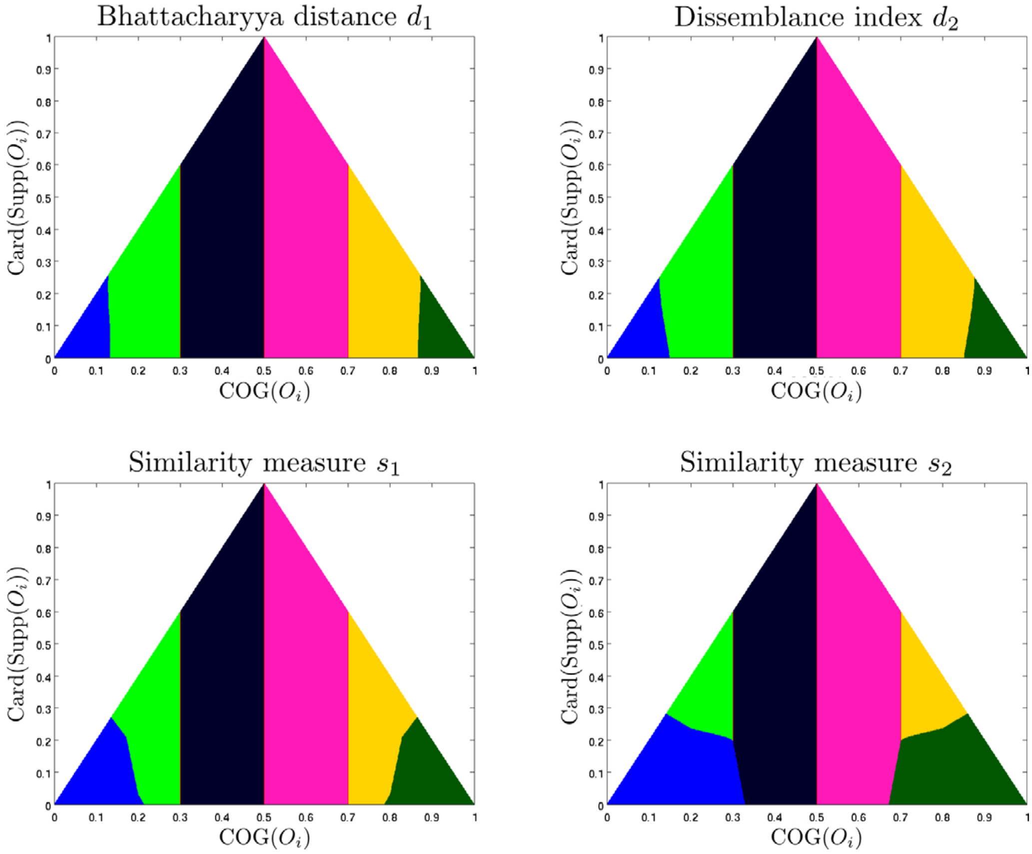


FIGURE 2
The results of linguistic approximation of the elements of O^{IN} represented by points with coordinates (COG(O_i), Card(Supp(O_i))) using a 6-element uniform linguistic scale $\mathcal{V}_6$. The meanings of the elementary terms of $\mathcal{V}_6$ are modeled as triangular fuzzy numbers. Only elementary linguistic terms are considered, each is represented by a different colour.

presented in Figure 4 into subplot, each subplots representing a different level of the derived terms. To investigate the results of linguistic approximation, each linguistic label available for linguistic approximation under the given linguistic variable (scale) is assigned a different colour (red is reserved for ambiguous cases) and the elements of O^{IN} are then presented in the plots in such colour that corresponds with the result of the linguistic approximation using the given distance/similarity measure. The frequencies of the assignment of each label are summarized in Tables 1 and 2. In tables considering elementary terms only, the comparison with the crisp benchmark linguistic approximation method is also available.

To asses the effect of odd/even number of elementary terms of the linguistic scale on linguistic approximation, two elementary linguistic scales $\mathcal{V}_6$, $\mathcal{V}_7$ are used in this paper: $\mathcal{T}(\mathcal{V}_6) = \{\mathcal{T}_1, \dots, \mathcal{T}_6\}$ with the respective meanings of these linguistic terms given as $\{T_1, \dots, T_6\} = \{(0, 0, 0.2), (0, 0.2, 0.4), (0.2, 0.4, 0.6), (0.4, 0.6, 0.8), (0.6, 0.8, 1), (0.8, 1, 1)\}$ and $\mathcal{T}(\mathcal{V}_7) = \{\mathcal{T}_1, \dots, \mathcal{T}_7\}$ with the meanings $\{T_1, \dots, T_7\} = \{(0, 0, 1/6), (0, 1/6, 1/3), (1/6, 1/3,$

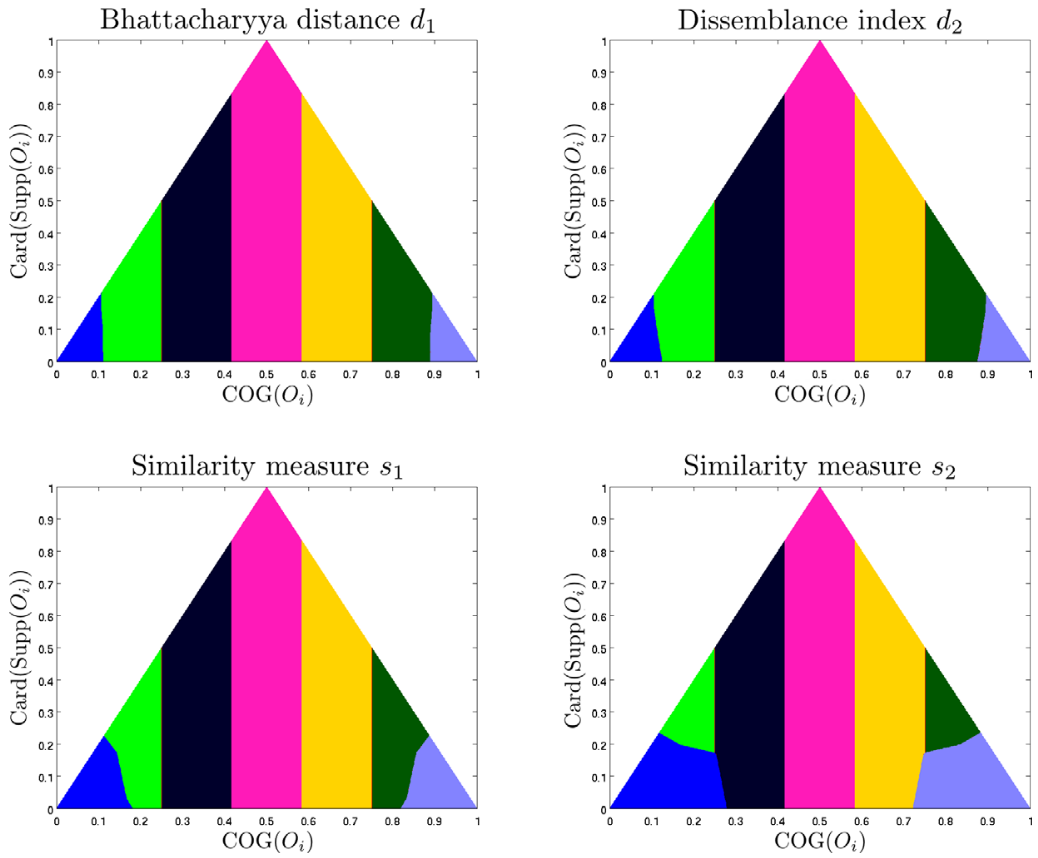


FIGURE 3
The results of linguistic approximation of the elements of O^{IN} represented by points with coordinates (COG(O_i), Card(Supp(O_i))) using a 7-element uniform linguistic scale $\mathcal{V}_7$. The meanings of the elementary terms of $\mathcal{V}_7$ are modeled as triangular fuzzy numbers. Only elementary linguistic terms are considered, each is represented by a different colour.

1/2), (1/3, 1/2, 2/3), (1/2, 2/3, 5/6), (2/3, 5/6, 1), (5/6, 1, 1)}. The results of the linguistic approximation for all the elements of O^{IN} are presented in Figures 2 and 3 and Tables 1 and 2. In the figures, each colour corresponds with one of the elementary terms, the figures therefore summarize the behavior of d_1, d_2, s_1 and s_2 in the linguistic approximation using the elementary terms of the linguistic scales and the "best-fit" approach. The tables then summarize the frequencies assignment of each particular linguistic label. A comparison with the naive crisp linguistic approximation technique specified above is also provided in the tables. The tables also include information concerning the number of ambiguous cases, when (1) has more than one solution and the linguistic label can not be unambiguously determined. The total number of fuzzy numbers evenly distributed in the feasible part of the COG-Card(Supp) space used for the computation of these results is 500 000.

Also, two *extended* linguistic scales $\mathcal{V}_{6e}, \mathcal{V}_{7e}$ obtained from linguistic scales $\mathcal{V}_6, \mathcal{V}_7$ are considered in our analysis. The corresponding results of the linguistic approximation for all the elements of O^{IN} are presented in Figures

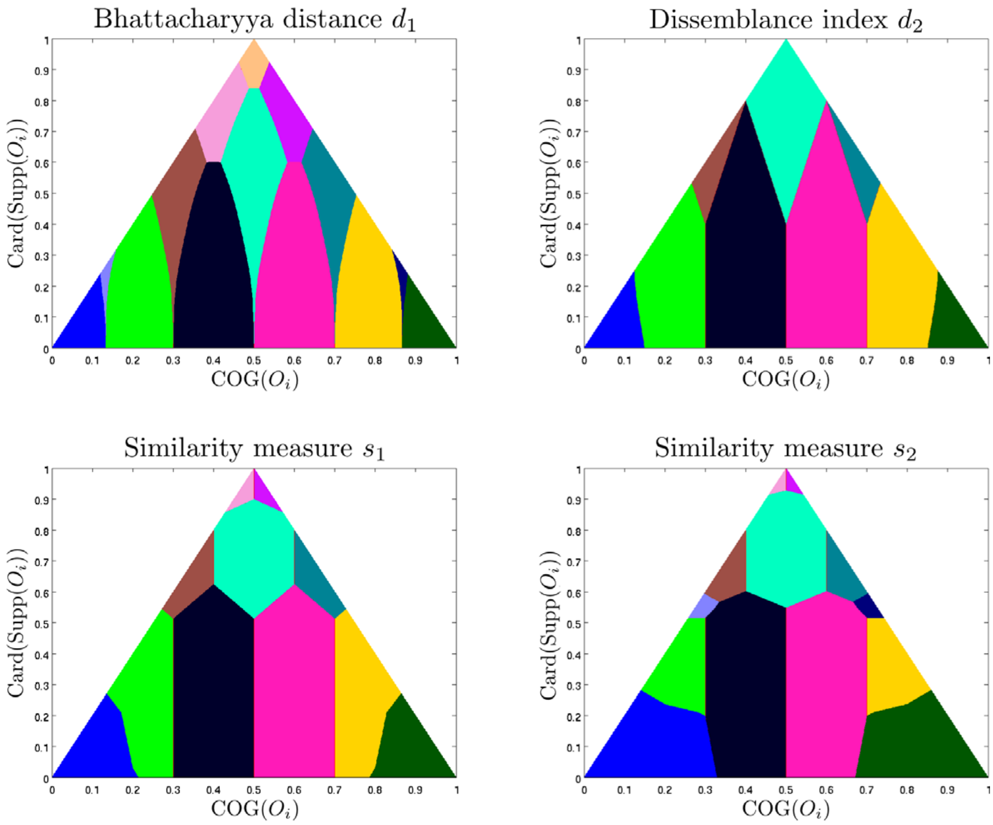


FIGURE 4
The results of linguistic approximation of the elements of O^{IN} represented by points with coordinates (COG(O_i), Card(Supp(O_i))) using a uniform linguistic scale $\mathcal{V}_{6e}$ with 6 elementary terms. The meanings of the elementary terms of $\mathcal{V}_{6e}$ are modeled as triangular fuzzy numbers. Elementary and derived linguistic terms are considered for the linguistic approximation, each elementary and derived term is represented by a different colour.

4 - 5 and the corresponding frequencies in Tables 3 - 4. We also present graphical summaries for the results obtained for $\mathcal{V}_{8e}$, $\mathcal{V}_{9e}$ and $\mathcal{V}_{10e}$ scales, since the analysis found that for s_1 and s_2 the areas of ambiguity start to appear, as presented in Figure 6.

5 DISCUSSION

The results presented in the previous section were obtained for the linguistic approximation of symmetrical triangular fuzzy numbers by the elementary (derived) linguistic terms of linguistic variables defined in the preliminaries section. The minimization of distance (maximization of similarity) is employed to find the best fitting linguistic approximation for a given fuzzy number, i.e. the “best fit” approach is adopted here. Although these assumptions may seem to be rather restrictive, there are still several relevant conclusions that can be made in terms of the performance of d_1, d_2, s_1 and s_2

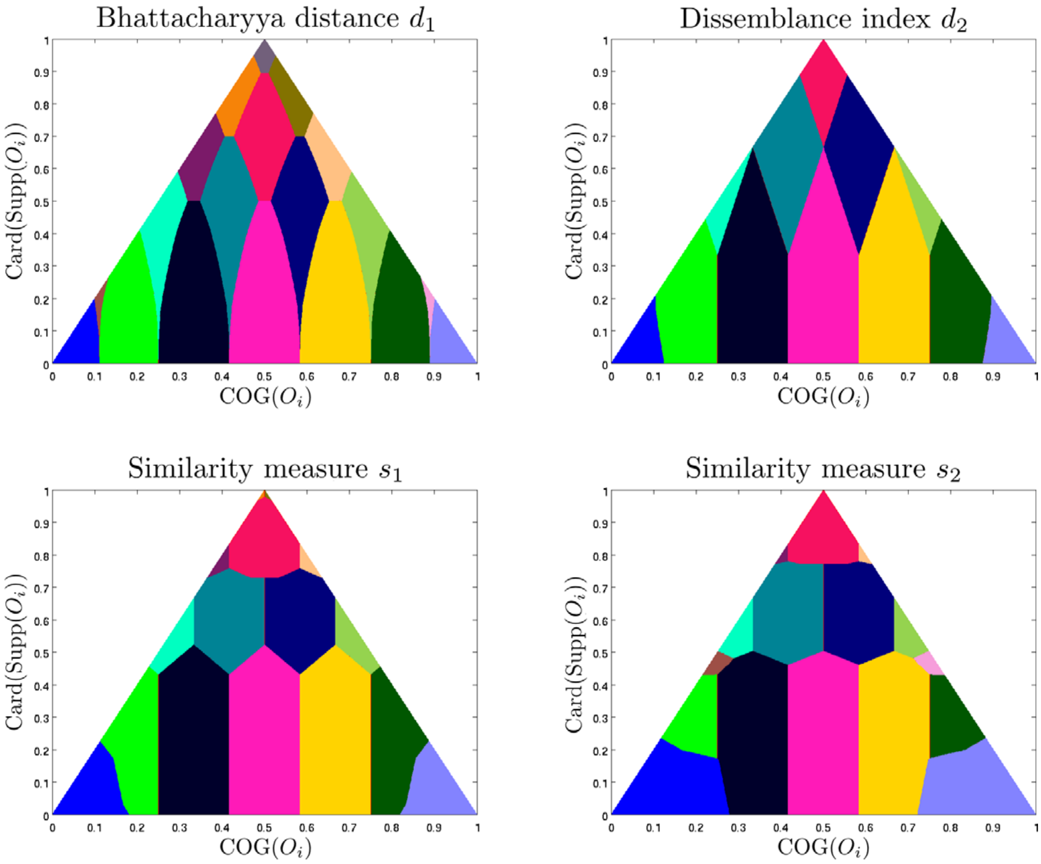


FIGURE 5
The results of linguistic approximation of the elements of O^{IN} represented by points with coordinates (COG(O_i), Card(Supp(O_i))) using a uniform linguistic scale $\mathcal{V}_{7e}$ with 7 elementary terms. The meanings of the elementary terms of $\mathcal{V}_{7e}$ are modeled as triangular fuzzy numbers. Elementary and derived linguistic terms are considered for the linguistic approximation, each elementary and derived term is represented by a different colour.

in linguistic approximation. The methodology for the investigation of performance of linguistic approximation of asymmetrical triangular fuzzy numbers by the elementary terms was outlined in [9] along with the discussion of possible issues connected with the graphical representation of its outputs. Let us first consider only the linguistic approximation using the elementary terms. It is apparent from Figures 2 and 3, that the outputs of linguistic approximation using d_1, d_2, s_1 and s_2 are similar as long as fuzzy numbers with higher uncertainty are considered (this holds approximately for Card(Supp(O_i)) > 0.4 for $\mathcal{V}_5$ - see Figure 1, Card(Supp(O_i)) > 0.3 for $\mathcal{V}_6$ - see Figure 2, Card(Supp(O_i)) > 0.25 for $\mathcal{V}_7$ - see Figure 3). In general there is an apparent trend of the methods getting more and more agreement concerning the linguistic approximation of fuzzy numbers with sufficient level of uncertainty, while the required uncertainty (proportional to Card(Supp(O_i))) gets lower with the number of elementary terms of the linguistic scale used for linguistic approximation. For fuzzy numbers of low uncertainty one of

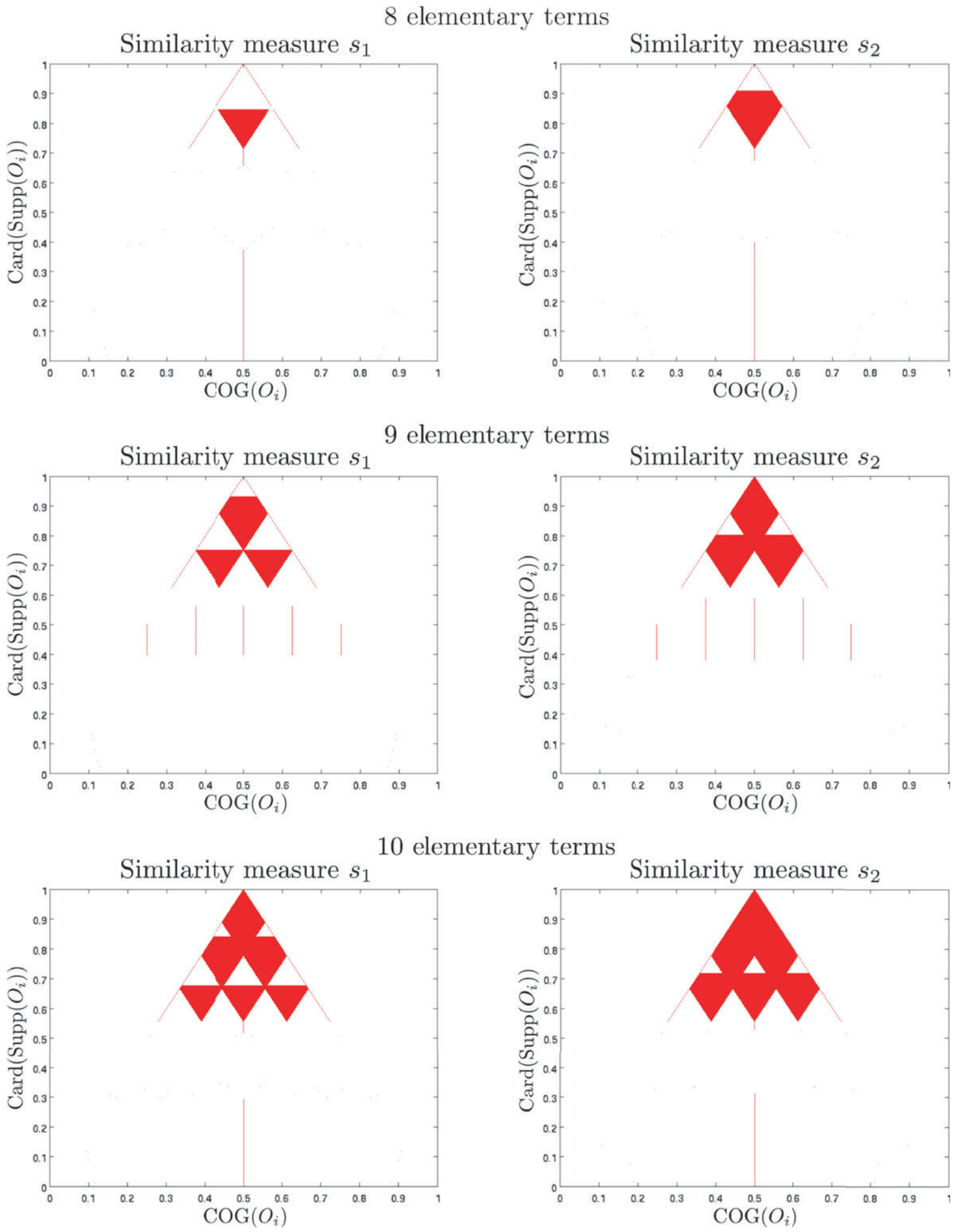


FIGURE 6
Elements of O^{IN} represented by points with coordinates (COG(O_i), Card(Supp(O_i))) using an extended linguistic scale $\mathcal{V}_{8e}$ with 8 elementary terms (top), $\mathcal{V}_{9e}$ with 9 elementary terms (middle) and $\mathcal{V}_{10e}$ with 10 elementary terms (bottom), for which the best-fit linguistic approximation using s_1 and s_2 is ambiguous (red areas).

the investigated similarity measures stands out. This measure is s_2, in case of which the linguistic terms $\mathcal{T}_2$ and $\mathcal{T}_{n-1}$ of a linguistic scale with n elementary terms will never be used as labels for low-uncertain fuzzy numbers. This could constitute a significant bias, since even in cases when the approximated fuzzy number is a subset of T_2 or T_{n-1}, it might not be assigned the linguistic label $\mathcal{T}_2$ or $\mathcal{T}_{n-1}$ respectively (i.e. some $O \subseteq T_2$ will not be linguistically

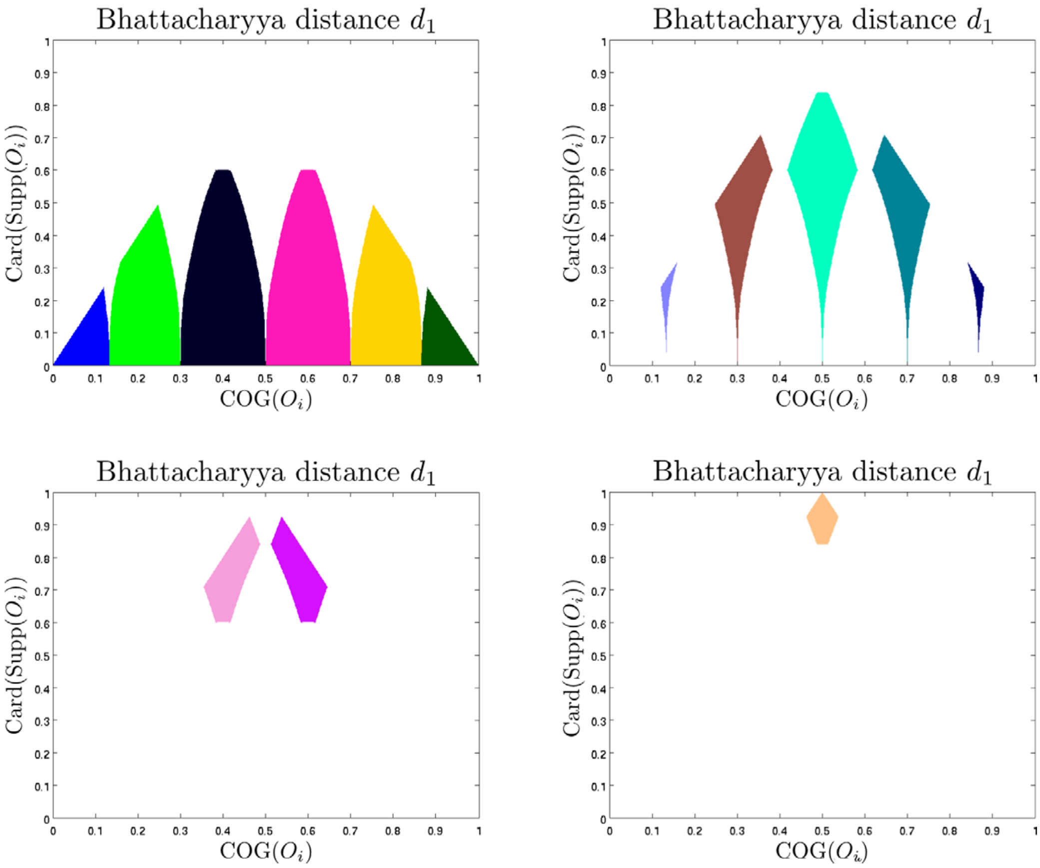


FIGURE 7
The results of linguistic approximation of the elements of O^{IN} represented by points with coordinates (COG(O_i), Card(Supp(O_i))) using a 6-element uniform linguistic scale $\mathcal{V}_6$ and the Bhattacharyya distance d_2 decomposed by levels: level-1 (top left), level-2 (top-right), level-3 (bottom left), level-4 (bottom right).

approximated by $\mathcal{T}_2$). These results are confirmed by the frequency analysis summarized in Tables 1 and 2 - note, that the frequency of use $\mathcal{T}_2$ and $\mathcal{T}_{n-1}$ in both tables is significantly lower than in all the other measures and the crisp benchmark approach. This can be attributed to the fact that the similarity measure s_2 stresses the shape of the fuzzy numbers much more strongly than the other investigated methods. If fuzzy numbers with relatively low uncertainty are expected to be linguistically approximated, s_2 might not be the best method to choose. Based on the figures we can also conclude, that the d_1 distance seems to rely mainly on the COG information in the assignment of linguistic labels, as long as only elementary terms of the linguistic scales are considered. However the results of linguistic approximation using d_1 differ from the results of the crisp benchmark (see Tables 1 and 2).

The results concerning the Bhattacharyya d_1 distance, however, change significantly when the derived linguistic terms are considered, i.e. when extended linguistic scales are used. Figures 4 and 5 clearly show that using d_1 it is possible to get a level-2 linguistic label even for very low-uncertain

	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$	$\mathcal{T}_6$	**Ambiguous**
d_1	17 091	72 606	159 200	159 200	72 606	17 091	2 206
d_2	18 177	71 501	159 200	159 200	71 501	18 177	2 244
s_1	30 673	59 022	159 200	159 200	59 022	30 673	2 210
s_2	59 985	32 646	156 461	156 461	32 646	59 985	1 816
Crisp	9 900	79 600	159 200	159 200	79 600	9 900	2 600

TABLE 1
Frequencies of assignment of each of the elementary linguistic terms $\mathcal{T}_1, \ldots, \mathcal{T}_6$ of $\mathcal{V}_6$ to the elements of O^{IN} in the "best-fit" linguistic approximation using d_1, d_2, s_1 and s_2 and the crisp approach. The ambiguous cases (when (1) has more than one solution) are calculated separately and are presented in the last column of the table.

	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$	$\mathcal{T}_6$	$\mathcal{T}_7$	**Ambiguous**
d_1	11 862	50 387	110 722	153 056	110 722	50 387	11 862	1 002
d_2	12 620	49 616	110 722	153 056	110 722	49 616	12 620	1 028
s_1	21 475	40 772	110 722	153 056	110 722	40 772	21 475	1 006
s_2	42 629	22 432	108 081	153 056	108 081	22 432	42 629	660
Crisp	6 972	55 278	110 722	153 056	110 722	55 278	6 972	1 000

TABLE 2
Frequencies of assignment of each of the elementary linguistic terms $\mathcal{T}_1, \ldots, \mathcal{T}_7$ of $\mathcal{V}_7$ to the elements of O^{IN} in the "best-fit" linguistic approximation using d_1, d_2, s_1 and s_2 and the crisp approach. The ambiguous cases (when (1) has more than one solution) are calculated separately and are presented in the last column of the table.

fuzzy numbers, that lie "between" the meanings of two neighboring elementary terms. This behavior can be described as preferring more uncertain linguistic terms the meanings of which are supersets to the approximated fuzzy number to the elementary terms of lower uncertainty, whose meanings, however, do not overlap with the approximated fuzzy number enough (see Figure 7 - top right subplot - for more details). This way when linguistic labels with meanings on different levels of uncertainty are allowed, d_1 embodies at least partially the requirement of subsethood - the approximating term that is a superset of the approximated fuzzy number (meaning-wise) is preferred to others, that might be closer in terms of shape (meaning-wise). Thanks to this inherent preference of more uncertain labels, d_1 can be observed to assign linguistic labels from higher levels, that are never even used when the other three measures are applied (see level 4 in Table 3 and level 5 in Table 4). Suggesting linguistic labels that are more general than the approximated output (while the output is a subset of the meaning of this more general label) might be a desired property in applications, where it is necessary to obtain a description that includes the output as a subcase (special-case or even as a representative). We can also note, that for any of the four investigated

Level 1	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$	$\mathcal{T}_6$
d_1	16 458	59 537	88 863	88 863	59 537	16 458
d_2	18 177	68 234	119 401	119 401	68 234	18 177
s_1	30 673	57 854	113 295	113 295	57 854	30 673
s_2	59 985	30 897	111 207	111 207	30 897	59 985

Level 2	$\mathcal{T}_{12}$	$\mathcal{T}_{23}$	$\mathcal{T}_{34}$	$\mathcal{T}_{45}$	$\mathcal{T}_{56}$
d_1	2 744	31 373	55 119	31 373	2 744
d_2	0	13 334	60 000	13 334	0
s_1	0	14 219	58 847	14 219	0
s_2	3 147	11 617	61 603	11 617	3 147

Level 3	$\mathcal{T}_{13}$	$\mathcal{T}_{24}$	$\mathcal{T}_{35}$	$\mathcal{T}_{46}$
d_1	0	19 858	19 858	0
d_2	0	0	0	0
s_1	0	3 532	3 532	0
s_2	0	1 511	1 511	0

Level 4	$\mathcal{T}_{14}$	$\mathcal{T}_{25}$	$\mathcal{T}_{36}$
d_1	0	7 163	0
d_2	0	0	0
s_1	0	0	0
s_2	0	0	0

Ambiguous
52
1 708
2 007
1 669

TABLE 3
Frequencies of assignment of each of the elementary linguistic terms $\mathcal{T}_1, \ldots, \mathcal{T}_6$ of $\mathcal{V}_{6e}$ and all the derived terms $\mathcal{T}_{ij} =$ "$\mathcal{T}_i$ to $\mathcal{T}_j$" to the elements of O^{IN} in the "best-fit" linguistic approximation using d_1, d_2, s_1 and s_2. Each level of the linguistic terms is presented in a separate subtable, the ambiguous cases (when (1) has more than one solution) are calculated separately and are presented in the last subtable. Linguistic terms from the highest two levels are not used.

distance/similarity measures, the derived linguistic terms of a linguistic scale with n elementary terms from level $n-1$ up are never used for $n \geq 4$.

Unlike d_1 and both the similarity measures, the dissemblance index d_2 tends to assign level-1 terms even to fuzzy numbers with considerably high uncertainty (see e.g. the height of the areas representing the middle two elementary terms in Figure 4 and the middle three in Figure 5) when extended linguistic scales are used. From this point of view, using d_2 there is a potential risk of uncertainty reduction.

When the similarity measures s_1 and s_2 are used in the extended linguistic scale setting, their performance is analogical to their performance when only elementary linguistic terms are allowed - that is as long as fuzzy numbers with lower uncertainty are approximated. For the approximation of fuzzy numbers with higher uncertainty, s_1 and s_2 seem to be less useful, mainly as the number of the elementary terms of the linguistic scale increases above 7. Figure 6 depicts the elements of O^{IN} for which the best-fit linguistic approximation using s_1 and s_2 using an extended linguistic scale with 8 elementary

Level 1	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$	$\mathcal{T}_6$	$\mathcal{T}_7$
d_1	11426	41326	61704	61743	61704	41326	11426
d_2	12620	47375	83000	83334	83000	47375	12620
s_1	21475	40042	79261	79690	79261	40042	21475
s_2	42629	21248	76811	80504	76811	21248	42629

Level 2	$\mathcal{T}_{12}$	$\mathcal{T}_{23}$	$\mathcal{T}_{34}$	$\mathcal{T}_{45}$	$\mathcal{T}_{56}$	$\mathcal{T}_{67}$
d_1	1919	21798	38255	38255	21798	1919
d_2	0	9297	46314	46314	9297	0
s_1	0	9623	42308	42308	9623	0
s_2	2657	7627	45535	45535	7627	2657

Level 3	$\mathcal{T}_{13}$	$\mathcal{T}_{24}$	$\mathcal{T}_{35}$	$\mathcal{T}_{46}$	$\mathcal{T}_{57}$
d_1	0	13791	33015	13791	0
d_2	0	0	18594	0	0
s_1	0	1963	29378	1963	0
s_2	0	827	23858	827	0

Level 4	$\mathcal{T}_{14}$	$\mathcal{T}_{25}$	$\mathcal{T}_{36}$	$\mathcal{T}_{47}$
d_1	0	10762	10762	0
d_2	0	0	0	0
s_1	0	167	167	0
s_2	0	0	0	0

Level 5	$\mathcal{T}_{15}$	$\mathcal{T}_{26}$	$\mathcal{T}_{37}$
d_1	0	3216	0
d_2	0	0	0
s_1	0	0	0
s_2	0	0	0

Ambiguous
64
860
1254
970

TABLE 4
Frequencies of assignment of each of the elementary linguistic terms $\mathcal{T}_1, \dots, \mathcal{T}_7$ of $\mathcal{V}_{7e}$ and all the derived terms $\mathcal{T}_{ij}$ = "$\mathcal{T}_i$ to $\mathcal{T}_j$" to the elements of O^{IN} in the "best-fit" linguistic approximation using d_1, d_2, s_1 and s_2 and the crisp approach. Each level of the linguistic terms is presented in separate subtable, the ambiguous cases (when (1) has more than one solution) are calculated separately and are presented in the last subtable. Linguistic terms from the highest two levels are not used.

terms (top subplots), 9 elementary terms (middle subplots) and 10 elementary terms (bottom subplots) is ambiguous (i.e. more than one linguistic label is assigned). The existence of these "ambiguity areas" in the plots represents a significant problem in the use of s_1 and s_2 for linguistic approximation in the above described setting. If the linguistic labels for the fuzzy numbers in ambiguity areas are chosen arbitrarily it is possible to assign different linguistic labels to two different fuzzy numbers in the same part of the ambiguity area (area suggesting the same two or more possible labels), which might be in direct contradiction with the natural ordering of these fuzzy numbers (i.e.

the greater fuzzy number might be assigned a label that precedes the label of the lesser fuzzy number in the ordering of the labels).

6 CONCLUSION

This paper continues in the investigation of behavior of the selected distance and similarity measures in linguistic approximation presented in [8–11]. We adopt the focus on symmetrical triangular fuzzy numbers on [0, 1], which allows easier interpretability of the results (the generalization to asymmetrical fuzzy numbers analogical to [9] is also possible). We extend the scope of the investigation of the performance of the methods to linguistic scales with odd and even number of elementary terms, and also to extended linguistic scales, which provide more possibilities to reflect the uncertainty of the approximated fuzzy numbers. This, however, as can be seen in the results, stresses the problem of ambiguity in linguistic approximation. When only elementary terms are used, it can happen that two linguistic approximations are suggested (because the distance/similarity of the approximated linguistic term to the meanings of two neighboring linguistic values is the same) - in this case the approximated object lies "directly in the middle" and a small shift will result in unambiguous assignment of a linguistic label (i.e. a random assignment of one of these linguistic labels is not a problem). When the extended linguistic scale is used, the small shift of the approximated fuzzy number no longer guarantees unambiguous linguistic approximation. Ambiguous areas appear under the use of some of the investigated methods, where two different linguistic labels (in terms of COG of their meanings), and if random assignment of a linguistic label is performed, it is possible to assign different linguistic labels to two different fuzzy numbers in the same ambiguity area, and the ordering of these labels might not correspond with the natural ordering of the two fuzzy numbers. A difference in the performance of the two distance measures with a clear implication in the possibilities of their use was also identified. The preference of more uncertain approximations by d_1 makes it more suitable in cases where uncertainty of the result plays an important role and should not be underestimated. Bhattacharyya distance can therefore be among the measures of choice when more general linguistic approximations (superterms) are desirable. An opposite behavior was identified in d_2, where the uncertainty of the approximated output can be reduced for some fuzzy numbers (compared to other investigated methods) in the process of linguistic approximation. Overall the results confirm, that the yet not completely mapped landscape of linguistic approximation deserves closer investigation, since specific features of the methods can be identified.

This might be the first step towards creating the “map of methods of choice” for particular application areas and contexts.

ACKNOWLEDGMENTS

This research was partially supported by the grant IGA PrF 2016 025 of the internal grant agency of the Palacký University, Olomouc.

REFERENCES

[1] F. J. Aherne, N. A. Thacker, and I. Rockett, Peter. (1998). The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 34(4):[363]–368.

[2] P. P. Bonissone. (1979). A pattern recognition approach to the problem of linguistic approximation in system analysis. In *Proceedings of the IEEE International Conference on Cybernetics and Society*, pages 793–798, Denver, Colorado.

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

[4] S. R. Hejazi, A. Doostparast, and S. M. Hosseini. (2011). An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers. *Expert Systems with Applications*, 38(8):9179–9185.

[5] A. Kaufmann and M. M. Gupta. (1991). *Introduction to fuzzy arithmetic: theory and applications*. Electrical-Computer Science and Engineering Series. Van Nostrand Reinhold Co.

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

[7] J. Stoklasa. (2014). *Linguistic models for decision support*. Lappeenranta University of Technology, Lappeenranta.

[8] T. Talášek and Stoklasa. (2016). Linguistic approximation under different distances/similarity measures for fuzzy numbers. In *Proceedings of the NSAIS'16 Workshop on Adaptive and Intelligent Systems 2016*, pages 49–52, Lappeenranta. LUT Scientific and expertise publications.

[9] T. Talášek and Stoklasa. (2016). The Role of Distance/Similarity Measures in the Linguistic Approximation of Triangular Fuzzy Numbers. In *Knowledge for Market Use 2016: Our Interconnected and Divided World*, pages 539–546, Olomouc.

[10] T. Talášek, J. Stoklasa, and J. Talašová. (2016). The role of distance and similarity in Bonissone's linguistic approximation method a numerical study. In *Proceedings of the 34th International Conference on Mathematical Methods in Economics*, pages 845–850, Liberec. Technical University of Liberec.

[11] T. Talášek, J. Stoklasa, M. Collan, and P. Luukka. (2015). Ordering of fuzzy numbers through linguistic approximation based on bonissone's two step method. In *CINTI 2015 16th IEEE International Symposium on Computational Intelligence and Informatics*, pages 285–290.

[12] S.H. Wei and S.M. Chen. (2009). A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *Expert Systems with Applications*, 36(1):589–598.

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

[14] R. R. Yager. (apr 2004). 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.

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

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

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

[18] R. Zwick, E. Carlstein, and D. V. Budescu. (1987). Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning*, 1(2):221–242.

Publication II

Stoklasa, J., Talášek, T. and Musilová, J.
Fuzzy approach - a new chapter in the methodology of psychology?

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.

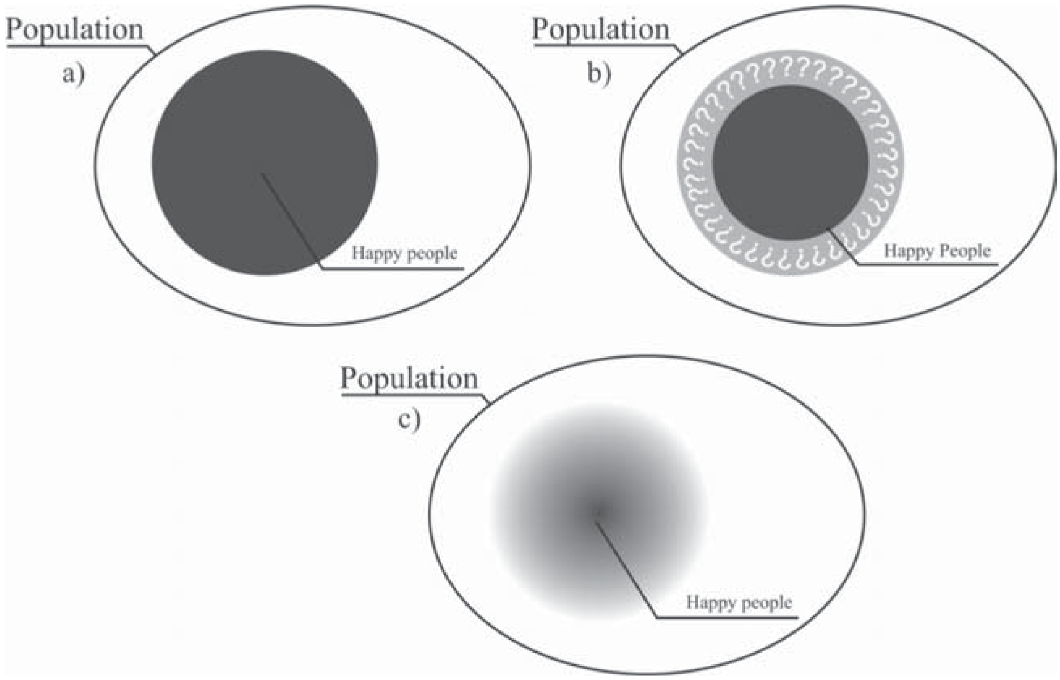


Figure 1. The concept of a fuzzy set: a) crisp set of happy people in the population—people are either happy or not happy; b) crisp set with borderline cases (grey area with question marks) where we cannot decide whether these people are happy or not; c) fuzzy set of happy people—people can be happy 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

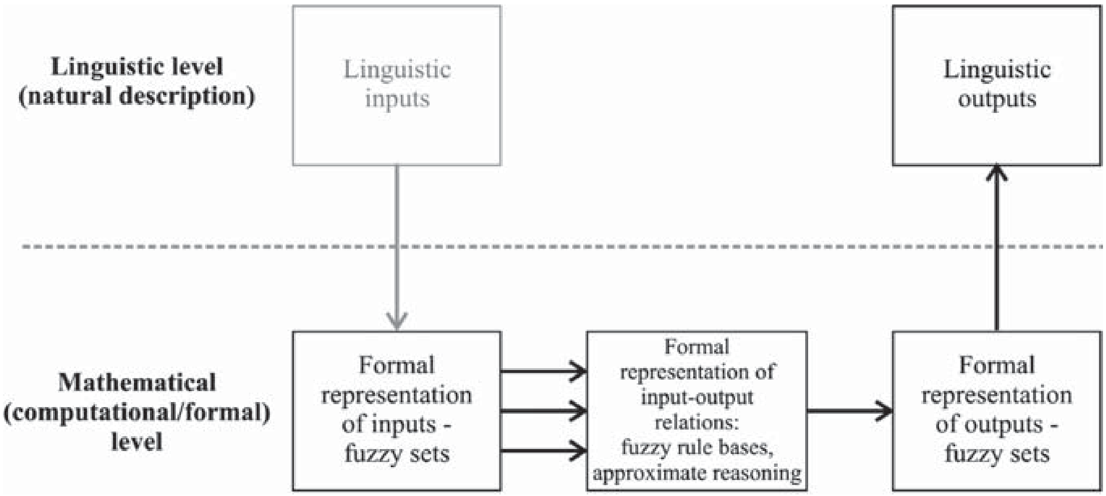


Figure 2. Scheme of the usual approach to mathematical modelling in psychology.

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

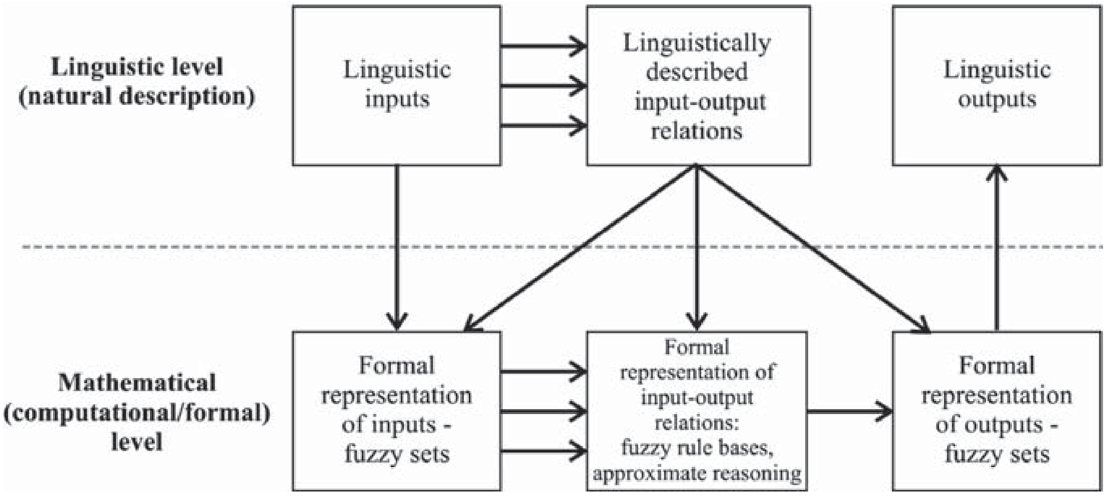


Figure 3. Scheme of the two-level linguistic approach to mathematical modelling suitable for the humanities.

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

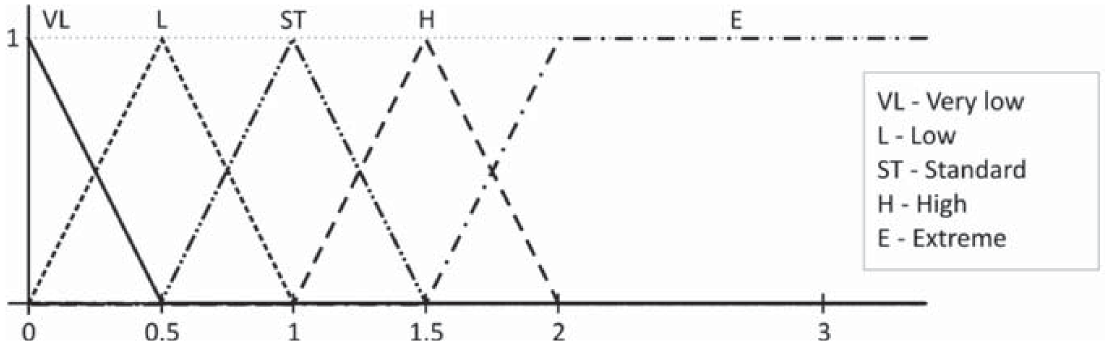


Figure 4. Linguistic scale for evaluating academic faculty in teaching used in IS HAP.

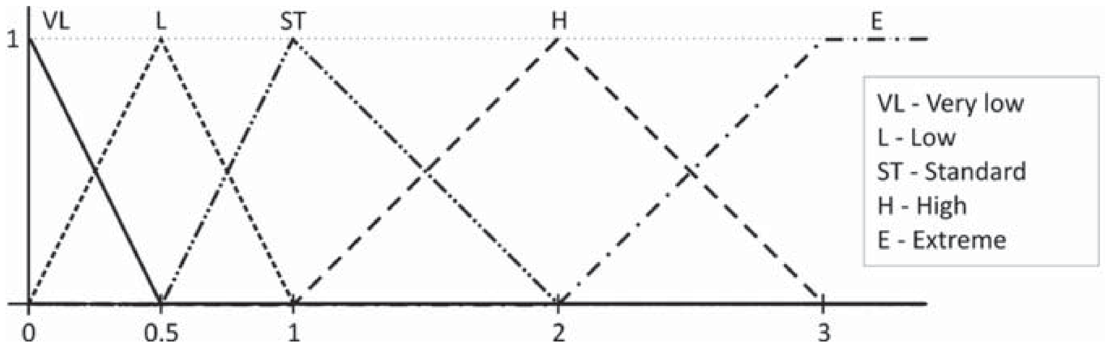


Figure 5. Linguistic scale for the evaluating academic faculty in research and development used in IS HAP—illustration of different meanings of the same linguistic terms (see Figure 4) in a different context.

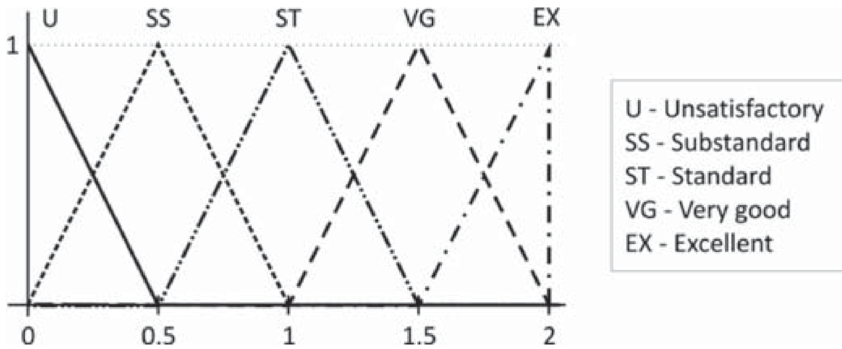


Figure 6. Linguistic scale for evaluating academic faculty used in IS HAP. The linguistic terms in this scale are used to describe outputs of the evaluation model to the users.

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

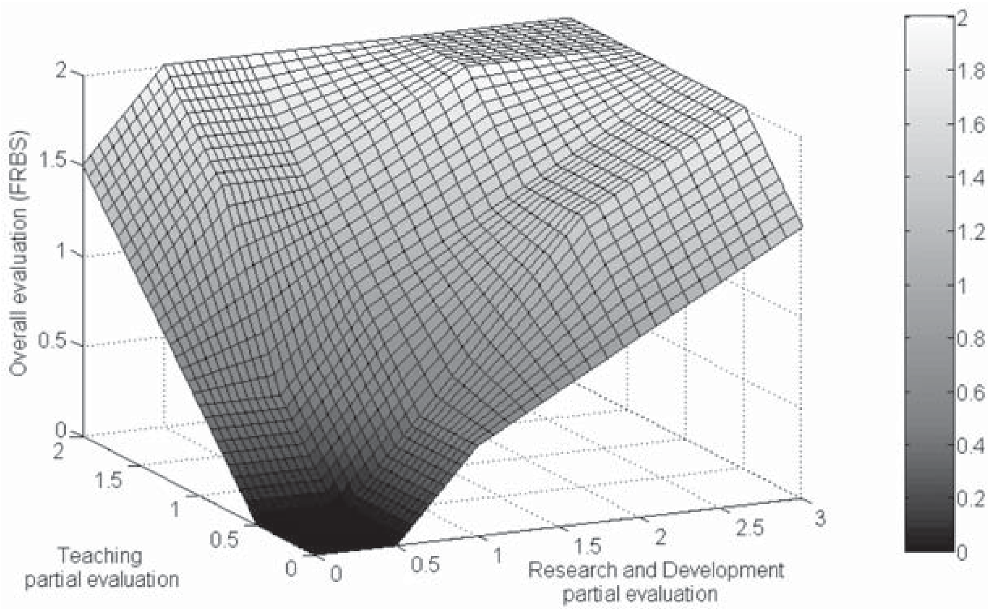
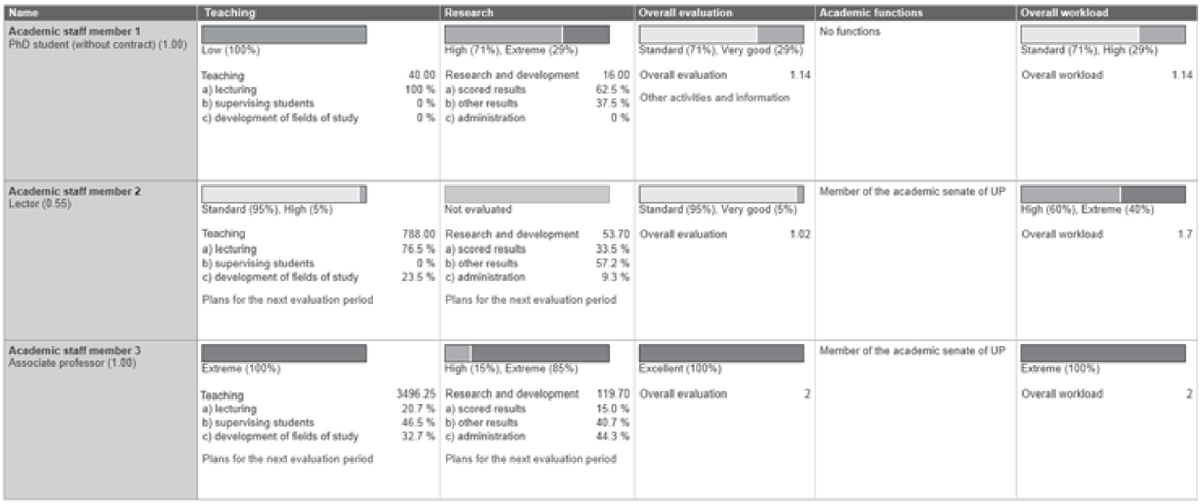


Figure 8. Plot of the evaluation function described by the fuzzy rule base from Figure 7.



Name	Teaching	Research	Overall evaluation	Academic functions	Overall workload
Academic staff member 1 PhD student (without contract) (1.00)	Low (100%) Teaching 40.00 a) lecturing 100 % b) supervising students 0 % c) development of fields of study 0 %	High (71%), Extreme (29%) Research and development 16.00 a) scored results 62.5 % b) other results 37.5 % c) administration 0 %	Standard (71%), Very good (29%) Overall evaluation 1.14 Other activities and information	No functions	Standard (71%), High (29%) Overall workload 1.14
Academic staff member 2 Lector (0.55)	Standard (95%), High (5%) Teaching 788.00 a) lecturing 76.5 % b) supervising students 0 % c) development of fields of study 23.5 % Plans for the next evaluation period	Not evaluated Research and development 53.70 a) scored results 33.5 % b) other results 57.2 % c) administration 9.3 % Plans for the next evaluation period	Standard (95%), Very good (5%) Overall evaluation 1.02	Member of the academic senate of UP	High (60%), Extreme (40%) Overall workload 1.7
Academic staff member 3 Associate professor (1.00)	Extreme (100%) Teaching 3496.25 a) lecturing 20.7 % b) supervising students 46.5 % c) development of fields of study 32.7 % Plans for the next evaluation period	High (15%), Extreme (85%) Research and development 119.70 a) scored results 15.0 % b) other results 40.7 % c) administration 44.3 % Plans for the next evaluation period	Excellent (100%) Overall evaluation 2	Member of the academic senate of UP	Extreme (100%) Overall workload 2

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

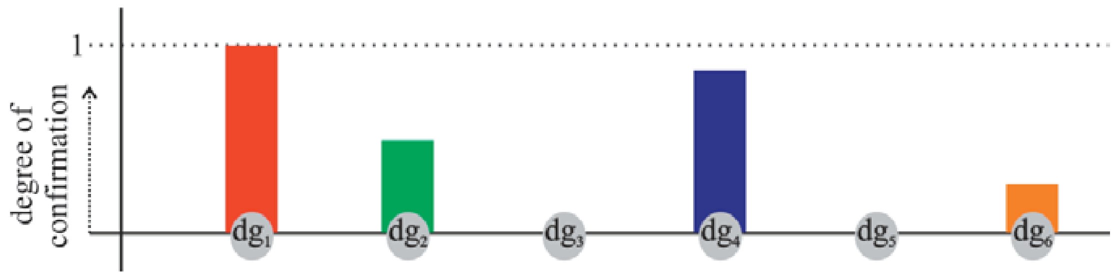


Figure 10. Example of a possible output of a fuzzy classification model—diagnostics (only confirmatory information for all diagnoses available).

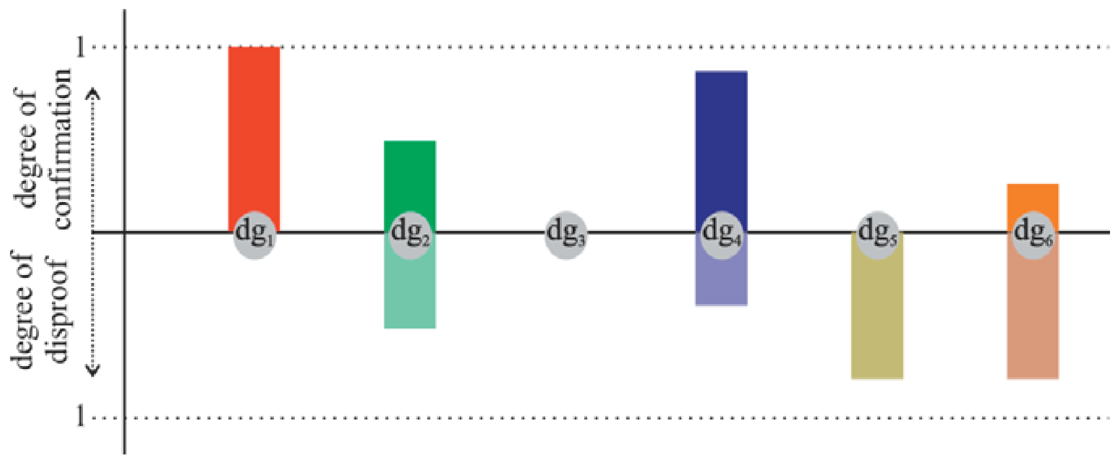


Figure 11. Example of a possible output of a fuzzy classification model—diagnostics (confirmatory information and information disproving each diagnosis available).

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

Publication III

Talášek, T., Stoklasa, J., Collan, M. and Luukka, P.
Ordering of fuzzy numbers through linguistic approximation based on Bonissone's two step method

Reprinted with the permission from
16^{th} *IEEE International Symposium on Computational Intelligence and Informatics 2015*,
pp. 285–290, 2015,

Ordering of fuzzy numbers through linguistic approximation based on Bonissone's two step method

Tomáš Talášek*†, Jan Stoklasa‡, Mikael Collan‡ and Pasi Luukka‡
*LUT Graduate School, Lappeenranta University of Technology
Skinnarilankatu 32, 53851, Lappeenranta, Finland
Email: tomas.talasek@gmail.com
†Department of Applied Economics, Palacký University, Olomouc
Křížkovského 12, 77180, Olomouc, Czech Republic
‡School of Business and Management, Lappeenranta University of Technology
Skinnarilankatu 32, 53851, Lappeenranta, Finland
Email: jan.stoklasa@lut.fi, mikael.collan@lut.fi, pasi.luukka@lut.fi

***Abstract*—Linguistic approximation is a suitable way of transforming mathematical outputs into words that can be easily used and understood by laymen. The methods for linguistic approximation range from simple distance based ones to more complex methods aspiring on finding high semantic match between the approximated output and its linguistic label. This paper builds on Bonissone's proposal of a two step method for linguistic approximation based on a pattern recognition approach. It suggests an algorithm for finding a partial ordering of fuzzy numbers utilizing the partial results from the two step method. As such it proposes a means of finding a partial ordering of fuzzy numbers through linguistic approximation. The proposed algorithm is showcased on several numerical examples and its performance is briefly discussed.**

I. INTRODUCTION

In practical applications the outputs of the mathematical models are frequently in the form of numbers or fuzzy sets. The users of mathematical models therefore must be trained how to interpret these outputs. Since both numbers and fuzzy sets are not a natural means of presenting information (knowledge of the mechanism that generated the results might be necessary for appropriate interpretation), there is a risk (especially for the new practitioners) that they misinterpret the output and react inappropriately. Reasonable way how to avoid the problem of misinterpretation is to present the output of the model in a more natural way for the users – in the linguistic form. The process of translation between fuzzy sets (or real numbers) and words is called linguistic approximation. This paper focuses on one particular linguistic approximation method proposed by Bonissone [1] and suggest a extension of this method that would allow direct comparisons of mathematical outputs through linguistic approximation (or using the information computed to obtain the linguistic approximation). Basic notions of the theory of fuzzy sets are defined in Section 2. Bonissone's method for linguistic approximation is presented in Section 3. In the next Section, the extension of Bonissone's method is proposed together with the ideas how this method could be used for the purposes of ordering fuzzy numbers. Behavior of the proposed algorithm for the ordering of fuzzy numbers is studied on three numerical examples in Section 4.

II. PRELIMINARIES

Let U be a nonempty set (the universe of discourse). 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 $A(.)$ is called a *membership function* of the fuzzy set A. $\text{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an *α-cut* of A for any $\alpha \in [0, 1]$, $\text{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\text{Ker}(A) \neq \emptyset$ (A is *normal*); (2) A_α are closed intervals for all $\alpha \in (0, 1]$ (this implies A is *unimodal*); (3) $\text{Supp}(A)$ is bounded. A family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on [a,b], if $\text{Supp}(A)$ is a subset of an interval $[a, b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \text{Cl}(\text{Supp}(A))$ and $[a_2, a_3] = \text{Ker}(A)$, where $\text{Cl}(\text{Supp}(A))$ denotes a closure of $\text{Supp}(A)$. An *union* of two fuzzy sets A and B (based on Lukasiewicz disjunction) is a fuzzy set $(A \cup_L B)$ defined as follows: $(A \cup_L B)(x) = \min\{1, A(x) + B(x)\}$, $\forall x \in U$. Fuzzy set A on U is a *subset* of fuzzy set B on U ($A \subseteq B$) if $A(x) \leq B(x)$, $\forall x \in U$.

The fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$; for such fuzzy numbers we will use a simplified notation $A = (a_1, a_2, a_3, a_4)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by ordered triplet $A = (a_1, a_2, a_4)$. More details on fuzzy numbers and computations with them can be found for example in [2].

A partial ordering of fuzzy numbers can be defined in the following way: Let A and B be fuzzy numbers on $[a, b]$ then $A > B$ if $A_\alpha > B_\alpha \forall \alpha \in (0, 1]$ that is if $([\inf A_\alpha, \sup A_\alpha] > [\inf B_\alpha, \sup B_\alpha]) \Leftrightarrow ((\inf A_\alpha > \inf B_\alpha \text{ and} \sup A_\alpha \geq$

978-1-4673-8520-6/15/$31.00©2015 IEEE

$\sup B_\alpha$) or ($\inf A_\alpha \geq \inf B_\alpha$ and $\sup A_\alpha > \sup B_\alpha$)). If $A > B$ and $B > A$, then $A = B$. Otherwise A and B are incomparable. This partial ordering is a natural one and can be well applied to the meanings of linguistic terms modeled by fuzzy numbers.

A *fuzzy scale* on $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on $[a, b]$, that form a Ruspini fuzzy partition (see [4]) of the interval $[a, b]$, i.e. for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. A *linguistic variable* ([10]) is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values (terms), X is a universe on which the meanings of the linguistic values are defined, G is an syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X. *Linguistic approximation* is the process that assign appropriate labels (known linguistic terms of a linguistic scale) to general fuzzy sets. From mathematical point of view it is a mapping from the set of all fuzzy sets on X to $\mathcal{T}(\mathcal{V})$.

A linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ is called a *linguistic scale* on $[a, b]$ if $X = [a, b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$ and $T_i = \mathcal{M}(\mathcal{T}_i)$, $i = 1, \ldots, s$, form a fuzzy scale on $[a, b]$. Terms $\mathcal{T}_1, \ldots, \mathcal{T}_s$ are called *elementary terms* of linguistic scale. The *extended linguistic scale* is linguistic scale, that besides elementary terms $\mathcal{T}_1, \ldots, \mathcal{T}_s$ contains also derived terms in the form $\mathcal{T}_i$ to $\mathcal{T}_j$, where $i < j$ and $i, j \in \{1, \ldots, s\}$ and $\mathcal{M}(\mathcal{T}_i \text{ to } \mathcal{T}_j) = T_i \cup_L T_{i+1} \cup_L \ldots \cup_L T_j$. The extended linguistic scale thus contains linguistic values of different levels of uncertainty – from the possibly least uncertain elementary terms $\{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$ to the most uncertain linguistic term $\mathcal{T}_{1s}$ (Uncertainty can be assessed by the cardinality of the meanings of these linguistic terms). Derived linguistic terms $\mathcal{T}_i$ to $\mathcal{T}_j$ are called *level* $j - i$ *terms* and can be also denoted by $\mathcal{T}_{ij}$. Elementary linguistic terms $\mathcal{T}_i$ are called *level 1 terms* and can be also denoted by $\mathcal{T}_{ii}$ (i.e. $\mathcal{T}_i = \mathcal{T}_{ii}$ to unify the notation). More details on linguistic scales and extended linguistic scales can be found for example in [6].

Linguistic hedge is a word (generally an adverb) that can be applied to a linguistic term to modify its meanings. From mathematical point of view a linguistic hedge is a function which modifies the membership functions of fuzzy sets - for example the linguistic hedge *very* (using Zadeh's [10] definition) applied to a fuzzy set A on U results is a fuzzy set *very* A with membership function defined $(\textit{very } A)(x) = A^2(x)$, $x \in U$.

We are frequently required to be able to represent fuzzy sets by real numbers, this procedure is called *defuzzification*. In applications an approximation of a fuzzy set A by its center of gravity (COG) t_A is frequently used. The *center of gravity* of a fuzzy set A defined on $[a, b]$ is defined by the formula $t_A = \int_a^b A(x)x\,dx / \int_a^b A(x)dx$. Other possible defuzzification methods are discussed in [3].

III. Two step method for linguistic approximation

In 1979 Bonissone [1] introduced his two step pattern recognition approach for linguistic approximation. This method consist of two steps – in the first step the set of m suitable linguistic terms of some linguistic variable (suitability is assessed according to a chosen set of features of the meanings of these linguistic terms) is found ("preselection step") and in the second step the most appropriate term from the preselected set is chosen based on Bhattacharyya distance between the output of mathematical model to be linguistically approximated and the meanings of the preselected linguistic terms.

Let $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ be the set of all linguistic terms of a linguistic variable $\mathcal{V}$ for which $T_i = \mathcal{M}(\mathcal{T}_i)$, $i = 1, \ldots, n$ are fuzzy sets defined on the same universe as a fuzzy set Out, where Out is an output of a mathematical model to be linguistically approximated by a linguistic term of $\mathcal{V}$.

In the first step (preselection) the set $\mathcal{T}(\mathcal{V})$ is reduced in the way, that m terms with characteristics most similar to the ones possessed by the output Out are preserved. For this purpose each T_i, $i = 1, \ldots, n$ is represented by four real numbers, that each represent one of the four features (selected by Bonissone [1]) of this fuzzy set: i) Cardinality of fuzzy set T_i (Card(T_i)); ii) Fuzziness of the fuzzy set T_i (nonprobalistic entropy, Entropy(T_i)); iii) Center of gravity of fuzzy sets T_i (COG(T_i)); iv) Skewness of fuzzy set T_i (SKEW(T_i)). These four features represent the fuzzy set (and the linguistic term $\mathcal{T}_i$ associated with it) in four-dimensional space. For preselection of linguistic terms the weighted Euclidean distance $d_1(T_i, Out)$ is computed between the fuzzy set Out and each of the fuzzy sets T_i, $i = 1, \ldots, n$ represented by a quadruplet of the numerical values of the chosen four features using Formula (1). A reordered set of linguistic terms $\{\mathcal{T}_{p1}, \ldots, \mathcal{T}_{pn}\}$ so that $d_1(T_{p1}, Out) \leq \ldots \leq d_1(T_{pn}, Out)$. This way we obtain the preselected term set $\mathcal{T}_p = \{\mathcal{T}_{p_1}, \ldots, \mathcal{T}_{p_m}\}$, where $1 \leq m \leq n$. For two fuzzy sets A and B represented by quadruplets of features $(a_1, \ldots, a_4)$ and $(b_1, \ldots, b_4)$ the Euclidean distance is computed by

$$d_1(A, B) = \sum_{i=1}^{4} w_i(a_i - b_i), \qquad (1)$$

where w_i are normalized real weights[1] (i.e. $\sum_{i=1}^{4} w_i = 1$ and $w_i \in [0, 1]$, $i = \{1, \ldots, 4\}$), and a_i (b_i) are computed by

$$\begin{aligned}
a_1 &= \text{Card}(A) = \int_U A(x)dx, && (2)\\
a_2 &= \text{Entropy}(A) = \int_U \text{S}(A(x))dx, && (3)\\
a_3 &= COG(A) = \int_U xA(x)dx / \text{Card}(A), && (4)\\
a_4 &= SKEW(A) = \int_U (x - \text{COG}(A))^3 A(x)dx, && (5)
\end{aligned}$$

where $S(y) = -y\ln(y) - (1 - y)\ln(1 - y)$.

In the second step, the linguistic approximation $\mathcal{T}^* \in \mathcal{T}_p$ of the fuzzy set Out is computed. Fuzzy set T^* (which is a mathematical meaning of a linguistic term $\mathcal{T}^*$) is computed

[1] The choice of weights is usually left with the user of the model and some features could be even optional. Wenstøp [7] for example proposed (in his method for linguistic approximation) to use only two features – cardinality and center of gravity.

as $T^* = \arg\min_{T_{p_j} \in T_p} d_2(T_{p_j}, Out)$ using the modified Bhattacharyya distance:

$$d_2(A,B) = \left[1 - \int_U (A^*(x) \cdot B^*(x))^{1/2} dx\right]^{1/2}, \qquad (6)$$

where $A^*(x) = A(x)/\text{Card}(A(x))$. This way the linguistic term $\mathcal{T}^*$ is found as the closest linguistic approximation among the preselected linguistic terms.

IV. Ordering of fuzzy sets through linguistic approximation based on the Two step method

Linguistic approximation is usually used in situations when the user of the mathematical model requires the results in an understandable form (i.e. in a linguistic form). To find a proper linguistic approximation lots of calculations may need to be done and substantial part of these result is not used for anything else (see the complexity of Bonissone's method described in the previous Section). This raises a question, whether it is possible in situations when we need to approximate more outputs of mathematical models (e.g. outputs of model representing evaluation of various alternatives in decision making problems – for more information see e.g. [5]) to use linguistic approximation (or partial results obtained through calculation of linguistic approximation) for the purposes of decision making – for example to order the alternatives with respect to their evaluations. In this paper we are therefore focusing on the possibility of using the information obtained in the process of linguistic approximation using the two step method to order the outputs of a mathematical model (fuzzy numbers on X). For an overview of other possible methods for ordering fuzzy numbers and their reasonable properties see [8], [9] .

In this paper the structure of an extended linguistic scale $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ is used. We suppose that the user of the mathematical model specifies (or at least approves) the meanings of its n the elementary linguistic terms of this scale. Linguistic terms of this extended linguistic scale can be ordered by the user directly (this could be difficult and time consuming) or can be partially ordered through the partial ordering of fuzzy numbers that represent the meanings of these linguistic terms, where $\mathcal{T}_{ij}$ is preferred to $\mathcal{T}_{kl}$, $\mathcal{T}_{ij} \succ \mathcal{T}_{kl} \Leftrightarrow M(\mathcal{T}_{ij}) > M(\mathcal{T}_{kl})$, where $\mathcal{T}_{ij}, \mathcal{T}_{kl} \in \mathcal{T}(\mathcal{V})$, $1 \le i < j \le n$, $1 \le k < l \le n$. Linguistic terms $\mathcal{T}_{ij}$ and $\mathcal{T}_{kl}$ are incomparable according to this ordering, if $i > k$ and $j < l$. In the case of incomparable fuzzy numbers, we can use a different method to obtain the ordering of these fuzzy numbers (e.g. a method based on the centers of gravity, where $\mathcal{T}_{ij} \succ_t \mathcal{T}_{kl} \Leftrightarrow t_{M(\mathcal{T}_{ij})} > t_{M(\mathcal{T}_{kl})}$, see Fig. 1). However, from our point of view it is not reasonable to present ordering obtained through the center of gravity method to the user due to possible information loss. It is reasonable to present such terms as incomparable (and eventually let the ordering of these term on the user of the model).

Methods for linguistic approximation usually use complex structures of linguistic terms (linguistic hedges may even be applied to generate the set of linguistic terms). In these cases there is a potential risk, that the user may not understand all these linguistic terms correctly. Therefore in our proposed method the elementary linguistic terms form a linguistic scale

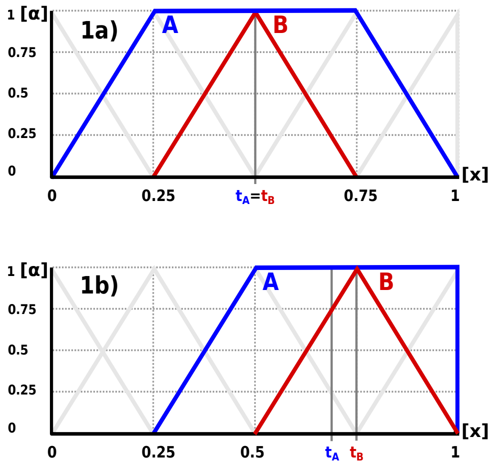


Fig. 1. Fuzzy numbers A and B that are incomparable through natural ordering of fuzzy numbers and center of gravity method (top Figure); fuzzy numbers A and B that are incomparable using the natural ordering method, but can be ordered based on their centers of gravity $B \succ_t A$ (bottom Figure).

and the user must confirm, that he/she understands each elementary linguistic terms correctly and that its fuzzy number meaning is appropriately defined. This in combination with the construction of the linguistic term set of the extended linguistic scale ensures, that the user understands correctly all possible outputs of linguistic approximation (output can be either one elementary term or a derived term represented as two elementary terms connected by "to"). Therefore the user works only with objects (linguistic terms, respectively fuzzy numbers representing their meaning) that he/she understands or with their Lukasiewicz union. The use of the extended linguistic scale ensures, that the output of linguistic approximation is a linguistic term, which meaning is modeled by a normal, unimodal fuzzy set (a fuzzy number).

Let us now take a closer look on the use of the two step method for linguistic approximation in ordering of fuzzy numbers. Let us consider r outputs of some mathematical model $Out_1, \ldots, Out_r$ in the form of fuzzy numbers on X. Let us consider an extended linguistic scale $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ with n elementary terms $\mathcal{T}_1, \ldots, \mathcal{T}_n$,. This linguistic scale will be used for the linguistic approximation of $Out_1, \ldots, Out_r$ and also to obtain an ordering of these outputs.

According to Bonissone in the first step of linguistic approximation it is our goal to reduce the set of all linguistic terms $\mathcal{T}(\mathcal{V})$ in the way, that m terms with meanings semantically closest to the "ideal" linguistic description of each approximated output are found, $\mathcal{T}_p^{Out_q} = \{\mathcal{T}_{p_1}^{Out_q}, \ldots, \mathcal{T}_{p_m}^{Out_q}\}$, where $\mathcal{T}_{p_i}^{Out_q} \in \mathcal{T}(\mathcal{V})$, $i = 1, \ldots, m$, $q = 1, \ldots, r$. Since the semantic context is provided by the universe on which the meanings of the linguistic values of the output linguistic variable are defined, it is sufficient to account for the position, fuzziness and shape of the fuzzy sets (that represent meanings of linguistic terms) to find a pair of semantically close ones.

In the second step the most appropriate linguistic approximation of each output is found from its preselected term set. In the process the Bhattacharyya distance $d_2(M(\mathcal{T}_{ij}), Out_q)$ between the meaning of each linguistic term $\mathcal{T}_{ij}$ from $\mathcal{T}_p^{Out_q}$ and each output $Out_q, q = 1, \ldots, r$ is computed. Therefore for the linguistic approximation of Out_q we obtain an ordered m-tuple of linguistic terms $\mathcal{T}_{app}^{Out_q}$=($\mathcal{T}_{approx_1}^{Out_q},\dots,\mathcal{T}_{approx_m}^{Out_q}$), where $\mathcal{T}_{approx_u}^{Out_q} \in \mathcal{T}_p^{Out_q}, d_2(M(\mathcal{T}_{approx_i}^{Out_q}), Out_q) \leq d_2(M(\mathcal{T}_{approx_j}^{Out_q}), Out_q)$, $\forall i < j$, $u = 1,\dots,m$, $q = 1,\dots,r$. Note that these results are used by Bonissone only to find the best approximation, we will show, how the complete m-tuples can be used to order $Out_1,\dots,Out_r$ (e.g. for the purposes of decision making problem mentioned above). $\mathcal{T}_{approx_1}^{Out_q}$ is linguistic approximation of fuzzy number Out_q according to Bonissone. Let us now consider two outputs of the mathematical model Out_1 and Out_2 in the form of fuzzy numbers on X. These two outputs can be (partially) ordered using the following algorithm:

1) Let $i = 1$.
2) We compute $\mathcal{T}_{app}^{Out_1}$ and $\mathcal{T}_{app}^{Out_2}$.
3) If $\mathcal{T}_{approx_i}^{Out_1} \succ \mathcal{T}_{approx_i}^{Out_2}$ then Out_1 is preferred to Out_2 or if $\mathcal{T}_{approx_i}^{Out_2} \succ \mathcal{T}_{approx_i}^{Out_1}$ then Out_2 is preferred to Out_1. END (ordering has been found). ELSE Go to 4.
4) If $\mathcal{T}_{approx_i}^{Out_1}$ and $\mathcal{T}_{approx_i}^{Out_2}$ are incomparable, leave the ordering of these two outputs to the user[2]. END (outputs are incomparable). ELSE Go to 5.
5) If $i < m$ and $\mathcal{T}_{approx_1}^{Out_1} = \mathcal{T}_{approx_1}^{Out_2}$, then increase i and Go to 3. ELSE Go to 6.
6) Find a pair of adjacent linguistic terms of the same level $\mathcal{T}_{ij}^{Out_1}, \mathcal{T}_{i+1,j+1}^{Out_1}$ and $\mathcal{T}_{ij}^{Out_2}, \mathcal{T}_{i+1,j+1}^{Out_2}$ $i,j = 1,\dots,n-1$. Let assume that $\mathcal{T}_{i+1,j+1} \succ \mathcal{T}_{i,j}$ then if $(d_2(M(\mathcal{T}_{ij}^{Out_1}), Out_1) > d_2(M(\mathcal{T}_{ij}^{Out_2}), Out_2)$ and $d_2(M(\mathcal{T}_{i+1,j+1}^{Out_1}), Out_1) < d_2(M(\mathcal{T}_{i+1,j+1}^{Out_2}), Out_2))$ then Out_1 is preferred to Out_2. ELSE if $(d_2(M(\mathcal{T}_{ij}^{Out_1}), Out_1) < d_2(M(\mathcal{T}_{ij}^{Out_2}), Out_2)$ and $d_2(M(\mathcal{T}_{i+1,j+1}^{Out_1}), Out_1) > d_2(M(\mathcal{T}_{i+1,j+1}^{Out_2}), Out_2))$ then Out_2 is preferred to Out_1. ELSE Go to the beginning of six and choose another pair. END (ordering has been found).

V. Examples of the Proposed Method:

Let $(\mathcal{V}, \mathcal{T}(\mathcal{V}), [0, 1], G, M)$ be a extended linguistic scale with five elementary linguistic terms $\mathcal{T}_1,\dots,\mathcal{T}_5$ with meanings specified in Table I. Meanings of elementary terms are depicted in Fig. 2. Derived linguistic terms are constructed in accordance with Chapter 2. Outputs of the mathematical model are fuzzy numbers on $[0, 1]$ (in examples we consider only triangular fuzzy numbers as approximated outputs). For the preselection step the value of m is equal to 5 (five linguistic terms are preselected).

In next three subsections we present three different examples and stress the strong and weak properties of the presented model for ordering of the outputs.

[2]This situation is depicted in Fig. 1. These situations involve the comparisons of fuzzy sets (one is a subset of the other) with different cardinalities. The natural ordering based on α-cuts of these fuzzy sets is non-existent. Since we are looking for easily interpretable results, we prefer at this point to present both outputs to the decision maker as incomparable and leave the choice of the better one with him/her.

TABLE I. Elementary linguistic terms of the extended linguistic scale used for linguistic approximation and their meanings.

$\mathcal{T}_i$	$T_i = M(\mathcal{T}_i)$	Associated linguistic term
$\mathcal{T}_1$	$(0, 0, 0.25)$	Very bad
$\mathcal{T}_2$	$(0, 0.25, 0.5)$	Bad
$\mathcal{T}_3$	$(0.25, 0.5, 0.75)$	Average
$\mathcal{T}_4$	$(0.5, 0.75, 1)$	Good
$\mathcal{T}_5$	$(0.75, 1, 1)$	Excellent

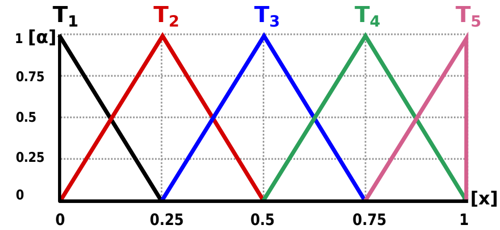


Fig. 2. Meanings of the elementary linguistic terms of the extended linguistic scale used for linguistic approximation.

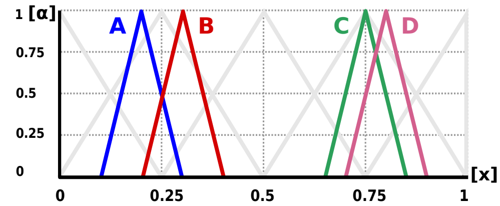


Fig. 3. Outputs of mathematical model used in Example 1 (A and B) and Example 2 (C and D).

A. Example 1

Let fuzzy numbers $A = (0.1, 0.2, 0.3)$ and $B = (0.2, 0.3, 0.4)$ (presented in Fig. 3) be two outputs of a mathematical model – evaluations of two different alternatives. Our goal is to choose a better alternative using the algorithm proposed in this paper.

At first the preselection of linguistic terms from $\mathcal{T}(\mathcal{V})$ is performed. Order of all the linguistic terms of $\mathcal{V}$ with their distances from the fuzzy set A (B) is presented in Table II. Five terms are preselected for the second step – $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_{12}$, $\mathcal{T}_3$ and $\mathcal{T}_{23}$ (these terms are the same for both outputs A and B). In the second step the terms are ordered with respect to Bhattacharyya distance and the results are presented in Table III.

As can be seen from the Table III, the ordering of the first three preselected terms is identical for both A and B. However the fourth suggested term for B ($\mathcal{T}_3$) is preferred to the fourth suggested term for A ($\mathcal{T}_1$). Hence B is better than A.

B. Example 2

Let fuzzy numbers $C = (0.6, 0.7, 0.8)$ and $D = (0.7, 0.8, 0.9)$ (presented in Fig. 3) be two outputs of a mathematical model we need to order.

Again five terms are preselected for the second step – $\mathcal{T}_4$, $\mathcal{T}_{45}$, $\mathcal{T}_{34}$, $\mathcal{T}_5$ and $\mathcal{T}_3$. These terms are the same for both outputs C and D and even their ordering is identical (see Table IV).

TABLE II. OUTPUTS OF THE FIRST STEP OF THE TWO STEP METHOD FOR FUZZY SETS A AND B (d_1 DISTANCES TO THE MEANINGS OF ALL LINGUISTIC TERMS OF $\mathcal{V}$); ORDERED.

	Preselection of linguistic terms for fuzzy number A				
Linguistic term	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_{12}$	$\mathcal{T}_3$	$\mathcal{T}_{23}$
$d_1(A, T_{ij})$	0.0149	0.0475	0.0763	0.1350	0.2131
Linguistic term	$\mathcal{T}_{13}$	$\mathcal{T}_4$	$\mathcal{T}_{34}$	$\mathcal{T}_{45}$	$\mathcal{T}_{35}$
$d_1(A, T_{ij})$	0.2899	0.3475	0.3631	0.4429	0.5099
Linguistic term	$\mathcal{T}_5$	$\mathcal{T}_{24}$	$\mathcal{T}_{14}$	$\mathcal{T}_{25}$	$\mathcal{T}_{15}$
$d_1(A, T_{ij})$	0.5149	0.5350	0.6591	0.7305	0.9100
	Preselection of linguistic terms for fuzzy number B				
Linguistic term	$\mathcal{T}_2$	$\mathcal{T}_1$	$\mathcal{T}_3$	$\mathcal{T}_{12}$	$\mathcal{T}_{23}$
$d_1(B, T_{ij})$	0.0475	0.0482	0.0850	0.0874	0.1881
Linguistic term	$\mathcal{T}_4$	$\mathcal{T}_{13}$	$\mathcal{T}_{34}$	$\mathcal{T}_{45}$	$\mathcal{T}_5$
$d_1(B, T_{ij})$	0.2475	0.2765	0.2881	0.3318	0.3815
Linguistic term	$\mathcal{T}_{35}$	$\mathcal{T}_{24}$	$\mathcal{T}_{14}$	$\mathcal{T}_{25}$	$\mathcal{T}_{15}$
$d_1(B, T_{ij})$	0.4232	0.4850	0.6210	0.6686	0.8600

TABLE III. PRESELECTED TERM SETS FOR EXAMPLE 1 AND THE BHATTACHARYYA DISTANCES OF THEIR MEANINGS TO A AND B RESPECTIVELY.

	Linguistic approximation of fuzzy number A				
Linguistic term	$\mathcal{T}_2$	$\mathcal{T}_{12}$	$\mathcal{T}_{23}$	$\mathcal{T}_1$	$\mathcal{T}_3$
$d_2(A, T_{ij})$	0.5131	0.5626	0.6890	0.7261	0.9802
	Linguistic approximation of fuzzy number B				
Linguistic term	$\mathcal{T}_2$	$\mathcal{T}_{12}$	$\mathcal{T}_{23}$	$\mathcal{T}_3$	$\mathcal{T}_1$
$d_2(B, T_{ij})$	0.5131	0.6273	0.6388	0.8159	0.9718

TABLE IV. PRESELECTED TERM SETS FOR EXAMPLE 2 AND THE BHATTACHARYYA DISTANCES OF THEIR MEANINGS TO C AND D RESPECTIVELY.

	Linguistic approximation of fuzzy number C				
Linguistic term	$\mathcal{T}_4$	$\mathcal{T}_{45}$	$\mathcal{T}_{34}$	$\mathcal{T}_5$	$\mathcal{T}_3$
$d_2(C, T_{ij})$	0.4782	0.5838	0.6551	0.8820	0.9181
	Linguistic approximation of fuzzy number D				
Linguistic term	$\mathcal{T}_4$	$\mathcal{T}_{45}$	$\mathcal{T}_{34}$	$\mathcal{T}_5$	$\mathcal{T}_3$
$d_2(D, T_{ij})$	0.5131	0.5626	0.6890	0.7261	0.9802

As Table IV suggests, Step 6 of the proposed algorithm needs to be applied in this case. That is first we need to find two adjacent linguistic values of the same level (pairs $\mathcal{T}_3, \mathcal{T}_4$; $\mathcal{T}_4, \mathcal{T}_5$ and $\mathcal{T}_{34}, \mathcal{T}_{45}$ can be used). Based on the pair $\mathcal{T}_4, \mathcal{T}_5$ output D is considered better than output C. The same result can be obtained for the pair $\mathcal{T}_{34}, \mathcal{T}_{45}$. Pair $\mathcal{T}_3, \mathcal{T}_4$ does not provide information based on which the ordering can be found.

C. Example 3

Let fuzzy numbers $A = (0.1, 0.4, 0.7)$, $B = (0.3, 0.4, 0.5)$ and $C = (0.3, 0.6, 0.9)$, $D = (0.5, 0.6, 0.7)$ (presented in Fig. 4) be two pair of outputs of a mathematical model we need to order.

Second step results are presented in Table V. The results suggest, that B is better than A while C is better than D. It is worth noting, that both cases are similar ($B \subseteq A$ and $D \subseteq C$; the numerical values of the Bhattacharyya distance to the ordered preselected terms are identical for A and C and for B and D, although different terms were preselected). However, in the first case the output with lower cardinality is considered better whereas in the second case the output with larger cardinality is preferred.

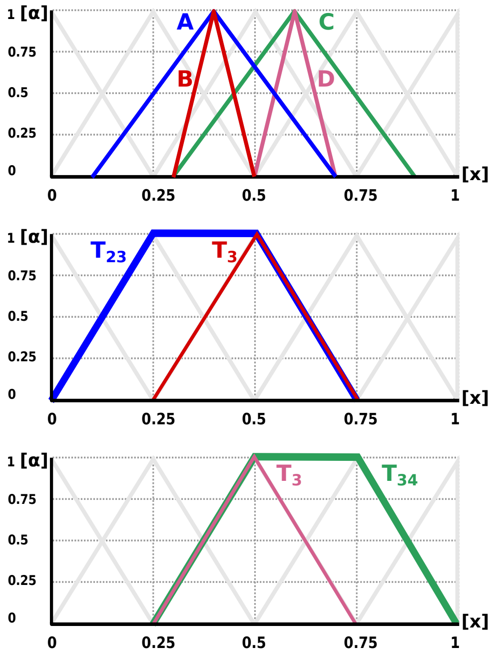


Fig. 4. Outputs of mathematical model used in Example 3.

TABLE V. PRESELECTED TERM SETS FOR EXAMPLE 3 AND THE BHATTACHARYYA DISTANCES OF THEIR MEANINGS TO A, B, C AND D RESPECTIVELY.

	Linguistic approximation of fuzzy number A				
Linguistic term	$\mathcal{T}_{23}$	$\mathcal{T}_3$	$\mathcal{T}_2$	$\mathcal{T}_{45}$	$\mathcal{T}_{12}$
$d_2(A, T_{ij})$	0.2096	0.3591	0.4960	0.5759	0.5969
	Linguistic approximation of fuzzy number B				
Linguistic term	$\mathcal{T}_3$	$\mathcal{T}_2$	$\mathcal{T}_{12}$	$\mathcal{T}_1$	$\mathcal{T}_4$
$d_2(B, T_{ij})$	0.5965	0.6970	0.7617	1.0000	1.0000
	Linguistic approximation of fuzzy number C				
Linguistic term	$\mathcal{T}_{34}$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_{23}$	$\mathcal{T}_{45}$
$d_2(C, T_{ij})$	0.2096	0.3591	0.4960	0.5759	0.5969
	Linguistic approximation of fuzzy number D				
Linguistic term	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_{45}$	$\mathcal{T}_5$	$\mathcal{T}_2$
$d_2(D, T_{ij})$	0.5965	0.6970	0.7617	1.0000	1.0000

VI. CONCLUSION

In the paper we have proposed a utilization of the information obtained by the two step method for linguistic approximation by Bonissone for the ordering of fuzzy numbers. This partial ordering is derived from the best candidates on appropriate linguistic approximation of the fuzzy numbers to be approximated. The extended linguistic scale has been proposed as a suitable linguistic variable for the approximation. The performance of the proposed algorithm is showcased on three numerical examples.

ACKNOWLEDGMENT

The research presented in this paper was partially supported by the grant IGA_FF_2015_014 of the internal grant agency of the Palacký University, Olomouc.

References

[1] Bonissone, P. P.: A pattern recognition approach to the problem of linguistic approximation in system analysis. *Proceedings of the IEEE International Conference on Cybernetics and Society*, (1979), 793–798.

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

[3] Leekwijck, W. V. and Kerre, E. E.: Defuzzification: criteria and classification. *Fuzzy sets and systems*, **108**(2) (1999), 159–178.

[4] Ruspini, E.: A New Approach to Clustering. *Inform. Control*, **15**(1), (1969), 22–32.

[5] Stoklasa, J.: *Linguistic models for decision support.* Lappeenranta: Lappeenranta University of Technology, 2014.

[6] Talašová, J. and Holeček, P.: *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, 681–686, 2009. Lappeenranta: Lappeenranta University of Technology, 2014.

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

[8] Wang, X. and Kerre, E. E.: Reasonable properties for the ordering of fuzzy quantities (I). *Fuzzy Sets and Systems*, **118**(3), (2001), 375–385.

[9] Wang, X. and Kerre, E. E.: Reasonable properties for the ordering of fuzzy quantities (II). *Fuzzy Sets and Systems*, **118**(3), (2001), 387–405.

[10] Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning I, II, III. *Inf. Sci.*, **8**, (1975), 199–257, 301–357, **9**, (1975), 43–80.

Publication IV

Talášek, T., Stoklasa, J. and Talašová, J.

Linguistic approximation using fuzzy 2-tuples in investment decision making

Reprinted with the permission from
Proceedings of the 33rd *International Conference on Mathematical Methods in Economics 2015*,
pp. 817–822, 2015,

Linguistic approximation using fuzzy 2-tuples in investment decision making

Tomáš Talášek[1], Jan Stoklasa [2], Jana Talašová[3]

Abstract. The paper focuses on linguistic approximation methods for the outputs of mathematical decision support models in the context of investment decision making. Since the decision makers in this context are frequently laymen in mathematics, the ability of the models to provide understandable and easily interpretable results is of great importance. We explore the possibilities of using the fuzzy 2-tuples concept introduced by Herrera et al. [2] for linguistic approximation and propose a method of linguistic approximation of fuzzy-number-type outputs suitable for the use in investment decision support. The performance of the proposed method is discussed on a practical example of mutual fund selection.

Keywords: Linguistic approximation, linguistic modelling, multiple criteria decision making, decision support, fuzzy 2-tuples, mutual funds.

JEL classification: C44
AMS classification: 90B50, 91B06, 91B74

1 Introduction

In linguistic modelling it is necessary to be able to assign appropriate labels (known values of a linguistic variable) to general fuzzy sets. In practical applications it is often required to obtain fuzzy numbers as outputs from mathematical models, subnormal or multimodal fuzzy sets on $\mathbb{R}$ (normal and unimodal fuzzy sets) are difficult to interpret and use by practitioners. This paper therefore concentrates on assigning linguistic labels to fuzzy numbers. This process is called *linguistic approximation*. From the mathematical point of view linguistic approximation is a mapping from a given class of fuzzy numbers on $\mathbb{R}$ to a set of linguistic values (labels) of a linguistic variable $\mathcal{V}$ whose mathematical meanings are modelled by fuzzy numbers on $\mathbb{R}$. In this paper we propose a new method for the linguistic approximation of fuzzy-number-evaluations that is based on the idea of fuzzy 2-tuples [2]. We showcase the proposed method on the outputs of a multi-stage decision support model for investment decision making proposed in [6] and further elaborated in [7].

2 Preliminaries

Let U be a nonempty set (the universe of discourse). 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 $A(.)$ is called a *membership function* of the fuzzy set A. $\text{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an *α-cut* of A for any $\alpha \in [0, 1]$, $\text{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\text{Ker}(A) \neq \emptyset$ (A is *normal*); (2) A_α are closed intervals for all $\alpha \in (0, 1]$ (this implies A is *unimodal*); (3) $\text{Supp}(A)$ is bounded. A family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on [a,b], if $\text{Supp}(A)$ is a subset of an interval $[a, b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \text{Cl}(\text{Supp}(A))$ and $[a_2, a_3] = \text{Ker}(A)$, where $\text{Cl}(\text{Supp}(A))$ denotes a closure of $\text{Supp}(A)$. Each fuzzy number A is determined by

[1]Palacký University, Olomouc, Faculty of Science, Dept. of Math. Analysis and Applications of Mathematics, 17. listopadu 1192/12, 77146, Olomouc, Czech Republic, tomas.talasek@upol.cz
[2]dtto, jan.stoklasa@upol.cz
[3]dtto, jana.talasova@upol.cz

$A = \{[\underline{a}(\alpha), \overline{a}(\alpha)]\}_{\alpha \in [0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall \alpha \in (0,1]$, and the closure of the support of A $\mathrm{Cl}(\mathrm{Supp}(A)) = [\underline{a}(0), \overline{a}(0)]$. An *intersection* of two fuzzy sets A and B is a fuzzy set $(A \cap B)$ defined as follows: $(A \cap B)(x) = \min\{A(x), B(x)\}$, $\forall x \in U$.

The fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$; for such fuzzy numbers we will use a simplified notation $A = (a_1, a_2, a_3, a_4)$. If $A \in \mathcal{F}_N(U)$ is a linear fuzzy number and c is a real number, then $A + c = (a_1 + c, a_2 + c, a_3 + c, a_4 + c)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by ordered triplet $A = (a_1, a_2, a_4)$. More details on fuzzy numbers and computations with them can be found for example in [1].

A *fuzzy scale* on $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on [a,b], that form a Ruspini fuzzy partition (see [4]) of the interval $[a, b]$, i.e. for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. A *linguistic variable* ([8]) is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values (terms), X is an universe on which the meanings of the linguistic values are defined, G is an syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X.

We are frequently required to be able to represent fuzzy sets by real numbers, this procedure is called *defuzzification.* In applications an approximation of a fuzzy number A by its center of gravity (COG) t_A is frequently used. The *center of gravity* of a fuzzy number A defined on $[a, b]$, except for fuzzy numbers where $a_1 = a_2 = a_3 = a_4$ (so called *fuzzy singletons*), is defined by the formula $t_A = \int_a^b A(x)x\,dx / \int_a^b A(x)dx$. The center of gravity of a fuzzy singleton is defined as $COG_A = a_1$. Other possible defuzzification methods are discussed in [3]. A *distance* of fuzzy numbers $A = \{[\underline{a}(\alpha), \overline{a}(\alpha)]\}_{\alpha \in [0,1]}$, $B = \{[\underline{b}(\alpha), \overline{b}(\alpha)]\}_{\alpha \in [0,1]}$ can be defined by the formula $d(A, B) = \int_0^1 |\underline{a}(\alpha) - \underline{b}(\alpha)| + |\overline{a}(\alpha) - \overline{b}(\alpha)|\, d\alpha$ (see e.g. [5] for other possible approaches).

3 Proposed method for linguistic approximation

Let us consider an evaluation scale $[a, b]$ and a mathematical model, the outputs of which are triangular fuzzy numbers $E_j \in \mathcal{F}_N([a, b])$, $j = 1, \ldots, m$ (more general types of fuzzy numbers can also be considered, but within this paper we restrict our investigation to triangular ones). These fuzzy numbers represent evaluations of some objects (alternatives). Such outputs can be obtained e.g. in situations where each expert provides a fuzzy evaluation in the form a triangular fuzzy number (all the experts are using the same evaluation scale) and an overall evaluation is computed by a fuzzy weighted average of these expert evaluations. Our aim is to propose a linguistic approximation method for these evaluations that would be understandable to the users of the outputs (decision makers/evaluators) and at the same time would provide more specific information than commonly used linguistic approximation methods (see e.g. [5]).

First we define a linguistic variable $(Eval, \mathcal{T}(Eval), [a, b], G, M)$. The set $\mathcal{T}_B(Eval) = \{\mathcal{T}_1, \ldots, \mathcal{T}_n\} \subset \mathcal{T}(Eval)$ constitutes a *basic term set* of $Eval$, consisting of all the terms the decision maker wants to use for evaluation purposes. $M(\mathcal{T}_i)$ is a triangular fuzzy number on $[a, b]$, for all $i = 1, \ldots, n$. The elements of the basic term set must be well understood by the decision maker and their fuzzy number meanings $M(\mathcal{T}_i) = T_i \in \mathcal{F}_N([a, b])$ for all $i = 1, \ldots, n$ specified in cooperation with him/her. For the purposes of this paper we will suppose that the fuzzy evaluations we need to linguistically approximate are not significantly more uncertain then the meanings of the elements of the basic terms set. Let us also consider that $T_1, \ldots, T_n$ form a uniform Ruspini fuzzy partition of $[a, b]$:

$$\begin{aligned} M(\mathcal{T}_1) &= (a, a, a + \Delta), \\ M(\mathcal{T}_i) &= (a + (i-2) \cdot \Delta, a + (i-1) \cdot \Delta, a + i \cdot \Delta),\ i = 2, \ldots, n-1, \\ M(\mathcal{T}_n) &= (b - \Delta, b, b), \end{aligned} \tag{1}$$

where $\Delta = (b - a)/(n - 1)$.

A fuzzy set representing feasible evaluations on $\mathbb{R}$ is defined as

$$FE(x) = \begin{cases} 1 & \text{if } x \in [a, b], \\ 0 & \text{if } x \notin [a, b]. \end{cases} \tag{2}$$

Non-uniform partitions and partitions by non-triangular fuzzy numbers will be the subject of further study and hence are left out of the scope of this paper.

Since the basic term set of $Eval$ contains only n elements, we might not be able to find a linguistic approximation of some of the possible fuzzy evaluations that would fit well enough (using just these n linguistic terms). We are however restricted to the use of the elements of the basic term set, as the linguistic terms it contains are the only ones that the decision maker clearly understands. To resolve this issue, we suggest to use the concept of fuzzy 2-tuples (which in this context translates into shifting the meanings of the basic terms to either side within the specified universe) and thus introduce the following syntactic rule to describe the results of such shifts and thus add *derived terms* to $\mathcal{T}_B(Eval)$:

$$\mathcal{T}(Eval) = \mathcal{T}_B(Eval) \cup \{\mathcal{T}_i^\beta;\ \beta \in [-0.5 \cdot \Delta, 0.5 \cdot \Delta),\ i = 1, \ldots, n\},$$

where $\mathcal{T}_i^\beta$ is "$\mathcal{T}_i$ *shifted by* β" (see Figure 1; the size of the shift can be described linguistically as well; it will be shown in the numerical example). $\mathcal{T}_i^\beta$ can be expressed as a 2-tuple $(\mathcal{T}_i, \beta)$, where $\mathcal{T}_i$, $i = 1, \ldots, n$ is an elementary linguistic term and $\beta \in [-0.5 \cdot \Delta, 0.5 \cdot \Delta)$ expresses the shift of $\mathcal{T}_i$, the sign of β indicates the direction of the shift. For $\beta = 0$, $\mathcal{T}_i^\beta$ coincides with $\mathcal{T}_i$, $i = 1, \ldots n$.

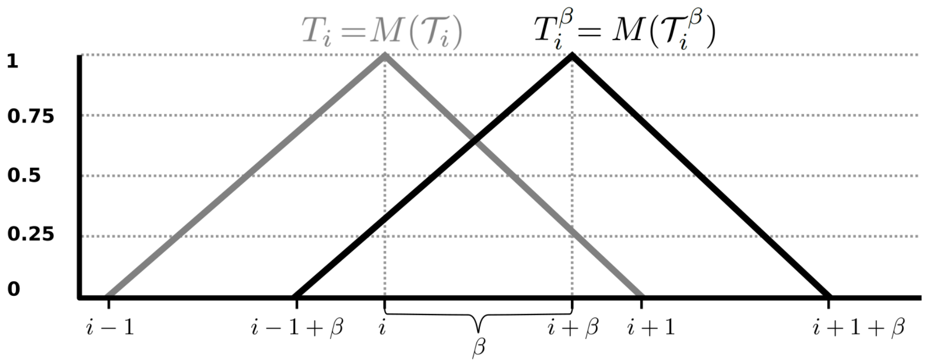


Figure 1 Shifting the meaning of a basic term set element $\mathcal{T}_i$ by β

To define a syntactic rule, we extend the universe $[a, b]$ to $[a - \Delta, b + \Delta]$ (this is a technicality to simplify the notation) and extend the meaning of $\mathcal{T}_1$ and $\mathcal{T}_n$ from T_1 and T_n to T_1' and T_n' respectively in the following way (note that the meanings of $\mathcal{T}_1$ and $\mathcal{T}_n$ remain unchanged on $[a, b]$).

$$\begin{aligned} T_1' &= (a - 0.5 \cdot \Delta, a - 0.5 \cdot \Delta, a, a + \Delta), \\ T_n' &= (b - \Delta, b, b + 0.5 \cdot \Delta, b + 0.5 \cdot \Delta), \end{aligned} \tag{3}$$

and for $\mathcal{T}_2, \ldots, \mathcal{T}_{n-1}$ we leave the meaning unchanged, that is

$$T_i' = T_i,\ i = 2, \ldots, n-1. \tag{4}$$

The meanings of the elements of the derived term set are now computed using the formula

$$M(\mathcal{T}_i^\beta) = (T_i' + \beta) \cap FE,\ i = 1, \ldots, n,\ \beta \in [-0.5 \cdot \Delta, 0.5 \cdot \Delta). \tag{5}$$

It is however not reasonable to move T_1 to the left and T_n to the right, hence for $\mathcal{T}_1$ we will consider $\beta \in [0, 0.5 \cdot \Delta)$ and for $\mathcal{T}_n$ we will consider $\beta \in [-0.5 \cdot \Delta, 0]$.

The linguistic approximation $\mathcal{T}^* \in \mathcal{T}(Eval)$ of a fuzzy evaluation $E_j \in \mathcal{F}_N([a, b])$ is computed by

$$M(\mathcal{T}_{i^*}^{\beta^*}) = \underset{\substack{i=1,\ldots,n \\ \beta \in [-0.5 \cdot \Delta, 0.5 \cdot \Delta)}}{\arg\min}\ d(E_j, T_i^\beta). \tag{6}$$

The result of this novel linguistic approximation method is a fuzzy 2-tuple $(\mathcal{T}_{i^*}, \beta^*)$. In case there are more solutions $(\mathcal{T}_{i^*}, \beta^*)$ to (6), all such fuzzy 2-tuples are presented to the decision maker. To obtain a fully linguistic description of the evaluation, β can be interpreted (described linguistically) using e.g. Table 1. For example $(Good, +0.1)$ translates into *slightly better than good.*

Negative β value	Linguistic description	Positive β value	Linguistic description
$[-0.05 \cdot \Delta, 0 \cdot \Delta)$	About	$(0 \cdot \Delta, 0.05 \cdot \Delta]$	About
$[-0.2 \cdot \Delta, -0.05 \cdot \Delta)$	Slightly worse than	$(0.05 \cdot \Delta, 0.2 \cdot \Delta]$	Slightly better than
$[-0.35 \cdot \Delta, -0.2 \cdot \Delta)$	Worse than	$(0.2 \cdot \Delta, 0.35 \cdot \Delta]$	Better than
$[-0.5 \cdot \Delta, -0.35 \cdot \Delta)$	Noticeably worse than	$(0.35 \cdot \Delta, 0.5 \cdot \Delta)$	Noticeably better than

Table 1 Linguistic labels for the interpretation of values of β.

4 Short example of the proposed linguistic approximation method

Let us consider a multiple-criteria decision support system proposed in [6]. The multi-stage model for mutual fund selection first assesses investor's investment aim and his/her investment horizon. In the next stage the risk profile of the investor is taken into account. Based on the information obtained in the first two stages an evaluation of each mutual fund under consideration is computed. This evaluation is obtained in form of a fuzzy number (triangular fuzzy numbers are used to approximate more complex outputs if necessary; see Figure 2). The best fit linguistic approximation of the fuzzy evaluations was performed. As a result the decision maker (investor) was provided with a linguistic description of the overall evaluation of each mutual fund and with a numerical value representing the center of gravity of the respective fuzzy evaluation. These results are summarized in Table 2. The linguistic approximation used in [6] assigns one of the five elements of the linguistic scale presented in Figure 3 (or a combination of these terms) to each overall fuzzy evaluation of a mutual fund. Much information is therefore lost in the process and the linguistic label may not fit well.

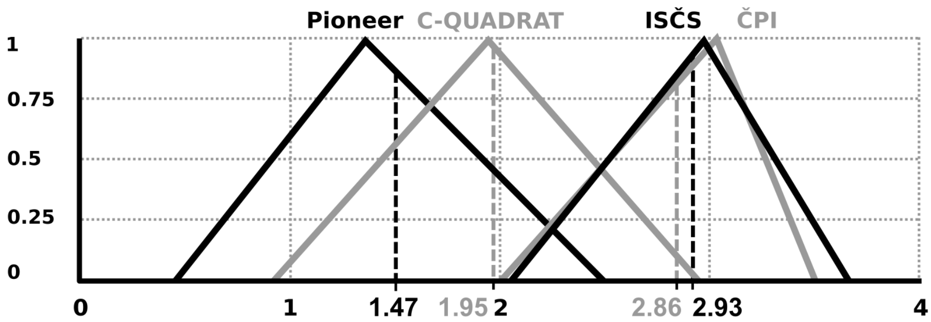


Figure 2 Overall evaluations of four mutual funds according to [6]

Alternatives	Overall evaluation	Output of mathematical model [6]		Proposed linguistic approximation	
		Linguistic approximation	COG	2-tuple	Linguistic approximation
Pioneer	(0.504, 1.404, 2.516)	Between Bad and Average	1.468	(Bad, +0.457)	Noticeably better than bad
ISČS	(2.116, 2.992, 3.656)	Good	2.932	(Good, -0.008)	About good
ČPI	(2.012, 3.036, 3.524)	Good	2.864	(Good, +0.014)	About good
C-QUADRAT	(0.928, 1.920, 2.960)	Average	1.948	(Average, -0.073)	Slightly worse than average

Table 2 Evaluations of four mutual funds. Results provided by [6] compared with the results obtained by the linguistic approximation method proposed in this paper.

Let us now approach the fuzzy evaluations obtained in [6] presented in Table 2 with the linguistic approximation method proposed in this paper. We can depart from the original evaluation scale used in [6], that is $(Eval_{ex}, \{Very\, bad,\ Bad,\ Average,\ Good,\ Excellent\}, [0, 4], G, M)$, where the meanings of the elements of the basic term set are summarized in Figure 3. Using the distance defined in preliminaries we obtain the linguistic approximations summarized in Table 2.

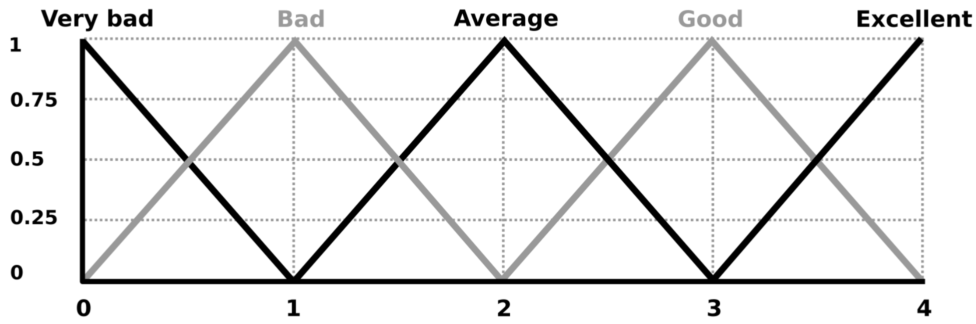


Figure 3 Meanings of the basic linguistic terms of $Eval_{ex}$ for $n = 5$ on the evaluation universe $[0, 4]$. These meanings correspond with the meanings of the values of the linguistic scale used in [6].

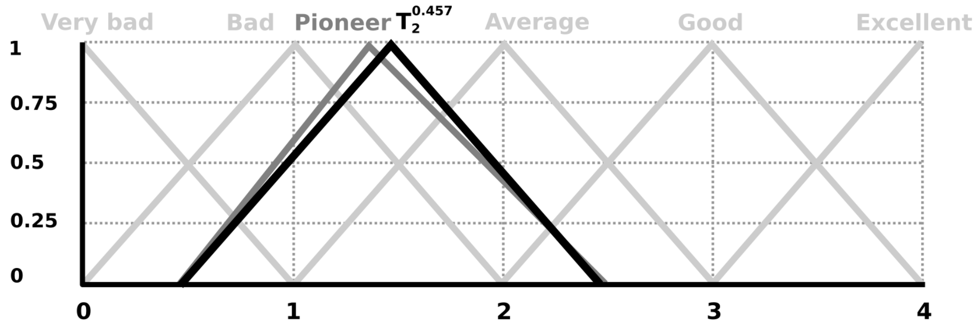


Figure 4 Illustration of the Pioneer mutual fund evaluation – its linguistic approximation by the fuzzy 2-tuple based method.

5 Discussion and conclusion

A novel linguistic approximation method for fuzzy numbers is presented in this paper. The method draws from the fuzzy 2-tuple concept, which allows it to provide a combination of linguistic term and its necessary shift (expressed numerically or linguistically). Using an initial linguistic terms set specified by the decision maker (the fuzzy number meanings of the linguistic terms are defined in cooperation with decision maker) it constructs a linguistic variable with substantially extended term set. The grammar and syntactic rule necessary to construct this linguistic variable are introduced in the paper. The linguistic approximation itself then uses a distance of fuzzy numbers to find the linguistic term that is closest to the approximated fuzzy number. The method performs particularly well when the approximated fuzzy numbers are of a similar shape and uncertainty as the meanings of the basic terms set. When the basic linguistic terms set is linearly ordered, the fuzzy 2-tuple linguistic approximation also provides means for ordering of the approximated fuzzy evaluations. This feature can prove useful both in decision making and evaluation (rankings of alternatives etc.). A short numerical example from the area of financial decision making showcasing the proposed method is also presented.

The example compares a classic "best-fit" approach to the linguistic approximation fuzzy outputs of mathematical models with the proposed fuzzy 2-tuple based method. We can clearly see, that although the same basic term set (with the same meanings) and the same distance was used for linguistic approximation in [6] and here, the linguistic approximations obtained by the methodology proposed in this paper provide more insights in the evaluations. The fuzzy 2-tuple based approximation operates with the basic terms well understood by the decision maker and uses small shifts of their meanings (numerically quantified; linguistic labelling of these shifts is also possible) to reflect differences between the fuzzy evaluations and the predefined meanings of the basic terms.

If we focus on the *Pioneer* fund from the example, we can see, that Talášek et al. [6] suggest a linguistic approximation involving two neighbouring basic linguistic terms – in fact it is suggested that the appropriate linguistic approximation lies somewhere in between these two basic terms. No indication is suggested as to which of the original basic terms is closer (consider that the basic terms are those that are well understood by the decision maker). The approach proposed in this paper provides a similar result. It however uses a single basic linguistic term and describes its necessary modification (shift of meaning). The decision maker is thus provided with a single linguistic label (which is similar to classic approaches to linguistic approximation) and provides an additional piece of information regarding the difference of the approximated fuzzy set and the meaning of the basic term which can be used if needed. The output of the proposed fuzzy 2-tuple based approximation is presented in Figure 4. The overall evaluation of the mutual fund Pioneer is depicted together with its linguistic approximation expressed by the fuzzy number $T_2^{0.457}$. The most promising feature of the proposed method is the use of a single well understood linguistic label along with a simple linguistic or numerical modifier of its meaning.

The effect of different distances of fuzzy numbers, different shapes of membership functions and different types of scales on the usefulness of the proposed method in practical applications, as well as the use of the proposed methof for the ranking of alternatives will be the subject of further study. Also empirical research concerning the understandability and intuitiveness of fuzzy 2-tuple based outputs for practitioners is planned.

Acknowledgements

Supported by the grant IGA PrF 2015 013 Mathematical models of the Internal Grant Agency of Palacký University, Olomouc and by the grant GA14-02424S Methods of operations research for decision support under uncertainty of the Grant Agency of the Czech Republic.

References

[1] Dubois, D., Prade, H.: *Fuzzy sets and systems: theory and applications*, Academic Press, 1980.

[2] Herrera, F., Martinez, L.: An approach for combining linguistic and numerical information based on the 2-tuple fuzzy linguistic representation model in decision-making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **8**(5), (2000), 539–562.

[3] Leekwijck, W. V., Kerre, E. E.: Defuzzification: criteria and classification. *Fuzzy sets and systems*, **108**(2) (1999), 159–178.

[4] Ruspini, E.: A New Approach to Clustering. *Inform. Control*, **15**, (1969), 22–32.

[5] Stoklasa, J.: *Linguistic models for decision support.* Lappeenranta: Lappeenranta University of Technology, 2014.

[6] Talášek, T., Bohanesová, E., Stoklasa, J., Talašová, J. Investment decision making using fuzzy scorecards–mutual funds selection, *Proceedings of the 32nd International Conference on Mathematical Methods in Economics 2014, September 2014*, Palacký University, Olomouc, 1021–1026.

[7] Talášek, T., Bohanesová, E., Talašová, J.: Two–step method for multiple criteria investment decision making: mutual funds selection using FuzzME software. *International Journal of Mathematics in Operational Research*, (accepted, to appear).

[8] Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning I, II, III. *Inf. Sci.*, **8**, (1975), 199–257, 301–357, **9**, (1975), 43–80.

Publication V

Talášek, T. and Stoklasa, J.
The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers

Reprinted with the permission from
Proceedings of the international scientific conference Knowledge for Market Use 2016,
pp. 539–546, 2016,

THE ROLE OF DISTANCE/SIMILARITY MEASURES IN THE LINGUISTIC APPROXIMATION OF TRIANGULAR FUZZY NUMBERS

TOMÁŠ TALÁŠEK, JAN STOKLASA

Palacký University Olomouc – Faculty of Arts – Department of Applied Economics, and Lappeenranta University of Technology – Graduate School, Lappeenranta, Finland; Palacký University Olomouc – Faculty of Arts – Department of Applied Economics

***Abstract**: Linguistic approximation is a common way of translating the outputs of mathematical models in the expressions in common language. These can then provide an easy-to-understand alternative to the numerical outputs of formal models. As such linguistically approximated outputs can facilitate the interpretability of the outputs of the models and reduce their possibility misuse. However, linguistic approximation remains an under-researched area and best practices in management, economics, decision support, social science and behavioural research are missing. The paper explores the performance of two selected distance measures of fuzzy numbers and two different fuzzy similarity measures in the context of linguistic approximation. Triangular fuzzy numbers are considered as the approximated entities, their symmetry is not required. We present the results of a numerical experiment performed to map the behaviour of the four distance/similarity measures under uniform output linguistic scales.*

***Keywords:** Linguistic approximation, fuzzy number, distance, similarity.*

***JEL classification:** C44*

***Grant affiliation:** This research was supported by the grant IGA FF 2015 014 Continuities and Discontinuities of Economy and Management in the Past and Present of the internal grant agency of the Palacký University Olomouc.*

1 Introduction

In practical applications, it is often suitable to represent outputs of mathematical models not only in mathematical form (as numerical values), but also to provide the users of these models with an easy-to-understand representation of these results or their summary (Yager, 2004). This can be done using linguistic labels that describe the outputs of the models in natural language. This way the results become more understandable to the people who are not sufficiently familiar with the mathematical background of the models (Stoklasa, 2014). The process that assigns linguistic labels to mathematical objects is called linguistic approximation and is discussed particularly in connection with fuzzy models and mathematical outputs that represent not only the values of the variables, but also their uncertainty. Usually a relatively small set of linguistic labels that are well understood by the user of the model is considered in linguistic approximation. Due to the fact that the set of possible linguistic labels contains only a small number of labels, the linguistic approximation can assign the same linguistic label to two different outputs of mathematical model (e.g. numbers 1.0 and 1.1 can be linguistically approximated as small). This is a natural consequence of the chosen granularity of the approximating term set. It is, however, clear, that the process of linguistic approximation has to be set up in accordance with the given application and its context. The scientific literature does not provide many insights into this topic. The goal of this paper is therefore to investigate, how different methods for linguistic approximation behave in cases, where the meaning of output of the mathematical model is close to two linguistic terms or lies directly "between" them.

In the following text we investigate how the choice of different distance/similarity measure affects the linguistic approximation of triangular fuzzy numbers using uniform linguistic scale. We recall the principles of linguistic approximation in Section 3 and describe four distance/similarity measures the performance of which is further analysed in Section 4 using a numerical experiment. We discuss the results of the experiment in Section 5 and draw conclusions in the last section.

2 Preliminaries

Let U be a nonempty set (the universe of discourse). 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 $A(.)$ is called a membership function of the fuzzy set A. $\mathrm{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a kernel of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an α-cut of A for any $\alpha \in [0,1]$, $\mathrm{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a support of A. Let A and B be fuzzy sets on the same universe U. We say that A is a fuzzy subset of $(A \subseteq B)$, if $A(x) \leq B(x)$ for all $x \in U$.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\mathrm{Ker}(A) \neq \emptyset$ (A is normal); (2) A_α are closed intervals for all $\alpha \in (0,1]$ (this implies A is unimodal); (3) $\mathrm{Supp}(A)$ is bounded. A family of all fuzzy numbers on U is denoted by $\mathrm{F}_N(U)$. A fuzzy number A is said to be defined on $[a,b]$, if $\mathrm{Supp}(A)$ is a subset of an interval $[a,b]$. The real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called significant values of the fuzzy number A if $[a_1, a_4] = \mathrm{Cl}(\mathrm{Supp}(A))$ and $[a_2, a_3] = \mathrm{Ker}(A)$, where $\mathrm{Cl}(\mathrm{Supp}(A))$ denotes a closure of $\mathrm{Supp}(A)$. Each fuzzy number A can be represented as $A = \{\underline{a}(\alpha), \overline{a}(\alpha)\}_{\alpha \in [0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall \alpha \in (0,1]$, and $[\underline{a}(0), \overline{a}(0)] = \mathrm{Cl}(\mathrm{Supp}(A))$. The cardinality of a fuzzy number A on $[a,b]$ is a real number $\mathrm{Card}(A)$ defined as $\mathrm{Card}(A) = \int_a^b A(x)\, dx$ and can be considered as a measure of uncertainty of the fuzzy number A.

The fuzzy number A is called linear if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$; for such fuzzy numbers we will use a simplified notation $A = (a_1, a_2, a_3, a_4)$. A linear fuzzy number A is said to be triangular if $a_2 = a_3$. We will denote triangular fuzzy numbers by an ordered triplet $A = (a_1, a_2, a_4)$. A triangular fuzzy number $A = (a_1, a_2, a_4)$ is called symmetric if $a_2 - a_1 = a_4 - a_2$. Otherwise it is called assymetric. More details on fuzzy numbers and computations with them can be found for example in Dubois & Prade (1980).

In real-life applications we often need to represent fuzzy numbers by real numbers. This process is called defuzzification. The most common method is to substitute fuzzy number by its centre of gravity (COG). Let A be a fuzzy number on $[a, b]$ for which $a_1 \neq a_4$. The centre of gravity of A is defined by the formula $\mathrm{COG}(A) = \int_a^b x\, A(x)\ dx/\mathrm{Card}(A)$.

A fuzzy scale on $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, \dots, T_s$ on $[a, b]$, that form a Ruspini fuzzy partition (Ruspini, 1969) of the interval $[a, b]$, i.e. for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$ and the T's are indexed according to their ordering. A linguistic variable (Zadeh, 1975) is defined as a quintuple (V, T(V), X, G, M), where V is the name of the variable, T(V) is the set of its linguistic values (terms), X is the universe on which the meanings of the linguistic values are defined, G is a syntactic rule for generating the values of V and M is a semantic rule which to every linguistic value A ∈ T(V) assigns its meaning $A = M(\mathrm{A})$ which is usually a fuzzy number on X. A linguistic variable is called a linguistic scale, if the meanings of its linguistic values form a fuzzy scale.

3 Linguistic approximation of fuzzy numbers

Let us now consider the task of finding an appropriate linguistic term from the set $\{\mathrm{T}_1, \dots, \mathrm{T}_s\}$ to represent the fuzzy set O on $[a, b]$, which is an output of a mathematical. Let us consider the linguistic terms are values of a linguistic scale V, i.e. $\mathrm{T(V)} = \{\mathrm{T}_1, \dots, \mathrm{T}_s\}$, and $T_i = M(\mathrm{T}_i),\ i = 1, \dots, s$ are fuzzy numbers on $[a, b]$. The linguistic approximation $\mathrm{T}_O \in \mathrm{T(V)}$ of the fuzzy set O is computed by

$$T_O = \arg \min_{i \in \{1,\dots,s\}} d(T_i, O) \qquad (1)$$

where $d(A, B)$ is a distance or similarity measure (in the case of similarity measure the $\arg\min$ function in formula (1) must be replaced by $\arg\max$ function) of two fuzzy numbers A, B. During the past forty years a large number of approaches were proposed for the computation of distance and similarity of fuzzy numbers (see e.g. Zwick et al. (1987)). It is necessary to keep in mind that the choice of distance/similarity measure will modify the behaviour of the linguistic approximation method. In the next chapter the following distances and similarity measures of fuzzy numbers A and B will be considered:

modified Bhattacharyya distance (Aherne et al., 1998):

$$d_1(A, B) = \sqrt{1 - \int_U \sqrt{\frac{A(x)}{\mathrm{Card}(A)} \cdot \frac{B(x)}{\mathrm{Card}(B)}}\, dx}$$

dissemblance index (Kaufman & Gupta, 1985)

$$d_2(A, B) = \int_0^1 \left|\underline{a}(\alpha) - \underline{b}(\alpha)\right| + \left|\overline{a}(\alpha) - \overline{b}(\alpha)\right| d\alpha$$

similarity measure (introduced by Wei & Chen, 2009)

$$s_{1(A,B)} = \left(1 - \frac{\sum_{j=1}^{4} \left|a_j - b_j\right|}{4}\right) \cdot \frac{\min\{\mathrm{Pe}(A), \mathrm{Pe}(B)\} + 1}{\max\{\mathrm{Pe}(A), \mathrm{Pe}(B)\} + 1}$$

Where $\mathrm{Pe}(A) = \sqrt{(a_1 - a_2)^2 + 1} + \sqrt{(a_3 - a_4)^2 + 1} + (a_3 - a_2) + (a_4 - a_1)$, $\mathrm{Pe}(B)$ is defined analogically

similarity measure (introduced by Hejazi et al., 2011)

$$s_2(A,B) = \left(1 - \frac{\sum_{j=1}^{4} |a_j - b_j|}{4}\right) \cdot \frac{\min\{\text{Pe}(A), \text{Pe}(B)\}}{\max\{\text{Pe}(A), \text{Pe}(B)\}} \cdot \frac{\min\{\text{Ar}(A), \text{Ar}(B)\} + 1}{\max\{\text{Ar}(A), \text{Ar}(B)\} + 1}$$

where $\text{Ar}(A) = 1/2\,(a_3 - a_2 + a_4 - a_1)$, $\text{Ar}(B)$ is defined analogically and $\text{Pe}(A)$ and $\text{Pe}(B)$ are computed identically as in the previous method.

These methods represent a sample of the distance and similarity measures of fuzzy sets used in the best-fit approaches to linguistic approximation (Yager, 2004, Stoklasa, 2014). In the next chapter we propose a numerical experiment to analyse and compare the performance of these measures in linguistic approximation.

4 Numerical experiment

For the purposes of the numerical experiment, 100 000 triangular fuzzy numbers on $[0,1]$ $O_h, h = 1, \dots, 100\,000$, were randomly generate. No restriction was posed on the symmetry of these fuzzy numbers (asymmetrical as well as symmetrical fuzzy numbers were generated). These fuzzy numbers were then linguistically approximated by a linguistic scale containing five linguistic terms $\text{T}_1, \dots, \text{T}_5$ with the respective meanings $T_1 = (0, 0, 0.25), T_2 = (0, 0.25, 0.5), T_3 = (0.25, 0.5, 0.75), T_4 = (0.5, 0.75, 1), T_5 = (0.75, 1, 1)$ using all four distance/similarity measures d_1, d_2, s_1 and s_2 (therefore for each output O_h four linguistic labels were obtained using $\arg\min_{i\in\{1,\dots,5\}} d_1(T_i, O_h)$, $\arg\min_{i\in\{1,\dots,5\}} d_2(T_i, O_h)$, $\arg\max_{i\in\{1,\dots,5\}} s_1(T_i, O_h)$ and $\arg\max_{i\in\{1,\dots,5\}} s_2(T_i, O_h)$ respectively).

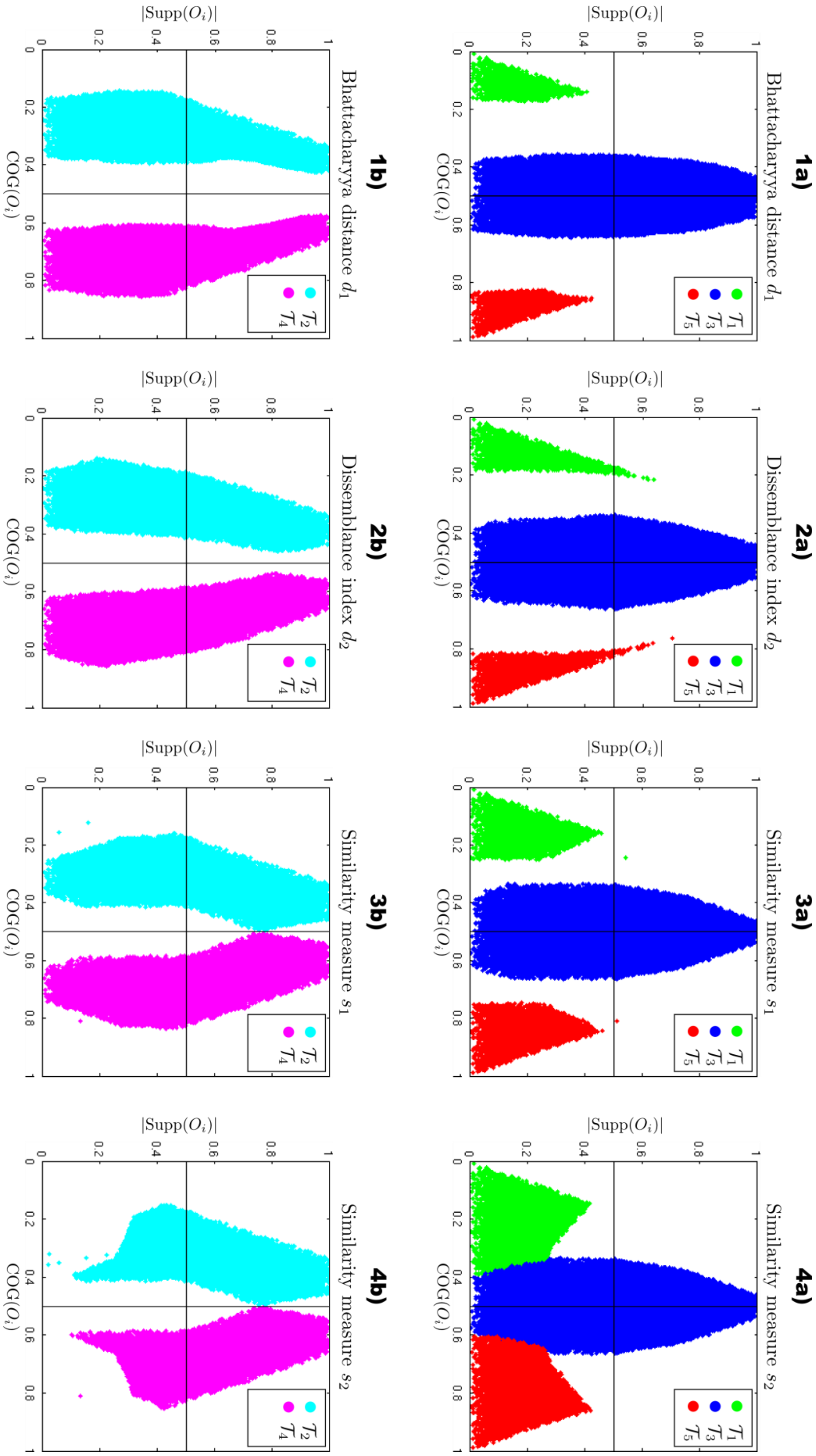
The results are summarized in Figure 1. Results for each distance/similarity measure are represented by two subfigures – the upper one depicts outputs approximated by linguistic terms $\text{T}_1, \text{T}_3, \text{T}_5$ and the second (bottom) one depicts outputs approximated by the remaining terms T_2 and T_4. Each linguistic term is represented by a different colour. The outputs to be approximated O_h are represented in two-dimensional space by the centre of gravity ($\text{COG}(O_h)$ on the x-axis) and length of the support ($|\text{Supp}(O_h)|$ on the y-axis). Results were split into two subfigures, because the coloured areas corresponding with the neighbouring linguistic terms partially overlap.

5 Discussion

Based on the results of the numerical experiment we have identified several important differences in the performance of the four studied distance/similarity measures in linguistic approximation. Note that since the symmetry of the generated triangular fuzzy numbers was not required, the results reported in this paper differ significantly from the results reported in (Talášek & Stoklasa, 2016), where only symmetrical fuzzy numbers were considered. Due to the random generation of triangular fuzzy numbers for the numerical experiment reported in this paper, the set of objects that are linguistically approximated here contains both symmetrical and asymmetrical fuzzy numbers (although asymmetrical fuzzy numbers constitute the majority of the objects). The results can therefore be considered to be more general and more relevant for the purposes of linguistic approximation of the outputs of fuzzy mathematical models where symmetry is not required or cannot be guaranteed.

First, it is interesting to note that under the considered linguistic variable the dissemblance index is the only measure (out of the four considered ones) that assigns the most extreme linguistic labels T_1 and T_5 to fuzzy numbers with the length of support higher than 0.5 (see Figure 1, Subfigures 2a and 2b, green and red points above the 0.5 horizontal line). This way even a highly uncertain fuzzy output O, $O = (o_1, o_2, o_3)$, $\text{Card}(O) > 0.25$ (i.e. $|\text{Supp}(O)| > 0.5$) can be approximated by T_1 or T_5

FIG. 1: Results of the numerical experiment. Linguistic approximations of randomly generated triangular fuzzy numbers $O_h, h = 1, \dots, 100\,000$ represented by points in two-dimensional space (centre of gravity (x-axis) and length of their support (y-axis)) computed using Bhattacharyya d_1 (subfigures 1a and 1b), d_2 (subfigures 2a and 2b), s_1 (subfigures 3a and 3b) and s_2 (subfigures 4a and 4b). Each linguistic term assigned as a linguistic approximation $T_1, \dots, T_5$ is represented by different colour.



even though these are the least uncertain linguistic terms ($\mathrm{Card}(T_1) = \mathrm{Card}(T_5) = 0.125$ and $\mathrm{Card}(T_2) = \mathrm{Card}(T_3) = \mathrm{Card}(T_4) = 0.25$). The resulting linguistic approximation can therefore distort the information concerning the uncertainty of the approximated fuzzy number and suggest it is lower than it really is (even though there are more uncertain linguistic terms available for the linguistic approximation).

Second, the similarity measure s_2 under our experimental setting tends to distort the information concerning the location of the approximated fuzzy number (represented by the centre of gravity). For fuzzy numbers with low uncertainty located close to the centre of the universe (in our case 0.5; see the right part of the green area and left part of the red area in Figure 1, Subfigures 4a and 4b) it is possible to obtain extreme linguistic approximation (T_1 or T_5). This means that e.g. a fuzzy number $P = (0.649, 0.65, 0.651)$, $\mathrm{Card}(P) = 0.001$, will be approximated by T_5. Note that P is very close to the real number 0.65 for which T_5 seems to be a counterintuitive (too extreme) linguistic approximation. This effect is caused by the focus of s_2 not only on the location of the approximated fuzzy number but also on its shape. However in the above mentioned example a linguistic approximation best fitting in terms of uncertainty is suggested.

In the case of the Bhattacharyya distance the centre of gravity of the approximated fuzzy number seems to be the most important piece of information in the determination of the linguistic approximation (using a uniform linguistic scale). The vertical boarders of the areas are almost perpendicular to the x-axis. Note that this feature of Bhattacharyya distance is even more apparent when symmetrical triangular fuzzy numbers are approximated. In this case the sole centre of gravity is a very good predictor of the result of linguistic approximation (Talášek & Stoklasa, 2016). The similarity measure s_2 also takes into account the uncertainty of the approximated fuzzy numbers – see the close-to-horizontal upper border of the green and red areas respectively in Figure 1, Subfigure 4a. Using s_2 fuzzy numbers with low uncertainty will almost never be approximated by T_2 or T_4. The similarity measure s_1 behaves in similar manner as s_2, the patterns in the data generated by the numerical experiment are just less distinct.

6 Conclusions

The analysis presented in this paper focuses on the performance of Bhattacharyya distance, dissemblance index and two selected similarity measures in linguistic approximation of triangular fuzzy numbers. More specifically a five-term uniform linguistic scale is considered to provide output values for the linguistic approximation. The results of a numerical experiment involving the random generation of 100 000 triangular fuzzy numbers and their subsequent linguistic approximation using the selected four distance/similarity measures are presented in Figure 1. The results confirmed that the measures perform differently depending on the requirements for linguistic approximation. The Bhattacharyya distance seems to be the method of choice when the most important piece of information is carried by centre of gravity (location of the output). On the other hand both similarity measures reflect also the uncertainty and shape of the approximated fuzzy numbers and hence tend to overemphasize the shape over location for low-uncertain fuzzy numbers. The dissemblance index is the only one of the analysed measures that assigns extreme linguistic approximations also for high uncertain fuzzy numbers. These findings are relevant for the design of model-user interfaces and appropriate presentation of data in application areas of fuzzy mathematical models, such as economics, finance, management, social sciences etc.

Literature:

Aherne, F., Thacker, N., & Rockett P. (1998). The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data. Kybernetika, 32(4), 363–368.

Dubois, D., & Prade, H. (1980). Fuzzy sets and systems: theory and applications. Orlando: Academic Press.

Hejazi, S. R., Doostparast, A., & Hosseini, S. M. (2011). An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers. Expert Systems with Applications, 38(8), 9179–9185.

Kaufman, A., & Gupta M. M. (1985). Introduction to Fuzzy Arithmetic. New York: Van Nostrand Reinhold.

Ruspini, E. (1969). A New Approach to Clustering. Information Control, 15, 22–32.

Stoklasa, J. (2014). Linguistic models for decision support. Lappeenranta: Lappeenranta University of Technology.

Talášek, T., & Stoklasa, J. (2016). Linguistic approximation under different distances/similarity measures for fuzzy numbers. In NSAIS´16 Workshop on Adaptive and Intelligent Systems, in press.

Wei, S. H., & Chen, S. M. (2009). A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. Expert Systems with Applications, 36(1), 589–598.

Yager, R. R. (2004). 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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15376863.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-I. Information Sciences, 8(3), 199–249. http://doi.org/10.1016/0020-0255(75)90036-5.

Zwick, R., Carlstein, E., & Budescu, D. V. (1987). Measures of similarity among fuzzy concepts: A comparative analysis. International Journal of Approximate Reasoning, 1(2), 221–242.

Contact:

Mgr. et Mgr. Jan Stoklasa, D.Sc. (Tech.)
Palacký University, Olomouc, Faculty of Arts, Department of Applied Economics
Křížkovského 8,
771 47 Olomouc
Czech Republic
jan.stoklasa@upol.cz

Mgr. Tomáš Talášek
Palacký University, Olomouc, Faculty of Arts, Department of Applied Economics
Křížkovského 8,
771 47 Olomouc
Czech Republic
tomas.talasek@upol.cz

Brief information about the author:

Tomáš Talášek received the MS degree in applications of mathematics in economy from the Palacký University, Olomouc, Czech Republic, in 2012. He has been working toward the PhD degree in applied mathematics at Palacký University Olomouc, Czech Republic and Lappeenranta University of Technology, Finland. His research interests include fuzzy sets, multiple criteria decision making, linguistic approximation and pattern recognition and their practical applications.

Jan Stoklasa received the MS degree in applied mathematics and MS degree in psychology from Palacký University, Olomouc, Czech Republic in 2009 and 2012 respectively. He received the Ph.D. and D.Sc. degree in applied mathematics from Palacký University, Olomouc, Czech Republic and Lappeenranta University of Technology, Finland respectively in 2014. He is currently an assistent professor at Palacký University, Olomouc and a research fellow at Lappeenranta University of Technology, School of Business and Management, Lappeenranta, Finland. His research interests include decision support models, multiple criteria decision making and evaluation and linguistic fuzzy models and their practical applications.

Publication VI

Talášek, T., Stoklasa, J. and Talašová, J.

The role of distance and similarity in Bonissone's linguistic approximation method – a numerical study

Reprinted with the permission from
Proceedings of the 34th *International Conference on Mathematical Methods in Economics 2016*,
pp. 845–850, 2016,

The role of distance and similarity in Bonissone's linguistic approximation method – a numerical study

Tomáš Talášek[1], Jan Stoklasa[2], Jana Talašová[3]

Abstract. Linguistic approximation is a common way of translating the outputs of mathematical models in the expressions in common language. These can then be presented to decision makers who have difficulties with interpretations of numerical outputs of formal models as an easy-to-understand alternative. Linguistic approximation is a tool to stress, modify or effectively convey meaning. As such it is an important yet neglected area of research in management science and decision support.

During the last forty years a large number of different methods for linguistic approximation were proposed. In this paper we investigate in detail the linguistic approximation method proposed by Bonissone (1979). We focus on its performance under different "fit" measures in its second step - we consider various distance and similarity measures of fuzzy sets to choose the most appropriate linguistic approximation. We conduct a numerical study of the performance of this linguistic approximation method, present its results and discuss the impact of a particular choice of a "fit" measure.

Keywords: Linguistic approximation, two-step method, fuzzy number, distance, similarity.

JEL classification: C44
AMS classification: 90B50, 91B06, 91B74

1 Introduction

In practical applications of decision support models that employ fuzzy sets it is often necessary to be able to assign a linguistic label (from predefined linguistic scale) to a fuzzy set (usually obtained as an output of some mathematical model). This process is called *linguistic approximation.* The main reason for applying linguistic approximation is to "translate" (abstract/formal) mathematical objects into the common language (recent research also suggests that the ideas of linguistic approximation can be used e.g. for ordering purposes - see [7]). This way the outputs of mathematical models can become easier to understand and use for the decision-makers. The process of linguistic approximation involves the selection of the best fitting linguistic term from a predefined term set as a representative of the given mathematical object (fuzzy set). Obviously, since the set of linguistic terms is finite (and usually contains only a few linguistic terms), the process distorts the actual output of the mathematical model to some extent (add or decrease uncertainty, shift the meaning in the given context etc. - hence *approximation*). The key to a successful linguistic approximation is to find an appropriate tradeoff between understandability and loss (distortion) of information (see e.g. [10, 6]). Linguistic approximation relies in many cases on distance and similarity of fuzzy sets, on the subsethood and the differences in relevant features of the output to be approximated and the meaning of its approximating linguistic term.

In this paper we focus on the Bonissone's two-step method for linguistic approximation [1], since it combines the idea of semantic similarity with the requirement of the closeness of meaning. In the first step, the method preselects a given amount of linguistic terms, that embody the semantic best fit (based on a specified set of features). In the second step the linguistic term whose meaning is the closest based on some distance/similarity measure is selected. We investigate the role of distance/similarity measure in the

[1]Palacký University, Olomouc, Křížkovského 8, Olomouc, Czech Republic and Lappeenranta University of Technology, Skinnarilankatu 34, Lappeenranta, Finland, tomas.talasek@upol.cz
[2]Palacký University, Olomouc, Křížkovského 8, Olomouc, Czech Republic, jan.stoklasa@upol.cz
[3]Palacký University, Olomouc, Křížkovského 8, Olomouc, Czech Republic, jana.talasova@upol.cz

second step of this method and on a numerical study we compare the performance of the Bhattacharyya distance suggested by Bonissone with the dissemblance index distance measure and two fuzzy similarity measures. Based on the results of a numerical study we analyze what features employed in the first step (namely position and uncertainty - the same features employed by Wenstøp [9] in his method) are emphasized and which are distorted by each of the distance/similarity measures. This way the paper strives to contribute to the scarce body of research on good practices in linguistic approximation.

2 Preliminaries

Let U be a nonempty set (the universe of discourse). 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 $A(.)$ is called a *membership function* of the fuzzy set A. $\text{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an *α-cut* of A for any $\alpha \in [0,1]$, $\text{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\text{Ker}(A) \neq \emptyset$ (A is *normal*); (2) A_α are closed intervals for all $\alpha \in (0,1]$ (this implies A is *unimodal*); (3) $\text{Supp}(A)$ is bounded. A family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on [a,b], if $\text{Supp}(A)$ is a subset of an interval $[a,b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \text{Cl}(\text{Supp}(A))$ and $[a_2, a_3] = \text{Ker}(A)$, where $\text{Cl}(\text{Supp}(A))$ denotes a closure of $\text{Supp}(A)$. Each fuzzy number A is determined by $A = \left\{[\underline{a}(\alpha), \overline{a}(\alpha)]\right\}_{\alpha \in [0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall \alpha \in (0,1]$, and the closure of the support of A $\text{Cl}(\text{Supp}(A)) = [\underline{a}(0), \overline{a}(0)]$. A *union* of two fuzzy sets A and B on U is a fuzzy set $(A \cup B)$ on U defined as follows: $(A \cup B)(x) = \min\{1, A(x) + B(x)\}$, $\forall x \in U$.

The fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$; for such fuzzy numbers we will use a simplified notation $A = (a_1, a_2, a_3, a_4)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by ordered triplet $A = (a_1, a_2, a_4)$. More details on fuzzy numbers and computations with them can be found for example in [2].

Let A be a fuzzy number on $[a,b]$ for which $a_1 \neq a_4$. Then A could be described by several real number characteristics, such as *cardinality*: $\text{Card}(A) = \int_{[a,b]} A(x)dx$; *center of gravity*: $\text{COG}(A) = \int_{[a,b]} xA(x)dx / \text{Card}(A)$; *fuzziness*: $\text{Fuzz}(A) = \int_{[a,b]} \text{S}(A(x))dx$, where $S(y) = -y\ln(y) - (1-y)\ln(1-y)$ and *skewness*: $\text{Skew}(A) = \int_{[a,b]} (x - \text{COG}(A))^3 A(x)dx$.

A *fuzzy scale* on $[a,b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on [a,b], that form a Ruspini fuzzy partition (see [5]) of the interval $[a,b]$, i.e. for all $x \in [a,b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. A *linguistic variable* ([11]) is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values (terms), X is an universe on which the meanings of the linguistic values are defined, G is an syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X. Linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ is called a *linguistic scale* on $[a,b]$ if $X = [a,b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$ and $M(\mathcal{T}_i) = T_i, i = 1, \ldots, s$ form a fuzzy scale on $[a,b]$. Terms $\mathcal{T}_i, i = 1, \ldots, s$ are called *elementary terms*. Linguistic scale on $[a,b]$ is called *extended linguistic scale*, if besides elementary terms contains also *delivered* terms in the form $\mathcal{T}_i$ to $\mathcal{T}_j$ where $i < j$, $i, j \in \{1, \ldots, n\}$ and $M(\mathcal{T}_i \text{ to } \mathcal{T}_j) = T_i \cup T_{i+1} \cup \cdots \cup T_j$.

3 Bonissone's two step method for linguistic approximation

Bonissone's two step approach for linguistic approximation [1] was proposed in 1979. In contrast to the majority of linguistic approximation approaches, Bonissone suggested to split the process into two steps – in the first step the set of suitable linguistic terms for the approximation of a given fuzzy number is found (this "pre-selection step" is done based on the semantic similarity), then in the second step the most appropriate term for the linguistic approximation is found from this set of suitable linguistic terms.

In the pre-selection step the set $\mathcal{P} = \{\mathcal{T}_{p_1}, \ldots, \mathcal{T}_{p_k}\}$ of k $(k \leq s)$ suitable linguistic terms from $\mathcal{T}(\mathcal{V})$

is formed in the way that the meaning of these pre-selected terms are similar to the fuzzy set O (an output of a mathematical model to be approximated) with respect to four characteristics (cardinality, center of gravity, fuzziness and skewness). These characteristics are assumed to capture the semantic value of a fuzzy set used to model the meaning of a linguistic term. The semantic value of a fuzzy set on a given universe can thus be represented by a quadruple of real numbers (values of 4 features in four-dimensional space). Let A be a fuzzy set on $[a, b]$. Then the respective characteristic quadruple is denoted as (a^1, a^2, a^3, a^4) where $a^1 = \mathrm{Card}(A)$, $a^2 = \mathrm{COG}(A)$, $a^3 = \mathrm{Fuzz}(A)$ and $a^4 = \mathrm{Skew}(A)$.

Let the fuzzy set O on $[a, b]$ be an output of a mathematical model that needs to be linguistically approximated by one linguistic term from the set $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \dots, \mathcal{T}_s\}$. $\mathcal{T}(\mathcal{V})$ is a linguistic term set of a linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), [a, b], G, M)$, such that $T_i = M(\mathcal{T}_i)$, $i = 1, \dots, s$ are fuzzy numbers on $[a, b]$. Linguistic terms $\{\mathcal{T}_1, \dots, \mathcal{T}_s\}$ are ordered with respect to the distance of their characteristic quadruples from the characteristic quadruple of O. The ordered set $\mathcal{N} = (\mathcal{T}_{p_1}, \dots, \mathcal{T}_{p_s})$ is thus obtained, such that $d(T_{p_1}, O) \leq d(T_{p_2}, O) \leq \dots \leq d(T_{p_s}, O)$ where

$$d(T_{p_i}, O) = \sum_{j=1}^{4} w_j |t^j_{p_i} - o^j|, \; i = 1 \dots, s, \tag{1}$$

and w_j, $j = 1, \dots, 4$ are normalized real weights (i.e. $\sum_{j=1}^4 w_j = 1,\ w_j \geq 0,\ j = 1, \dots, 4$). The choice of weights is usually left with the user of the model and some features could be even optional. Wenstøp [9] for example proposed (in his method for linguistic approximation) to use only two features – cardinality (uncertainty) and center of gravity (position). First k linguistic terms (the parameter k is specified by the decision maker) from the ordered set $\mathcal{N}$ are stored in the set $\mathcal{P}$ and the pre-selection step is finished.

In the second step, the linguistic approximation $\mathcal{T}_O \in \mathcal{P}$ of the fuzzy set O is computed. The fuzzy set $T_O = M(\mathcal{T}_O)$ is computed as

$$T_O = \arg \min_{\mathcal{T}_{p_i} \in \mathcal{P}} d_1(T_{p_i}, O) \tag{2}$$

using the modified Bhattacharyya distance:

$$d_1(A, B) = \left[1 - \int_U (A^*(x) \cdot B^*(x))^{1/2} dx\right]^{1/2}, \tag{3}$$

where $A^*(x) = A(x)/\mathrm{Card}(A(x))$ and $B^*(x) = B(x)/\mathrm{Card}(B(x))$. This way the linguistic term $\mathcal{T}_O$ is found as the closest linguistic approximation among the pre-selected linguistic terms.

The Bhattacharyya distance (3) can be substituted by different distances or similarity measures[1] of fuzzy numbers – this step will, however, modify the behaviour of the linguistic approximation method. In the next section the following distance and similarity measures of fuzzy numbers are considered:

- A *dissemblance index* (introduced by Kaufman and Gupta [4]) of fuzzy numbers A and B is defined by the formula

$$d_2(A, B) = \int_0^1 |\underline{a}(\alpha) - \underline{b}(\alpha)| + |\overline{a}(\alpha) - \overline{b}(\alpha)|\, d\alpha, \tag{4}$$

- A *similarity measure* (introduced by Wei and Chen in [8]) of fuzzy numbers A and B is defined by the formula

$$s_1(A, B) = \Big(1 - \frac{\sum_{i=1}^4 |a_i - b_i|}{4}\Big) \cdot \frac{\min\{Pe(A), Pe(B)\} + \min\{\mathrm{hgt}(A), \mathrm{hgt}(B)\}}{\max\{Pe(A), Pe(B)\} + \max\{\mathrm{hgt}(A), \mathrm{hgt}(B)\}}, \tag{5}$$

  where $Pe(A) = \sqrt{(a_1 - a_2)^2 + (\mathrm{hgt}(A))^2} + \sqrt{(a_3 - a_4)^2 + (\mathrm{hgt}(A))^2} + (a_3 - a_2) + (a_4 - a_1)$, $Pe(B)$ is defined analogically

- A *similarity measure* (introduced by Hejazi and Doostparast in [3]) of fuzzy numbers A and B can be defined by the formula

$$s_2(A, B) = \Big(1 - \frac{\sum_{i=1}^4 |a_i - b_i|}{4}\Big) \cdot \frac{\min\{Pe(A), Pe(B)\}}{\max\{Pe(A), Pe(B)\}} \cdot \frac{\min\{Ar(A), Ar(B)\} + \min\{\mathrm{hgt}(A), \mathrm{hgt}(B)\}}{\max\{Ar(A), Ar(B)\} + \max\{\mathrm{hgt}(A), \mathrm{hgt}(B)\}}, \tag{6}$$

  where $Ar(A) = \frac{1}{2}\mathrm{hgt}(A)(a_3 - a_2 + a_4 - a_1)$, $Ar(B)$ is defined analogically and $Pe(A)$ and $Pe(B)$ are computed identically as in the previous method.

[1] In the case of similarity measure the *arg min* function in the formula (2) must be changed to *arg max* function.

4 Numerical experiment

We restrict ourselves for the purpose of this paper to the linguistic approximation of triangular fuzzy numbers. We assess the performance of the distance measures d_1 and d_2 and of the two similarity measures s_1 and s_2 in the context of Bonisonne's linguistic approximation method by the following numerical experiment. We randomly generate 100 000 triangular fuzzy numbers on $[0,1]$ to be linguistically approximated (denoted $\{O_1,\ldots,O_{100000}\}$) and compute their cardinalities $\{o_1^1,\ldots,o_{100000}^1\}$ and their centers of gravity $\{o_1^2,\ldots,o_{100000}^2\}$. We assume that for all these generated fuzzy numbers the hypothetical result of the first phase of Bonisonne's method is the set of linguistic terms $\mathcal{P}=\{\mathcal{T}_{p_1},\ldots,\mathcal{T}_{p_k}\}$ - in our numerical study this set is the linguistic term set of an extended linguistic scale constructed from a uniform Ruspini fuzzy partition of the universe $[0,1]$ with 5 triangular fuzzy numbers. This way we obtain the linguistic term set $\mathcal{P}=\{\mathcal{T}_{p_1},\ldots,\mathcal{T}_{p_{15}}\}$, the meanings of these linguistic terms are $\{T_1,\ldots,T_{15}\}$, with cardinalities $\{t_1^1,\ldots,t_{15}^1\}$ and centers of gravity $\{t_1^2,\ldots,t_{15}^2\}$. Using each distance and similarity measure we find the linguistic approximation of each generated output applying the second step of Bonisonne's method - this way we obtain $\{\mathcal{T}_{O_1}^{d_1},\ldots,\mathcal{T}_{O_{100000}}^{d_1}\}$, $\{\mathcal{T}_{O_1}^{d_2},\ldots,\mathcal{T}_{O_{100000}}^{d_2}\}$, $\{\mathcal{T}_{O_1}^{s_1},\ldots,\mathcal{T}_{O_{100000}}^{s_1}\}$ and $\{\mathcal{T}_{O_1}^{s_2},\ldots,\mathcal{T}_{O_{100000}}^{s_2}\}$ as the linguistic approximations of the generated triangular fuzzy numbers using d_1, d_2, s_1 and s_2 respectively.

Figure 1 plots the cardinalities of the approximated fuzzy numbers (horizontal axis) against the cardinality of the meaning of the respective linguistic approximation for all the distance/similarity measures. We can clearly see from the plots, that all four measures provide linguistic approximations with both higher cardinality (points above the main diagonal) and with lower cardinality. It, however, seems, that higher cardinality case is more frequent (points in the left upper corner of the plots). This can be reasonable, since even in common language we tend to use super-categories to generalize the meaning. In all the methods it is possible to also get a linguistic approximation with a lower cardinality (i.e. the meaning of the linguistic approximation is less uncertain than the original output of the model). Note, that since we have generated triangular fuzzy numbers on $[0,1]$, the maximum possible cardinality of any generated fuzzy number was 0.5. The Bhattacharyya distance is the only one from the investigated measures, that provides very highly uncertain approximations. This behavior could be tolerated only if the reason for the addition of uncertainty is the tendency of the measure to achieve a linguistic approximation that is more general than the approximated fuzzy set. Table 1 summarizes in how many cases the kernel of the resulting linguistic approximation is a superset of the kernel of the approximated results - in these cases the "typical representatives" of the output are also the "typical representatives" of the approximated linguistic term. We can see that Bhattacharyya distance focuses on this aspect much more than the other investigated methods.

The situation for the centers of gravity is summarized analogically in Figure 2. Here the desired state can be no presence of a systematic bias of the approximation. This corresponds with the points being close to the main diagonal in the respective plot, or evenly distributed to the left and to the right. We can see that with respect to this requirement the Bhattacharyya distance performs rather well. Both similarities perform in most cases in the following way: i) in case of lower centers of gravity of the approximated result they shift the center of gravity of the meaning of the linguistic approximation lower than the original center of gravity of the approximated results, ii) in case of higher centers of gravity of the approximated result they shift the center of gravity of the meaning of the linguistic approximation higher than the original center of gravity of the approximated results. Similarities seem to have an amplifying effect on the center of gravity - shifting the center of gravity to the endpoints of the universe. This can be a desirable property in cases, when such an amplification of meaning is needed.

k	d_1	d_2	s_1	s_2
$\frac{\text{Card}\{O_i \mid \text{Ker}(O_i)\subseteq\text{Ker}(T_{O_i}^k), i=1,\ldots,100000\}}{100000}$	0.2868	0.1559	0.1778	0.1632

Table 1: The relative count of cases when the $\text{Ker}(O_i) \subseteq \text{Ker}(T_{O_i}^k)$, $k \in \{d_1, d_2, s_1, ss_2\}$ out of the given 100 000.

5 Conclusion

In the paper we have investigated the role of different distance and similarity measures of fuzzy numbers in the second step of Bonissone's linguistic approximation method. We focused on the cardinality and

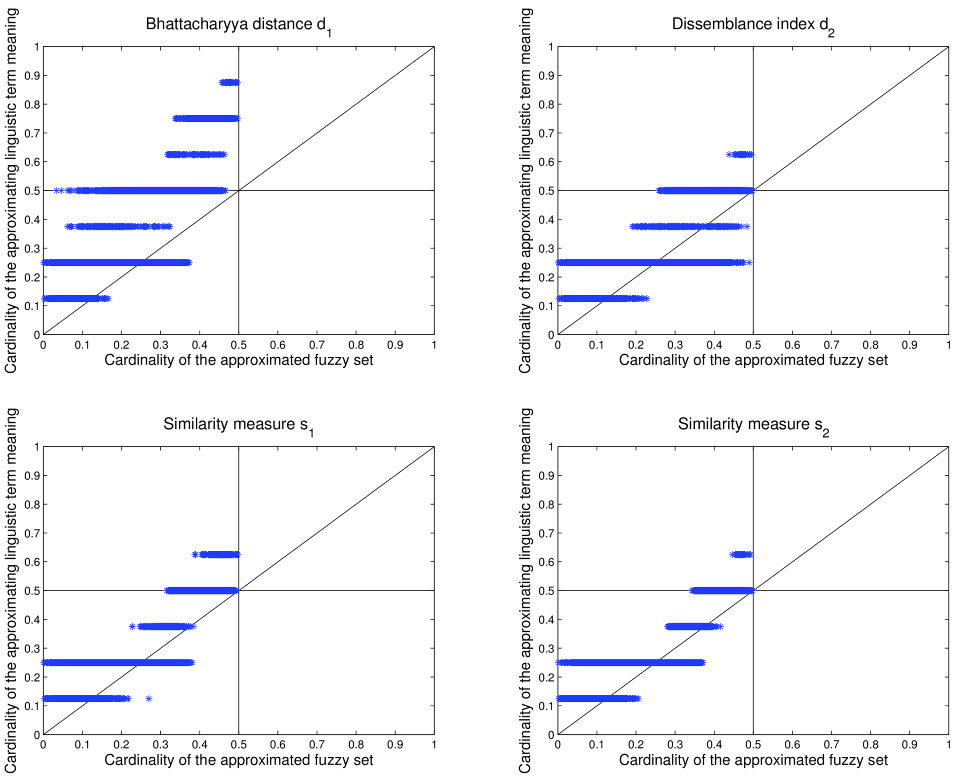


Figure 1: Comparison of cardinalities of the approximated outputs (horizontal axis) and the meanings of their linguistic approximations (vertical axis) for d_1, d_2, s_1 and s_2.

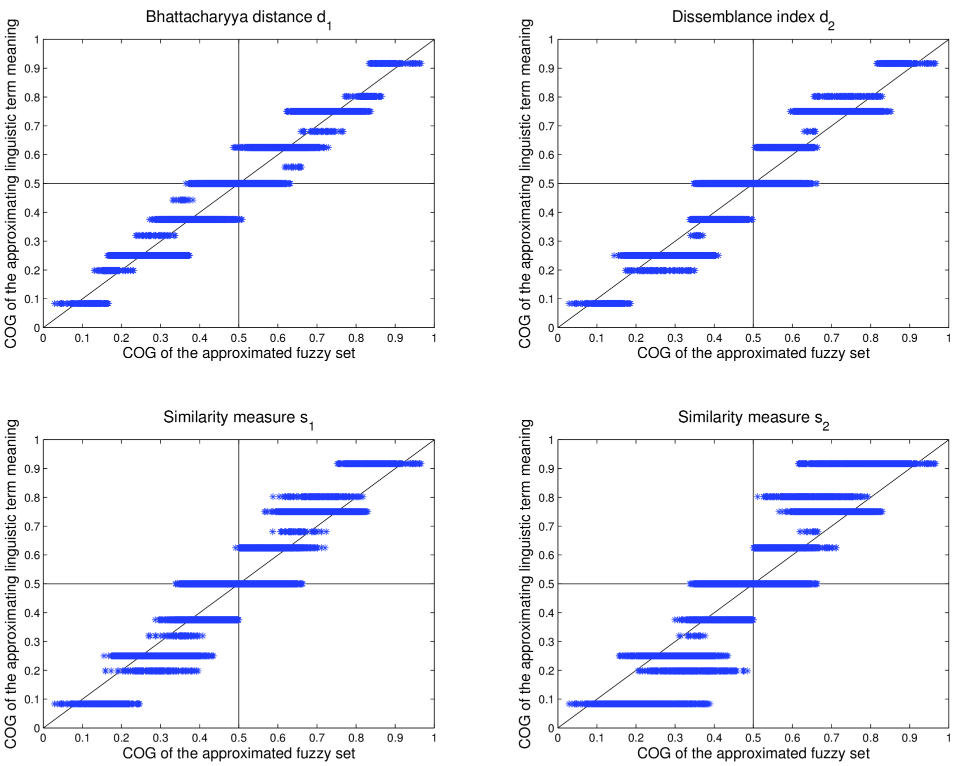


Figure 2: Comparison of centers of gravity of the approximated outputs (horizontal axis) and the meanings of their linguistic approximations (vertical axis) for d_1, d_2, s_1 and s_2.

center of gravity characteristics of fuzzy numbers. We have performed a numerical experiment which investigated the differences between the chosen characteristics of randomly generated triangular fuzzy numbers on the interval $[0, 1]$ and the characteristics of the meanings of their linguistic approximations computed by Bonissone's method. This served as a basis for the analysis of the performance of two different distance measures and two similarity measures of fuzzy numbers in the linguistic approximation context.

The results of the numerical experiment suggest, that the Bhattacharyya distance tends to provide more uncertain approximations than the other methods and is more likely to provide approximating linguistic term that "catch" the typical representatives (the kernel of the approximating linguistic term meaning is a superset to the kernel of the approximated fuzzy number). Both presented similarity methods have amplifying effect on the center of gravity – they shift the center of gravity of the approximating linguistic term meaning to the endpoints of the universe.

Acknowledgements

Supported by the grant by the grant GA 14-02424S Methods of operations research for decision support under uncertainty of the Grant Agency of the Czech Republic and partially also by the grant IGA PrF 2016 025 of the internal grant agency of the Palacký University, Olomouc.

References

[1] Bonissone, P. P.: A pattern recognition approach to the problem of linguistic approximation in system analysis. In: *Proceedings of the IEEE International Conference on Cybernetics and Society*, 1979, 793–798.

[2] Dubois, D., and Prade, H.: *Fuzzy sets and systems: theory and applications*, Academic Press, 1980.

[3] Hejazi, S. R., Doostparast, A., and Hosseini, S. M.: An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers. *Expert Systems with Applications*, **38**, 8 (2011), 9179–9185.

[4] Kaufman, A., and Gupta, M. M.: *Introduction to Fuzzy Arithmetic*, Van Nostrand Reinhold, New York, 1985.

[5] Ruspini, E.: A New Approach to Clustering. *Information and Control*, **15** (1969), 22–32.

[6] Stoklasa, J.: *Linguistic models for decision support.* Lappeenranta University of Technology, Lappeenranta, 2014.

[7] Talášek, T., Stoklasa, J., Collan, M., and Luukka, P.: Ordering of Fuzzy Numbers through Linguistic Approximation Based on Bonissone's Two Step Method. In: *Proceedings of the 16th IEEE International Symposium on Computational Intelligence and Informatics.* Budapest, 2015, 285–290.

[8] Wei, S. H., and Chen, S. M.: A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *Expert Systems with Applications*, **36**, 1 (2009), 589–598.

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

[10] Yager, R. R.: On the retranslation process in Zadeh's paradigm of computing with words. *IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics*, **34**, 2 (2004), 1184–1195.

[11] Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning I, II, III. *Information Sciences*, **8** (1975), 199–257, 301–357, **9** (1975), 43–80.

Publication VII

Stoklasa, J. and Talášek, T.
Linguistic approximation of values close to the gain/loss threshold

Reprinted with the permission from
Proceedings of the 35th *International Conference on Mathematical Methods in Economics 2017*,
pp. 726–731, 2017,

Linguistic approximation of values close to the gain/loss threshold

Jan Stoklasa[1], Tomáš Talášek[2]

Abstract. Linguistic approximation (LA) is a natural last step of linguistic fuzzy modelling, providing linguistic labels (with their meaning known to the decision makers and understood well by them). Linguistic approximation techniques are based on approximation and hence the nature of the approximated output of mathematical model can be altered a bit by the application of these methods. LA can be considered beneficial in linguistic fuzzy modelling, as long as the interpretability and understandability of the provided linguistic outputs outweighs the possible loss/distortion of information. In many cases the distortion of information might be small and as such completely acceptable. Recently, however, Stoklasa and Talášek (2015) pointed out that when specific thresholds are of importance in the decision-making situation (e.g. the border between gains and losses), LA can distort the outcome of the decision-making situation by providing a loss label for a gain and vice-versa. In this paper, we investigate the phenomenon under different linguistic scales used for the approximation and provide a thorough discussion of this phenomenon in the context of linguistic approximation.

Keywords: Linguistic approximation, gains, losses, threshold, distance, linguistic scale.

JEL classification: D81, C44
AMS classification: 90B50, 91B06

1 Introduction

Mathematical models for economic practice and for managerial decision support (including e.g. investment decision support models, evaluation models) require a suitable interface to facilitate the exchange of information between the model and its users. Linguistic fuzzy modelling provides such an interface in terms of presenting the model and its outputs in terms of natural language [9]. To build a linguistic fuzzy model capable of providing understandable linguistic outputs to its users, we need to be able to transform the mathematical objects computed by the model into natural language. The process of transformation of the mathematical outputs of models into natural language is called linguistic approximation. There are various approaches to linguistic approximation (see e.g. [23] for an overview and [10, 15] for additional analysis of some of the methods). The majority of the methods of linguistic approximation is based on finding the fuzzy object (usually a fuzzy number) with a known linguistic label - e.g. methods finding the fuzzy set with a known linguistic label which is the closest (w.r.t. some distance measure, see e.g. [3] or) or the most similar (w.r.t. some similarity measure) to the approximated object. The performance of different similarity and distance measures has been recently studied in several papers (see e.g. [16, 17, 20]). Alternatively, there are also methods that use linguistic hedges and connectives to combine fuzzy sets with a known linguistic label to create and object close or similar enough to the approximated one (see e.g. [1, 5, 20, 22]). New methods for linguistic approximation are also being developed [19] and alternative uses for linguistic approximation are being considered (e.g. ordering of fuzzy numbers in [18], or conveying/stressing of specific pieces of information [23], the importance of the linguistic level of models has recently been discussed also in [2, 8, 9, 11, 12, 13, 14]). Clearly, linguistic modelling and linguistic approximation are topics that currently deserve the attention of researchers.

Although research into the behavioral aspects of linguistic approximation has already started, there are still several issues that need attention. For one the reliance on the distance or similarity measures in linguistic approximation to find the best fitting approximating linguistic term (in terms of the distance or similarity of its fuzzy-set-meaning to the approximated object) can prove problematic, since low distance and semantic closeness might not always be the same thing. Stoklasa and Talášek [10, p. 965, Figure 4] discuss the existence of a possible drawback of the use of linguistic approximation based on distance or similarity measures in the context of

[1]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic and Lappeenranta University of Technology, School of Business and Management, Skinnarilankatu 34, Lappeenranta, Finland, jan.stoklasa@upol.cz.

[2]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic, tomas.talasek@upol.cz.

economic decision-making concerning gains and losses. As already identified by Kahneman and Tversky e.g. in the context of prospect theory [6], the decision making and attitude to risk might be different based on the framing of a particular value as loss or gain by a specific decision maker. In this paper, we aim to investigate the possible drawbacks of the use of distance and similarity in linguistic approximation of the outputs of mathematical models in financial units, provide a closer-to-real-life example of the possible problems and analyze the performance of a frequently used distance measure and its possible alternative under different linguistic variables used for the linguistic approximation.

The paper therefore continues by a chapter summarizing the necessary theory and notation for linguistic fuzzy modelling including linguistic variables, linguistic scales and the basic idea of linguistic approximation. The next section specifies the problem under investigation, introduces the distance measures the performance of which will be investigated in this paper in the context of linguistic approximation of gains and losses and also specifies the linguistic variables that will be studied. A prototype example of the problem is also presented in this section. The next section summarizes the results of a numerical analysis of the performance of the selected methods and discusses the results and the last section draws conclusions for the paper.

2 Preliminaries

Let U be a nonempty set (the universe of discourse). 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 $A(.)$ is called the *membership function* of the fuzzy set A. Ker$(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes the α*-cut* of A for any $\alpha \in [0,1]$, Supp$(A) = \{x \in U | A(x) > 0\}$ denotes the *support* of A. The *cardinality* of a fuzzy set A is computed as Card$(A) = \int_U A(x)dx$. A real-number characteristic representing the location of the fuzzy set A in the universe of discourse U is called the *center of gravity*: COG$(A) = \int_U xA(x)dx/\text{Card}(A)$.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: a) Ker$(A) \neq \emptyset$ (A is *normal*); b) A_α are closed intervals for all $\alpha \in (0,1]$ (this implies A is *unimodal*); c) Supp(A) is bounded. The family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on $[a,b]$, if Supp(A) is a subset of an interval $[a,b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_2, a_3] = \text{Ker}(A)$ and $[a_1, a_4] = \text{Cl(Supp}(A))$, where Cl(Supp$(A)$) denotes a closure of Supp(A). Each fuzzy number A can be also represented in the form of $A = \left\{[\underline{a}(\alpha), \overline{a}(\alpha)]\right\}_{\alpha\in[0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall\alpha \in (0,1]$, and the closure of the support of A, Cl(Supp(A)) $= [\underline{a}(0), \overline{a}(0)]$. A fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$; for such fuzzy numbers we will use a simplified notation $A = (a_1, a_2, a_3, a_4)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by ordered triplet $A = (a_1, a_2, a_4)$. More details on fuzzy numbers and computations with them can be found for example in [4].

A *fuzzy scale* on $[a,b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on $[a,b]$, where for all $x \in [a,b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. A *linguistic variable* [24] is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is the name of the variable, $\mathcal{T}(\mathcal{V})$ is the set of its linguistic values (terms), X is the universe on which the meanings of the linguistic values are defined, G is a syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X. Linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ is called a *linguistic scale* on $[a,b]$ if $X = [a,b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$ and $\mathcal{M}(\mathcal{T}_i) = T_i, i = 1, \ldots, s$ form a fuzzy scale on $[a,b]$.

3 Definition of the problem - distance based linguistic approximation in the gain/loss domain

Kahneman and Tversky (see e.g. [7, 21]) suggested and subsequently experimentally proved, that the carrier of decision-power in real life situations concerning e.g. sums of money is not the absolute value, but its reframing into gain or loss. They also postulate, that people deal differently with gains and losses (willingness to take risk might change, see [6] for more). The purpose of linguistic approximation is to find the best linguistic label for a given mathematical output. If we assume a fuzzy number O to be linguistically approximated by one of the linguistic values of a linguistic scale $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$, then a distance based approach to linguistic approximation translates into (1), i.e. into finding such an element in $T(V)$, for which the distance d of its fuzzy-number meaning to the approximated fuzzy output O is minimal.

$$T_O = \arg \min_{\mathcal{T}_i \in \mathcal{T}(\mathcal{V})} d(T_i, O) \tag{1}$$

We need to stress here, that the linguistic approximation is not always able to preserve all the information carried out by the approximated output (hence "approximation"). We, however, need to make sure, that the most important characteristics of the approximated objects are not distorted too much. In the context of gains/losses, we would at least expect a clear loss not to be assigned a "gain" label and vise-versa. The outcome of the linguistic approximation obviously depends on the linguistic variable used in the process and on the definition of the meaning of its linguistic values. In this paper, we assume two different general types of linguistic scales for the purpose (the meanings of the linguistic values of both of them are summarized in Figure 1). The first linguistic scale assumes a decision maker not distinguishing in the loss domain, while the other one assumes that losses and gains are partitioned in a similar manner, the red lines represent the loss/gain threshold.

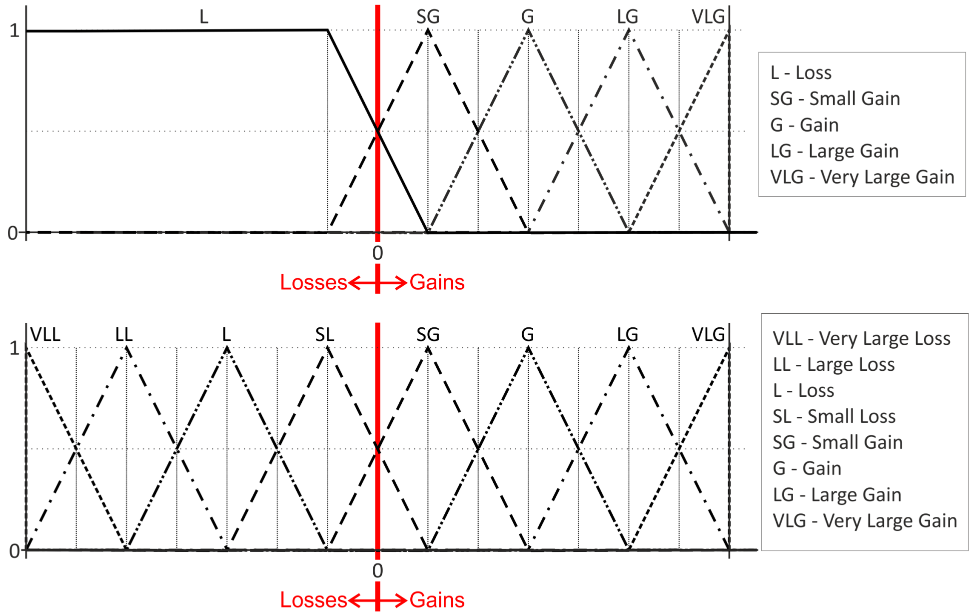


Figure 1: The scales used for linguistic approximation of the outputs of mathematical models representing financial values (e.g. NPV of a project, etc.) or future cash flow estimates. The case represented in the top figure does not differentiate in the area of losses, the bottom linguistic scale differentiates in the area of losses in the same way as in gains.

Obviously the other crucial factor influencing the outcome of the approximation is the distance measure used. One of the frequently used distance measures of fuzzy numbers is the *dissemblance index* of fuzzy numbers A and B, $d_1(A, B)$, defined by the formula (2). The dissemblance index requires both A and B to be fuzzy numbers, which is not a problem, since the meanings of the linguistic values of the approximating linguistic variable are usually represented by fuzzy numbers and the approximated object can be expected to be a fuzzy number as well. Without any loss of generality we use the dissemblance index in a non-normalized form, if needed, it can be normalized so that its value lies within the $[0, 1]$ interval, i.e. $d_1(A, B)/2\,(b - a) \in [0, 1]$ for $A, B \in \mathcal{F}_N([a, b])$. Note, that in the gain/loss domain, we are expecting the outputs of the mathematical models to be fuzzy quantities (e.g. represented by triangular fuzzy numbers).

$$d_1(A, B) = \int_0^1 |\underline{a}(\alpha) - \underline{b}(\alpha)| + |\overline{a}(\alpha) - \overline{b}(\alpha)|\, d\alpha, \tag{2}$$

Using d_1 and the top linguistic scale in Figure 1, we can obtain very counterintuitive results of linguistic approximations. An example of such a problematic result is presented in Figure 2, where a clear "loss" represented by the fuzzy number Out is linguistically approximated by the label "small gain". Such a mislabelling of an output can have serious consequences in decision support, since a "gain" label can motivate a different reaction of the decision maker than would be required for an actual loss Out. Note, that in Figure 2, we have $d_1(Out, SG) = a + b < \int_0^1 |\underline{l}(\alpha) - \underline{Out}(\alpha)| d\alpha < d_1(L, Out) = \int_0^1 |\underline{l}(\alpha) - \underline{Out}(\alpha)| + |\overline{l}(\alpha) - \overline{Out}(\alpha)| d\alpha$.

We have thus identified an even clearer example of the possible problem with distance-based linguistic approximation, where even though a mathematically sound distance measure and a reasonable linguistic scale is used, the resulting approximation can completely change the nature of (information carried by) the actual approximated output. In the next section, we investigate how serious this problem is for the dissemblance index and compare the performance of this distance measure with another distance measure - namely the modified Bhattacharyya distance

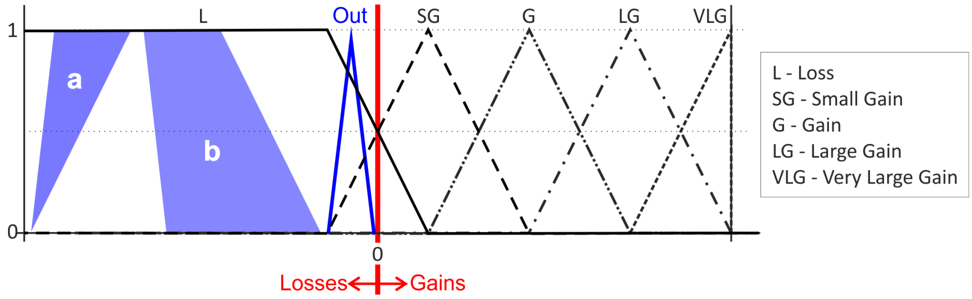


Figure 2: A graphical representation of $d_1(Out, SG)$ represented by the blue shapes a) and b). Note, that $a = \int_0^1 |\underline{Out}(\alpha) - \underline{sg}(\alpha)|d\alpha$, $b = \int_0^1 |\overline{Out}(\alpha) - \overline{sg}(\alpha)|d\alpha$ and $a + b = d_1(Out, SG)$. Clearly $d_1(Out, SG) < d_1(Out, L)$ and thus "small gain" is considered to be a better linguistic approximation for Out than "loss", even though Out is completely in the loss domain.

of fuzzy numbers d_2, which can be computed in the following way:

$$d_2(A, B) = \left[1 - \int_U (A^*(x) \cdot B^*(x))^{1/2} dx\right]^{1/2}, \tag{3}$$

where $A^*(x) = A(x)/\text{Card}(A)$ and $B^*(x) = B(x)/\text{Card}(B)$. We also investigate how a change in the linguistic scale used for the approximation influences the results of the approximation and the performance (and appropriateness) of both distance measures.

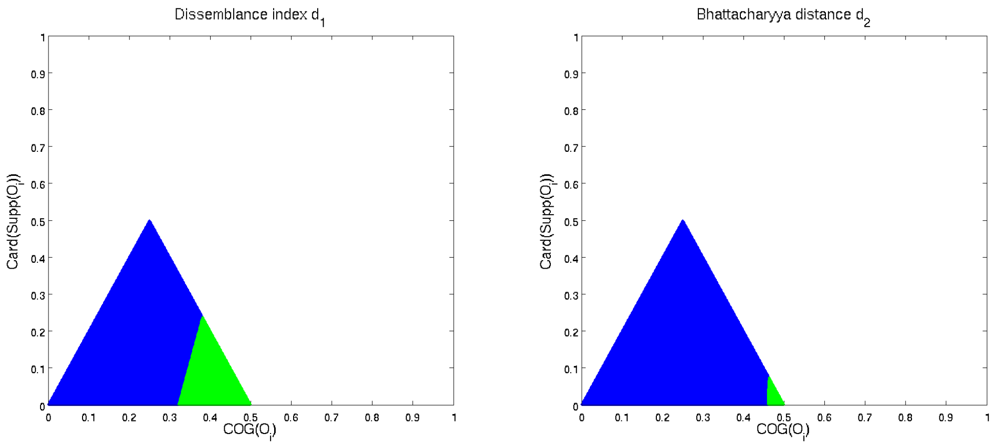


Figure 3: Results of the numerical experiment for the linguistic scale not differentiating in the loss domain. Each point represents one symmetrical triangular fuzzy number O_i, $\text{Supp}(O_i) \subseteq [-r, 0], i = 1, \ldots, 125\,000$, characterized by its center of gravity (x-coordinate, the $[-r, r]$ universe is just linearly transformed to $[0, 1]$) and the cardinality of its support (y-coordinate). The colour represents the result of the linguistic approximation: blue for *loss* and green for *small gain*. Results are presented for the linguistic approximation using d_1 (left plot, 103 758 fuzzy numbers approximated correctly as losses, 21 242 incorrectly as gains) and d_2 (right plot, 123 484 fuzzy numbers approximated correctly as losses, 1 516 incorrectly as gains).

4 Numerical analysis and discussion of the results

To stress the magnitude of the problem of possible mislabelling of "losses" by a "gain" label, we will consider only fuzzy-number outputs of the mathematical model to be approximated which are completely in the domain of losses, i.e. for which the whole support lies in the loss domain. To simplify the analysis, we will also assume the approximated objects are symmetrical triangular fuzzy numbers. Using the same approach as in [17], a total of 125 000 symmetrical triangular fuzzy numbers O_i were generated, $i = 1, \ldots, 125\,000$, which uniformly cover the $[-r; 0]$ universe, where r represents the maximum expected gain and $-r$ the maximum expected loss. The results of the numerical experiment using the top linguistic scale from Figure 1 are presented in Figure 3. Almost 17% of the triangular fuzzy numbers representing a clear loss are mislabelled as gains using the dissemblance index. Note also, that low-uncertain fuzzy numbers can still be labelled as gains, even though their COG is close to the middle of $[-r; 0]$ interval. The dissemblance index clearly is not a good choice with a linguistic scale which treats gains and losses asymmetrically, since the large cardinality of the L fuzzy number distorts the computations. On the other hand the use of Bhattacharyya distance in this case can significantly reduce the risk of mislabeling - see that only about 1.5% of the clear loss fuzzy outputs were labelled as gains - all of them with low cardinality and COG close to the loss/gain threshold.

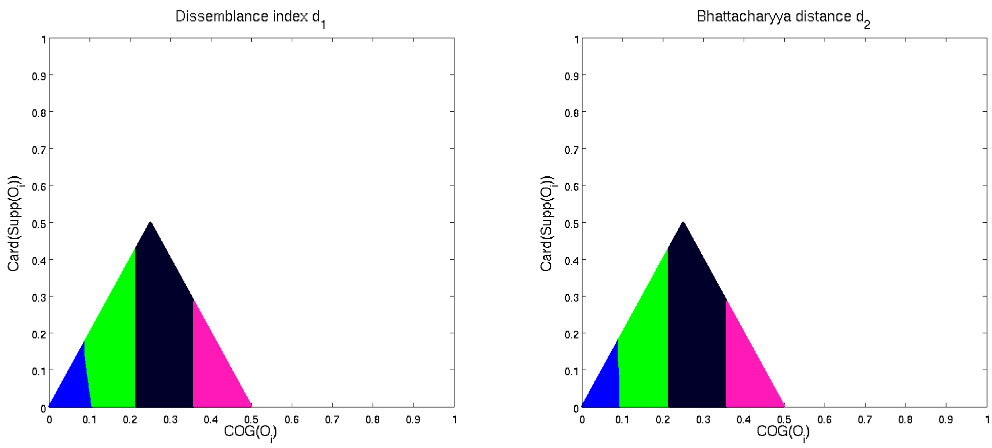


Figure 4: Results of the numerical experiment for the linguistic scale differentiating in the loss domain in the same way as in the gains domain. Each point represents one symmetrical triangular fuzzy number O_i, Supp$(O_i) \subseteq [-r, 0], i = 1, \ldots, 125\,000,$ characterized by its center of gravity (x-coordinate, the $[-r, r]$ universe is just linearly transformed to [0,1]) and the cardinality of its support (y-coordinate). The colour represents the result of the linguistic approximation: blue for *very large loss*, green for *large loss*, black for *loss* and purple for *small loss*. Results are presented for the linguistic approximation using d_1 (left plot, 12 620 times VLL, 49 616 times LL, 55 278 times L , 6 972 times SL and in 514 cases two loss labels were suggested with the same distance to O_i) and d_2 (right plot, 11 862 times VLL, 50 387 times LL, 55 278 times L , 6 972 times SL and in 501 cases two loss labels were suggested with the same distance to O_i).

The same analysis was also performed for the symmetrical linguistic scale presented in the bottom part of Figure 1. The results are presented in Figure 4. For a symmetrical underlying linguistic scale the differences between the distance measures are almost nonexistent. Also note, that a gain label was never assigned for a symmetrical triangular fuzzy number representing a clear loss.

We can clearly see that both the selection of the linguistic scale and the selection of the distance method can significantly influence the results of the linguistic approximation. As Bhattacharyya distance favours supersets, it seems to be a method of choice for the use with linguistic scales which are not symmetrical with respect to the loss/gain threshold. On the other hand the selection of a symmetrical linguistic scale can get rid of the mislabelling problem between gains and losses and renders the performance of the two investigated distance measures almost identical.

5 Conclusion

This paper investigates the performance of two different distance measures of fuzzy numbers in the distance-based linguistic approximation of fuzzy numbers representing uncertain sums of money in the gain/loss framing. It provides a clear example of possible mislabeling problem, where as a result of the choice of a selection of an improper distance measure losses can be linguistically labelled as gains (and by the same logic gains as losses). In the context of the findings of prospect theory, this presents a significant problem in decision support, since gains and losses can motivate different decision strategies. A numerical analysis of this problem is performed and two possible solutions of the problem - the use of symmetrical linguistic scales or the use of Bhattacharyya distance method are suggested. The paper presents a first step in the investigation of the performance of linguistic approximation methods in the gain/loss domain, the investigation of the role of other distance and similarity measures as well as the implications of different formats of linguistic scales will be the natural next steps of this research stream.

Acknowledgements

Supported by the grant No. IGA_FF_2017_011 of the Internal Grant Agency of Palacký University Olomouc.

References

[1] Bonissone, P. P.: A fuzzy sets based linguistic approach: theory and applications. In: *Proceedings of the 12th conference on Winter simulation*, 1980. 99–111.

[2] 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** (2014), 437–455.

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

[4] Dubois, D., and Prade, H., eds.: *Fundamentals of Fuzzy Sets*. Kluwer Academic Publishers, Massachusetts, 2000.

[5] Eshragh, F., and Mamdani, E. H.: A general approach to linguistic approximation. *International Journal of Man-Machine Studies* **11** (1979), 501–519.

[6] Kahneman, D., and Tversky, A.: Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47** (1979), 263–292.

[7] Kahneman, D., and Tversky, A.: Choices , Values , and Frames. *American Psychologist* **39** (1984), 341–350.

[8] Krejčí, J., and Stoklasa, J.: Fuzzified AHP in the evaluation of scientific monographs. *Central European Journal of Operations Research* **24** (2016), 353–370.

[9] Stoklasa, J.: *Linguistic models for decision support*. Lappeenranta University of Technology, Lappeenranta, 2014.

[10] Stoklasa, J., and Talášek, T.: Jazykově orientované modelování pro ekonomickou a manažerskou praxi některé otevené problémy. In: *Znalosti pro tržní praxi 2015*. 959–969.

[11] Stoklasa, J., and Talášek, T.: On the use of linguistic labels in AHP: calibration, consistency and related issues. In: *Proceedings of the 34th International Conference on Mathematical Methods in Economics*. Technical University of Liberec, Liberec, 785–790.

[12] Stoklasa, J., Talášek, T., Kubátová, J., and Seitlová, K.: Likert scales in group multiple-criteria evaluation. *Journal of Multiple-Valued Logic and Soft Computing* (2017), (in press).

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

[14] Stoklasa, J., Talášek, T., and Stoklasová, J.: Semantic differential and linguistic approximation - identification of a possible common ground for research in social sciences. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016*. Societas Scientiarum Olomucensis II, Olomouc, 495–501.

[15] Talášek, T., and Stoklasa, J.: Jazykově orientované modelování pro ekonomickou a manažerskou praxi - vybrané metody pro jazykovou aproximaci. In: *Znalosti pro tržní praxi 2015*. 977–989.

[16] Talášek, T., and Stoklasa, J.: The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016*. Societas Scientiarum Olomucensis II, Olomouc, 539–546.

[17] Talášek, T., and Stoklasa, J.: A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation Under Different Linguistic Scales. *Journal of Multiple-Valued Logic and Soft Computing* (2017), (in press).

[18] Talášek, T., Stoklasa, J., Collan, M., and Luukka, P.: Ordering of fuzzy numbers through linguistic approximation based on Bonissone's two step method. In: *16th IEEE International Symposium on Computational Intelligence and Informatics*. 285–290.

[19] Talášek, T., Stoklasa, J., and Talašová, J.: Linguistic approximation using fuzzy 2-tuples in investment decision making. In: *Proceedings of the 33rd International Conference on Mathematical Methods in Economics*. 817–822.

[20] Talášek, T., Stoklasa, J., and Talašová, J.: The role of distance and similarity in Bonissone's linguistic approximation method a numerical study. In: *Proceedings of the 34th International Conference on Mathematical Methods in Economics*. Technical University of Liberec, Liberec, 845–850.

[21] Tversky, A., and Kahneman, D.: Rational Choice and the Framing of Decisions. *The Journal of Business* **59** (1986), S251–S278.

[22] Wenstøp, F.: Quantitative analysis with linguistic values. *Fuzzy Sets and Systems* **4** (1980), 99–115.

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

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

Publication VIII

Talášek, T. and Stoklasa, J.
Distance-based linguistic approximation methods: graphical analysis and numerical experiments

Reprinted with the permission from
Proceedings of the 35^{th} *International Conference on Mathematical Methods in Economics 2017*,
pp. 777–782, 2017,

Distance-based linguistic approximation methods: graphical analysis and numerical experiments

Tomáš Talášek[1], Jan Stoklasa[2]

Abstract. Linguistic approximation (LA) as a tool for converting the outputs of mathematical models into linguistic terms or expressions is a crucial tool in linguistic fuzzy modelling. The success of the models depends significantly on the ability of the users of these models to understand well enough the outputs provided by the models. Linguistic approximation offers a natural language for conveying information. On the other hand it is still an approximation of the original results and as such, there is information distortion taking place.
In this paper we study several distance-based linguistic approximation methods and analyse their performance in terms of LA for Mamdani-type outputs of mathematical models using a numerical experiment. We provide graphical summaries of the performance of these distance measures in LA as well as the frequencies of choosing specific linguistic labels considered to be the values of an extended linguistic scale. We discuss the differences in the focus of these methods and its implications for their usability. The paper strives to increase understanding of the LA methods and to contribute to the creation of a LA road map for practical use.

Keywords: Linguistic approximation, numerical experiment, distance-based methods, graphical analysis, Mamdani, fuzzy.

JEL classification: D81, C44
AMS classification: 90B50, 91B06

1 Introduction

Mathematical models for decision support and mathematical models representing expert knowledge frequently provide outputs in the form of fuzzy numbers or intervals (see e.g. [3, 7, 11, 15, 22]). In such cases when uncertainty is present, it might be convenient to provide the decision makers also with linguistic summaries of these results. Linguistic fuzzy modelling applying appropriate linguistic approximation (see e.g. [2, 6, 12, 25, 27] for some linguistic approximation techniques examples) can thus help to enhance the understandability of the outputs. Since the process of linguistic approximation can distort the information that is being approximated - note that the most fitting linguistic label from a usually small set of available well understood labels is assigned - a thorough investigation of the process of LA is needed. The research in this area has already started from the theoretical [4, 26] and behavioral [13, 16, 24] perspective and also from the perspective of the performance of LA methods in various contexts [18, 23, 19, 20]. Also alternative uses of LA were studied recently [14, 17, 21]. Most of the recent studies of LA methods and their applicability focus on the approximation of rather simple objects - i.e. fuzzy numbers (and frequently of the triangular or rectangular type). More general types of outputs, such as the outputs of Mamdani fuzzy inference [8] remain unaddressed. This paper strives to suggest the first step toward the analysis of the performance of LA methods in connection with Mamdani-type outputs.

2 Preliminaries

Let U be a nonempty set (the universe of discourse). A *fuzzy set* A on U is defined by the mapping $A : U \to [0, 1]$. A family of all fuzzy sets on U is denoted by $\mathcal{F}(U)$. 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. $\text{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an α*-cut* of A for any $\alpha \in [0, 1]$, $\text{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A. Let A and B be fuzzy sets on the same universe U. We say that A is a *fuzzy subset* of B ($A \subseteq B$), if $A(x) \leq B(x)$ for all $x \in U$.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\text{Ker}(A) \neq \emptyset$ (A is *normal*); (2) A_α are closed intervals for all $\alpha \in (0, 1]$ (this implies A is *unimodal*); (3) $\text{Supp}(A)$

[1]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic, tomas.talasek@upol.cz.
[2]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic and Lappeenranta University of Technology, School of Business and Management, Skinnarilankatu 34, Lappeenranta, Finland, jan.stoklasa@upol.cz.

is bounded. A family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on [a,b], if Supp(A) is a subset of the interval $[a, b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \text{Cl(Supp}(A))$ and $[a_2, a_3] = \text{Ker}(A)$, where Cl(Supp($A$)) denotes a closure of Supp(A). Each fuzzy number A is determined by $A = \left\{[\underline{a}(\alpha), \overline{a}(\alpha)]\right\}_{\alpha\in[0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall\alpha \in (0, 1]$, and the closure of the support of A Cl(Supp(A)) $= [\underline{a}(0), \overline{a}(0)]$. The *cardinality* of fuzzy number A on $[a, b]$ is a real number Card(A) defined as follows: Card(A) $= \int_a^b A(x)dx$. A *union* of two fuzzy sets A and B on U is a fuzzy set $(A \cup B)$ on U defined as folllows: $(A \cup B)(x) = \min\{A(x), B(x)\}$ and a *Łukasiewicz union* of two fuzzy sets A and B on U is a fuzzy set $(A \cup_L B)$ on U defined as follows: $(A \cup_L B)(x) = \min\{1, A(x) + B(x)\}$, $\forall x \in U$. Let $A_1, \ldots, A_n$ be a fuzzy sets on $U_1, \ldots, U_n$ respectively. The *Cartesian product* of $A_1, \ldots, A_n$ is a fuzzy set $(A_1 \times \cdots \times A_n)$ on $U_1 \times \cdots \times U_n$ with membership function $(A_1 \times \cdots \times A_n)(x_1, \ldots, x_n) = \min\{A_1(x_1), \ldots, A_n(x_n)\}$, $\forall x_i \in U_i$. A fuzzy set R on $U_1 \times \cdots \times U_n$ is called an *n*-ary fuzzy relation. Let R be a fuzzy relation on $U \times V$ and S be a fuzzy relation on $V \times W$. The composition $(R \circ S)$ is a fuzzy set on $U \times W$ a with membership function $(R \circ S)(x, z) = \sup_{y\in V} \min\{R(x, y), S(y, z)\}$, $\forall x \in U, z \in W$.

The fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$ and $a_1 \neq a_4$; for such fuzzy numbers we will use a simplified notation $A = (a_1, a_2, a_3, a_4)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by ordered triplet $A = (a_1, a_2, a_4)$. Triangular fuzzy number $A = (a_1, a_2, a_4)$ is called *symmetric triangular fuzzy number* if $a_2 - a_1 = a_4 - a_2$. More details on fuzzy numbers and computations with them can be found for example in [5].

A *fuzzy scale* on $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on [a,b], that form a Ruspini fuzzy partition (see [10]) of the interval $[a, b]$, i.e. for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. A *linguistic variable* ([27]) is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values (terms), X is an universe on which the meanings of the linguistic values are defined, G is an syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X. Linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ is called a *linguistic scale* on $[a, b]$ if $X = [a, b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ and $T_i = M(\mathcal{T}_i)$, $i = 1, \ldots, n$ form a fuzzy scale on $[a, b]$. Fuzzy scale is called *uniform* when Card(Supp(T_i)) $= 2 \cdot (b - a)/(n - 1)$ for all $i = 2, ..., n - 1$, Card(Supp(T_i)) $= (b - a)/(n - 1)$ for $i = 1$ and $i = n$, T_i form a Ruspini fuzzy partition of U, and $T_2, \ldots, T_{n-1}$ are symmetrical triangular fuzzy numbers. Linguistic terms $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ of linguistic scale $\mathcal{T}(\mathcal{V})$ are called *elementary (level 1) terms* of linguistic scale. Linguistic scale using additional linguistic terms $\mathcal{T}_i$ to $\mathcal{T}_j$ where $i = 1, \ldots, n - 1$, $j = 2, \ldots, n$ and $i < j$ (called *derived linguistic terms*) is called *extended linguistic scale*; $M(\mathcal{T}_i \text{ to } \mathcal{T}_j) = T_i \cup_L T_{i+1} \cup_L \cdots \cup_L T_j$. The extended linguistic scale thus contains linguistic values of different levels of uncertainty – from the possibly least uncertain elementary terms $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ to the most uncertain linguistic term $\mathcal{T}_1$ to $\mathcal{T}_n$ (uncertainty can be assessed by the cardinality of the meanings of these linguistic terms). Derived linguistic terms $\mathcal{T}_i$ to $\mathcal{T}_j$ are called *level* $j - i + 1$ *terms* and can be also denoted by $\mathcal{T}_{ij}$.

Let Out be a fuzzy number on $[a, b]$ and $(\mathcal{V}, \mathcal{T}(\mathcal{V}), [a, b], G, M)$ be a linguistic variable such that $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$ and $T_i = M(\mathcal{T}_i)$, $i = 1, \ldots, s$, are fuzzy numbers on $[a, b]$. The *linguistic approximation* of fuzzy number Out is a process of searching for a suitable linguistic term $\mathcal{T}_{Out}$ from $\mathcal{T}(\mathcal{V})$ which describes the meaning of the fuzzy number Out the best. One of the most popular approaches to finding the linguistic term $\mathcal{T}_{Out}$ is using the "best-fit" approach:

$$T_{Out} = \arg \min_{i\in\{1,\ldots,s\}} d(T_i, Out), \tag{1}$$

where $d(A, B)$ is a distance measure[3] of two fuzzy numbers A and B.

3 Definition of a problem

As was outlined in the introduction, our investigation aims on the Mamdani-type outputs and their linguistic approximation. More specifically we aim on the linguistic approximation of the outputs of Mamdani fuzzy inference. Let us consider m linguistic scales $(\mathcal{V}_j, \mathcal{T}(\mathcal{V}_j), X_j, G_j, M_j)$, $j = 1, \ldots, m$, representing the inputs of the fuzzy inference system and an output linguistic scale $(\mathcal{W}, \mathcal{T}(\mathcal{W}), Y, G_{\mathcal{W}}, M_{\mathcal{W}})$. Let us also consider a collection of n rules representing the relationships between the input and output variables in the form:

If $\mathcal{V}_1$ is $\mathcal{A}_{i1}$ and $\ldots$ and $\mathcal{V}_m$ is $\mathcal{A}_{im}$, then $\mathcal{W}$ is $\mathcal{B}_i$,

where $\mathcal{A}_{j1} \in \mathcal{T}(\mathcal{V}_j)$ and $\mathcal{B}_i \in \mathcal{T}(\mathcal{W})$, $M_j(\mathcal{A}_{ij}) = A_{ij}$ and $M_{\mathcal{W}}(\mathcal{B}_i) = B_i$ for $i = 1, \ldots, n$, $j = 1 \ldots, m$. The output Out of Mamdani fuzzy inference computed for the input $(A'_1 \times \cdots \times A'_m)$, $A'_j \in \mathcal{F}(X_j), j = 1, \ldots, m$,

[3] Alternatively a similarity measure of two fuzzy numbers can be used. In this case, the $\arg\min$ function in formula (1) is replaced by $\arg\max$.

using the fuzzy rule base R consisting of n fuzzy rules, $R = \bigcup_{i=1}^{n}(A_{i1} \times \cdots \times A_{im} \times B_i)$ is computed by (2).

$$Out = (A'_1 \times \cdots \times A'_m) \circ \bigcup_{i=1}^{n}(A_{i1} \times \cdots \times A_{im} \times B_i). \quad (2)$$

A sample output of this type is presented in Figure 1. Clearly Out does not need to be a convex fuzzy set any more. In this paper only normal Mamdani-type outputs are considered. The convexity of Out is not required. To find the linguistic approximation of Out, two distances of fuzzy sets that do not require convexity will be compared in terms of their performance in LA:

- *modified Bhattacharyya distance* [1]:

$$d_1(A, B) = \left[1 - \int_a^b (A^*(x) \cdot B^*(x))^{1/2} dx\right]^{1/2}, \quad (3)$$

  where $A^*(x) = A(x)/\text{Card}(A)$ and $B^*(x) = B(x)/\text{Card}(B)$,
- *Fuzzy distance* [9]:

$$d_2(A, B) = \frac{\int_a^b \mid A(x) - B(x) \mid}{\text{Card}(A) + \text{Card}(B)}. \quad (4)$$

4 Numerical analysis and discussion of the results

The numerical analysis of the performance of d_1 and d_2 in linguistic approximation of Mamdani-type outputs will be carried out under the assumption of a 5-element uniform linguistic scale representing the meanings of the elementary terms of the output variable by triangular fuzzy numbers $\{B_1, \ldots, B_5\} = \{(0, 0, 0.25), (0, 0.25, 0.5), (0.25, 0.5, 0.75), (0.5, 0.75, 1), (0.75, 1, 1)\}$ as presented in Figure 1. This assumption results in no loss of generality, as the results are generalisable for any number of elementary terms larger than 4, as long as their meanings form a uniform Ruspini partition on the given universe. More specifically an extended scale defined using elementary linguistic terms $\{\mathcal{B}_1, \ldots, \mathcal{B}_5\}$ is assumed for the purposes of LA. We also restrict our investigation to the linguistic approximation of the outputs the type $Out = h_L \cdot B_{k-1} \cup B_k \cup h_R \cdot B_{k+1}$ for $h_L, h_R \in [0, 1]$ and $k = 3, \ldots, t-2$, where t is the number of elementary terms in $\mathcal{W}$.

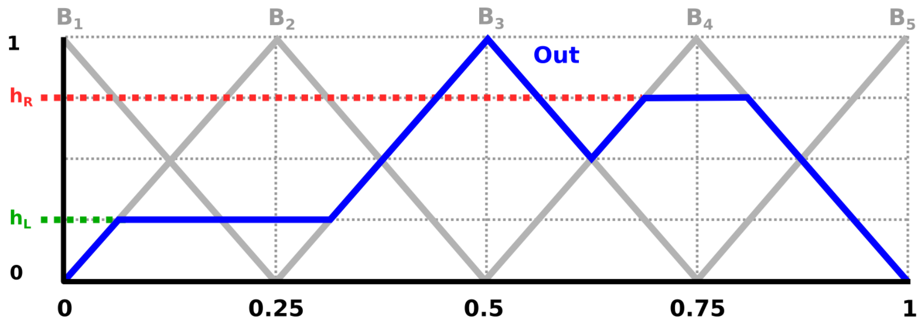


Figure 1: An example of a Mamdani-type output. This type of outputs of Mamdani fuzzy inference models is considered to be linguistically approximated in this paper. $B_1, \ldots, B_5$ are the meanings of the elementary linguistic values of the output linguistic variable used for the linguistic approximation.

For the purpose of systematic investigation of the behavior of LA under fuzzy distances d_1 and d_2, 501 uniformly distributed values of h_L and h_R from interval $[0, 1]$ were generated. This way, 251 001 Mamdani-type fuzzy sets $\{Out_1, \ldots, Out_{251001}\}$ were generated and linguistically approximated. Note, that each fuzzy set Out_i, $i = 1, \ldots, 251\,001$ can be unambiguously described by a 2-tuple (h_L, h_R). The results of the linguistic approximation using fuzzy distances d_1 and d_2 are depicted in Figures 2 and 3 respectively. The result of the linguistic approximation (i.e. the resulting element of the extended scale) for each approximated Mamdani-type output represented as a point with coordinates (h_L, h_R), is represented by a specific color.

It can by clearly seen from Figures 2 and 3 that although both fuzzy distances assign only four linguistic terms $\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}$ and $\mathcal{B}_{24}$, the results of LA are significantly different for each fuzzy distance. The Bhattacharyya distance favors linguistic terms the meanings of which are supersets to the linguistically approximated fuzzy set Out_i, $i = 1, \ldots, 251\,001$; only 1.33% of fuzzy sets are linguistically approximated by the elementary term $\mathcal{B}_3$ using the Bhattacharyya distance d_1 (in the case of fuzzy distance d_2 it is more than 15%). Moreover, when the value of h_L or h_R is higher than 0.13, the term $\mathcal{B}_3$ is never assigned based on d_1. Higher uncertainty of Out thus implies that an at least level-2 term will be assigned as a linguistic approximation using the Battacharyya distance. Also linguistic terms $\mathcal{B}_{23}, \mathcal{B}_{34}$ are not used often – each of them is used in less than 9% of the cases. That is caused by the fact, that when both h_L and h_R are higher than 0.135 simultaneously, the only possible outcome of LA is $\mathcal{B}_{24}$ (i.e. a level-3 term). Actually, more than 80.7% of the approximated fuzzy sets are approximated by the term $\mathcal{B}_{24}$. This only confirms the findings obtained in [20] that Bhattacharyya distance tends to suggest such approximating labels that tend to be supersets of the approximated object meaning-wise.

On the contrary the fuzzy distance d_2 divided the space of the approximated fuzzy numbers Out_i, $i = 1, \ldots, 251\,001$ into four rectangular-like areas (see Figure 3) with respect to the result of the linguistic approximation. Note that all four possible outcomes of the linguistic approximation are suggested with similar frequencies. Linguistic terms $\mathcal{B}_{23}, \mathcal{B}_{34}$ are both assigned to almost 25% of the approximated fuzzy sets. Linguistic term $\mathcal{B}_{24}$ (the most uncertain term that is used) is assigned to 35% of approximated fuzzy sets. Note that for $h_L = h_R \approx 0.366$ we obtain a fuzzy set that can be linguistically approximated by all four linguistic terms $\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}$ and $\mathcal{B}_{24}$ (the fuzzy distance d_2 between this fuzzy set and the meaning of each of the four linguistic terms considered is the same). This could be a potential handicap of this fuzzy distance due to the fact, that just a small change of h_L and/or h_R around this point could result in a different linguistic label from the set $\{\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}, \mathcal{B}_{24}\}$. Only limited number of approximated fuzzy sets results into ambiguous cases, when the LA was unable to assign proper linguistic term (10 fuzzy sets in the case of d_1 and 22 in the case of d_2, see Table 1). These ambiguous cases are depicted in the Figures 2 and 3 by yellow color. Note, that in fact the borders between the pairs of differently colored areas are always constituted by ambiguous cases. Due to the chosen mesh just some of the ambiguous cases manifested themselves in the numerical analysis.

Thus if the meaning of the linguistic approximation is required to be a superset to the approximated fuzzy set, it is more reasonable to choose Bhattacharyya distance d_1 over d_2. The requirement of supersets is a “safe” approach to linguistic approximation, since the reduction of uncertainty is not significant (in fact the approximating fuzzy set is frequently more uncertain). This can, however, lead to the situations where most of the approximated fuzzy sets will be labeled by the same linguistic term. The fuzzy distance d_2 uses the available linguistic terms of the approximating extended linguistic scale more uniformly, which results in low-uncertain outputs being approximated by a low uncertain linguistic term (but under a larger risk that a part of the approximated fuzzy output will not be covered well by the selected linguistic approximation, i.e. the intersection of the meaning of the approximating term and the output will be nonempty).

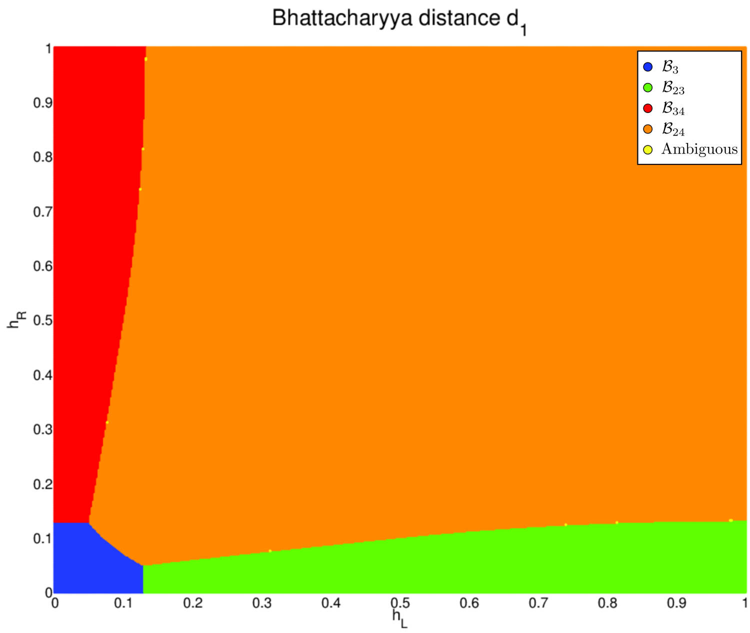


Figure 2: The results of the linguistic approximation of fuzzy sets $\{Out_1, \ldots, Out_{251001}\}$ using Bhattacharyya distance d_1. Approximated Mamdani-type fuzzy sets are represented by points with coordinates (h_L, h_R). The color of each point represents the linguistic approximation of the corresponding fuzzy set. Ambiguous cases (when more than one linguistic term is suggested based on d_1) are depicted using yellow color.

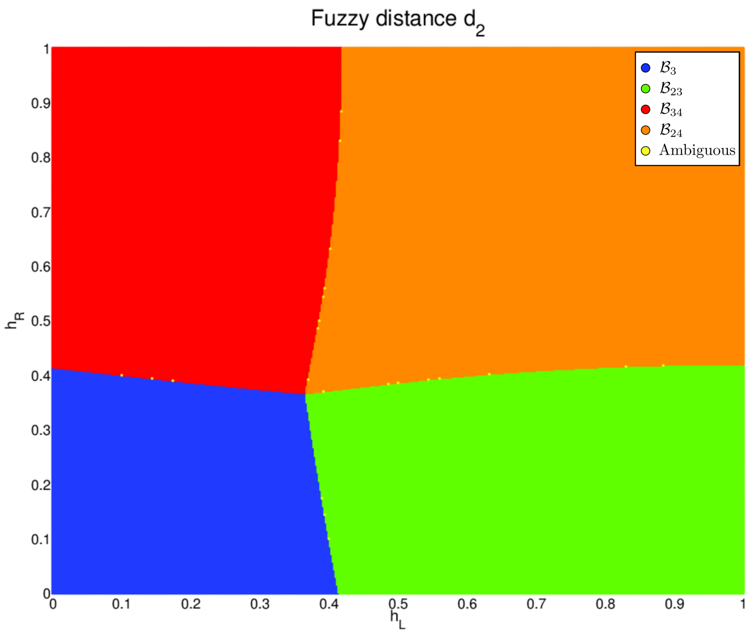


Figure 3: The results of the linguistic approximation of fuzzy sets $\{Out_1, \ldots, Out_{251001}\}$ using Fuzzy distance d_2. Approximated Mamdani-type fuzzy sets are represented by points with coordinates (h_L, h_R). The color of each point represents the linguistic approximation of the corresponding fuzzy set. Ambiguous cases (when more than one linguistic term is suggested based on d_2) are depicted using yellow color.

	$\mathcal{B}_3$	$\mathcal{B}_{23}$	$\mathcal{B}_{34}$	$\mathcal{B}_{24}$	**Ambiguous**
d_1	3 329	22 541	22 541	202 580	10
d_2	37 948	61 806	61 806	89 419	22

Table 1: Frequencies of assignment of linguistic terms $\{\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}, \mathcal{B}_{24}\}$ and Ambiguous cases obtained for fuzzy sets $\{Out_1, \ldots, Out_{251001}\}$ in the linguistic approximation using d_1 and d_2. Unlisted linguistic terms (e.g. $\mathcal{B}_1, \mathcal{B}_{13}, \ldots$) were not assigned to any approximated Mamdani-type fuzzy set.

5 Conclusion

Mamdani-type fuzzy sets are most frequently obtained through Mamdani fuzzy inference systems. These systems are represented by a set of fuzzy IF-THEN rules which can be formulated linguistically. As such these systems present a useful tool for the descrtiption and investigation of economical phenomena and rather complex systems. This paper investigates the performance of linguistic approximation of Mamdani-type outputs [8] under two different fuzzy distance measures. The results of this numerical analysis clearly show, that the choice of the fuzzy distance measure can significantly affect the results of linguistic approximation. In the paper we have identified several important properties that could help the users of mathematical models choose the most suitable fuzzy distance measure. First, although it is not frequently studied in the literature, linguistic approximation of Mamdani-type fuzzy sets can be done, even within the "best-fit" LA context. Second, performance of the LA can be customized by an appropriate choice of the distance measure - d_1 is more prone to suggest a superset meaning-wise, d_2 suggests LA closer in terms of cardinality. This paper is the initial step of exploring the results of LA of Mamdani-type outputs under different fuzzy distance/similarity measures.

Acknowledgements

Supported by the grant No. IGA_FF_2017_011 of the Internal Grant Agency of Palacký University Olomouc.

References

[1] Aherne, F., Thacker, N., and Rockett, P.: The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data. *Kybernetika* **32** (1998), 363–368.

[2] Bonissone, P. P.: A fuzzy sets based linguistic approach: theory and applications. In: *Proceedings of the 12th conference on Winter simulation*, 1980. 99–111.

[3] 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** (2014), 437–455.

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

[5] Dubois, D., and Prade, H., eds.: *Fundamentals of Fuzzy Sets*. Kluwer Academic Publishers, Massachusetts, 2000.

[6] Eshragh, F., and Mamdani, E. H.: A general approach to linguistic approximation. *International Journal of Man-Machine Studies* **11** (1979), 501–519.

[7] Krejčí, J., and Stoklasa, J.: Fuzzified AHP in the evaluation of scientific monographs. *Central European Journal of Operations Research* **24** (2016), 353–370.

[8] Mamdani, E., and Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies* **7** (1975), 1 – 13.

[9] Pappis, C. P., and Karacapilidis, N. I.: A comparative assessment of measures of similarity of fuzzy values. *Fuzzy Sets and Systems* **56** (1993), 171 – 174.

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

[11] Stoklasa, J.: Multiphase linguistic fuzzy model for the Czech emergency medical rescue services. In: *Proceedings of the 28th International Conference on Mathematical Methods in Economics*, 2010. University of South Bohemia in České Budějovice, 590–595.

[12] Stoklasa, J.: *Linguistic models for decision support*. Lappeenranta University of Technology, Lappeenranta, 2014.

[13] Stoklasa, J., and Talášek, T.: Jazykově orientované modelování pro ekonomickou a manažerskou praxi některé otevené problémy. In: *Znalosti pro tržní praxi 2015*. 959–969.

[14] Stoklasa, J., and Talášek, T.: On the use of linguistic labels in AHP: calibration, consistency and related issues. In: *Proceedings of the 34th International Conference on Mathematical Methods in Economics*. Technical University of Liberec, Liberec, 785–790.

[15] Stoklasa, J., Talášek, T., Kubátová, J., and Seitlová, K.: Likert scales in group multiple-criteria evaluation. *Journal of Multiple-Valued Logic and Soft Computing* (2017), (in press).

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

[17] Stoklasa, J., Talášek, T., and Stoklasová, J.: Semantic differential and linguistic approximation - identification of a possible common ground for research in social sciences. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016*. Societas Scientiarum Olomucensis II, Olomouc, 495–501.

[18] Talášek, T., and Stoklasa, J.: Jazykově orientované modelování pro ekonomickou a manažerskou praxi - vybrané metody pro jazykovou aproximaci. In: *Znalosti pro tržní praxi 2015*. 977–989.

[19] Talášek, T., and Stoklasa, J.: The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016*. Societas Scientiarum Olomucensis II, Olomouc, 539–546.

[20] Talášek, T., and Stoklasa, J.: A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation Under Different Linguistic Scales. *Journal of Multiple-Valued Logic and Soft Computing* (2017), (in press).

[21] Talášek, T., Stoklasa, J., Collan, M., and Luukka, P.: Ordering of fuzzy numbers through linguistic approximation based on Bonissone's two step method. In: *16th IEEE International Symposium on Computational Intelligence and Informatics*. 285–290.

[22] Talášek, T., Stoklasa, J., and Talašová, J.: Linguistic approximation using fuzzy 2-tuples in investment decision making. In: *Proceedings of the 33rd International Conference on Mathematical Methods in Economics*. 817–822.

[23] Talášek, T., Stoklasa, J., and Talašová, J.: The role of distance and similarity in Bonissone's linguistic approximation method a numerical study. In: *Proceedings of the 34th International Conference on Mathematical Methods in Economics*. Technical University of Liberec, Liberec, 845–850.

[24] Talašová, J., Stoklasa, J., and Holeček, P.: HR management through linguistic fuzzy rule bases - a versatile and safe tool? In: *Proceedings of the 32nd International Conference on Mathematical Methods in Economics*. 1027–1032.

[25] Wenstøp, F.: Quantitative analysis with linguistic values. *Fuzzy Sets and Systems* **4** (1980), 99–115.

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

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

Publication IX

Taláček, T. and Stoklasa, J.
Selection of tools for managerial decision support – the identification of methods of choice in linguistic approximation

Reprinted with the permission from
Proceedings of the international scientific conference Knowledge for Market Use 2017,
pp. 606–613, 2017,

SELECTION OF TOOLS FOR MANAGERIAL DECISION SUPPORT – THE IDENTIFICATION OF METHODS OF CHOICE IN LINGUISTIC APPROXIMATION

TOMÁŠ TALÁŠEK, JAN STOKLASA

Palacký University Olomouc, Faculty of Arts, Department of Applied Economics; Palacký University Olomouc, Faculty of Arts, Department of Applied Economics AND Lappeenranta University of Technology, School of Business and Management

Abstract: *Linguistic approximation is a mathematical tool that transforms outputs of mathematical models (in the form of real numbers, intervals, fuzzy sets, etc.) into natural language. It therefore allows for the presentation of results of decision-support models in natural language in the final step of decision support. This is a crucial step especially for complex mathematical models whose results can be hard-to-interpret for non-experienced users (managers, laymen). For them the linguistic approximation provides easy-to-understand alternative outputs of mathematical models. The paper strives to suggest an analytical framework to select the appropriate method for linguistic approximation based on numerical experimentation and graphical summaries of outputs. One distance measure of fuzzy sets is selected to show the applicability of the proposed analytical framework. The goal is to provide managers and practitioners in economics and finance with results they can understand and apply.*

Keywords: *Linguistic approximation, fuzzy ideal, fuzzy number, dissemblance index, distance.*

JEL classification: *C44*

Grant affiliation: *This research was supported by the grant IGA FF 2017 011 Continuities and Discontinuities of Economy and Management in the Past and Present 3 by the internal grant agency of Palacký University Olomouc.*

1. Introduction

Linguistic approximation is a process of transforming mathematical outputs (i.e. numbers, intervals, fuzzy numbers) into a more natural form understandable by managers/laymen - natural language. This step may be crucial especially in some decision-support models, where the outputs of mathematical models are complex and cannot be interpreted by non-experienced users without the risk of misunderstanding or misinterpretation. In these cases, a proper linguistic label is often more understandable than numbers. In other words, linguistic summaries might be better understandable for laymen users and hence more appropriate for them - even though the transition into natural language might introduce some uncertainty and/or bias into the situation (Stoklasa, Talášek and Musilová). However, the outcome of linguistic approximation may depend heavily on the selection of approximating method or the distance/similarity measure within the method. Therefore it is necessary to examine the appropriateness of the selected linguistic approximation methods for the selected decision-support model in sufficient detail. Studies in this topic have started to appear recently (Talášek and Stoklasa, 2016; Talášek and Stoklasa, 2017; Talášek, Stoklasa and Talašová, 2016) , none of them, however, addresses directly the issue of the definition of an ideal and its possible impact on the appropriateness of linguistic approximation methods or their results. It is the aim of this paper to fill this gap with a suggestion of an analytical framework and a case study of one linguistic approximation method under a specific fuzzy distance measure.

More specifically, we introduce the analytical framework for the examination of behaviour of the *best-fit* linguistic approximation method, which is based on distance/similarity measures (Stoklasa, 2014; Yager, 2004). In contrast with other approaches to the analysis of linguistic approximation methods (Talášek and Stoklasa, 2016; Talášek and Stoklasa, 2017), this paper does not consider linguistic variables or linguistic scales (Zadeh, 1975) to provide values for the linguistic approximation. Instead, it focuses on the appropriateness of assigning a single, but very relevant and frequently used linguistic label: "*THE BEST*" *or "IDEAL"* (i.e. it considers such situations, where an ideal is defined and used as a benchmark). As such, the results are relevant not only for the purpose of the analysis of appropriateness of particular fuzzy distances or similarities in linguistic approximation, but also in the context of evaluation based on the distance from ideal (such as e.g. TOPSIS). For the simplicity and without loss of generality, the framework will be illustrated on one of the popular distances used in linguistic approximation - the *dissemblance index (*Kaufman and Gupta, 1985).

2. Preliminaries

Let U be a nonempty set (the universe of discourse). 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 $A(.)$ is called a membership function of the fuzzy set A. $\mathrm{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a kernel of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an α-cut of A for any $\alpha \in [0,1]$, $\mathrm{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a support of A. Let A and B be fuzzy sets on the same universe U. We say that A is a fuzzy subset of $(A \subseteq B)$, if $A(x) \leq B(x)$ for all $x \in U$.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\mathrm{Ker}(A) \neq \emptyset$ (A is normal); (2) A_α are closed intervals for all $\alpha \in (0,1]$ (this implies A is unimodal); (3) $\mathrm{Supp}(A)$ is bounded. A fuzzy number A is said to be defined on $[a,b]$, if $\mathrm{Supp}(A)$ is a subset of an interval $[a,b]$. The real numbers $a^1 \leq a^2 \leq a^3 \leq a^4$ are called significant values of the fuzzy number A if $[a^1, a^4] = \mathrm{Cl}\big(\mathrm{Supp}(A)\big)$ and $[a^2, a^3] = \mathrm{Ker}(A)$, where $\mathrm{Cl}(\mathrm{Supp}(A))$ denotes a closure of $\mathrm{Supp}(A)$. Each fuzzy number A can be represented as $A = \{\underline{a}(\alpha), \overline{a}(\alpha)\}_{\alpha \in [0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall \alpha \in (0,1]$, and $[\underline{a}(0), \overline{a}(0)] = \mathrm{Cl}(\mathrm{Supp}(A))$. The centre of gravity of a fuzzy number A on $[a,b]$ (if $a < b$) is defined by the formula $\mathrm{COG}(A) = \int_a^b x\, A(x)\ dx / \mathrm{Card}(A)$.

The fuzzy number A is called triangular if its membership function is linear on $[a^1, a^2]$ and $[a^3, a^4]$ and $a^2 = a^3$. Triangular fuzzy numbers will be denoted by $A = (a^1, a^2, a^4)$. A triangular fuzzy number $A = (a^1, a^2, a^4)$ is called symmetric if $a^2 - a^1 = a^4 - a^2$. Otherwise it is called assymetric. More details on fuzzy numbers and computations with them can be found for example in Dubois and Prade (1980).

3. Analytical framework to select the appropriate method for linguistic approximation

The process of linguistic approximation of the output O of a mathematical model can be divided into two steps – in the first step, the set of proper linguistic terms/labels (words in natural language) describing the possible outputs of models is selected and their meaning (usually in the form of fuzzy numbers) is established. The crucial part of this process is that the future users of the model are part of this process. In the second step, the linguistic approximation method is selected and the most suitable linguistic label is selected (from the set of linguistic terms from first part). In our case, a single label "IDEAL" is considered with a meaning represented by a triangular fuzzy number F on $[0,1]$, $F = (k, 1,1)$, where $k \in [0,1)$.

One of the most common linguistic approximation methods is the so called *best-fit approach* which employs distance/similarity measure of fuzzy sets. In this approach the distances (or similarity) between the output O and the meaning of each possible linguistic label are computed and then the label which meaning is closest (or is the most similar in the case of similarity measure) to the output O is selected as a result from the linguistic approximation (in cases when more than one label are selected, further investigation must be done).

The choice of proper distance/similarity method for the best-fit approach in the second step is crucial for the results of linguistic approximation. Different measures have different properties and (from the practical point of view) suggest different types of linguistic labels. Therefore it is necessary to properly investigate the possible impact of the selected distance/similarity measure on our model and its outputs. The presented framework examines the distance of different fuzzy numbers (outputs of models) from the fuzzy number F representing the "ideal value" of the output of mathematical model. This approach differs from other approaches (i.e. Talášek and Stoklasa, 2016; Talášek and Stoklasa, 2017) which usually focus on the final result of linguistic approximation, not the distance from one particular linguistic term. This approach, however, enables to study the effect of selected distance/similarity measure more deeply and offers also evaluation-oriented interpretation. Note, that the distance from the ideal solution can be used to define the best alternative or to order the alternatives in terms of their closeness to the ideal (provided that the distance/similarity measure chosen for this purpose is a suitable one). In the evaluation framework, a reasonable requirement for the distance measure DM could be, that if an alternative A is clearly better than alternative B, then $DM(A, F) \leq DM(B, F)$. A violation of such assumption could indicate either a wrong choice of the distance measure, or a wrong definition of the ideal and its representation F.

3.1. Analytical framework

The presented analytical framework is based on graphical representation of the values of distances between different triangular symmetric fuzzy numbers and a fuzzy ideal F representing an "ideal evaluation" or a "most desired value". All fuzzy numbers are defined on the $[0,1]$ interval (restriction on $[0,1]$ interval is chosen for the reader's convenience, however the framework can be used on any interval without the loss of generality) . The fuzzy ideal in our framework is a triangular asymmetric fuzzy number $F = (k, 1,1)$, where $k \in [0, 1]$. Usually the values of k are chosen as 0.9, 0.95 or 0.99 and therefore the fuzzy ideal represents a slightly uncertain fuzzy number close to 1. The concept of fuzzy ideal is well known and used in decision making methods such as TOPSIS (see e.g. Collan, Fedrizzi, and Luukka, 2013) or fuzzy MCDM methods (see e.g. Stoklasa, Talášek and Luukka, 2018; Stoklasa, Talášek, Kubátová and Seitlová, 2017).

One of the reasons why symmetric triangular fuzzy numbers are chosen is because they can be unambiguously identified by an ordered 2-tuple representing their centre of gravity and the length of their support. Symmetric triangular fuzzy numbers represent uncertain quantities or imprecise measurements and as such are frequent objects in mathematical models. In this paper m uniformly distributed symmetric triangular fuzzy numbers $O_i, i = 1, \dots, m$ representing the possible outputs of mathematical model are generated analogously to Talášek and Stoklasa (2017). Cartesian product of n uniformly distributed values of centre of gravity from $[0,1]$ interval and n uniformly distributed values of the length of support from $[0,1]$ is obtained. This Cartesian product is a set of n^2 2-tuples representing symmetric triangular fuzzy numbers. However, the supports of some of these fuzzy numbers are not subsets of the $[0,1]$ interval and therefore these fuzzy numbers must be removed from the set; the remaining 2-tuples represent the fuzzy numbers $O_i, i = 1, \dots, m$.

After the fuzzy ideal F is chosen and m symmetric triangular fuzzy numbers are generated, the investigated distance/similarity measure is selected and the distance/similarity between F and each $O_i, i \in 1, \dots, m$ is computed. Then the results are plotted in a 3D graph, where each point represents the distance (z-axis) of point O_i unambiguously represented by the centre of gravity (x-axis) and length of its support (y-axis) from the fuzzy ideal F. Moreover the colour of each point represents the distance of this point from fuzzy ideal. Due to this property, the graph can be plotted from the TOP view – i.e. the graph could be constructed as 2D graph where z-axis is represented by colours. However, it is more convenient to use both graphs sidewise instead of choosing only one of them.

After the graphs are constructed, the author of the model should inspect the properties of the distance/similarity measure and then decide if it is reasonable to use this measure in this particular mathematical model. How to inspect the properties is shown in the next chapter using a numerical example.

4. Numerical example and discussion

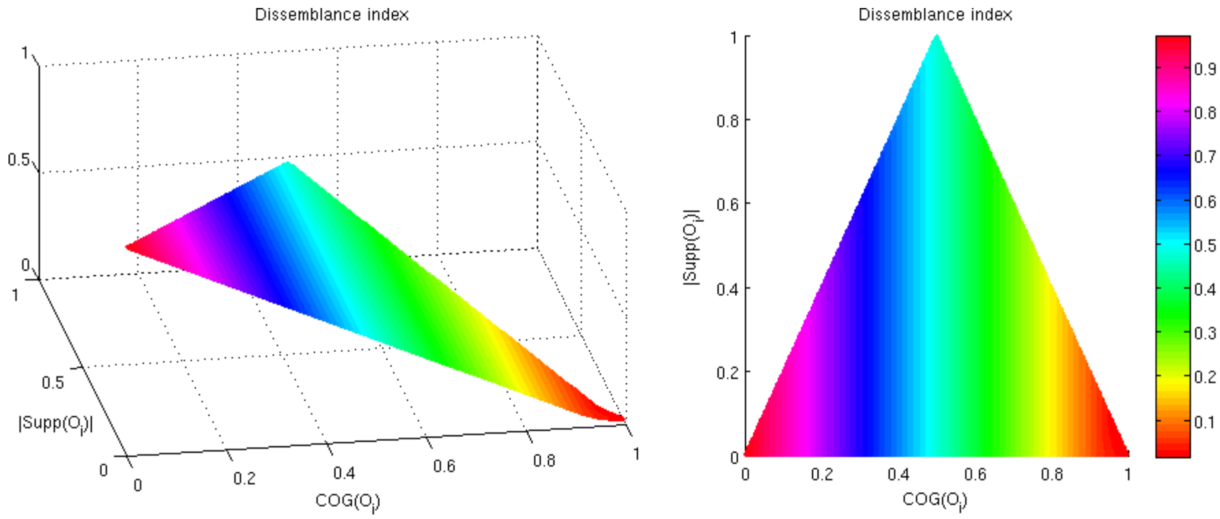
In this chapter, the usage of the proposed framework will be shown. For this purpose, the fuzzy ideal F and the distance measure that will be investigated must be chosen. The distance measure called dissemblance index (Kaufman and Gupta, 1985) was chosen as a representative measure, analogical analysis can be done for other distance/similarity measures of fuzzy numbers. The dissemblance index d of fuzzy numbers A and B is defined as:

$$d(A, B) = \int_0^1 |\underline{a}(\alpha) - \underline{b}(\alpha)| + |\overline{a}(\alpha) - \overline{b}(\alpha)| d\alpha.$$

Fuzzy ideal $F = (0.9,1,1)$ was chosen due to the fact, that this ideal is more uncertain and therefore more reflect fuzziness. Finally, 80 000 uniformly distributed symmetrical triangular fuzzy numbers $O_i, i \in 1, \dots, 80\,000$, is generated and the distance between each O_i and F is computed.

The result of this approach is depicted in Figure 1. From the right subfigure of Figure 1 can be clearly seen, that the length of the support does not affect the distance from the ideal F, i.e. the distance from F is dependent on the COG only (note, that this is the result of the use of symmetrical triangular fuzzy numbers). From the left subfigure we can clearly see, that the distance of O_i from the fuzzy ideal F, measured by the dissemblance index, changes linearly with respect to the centre of gravity of the investigated fuzzy number. The linearity is disrupted only for fuzzy numbers whose centre of gravity is situated close to 1 (see the red part of the left subfigure where the linearity is disrupted).

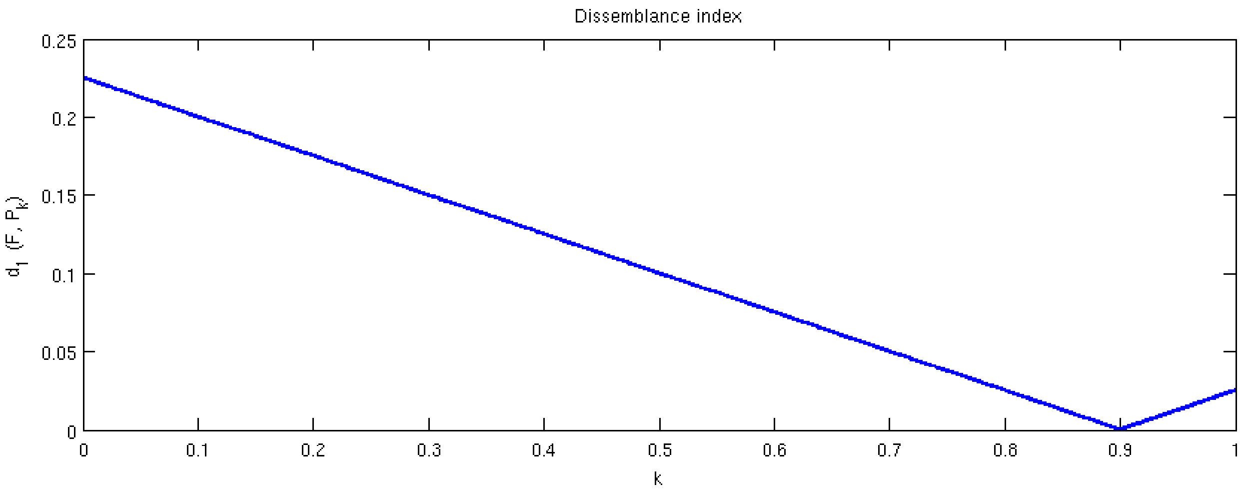
FIG. 1: Results from the numerical example, where the distance between 80 000 symmetric triangular fuzzy numbers $O_i, i = 1, \dots, 80\,000$ and fuzzy ideal $F = (0.9, 1, 1)$ is depicted. Each point O_i is represented by its centre of gravity (x-axis) and the length of support (y-axis). The distances between points are represented by colours and in the case of the left subgraph also by the z-axis.



To investigate the disruption of linearity, an additional analysis was performed. The symmetric triangular fuzzy numbers O_i were replaced by 1 000 asymmetrical triangular fuzzy numbers $\{P_0, P_{0.001}, P_{0.002}, \dots, P_{0.999}, P_1\}$, where $P_k = (k,1,1)$. In fact, these fuzzy numbers represents different alternative definitions of fuzzy ideals. The distance d between these fuzzy numbers and the fuzzy ideal was computed and the result is plotted in Figure 2.

From Figure 2 we can clearly see that the closer the value of k is to the value 0.9, the lower the distance between P_k and F is. Please note, that for the $k = 0.9$ the distance is equal to 0. This is expected, because in this case the fuzzy number $P_{0.9}$ is equivalent to F. However, if the value of k is higher than 0.9, the distance will become positive again. This can be counterintuitive, because now the fuzzy number P_k is a subset of F, i.e. the fuzzy number P_k is even closer to number 1 (real-valued ideal result) than the fuzzy ideal F but it is evaluated as having a nonzero distance from the fuzzy ideal (in other words worse than the fuzzy ideal). We can thus conclude that the dissemblance index works well as long as the supports of the outputs of the mathematical model have no intersection with the support of the fuzzy-number representation of the ideal. When low-uncertain fuzzy values close to 1 are considered, the distance measure starts to provide counterintuitive results. The reason is, however, not the distance measure, but the definition of the fuzzy ideal. Note, that defining the ideal as $(0.9,1,1)$ we expect the ideal not to be 1 and to be partially uncertain. As long as no value is close to the crisp 1, this is not a problem. If, however, values close to 1 are frequent, then the choice of the dissemblance index in combination with the definition of the fuzzy ideal is not a good one. This very simple combination of a numerical experiment with graphical outputs can provide easy-to-interpret insights into the ideal-distance pair choice appropriateness and help identify possible problems beforehand.

FIG. 2: Results from the numerical analysis, where the distance between 1 000 asymmetric triangular fuzzy numbers $\{P_0, P_{0.001}, P_{0.002}, \dots, P_{0.999}, P_1\}$, where $P_k = (k, 1, 1)$ and fuzzy ideal $F = (0.9, 1, 1)$ is depicted. On the x-axis the values of k of each fuzzy number P_k are depicted, whereas the y-axis represents the distance from the fuzzy ideal F.



5. Conclusion

In the paper a new analytical framework for selection of appropriate distance/similarity measure for linguistic approximation was proposed. This framework differ from standard framework in a way, that it compares the resulting distance between different outputs of mathematical models and so called fuzzy ideal – fuzzy number representing the goal. With this framework the author of mathematical model could examine the behavior of fuzzy measure more deeply and check if the measure possesses the required properties. The applicability of the framework is explained on a numerical example where one distance measure is examined and the found results are described. Also an additional analysis of found results is presented.

Literature:

Collan, M., Fedrizzi, M., & Luukka, P. (2013). A multi-expert system for ranking patents: An approach based on fuzzy pay-off distributions and a TOPSIS-AHP framework. *Expert Systems with Applications*, 40(12), 4749–4759.

Dubois, D. & Prade, H. (1980). *Fuzzy sets and systems: theory and applications*. Orlando: Academic Press.

Kaufman, A. & Gupta M. M. (1985). *Introduction to Fuzzy Arithmetic.* New York: Van Nostrand Reinhold.

Stoklasa, J. (2014). *Linguistic models for decision support*. Lappeenranta: Lappeenranta University of Technology.

Stoklasa, J., Talášek, T., Kubátová, J., & Seitlová, K. (2017). Likert scales in group multiple-criteria evaluation. *Journal of Multiple-Valued Logic and Soft Computing*, (in press).

Stoklasa, J., Talášek, T., & Luukka, P. (2018). Fuzzified likert scales in group multiple-criteria evaluation. In M. Collan & J. Kacprzyk (Eds.), *Soft Computing Applications for Group Decision-making and Consensus Modeling* (Vol. 357, pp. 165–185).

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

Talášek, T. & Stoklasa J. (2016). The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers. In *Proceedings of the international scientific conference Knowledge for Market Use 2016*, 539–546.

Talášek, T. & Stoklasa, J. (2017). Linguistic approximation under different distances/similarity measures for fuzzy numbers. *Journal of Multiple-Valued Logic and Soft Computing*, (in press).

Talášek, T., Stoklasa, J. & Talašová, J. (2016). The role of distance and similarity in Bonissone's linguistic approximation method – a numerical study. In *Proceedings of the 34th International Conference on Mathematical Methods in Economics*, 845–850.

Yager, R. R. (2004). 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.

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

Contact:

Mgr. Tomáš Talášek
Palacký University, Olomouc, Faculty of Arts, Department of Applied Economics
Křížkovského 8,
771 47 Olomouc Czech Republic
E-mail: tomas.talasek@upol.cz

Mgr. et Mgr. Jan Stoklasa, D.Sc. (Tech.)
Palacký University, Olomouc, Faculty of Arts, Department of Applied Economics
Křížkovského 8
771 47 Olomouc Czech Republic
E-mail: jan.stoklasa@upol.cz

and

Lappeenranta University of Technology, School of Business and Management
P.O.Box 20,
FI-53851 Lappeenranta
Finland
Email: jan.stoklasa@lut.fi

Brief information about the author:

Tomáš Talášek received the MS degree in applications of mathematics in economy from the Palacký University, Olomouc, Czech Republic, in 2012. He has been working toward the PhD degree in applied mathematics at Palacký University Olomouc, Czech Republic and Lappeenranta University of Technology, Finland. His research interests include economic and financial applications of the theory of fuzzy sets and linguistic fuzzy modelling, multiple criteria decision making, linguistic approximation and pattern recognition and the development of associated methodology and models.

Jan Stoklasa received the MS degree in applied mathematics and MS degree in psychology from Palacký University, Olomouc, Czech Republic in 2009 and 2012 respectively. He received the Ph.D. and D.Sc. degree in applied mathematics from Palacký University, Olomouc, Czech Republic and Lappeenranta University of Technology, Finland respectively in 2014. He is currently an assistant professor at Palacký University, Olomouc and a research fellow at Lappeenranta University of Technology, School of Business and Management, Lappeenranta, Finland. His research interests include decision support models, multiple criteria decision making and evaluation and linguistic fuzzy models and their practical applications in economics and social sciences.

Publication X

Talášek, T. and Stoklasa, J.
Ordering of fuzzy quantities with respect to a fuzzy benchmark – how the shape of the fuzzy benchmark and the choice of distance/similarity affect the ordering

Reprinted with the permission from
Proceedings of the 36th *International Conference on Mathematical Methods in Economics 2018*,
pp. 573–578, 2018,

Ordering of fuzzy quantities with respect to a fuzzy benchmark – how the shape of the fuzzy benchmark and the choice of distance/similarity affect the ordering

Tomáš Talášek[1], Jan Stoklasa[2]

Abstract. To order several outputs of a model represented by fuzzy numbers, we can define a reference outcome of the model called benchmark (e.g. a fuzzy singleton when ideals are used as benchmarks). Then the distance (or similarity) between this reference outcome and each of the fuzzy outputs is used for the ordering of the outputs of a model. In many cases, however, the benchmark is represented by a fuzzy number i.e. when an expert estimate of the benchmark is given, or when predictions of future values are considered.

This paper investigates the consequences of using fuzzy benchmarks for the ordering of fuzzy numbers. The paper studies if and how the use of a fuzzy benchmark with different cardinality may affect the final ordering of fuzzy numbers with respect to a chosen distance/similarity of fuzzy numbers. Different sets of fuzzy numbers representing different outputs of models (e.g. fuzzy net present values) are ordered by several distances/similarities of fuzzy numbers while the definition of the fuzzy benchmark changes cardinality-wise. Based on the analysis of the results and their graphical summaries we identify distances/similarities suitable for use with fuzzy benchmarks.

Keywords: Ordering, fuzzy numbers, fuzzy benchmark, similarity, distance.

JEL classification: D81, C44
AMS classification: 90B50, 91B06

1 Introduction

In fuzzy multiple-criteria evaluation it might be difficult to select the best alternative, since the evaluations of the alternatives are frequently represented by fuzzy numbers (or even collections of fuzzy numbers). A comparison with an "ideal" value (see e.g. [2]) could be used to obtain the ordering of the alternatives. Yet for this the "ideal" has to be defined and an appropriate measure of the distance of the fuzzy-valued evaluation from (or its similarity with) the "ideal evaluation" has to be chosen. Even though a non-fuzzy ideal value seems to be a natural choice, many authors decide to use fuzzy ideals (i.e. ideals represented by a fuzzy number) in various mathematical methods including TOPSIS [5] or selection of human resources [7]. The reasons for this choice vary from the inability of some distance/similarity measures of fuzzy numbers to work with fuzzy singletons to the simple statement that since all the evaluations are fuzzy, the ideal should be fuzzy as well. We are not analyzing the reasonability of this choice in the present paper, though. Our aim is to investigate what are the possible consequences of the use of fuzzy ideal in the process of determination of the ordering of alternatives represented by fuzzy evaluations. We depart from the widely accepted alpha-cut ordering of fuzzy numbers and show that under some circumstances even this "natural" ordering of fuzzy numbers can be contradicted by the results of a ordering procedure based on the similarity of fuzzy numbers and using a fuzzy ideal. Even more importantly we show that under the same similarity of fuzzy numbers the ordering of the given fuzzy evaluations can change simply as a result of the change of the uncertainty of the fuzzy ideal.

2 Preliminaries

Let U be a nonempty set (the universe of discourse). 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 $A(.)$ is called a *membership function* of the fuzzy set A. $\text{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an *α-cut* of A for any $\alpha \in [0, 1]$, $\text{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1)

[1]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic, tomas.talasek@upol.cz.

[2]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic and Lappeenranta University of Technology, School of Business and Management, Skinnarilankatu 34, Lappeenranta, Finland, jan.stoklasa@upol.cz.

$\text{Ker}(A) \neq \emptyset$ (A is *normal*); (2) A_α are closed intervals for all $\alpha \in (0,1]$ (this implies A is *unimodal*); (3) $\text{Supp}(A)$ is bounded. A fuzzy number A is said to be defined on $[a,b]$, if $\text{Supp}(A)$ is a subset of an interval $[a,b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1,a_4] = \text{Cl}(\text{Supp}(A))$ and $[a_2,a_3] = \text{Ker}(A)$, where $\text{Cl}(\text{Supp}(A))$ denotes a closure of $\text{Supp}(A)$. Each fuzzy number A is determined by $A = \{[\underline{a}(\alpha), \overline{a}(\alpha)]\}_{\alpha\in[0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall \alpha \in (0,1]$, and the closure of the support of A $\text{Cl}(\text{Supp}(A)) = [\underline{a}(0), \overline{a}(0)]$. Given two fuzzy numbers A and B on the same universe U, their natural ordering can be defined based on their α-cuts in the following way: if $A_\alpha \leq B_\alpha$ for all $\alpha \in (0,1]$, then A is less or equal to B, denoted $A \leq_\alpha B$.

The fuzzy number A is called *linear* if its membership function is linear on $[a_1,a_2]$ and $[a_3,a_4]$; for such fuzzy numbers we will use a simplified notation $A = (a_1,a_2,a_3,a_4)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$.

Let A be a fuzzy number on $[a,b]$. Then the *cardinality* of A, denoted by $\text{Card}(A)$ is computed by $\text{Card}(A) = \int_a^b A(x)dx$. If $a_1 \neq a_4$, then the *center of gravity* of A denoted by $\text{COG}(A)$ is computed by $\text{COG}(A) = \int_a^b xA(x)dx/\text{Card}(A)$. More details on fuzzy numbers and computations with them can be found for example in [3].

3 Ordering of outputs of mathematical model with respect to a benchmark

Let $O = \{O_1, \ldots, O_n\}$ be a set of n outputs of mathematical model (evaluations), where O_i, $i = 1, \ldots, n$, is a fuzzy number and let the fuzzy number B be a reference outcome of the model called benchmark (i.e. the desired output of the model, or in other words a (fuzzy) ideal value). The goal is to order the outputs O_i of the mathematical model with respect to the chosen benchmark B. To reach this goal a distance $d(B,O_i)$ (or similarity $s(B,O_i)$) between the benchmark B and each output O_i must be computed. Then, the outputs of the model can be ordered with respect to the computed distances in the ascending order (in the case of similarity, the outputs are ordered in the descending order).

It is important to keep in mind that the selected distance/similarity of fuzzy numbers affects the output of the final ordering, because different distances/similarities focus on different attributes of the fuzzy numbers to be ordered. The definition of the benchmark also plays a significant role (see the following sections). For the purpose of this paper we investigate two distances of fuzzy numbers (d_1 and d_2) and two similarities of fuzzy numbers (s_1 and s_2) - see their definitions for the case of fuzzy numbers C and D defined on the same universe (analogous measures were analyzed e.g. in [8, 10]):

- A *modified Bhattacharyya distance* [1]

$$d_1(C,D) = \left[1 - \int_U (C^*(x) \cdot D^*(x))^{1/2} dx\right]^{1/2}, \tag{1}$$

  where $C^*(x) = C(x)/\text{Card}(C(x))$ and $D^*(x) = D(x)/\text{Card}(D(x))$. Note that this distance measure requires $\text{Card}(C(x)) \neq 0$ and $\text{Card}(D(x)) \neq 0$, i.e. neither the fuzzy evaluation nor the fuzzy ideal can be represented by a fuzzy singleton.
- A *dissemblance index* [6]

$$d_2(C,D) = \int_0^1 |\underline{c}(\alpha) - \underline{d}(\alpha)| + |\overline{c}(\alpha) - \overline{d}(\alpha)| \, d\alpha, \tag{2}$$

- A Weis and Chens *similarity measure* [11]

$$s_1(C,D) = \Big(1 - \frac{\sum_{i=1}^4 |c_i - d_i|}{4}\Big) \cdot \frac{\min\{Pe(C), Pe(D)\} + \min\{\text{hgt}(C), \text{hgt}(D)\}}{\max\{Pe(C), Pe(D)\} + \max\{\text{hgt}(C), \text{hgt}(D)\}}, \tag{3}$$

  where $Pe(C) = \sqrt{(c_1 - c_2)^2 + (\text{hgt}(C))^2} + \sqrt{(c_3 - c_4)^2 + (\text{hgt}(C))^2} + (c_3 - c_2) + (c_4 - c_1)$, $Pe(D)$ is defined analogically,
- A Hejazi and Doostparast *similarity measure* [4]

$$s_2(C,D) = \Big(1 - \frac{\sum_{i=1}^4 |c_i - d_i|}{4}\Big) \cdot \frac{\min\{Pe(C), Pe(D)\}}{\max\{Pe(C), Pe(D)\}} \cdot \frac{\min\{Ar(C), Ar(D)\} + \min\{\text{hgt}(C), \text{hgt}(D)\}}{\max\{Ar(C), Ar(D)\} + \max\{\text{hgt}(C), \text{hgt}(D)\}}, \tag{4}$$

  where $Ar(C) = \frac{1}{2}\text{hgt}(C)(c_3 - c_2 + c_4 - c_1)$, $Ar(D)$ is defined analogically and $Pe(C)$ and $Pe(D)$ are computed identically as in the previous method.

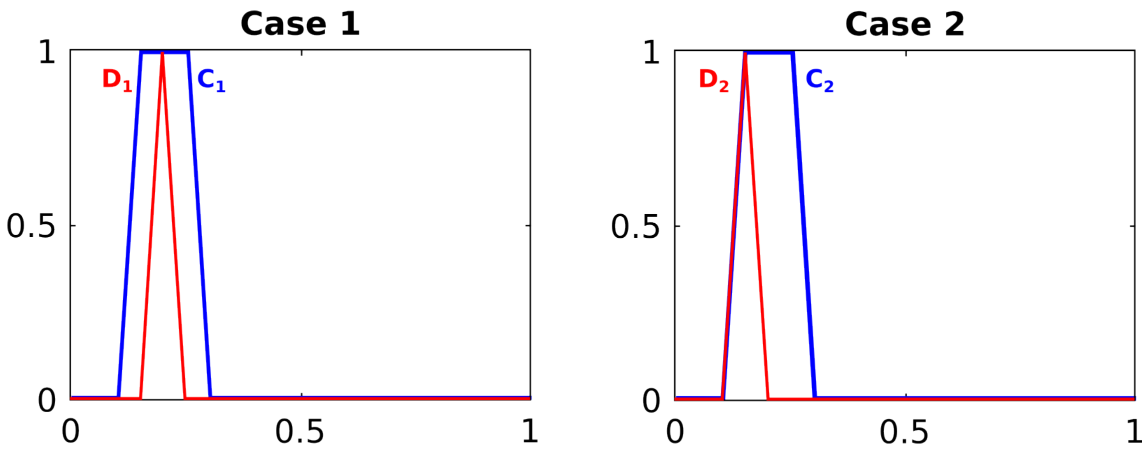


Figure 1 The selected pairs of fuzzy evaluations to be ordered with respect to a chosen fuzzy ideal. Left subfigure represents a case where the two fuzzy evaluations cannot be ordered based on the natural α-cut ordering, Right subfigure represents a case where the $D_2 \leq_\alpha C_2$. In both cases the shape of the red fuzzy number is the same (it is a symmetrical triangular fuzzy number with the same cardinality), just the location (i.e. center of gravity) differs. The blue fuzzy numbers are identical in both subfigures.

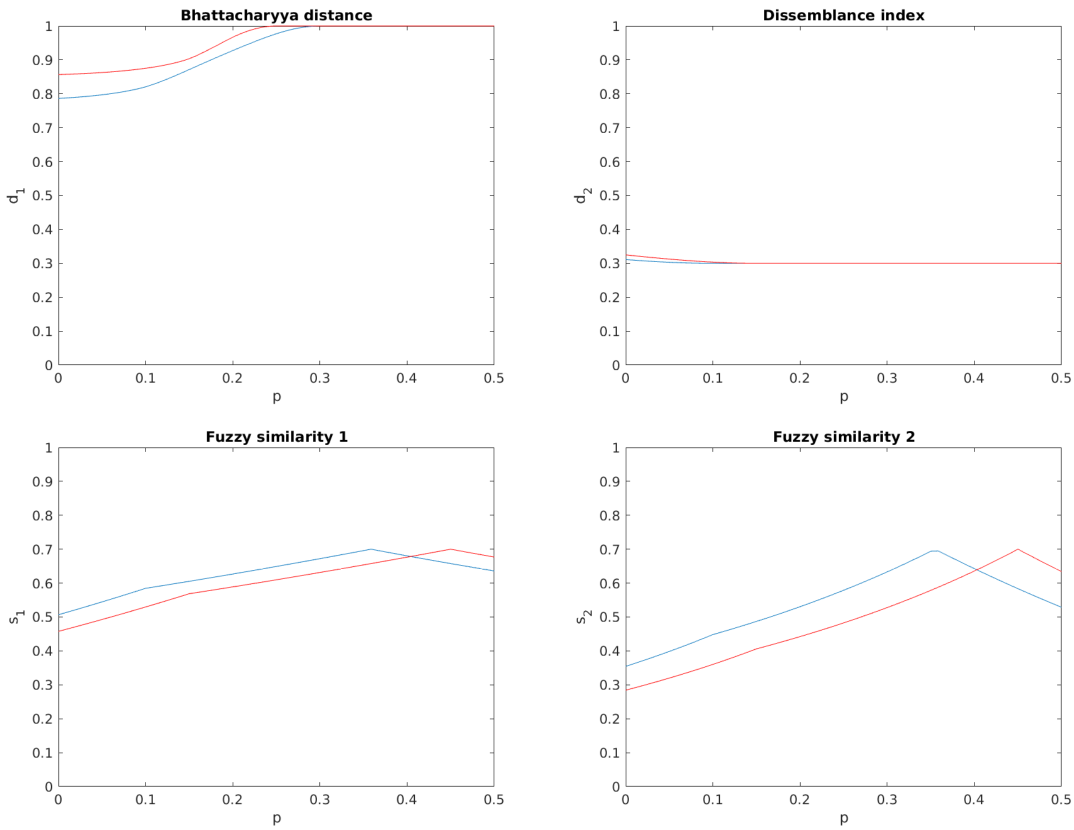


Figure 2 Results of the analysis for C_1 (blue curves) and D_1 (red curves). Top left subfigure summarizes $d_1(C_1, B_p)$ in blue and $d_1(D_1, B_p)$ in red. Top right subfigure summarizes $d_2(C_1, B_p)$ in blue and $d_2(D_1, B_p)$ in red. In both cases whichever curve is the lowest represents the preferred alternative/evaluation w.r.t. the given value of p. Bottom left subfigure summarizes $s_1(C_1, B_p)$ in blue and $s_1(D_1, B_p)$ in red. Bottom right subfigure summarizes $s_2(C_1, B_p)$ in blue and $s_2(D_1, B_p)$ in red. In both cases whichever curve is the highest represents the preferred alternative/evaluation w.r.t. the given value of p.

Let us, for the sake of simplicity, consider just two pairs of fuzzy evaluations to be ordered. The selected two pairs are depicted in Figure 1 (similar pairs of fuzzy numbers are analyzed e.g. in [11]). In the next section we apply the four above-mentioned distance/similarity measures of fuzzy numbers to obtain the ordering of C_1 and D_1 and C_2 and D_2 respectively.

As a benchmark we choose to use a parameterized symmetrical triangular fuzzy number $B_p = (p, 0.5, 0.5, 1 - p), p \in [0, 0.5)$ with a center of gravity larger than any of the significant values of the fuzzy numbers to be ordered; without any loss of generality we set COG$(B_p) = 0.5$. The cardinality of this fuzzy benchmark is dependent on the choice of the value of the parameter p (another analysis suggesting to use parametrized fuzzy number to explore the behaviour of fuzzy distance can be found in [9]). In the next chapter we analyze what ordering is suggested by (1) - (4) for C_1 and D_1 and C_2 and D_2 for any choice of $p \in [0, 0.5)$.

4 Analysis of the performance of the selected distance/similarity measures of fuzzy numbers

Two pairs of fuzzy numbers representing the expected outputs of mathematical model were chosen (see Figure 1): $C_1 = (0.10, 0.15, 0.25, 0.3)$, $D_1 = (0.15, 0.20, 0.20, 0.25)$ in its left subfigure and $C_2 = (0.10, 0.15, 0.25, 0.3)$, $D_2 = (0.10, 0.15, 0.15, 0.20)$ in its right subfigure. Please note, that fuzzy numbers C_1 and C_2 represent the same trapezoidal fuzzy number and fuzzy numbers D_1 and D_2 have same shape but different center of gravity. Also note, that fuzzy numbers from the first pair have identical centers of gravity. The left subfigure of Figure 1 thus represents a case where the two fuzzy evaluations cannot be ordered based on their α-cuts, while in the right subfigure we clearly have $D_2 \leq_\alpha C_2$. We would thus expect that all four chosen distance/similarity measures will thus provide the same ordering for C_2 and D_2, even regardless of the definition of the benchmark B_p. We have calculated the distances/similarities of the fuzzy evaluations to all the values of the fuzzy ideal $B_p, p \in [0, 0.5)$ for both cases from Figure 1 and present the results graphically in Figures 2 and 3. Note, that for the d_1 and d_2 distances (top 2 subfigures in both figures) the lower curve represents the preferred evaluation, while for the similarities s_1 and s_2 the higher curve represents the preferred evaluation for the given value of p.

In the Case 1 (fuzzy evaluations C_1 and D_1) both distances of fuzzy numbers favor the trapezoidal fuzzy number C_1 before triangular fuzzy number D_1. However, this applies only for the cases, where the benchmark B_p has higher cardinality (p is close to 0). When the value of p crosses a certain threshold (different for both distances), both distances stop discriminating between C_1 and D_1. In other words when the cardinality of B_p is low, which results in an empty intersection of B_p with C_1 and D_1, both fuzzy evaluations are considered equally distant from the ideal. This is especially evident in the case of the dissemblace index (d_2). We, however, need to keep in mind that C_1 and D_1 have the same center of gravity and are the very prototype of a pair of fuzzy numbers that cannot be ordered based on the α-cut ordering. On the other hand similarity s_1 and s_2 distinguish between the fuzzy evaluations C_1 and D_1 even in the cases when p is close to 0.5. However, in both cases there is a value of p (around 0.4), when the ordering of C_1 and D_1 switches. For p lower then this threshold, both methods prefer the trapezoidal fuzzy number C_1 (as well as in the case of d_1 and d_2), but when the cardinality of the benchmark becomes lower, both method start to prefer D_1 over C_1. This behaviour of the similarity measures can be attributed to their focus on the shape of the fuzzy numbers being compared - when the cardinality of the fuzzy ideal B_p gets lower, its shape gets closer to the shape of the triangular C_1 evaluation. This would suggest that the similarity measures focusing on the shape are not the best choice for the ordering of fuzzy evaluations. Their use can be, however, justified, as long as cardinality (or the amount of uncertainty) of the fuzzy ideal should be well matched by the best alternative.

For the Case 2 which is represented by the fuzzy numbers C_2 and D_2 to be ordered, the modified Bhattacharyya distance d_1 provides similar results as in the previous case – the trapezoidal fuzzy number C_2 is preferred over the triangular fuzzy number D_2. Unlike in the first case the dissemblace index now favors the trapezoidal fuzzy number C_2 for any value of p. This behaviour is expected because D_2 is a subset of fuzzy number C_2 and COG$(D_2) <$ COG(C_2). The C_2 is also preferred over D_2 if the similarity s_1 is used. And again this preference is independent on the value of p. However, for the similarity s_2 there again exists a threshold, where the preference is switched. In other words from a certain value of p the similarity s_2 gives results directly opposite to those suggested by the α-cut ordering of the fuzzy evaluations C_2 and D_2. This is very counterintuitive, but it can still be explained by the focus of s_2 on the shape of the fuzzy numbers that are being compared.

From the above text can be conclude, that the Bhattacharyya distance d_1 exhibits consistent behaviour, but can only order fuzzy numbers, if the benchmark B_p has higher cardinality (p is close to zero). In fact d_1 discriminates (returns a value different from 1) only if the intersection of the fuzzy evaluation with the fuzzy ideal is nonempty. This might not be a desired property. The Dissemblance index d_2 may overcome this disadvantage of d_1 in some cases, but it is still possible that the distances may coincide. However, as shown in Case 1, this happens in the “difficult to resolve” cases where the inability to distinguish two evaluations might actually be the correct answer we are looking for (in some cases the values simply cannot be ordered without additional information).

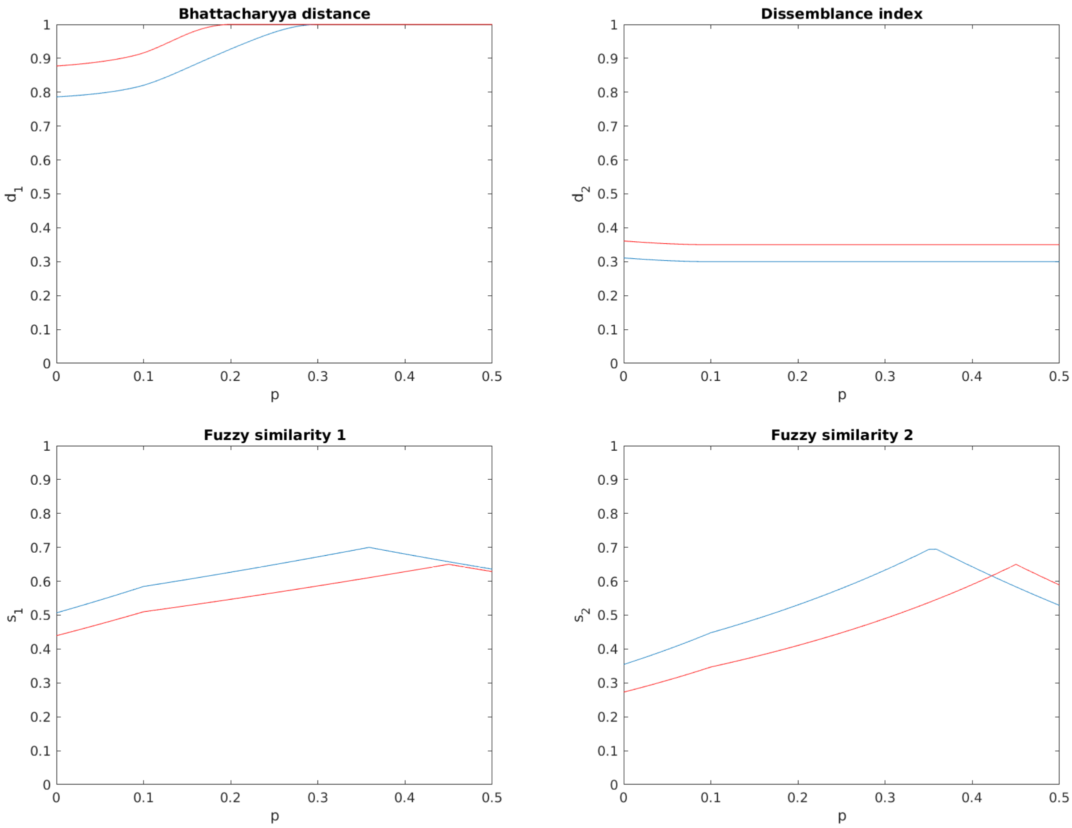


Figure 3 Results of the analysis for C_2 and D_2. Top left subfigure summarizes $d_1(C_2, B_p)$ in blue and $d_1(D_2, B_p)$ in red. Top right subfigure summarizes $d_2(C_2, B_p)$ in blue and $d_2(D_2, B_p)$ in red. In both cases whichever curve is the lowest represents the preferred alternative/evaluation w.r.t. the given value of p. Bottom left subfigure summarizes $s_1(C_2, B_p)$ in blue and $s_1(D_2, B_p)$ in red. Bottom right subfigure summarizes $s_2(C_2, B_p)$ in blue and $s_2(D_2, B_p)$ in red. In both cases whichever curve is the highest represents the preferred alternative/evaluation w.r.t. the given value of p.

If the similarity of fuzzy numbers is used, it is important to take into consideration, that for different cardinality of benchmark B_p the ordering could be different; it can even be in conflict with the natural ordering of fuzzy numbers based on the α-cuts. This is caused due to the construction of these similarities, because they take into consideration the shape of fuzzy numbers. This is especially visible in the case of s_2.

Overall the chosen distance/similarity measures seem to perform better for low values of p, i.e. for fuzzy ideals with high cardinality. The less uncertain the fuzzy ideal is, the more problems seem to appear - ranging from the loss of discrimination power in d_1 to a complete switch in the ordering of the alternatives when s_2 is used. This is an interesting finding - mainly because fuzzy ideals are often derived by fuzzification from crisp (non-fuzzy) ideal values just to allow for the use of distances/similarities of fuzzy numbers. Our findings suggest that just adding a little uncertainty can be counterproductive as long as the ordering of alternatives based on their fuzzy evaluations is needed. Out of the four analyzed distance/similarity measured the dissemblance index (d_2) seems to be performing best and seems to be the least dependent on the definition of the fuzzy ideal.

5 Conclusion

In the paper we investigate the reasonability of using fuzzy ideals in connection with two chosen distances of fuzzy numbers and two chosen similarities of fuzzy numbers for the purpose of ordering of fuzzy evaluations. We show that in many circumstances the resulting ordering depends not only on the choice of the distance/similarity measure, but also on the definition of the fuzzy ideal. Surprisingly, low-uncertain fuzzy ideals do not seem to provide appropriate results. The only distance/similarity measure that did not exhibit significant drawbacks under any definition of the fuzzy ideal is the dissemblance index. Even though this study focuses on just four chosen distance/similarity measures of fuzzy numbers, it still provides valuable insights into the use of fuzzy ideals in the evaluation setting.

Acknowledgements

Supported by the grant No. IGA_FF_2018_002 of the Internal Grant Agency of Palacký University Olomouc.

References

[1] Bonissone, P. P.: A pattern recognition approach to the problem of linguistic approximation in system analysis. *Proceedings of the IEEE International Conference on Cybernetics and Society* (1979), 793–798.

[2] Deng, H.: Comparing and ranking fuzzy numbers using ideal solutions. *Applied Mathematical Modelling* **38** (2014), 1638–1646.

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

[4] Hejazi, S., Doostparast, A., and Hosseini, S.: An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers. *Expert Systems with Applications* **38** (2011), 9179–9185.

[5] Javadian, N., Kazemi, M., Khaksar-Haghani, F., Amiri-Aref, M., and Kia, R.: A general fuzzy TOPSIS based on new fuzzy positive and negative ideal solution. In: *2009 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, Hong Kong, China, 2271–2274.

[6] Kaufmann, A., and Gupta, M. M.: *Introduction to Fuzzy Arithmetic: Theory and Applications*. Electrical-Computer Science and Engineering Series. Van Nostrand Reinhold, New York, 1985.

[7] Luukka, P., and Collan, M.: Fuzzy Scorecards, FHOWA, and a New Fuzzy Similarity Based Ranking Method for Selection of Human Resources. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 601–606.

[8] Talášek, T., and Stoklasa, J.: A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation under Different Linguistic Scales. *Journal of Multiple-Valued Logic and Soft Computing* **29** (2017), 485–503.

[9] Talášek, T., and Stoklasa, J.: Selection of tools for managerial decision support the identification of methods of choice in linguistic approximation. In: *Proceedings of the international scientific conference Knowledge for Market Use 2017* (Slavíčková, P., ed.). Palacký University, Olomouc, Olomouc, 606–613.

[10] Talášek, T., Stoklasa, J., and Talašová, J.: The role of distance and similarity in Bonissone's linguistic approximation method a numerical study. In: *Proceedings of the 34th International Conference on Mathematical Methods in Economics* (Kocourek, A., and Vavroušek, M., eds.). Technická Univerzita v Liberci, Liberec, 845–850.

[11] Wei, S.-H., and Chen, S.-M.: A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *Expert Systems with Applications* **36** (2009), 589–598.

Publication XI

Talášek, T. and Stoklasa, J.

Three-dimensional histogram visualization of the performance of linguistic approximation of asymmetrical triangular fuzzy numbers

Reprinted with the permission from
Proceedings of the international scientific conference Knowledge for Market Use 2018,
pp. 445–451, 2018,

THREE-DIMENSIONAL HISTOGRAM VISUALIZATION OF THE PERFORMANCE OF LINGUISTIC APPROXIMATION OF ASYMMETRICAL TRIANGULAR FUZZY NUMBERS

TOMÁŠ TALÁŠEK - JAN STOKLASA

Palacký University Olomouc - Faculty of Arts - Department of Applied Economics; Palacký University Olomouc - Faculty of Arts - Department of Applied Economics AND Lappeenranta University of Technology, School of Business and Management

Abstract: *The linguistic approximation transforms outputs of mathematical models (real numbers, intervals, fuzzy numbers,…) into natural language. This is especially useful in situations, when the users lack proper mathematical background (managers, laymen) or need to make decisions quickly. However, the performance of linguistic approximation depends heavily on the chosen distance of fuzzy numbers. Verification of the appropriateness of the chosen distance measure is thus crucial. Yet methods for the analysis of linguistic approximation methods for asymmetrical fuzzy numbers are scarce at best. In the paper a new visualization method for the analysis is proposed and demonstrated on a chosen distance of fuzzy numbers. The proposed method uses three-dimensional histograms to provide understandable overview of the properties of the selected linguistic approximation method and visualizes the outcomes of the approximation.*

Keywords: *linguistic approximation, fuzzy numbers, asymmetry, distance, histogram*

JEL classification: *C44*

Grant affiliation: *This research was supported by the grant IGA_FF_2018_002 Blending of eastern and western thinking in global economics and management by the internal grant agency of Palacký University Olomouc.*

Introduction

The issue of linguistic approximation (LA) has recently received some well-deserved attention in various papers (Yager, 2004; Talášek & Stoklasa, 2016). Linguistic approximation can be seen as the basis for the ability of transforming mathematical outputs (particularly fuzzy ones) into the expressions in common language. As such it opens the doors for the communication of model users with the model in a language they understand well. Replacing the formal description of the outputs of mathematical models with a natural one provides many benefits (understandability, usability by laymen, etc.), but there are also several shortcomings of this transition to "natural language" description – the most important of which might be the distortion of information originally present. Unfortunately, even the methods recently proposed for the analysis of the appropriateness of LA in various contexts and using various distance/similarity measures may provide outputs that are difficult to interpret by inexperienced users. This is true particularly when LA of asymmetrical fuzzy numbers is considered. In this case the visualization of the results of the analysis can be tricky, since the areas of objects approximated by different linguistic terms may be overlapping (Talášek & Stoklasa, 2016). In this paper we suggest a 3-D histogram representation of the outputs of LA performance analysis for asymmetrical fuzzy numbers. Our goal is to facilitate easy analysis of the appropriateness of LA and to make the results of LA-analytical methods easily interpretable for less-experienced users.

1 Preliminaries

Fuzzy sets were first introduced by Zadeh (1965). This section presents the basic notations of the theory of fuzzy sets. Let U be a nonempty set (the universe of discourse). 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 $A(.)$ is called a *membership function* of the fuzzy set A. $\mathrm{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an *α-cut* of A for any $\alpha \in [0,1]$, $\mathrm{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\mathrm{Ker}(A) \neq \emptyset$ (A is *normal*); (2) A_α are closed intervals for all $\alpha \in (0, 1]$ (this implies A is *unimodal*); (3) $\mathrm{Supp}(A)$ is bounded. A family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on $[a, b]$, if $\mathrm{Supp}(A)$ is a subset of an interval $[a, b]$. The real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \mathrm{Cl}\big(\mathrm{Supp}(A)\big)$ and $[a_2, a_3] = \mathrm{Ker}(A)$, where $\mathrm{Cl}(\mathrm{Supp}(A))$ denotes a closure of $\mathrm{Supp}(A)$.

The *cardinality* of a fuzzy number *A* on $[a, b]$ is a real number $\mathrm{Card}(A)$ defined as $\mathrm{Card}(A) = \int_a^b A(x)\,dx$ and can be considered as a measure of uncertainty of the fuzzy number A. The *centre of gravity* of a fuzzy number *A* on $[a, b]$ for which $a_1 \neq a_4$ is defined by the formula $\mathrm{COG}(A) = \int_a^b x\,A(x)\ dx/\mathrm{Card}(A)$. For a fuzzy singleton (*A* for which $a_1 = a_4$) we can define $\mathrm{COG}(A) = a_1$.

The fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$; for such fuzzy numbers we will use a simplified notation $A \sim (a_1, a_2, a_3, a_4)$. A linear fuzzy number A is said to be *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by an ordered triplet $A \sim (a_1, a_2, a_4)$. A triangular fuzzy number $A \sim (a_1, a_2, a_4)$ is called *symmetrical* if $a_2 - a_1 = a_4 - a_2$. Otherwise it is called *asymmetrical*. More details on fuzzy numbers and computations with them can be found for example in Dubois and Prade (1980) or Klir and Yuan (1995).

A *fuzzy scale* on $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, \dots, T_s$ on $[a, b]$, that form a Ruspini fuzzy partition (Ruspini, 1969) of the interval $[a, b]$, i.e. for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$ and the T's are indexed according to their ordering. A *linguistic variable* (Zadeh, 1975) is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is the name of the variable, $\mathcal{T}(\mathcal{V})$ is the set of its linguistic values (terms), X is the universe on which the meanings of the linguistic values are defined, G is a syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X. A linguistic variable is called a *linguistic scale*, if the meanings of its linguistic values form a fuzzy scale.

Linguistic scales are frequently chosen as the linguistic variables for LA. LA in general means assigning a linguistic value of a given linguistic scale to a given fuzzy object. This is frequently done based on the distance or similarity of fuzzy sets – the linguistic value (either of the linguistic scale in its original or extended form, or of a subset of its linguistic terms selected in accordance with some criteria) with the meaning closest or most similar to the approximated fuzzy object is assigned as the linguistic approximation.

2 Linguistic approximation of triangular fuzzy numbers

The linguistic approximation of fuzzy outputs of mathematical model consists of two steps. In the first step the set of $\boldsymbol{n} \geq \mathbf{2}$ appropriate linguistic terms $\mathcal{T}_{\mathbf{1}}, \dots, \mathcal{T}_{\boldsymbol{n}}$ (and their meaning represented by fuzzy numbers) that suitably represent the possible outputs of mathematical model must be specified by the user of the model (therefore it is ensured that the user will understand to final result of linguistic approximation). Usually a fuzzy scale is assumed as the underlying structure. In the second step the linguistic term $\mathcal{T}^*$ (from the set $\mathcal{T}_{\mathbf{1}}, \dots, \mathcal{T}_{\boldsymbol{n}}$) which represents the output of the model best is chosen. The process of choosing the term $\mathcal{T}^*$ is studied in the further text.

One of the frequently used methods of linguistic approximation is the so called "*best-fit*" approach (Talášek and Stoklasa, 2017). In this approach, each linguistic term $\mathcal{T}_i, i = 1, \dots, n$ is represented by a fuzzy number T_i and the distance between this fuzzy number and the output of the model Out is computed. The fuzzy number T^* that is the closest to the Out is found and the respective linguistic term $\mathcal{T}^*$ is selected as the result of LA. Instead of the distance it is also possible to use a similarity measure – in that case the fuzzy number that is the most similar to Out is found). The problem is that there exist various distance measures of fuzzy numbers with various properties and the choice of the distance affects the result of linguistic approximation. Therefore, it is critical for the purposes of linguistic approximation to know the properties of considered distance measures (and their performance in LA) before one of them is selected.

Talášek et al. (2017) propose a framework for the visualization of the performance of linguistic approximation (with chosen distance/similarity measure) of symmetrical triangular fuzzy numbers. In this framework, each symmetrical triangular fuzzy number is unambiguously represented by a point in the two-dimensional space by its centre of gravity (on the x-axis) and length of the support (on the y-axis). The colour of the point represents the resulting linguistic approximation (linguistic term).

Identical framework is used by Talášek and Stoklasa (2016) for the visualization of the performance of linguistic approximation of asymmetrical triangular fuzzy numbers. The problem is that the representation is not unambiguous (i.e. fuzzy numbers $A \sim (1,1,4)$ and $B \sim (0,3,3)$ have same centre of gravity and length of support and therefore are represented by the same point in the two-dimensional space). The authors suggested to split the results into several subfigures (splitting is based on the resulted linguistic terms) to ensure that there is no information distortion.

In the next chapter we propose a different approach to the visualization of the performance of LA on asymmetrical fuzzy numbers on a specific example of the Bhattacharyya distance and triangular fuzzy numbers. The added value of the proposed approach lies in its ability to estimate relative frequencies of assigning specific linguistic terms to fuzzy objects belonging to the given 3-D histogram bin (based on its COG and length of support). These estimates can be obtained using any reasonable method for the random generation of asymmetrical fuzzy numbers (as long as the method ensures that any asymmetrical triangular fuzzy number may be generated; given the specified precision of computations). The precision of the estimation is dependent on the number of generated fuzzy numbers. Generalization to different distance and similarity measures and to different types of fuzzy numbers is straightforward.

3 The proposed 3-D histogram visualization method presented on an artificial example

The presented method will be demonstrated on the distance of fuzzy numbers called *modified Bhattacharyya distance* (Aherne, Thacker & Rockett, 1998):

$$d(A,B) = \sqrt{1 - \int_U \sqrt{A^*(x) \cdot B^*(x)} dx}$$

where $A^*(x) = A(x)/\mathrm{Card}(A)$ and $B^*(x) = B(x)/\mathrm{Card}(B)$. In other words, given a linguistic variable the most appropriate LA will be selected as the linguistic value of the variable, for which the Bhattacharyya distance of its fuzzy-number meaning to the approximated fuzzy set (fuzzy numbers) is the lowest.

For the purpose of the explanation of the proposed visualization technique a set $O = \{O_1, \ldots, O_{100000}\}$ of 100 000 triangular asymmetrical fuzzy numbers on $[0,1]$ representing the outputs of a mathematical model was randomly generated. The generation of each fuzzy number was done in two steps – first, three real numbers from $[0,1]$ interval were randomly generated. Subsequently, those numbers were ordered in ascending order and this triplet is used as a base for significant values of triangular fuzzy number. This approach to the generation does not guarantee a uniform coverage of the universe of all possible triangular asymmetrical fuzzy numbers (fuzzy numbers with COG close to 0.5 and with larger supports are more frequent – see FIG. 1). This is, however, not a problem in the analysis and graphical representation.

Two characteristics were chosen to represent the generated fuzzy numbers in a 2-dimensional space: COG and the length of support. The first being a representation of location (even though less-informative in the case of asymmetrical fuzzy numbers) and the second being a measure of uncertainty. The universe of COGs of the generated fuzzy numbers (i.e. $[0,1]$) and the universe of the possible lengths of the supports of these fuzzy numbers (i.e. $[0,1]$) were both uniformly divided into 25 parts. This results in a 25 times 25 bin 3-D histogram presented in FIG. 1.

FIG. 1: Three-dimensional histograms of the characteristics of the 100 000 randomly generated fuzzy numbers used for the analysis. 25 times 25 bin representation (left), 10 times 10 bin representation (right).

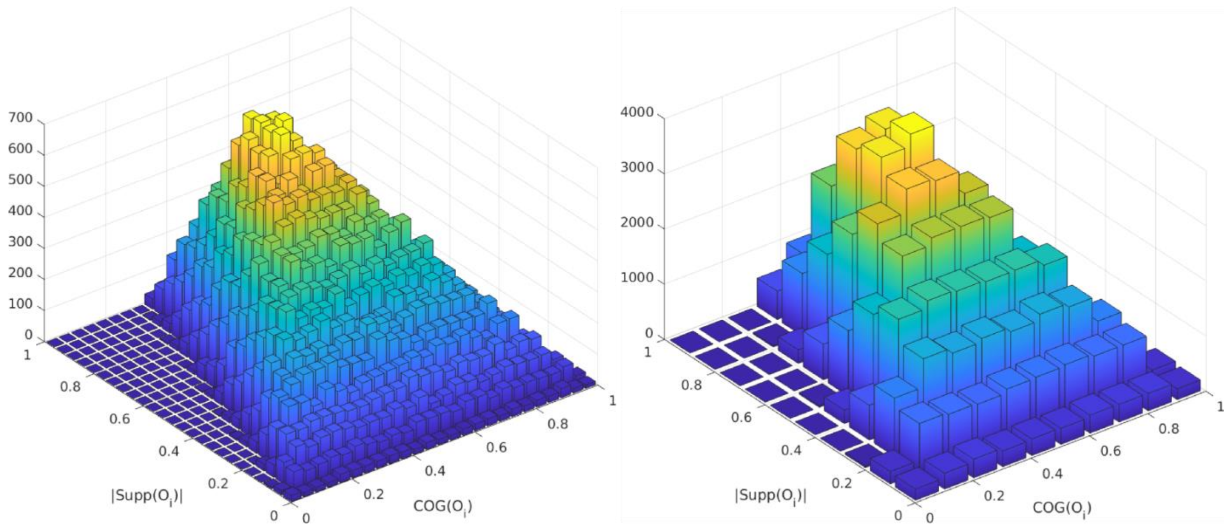
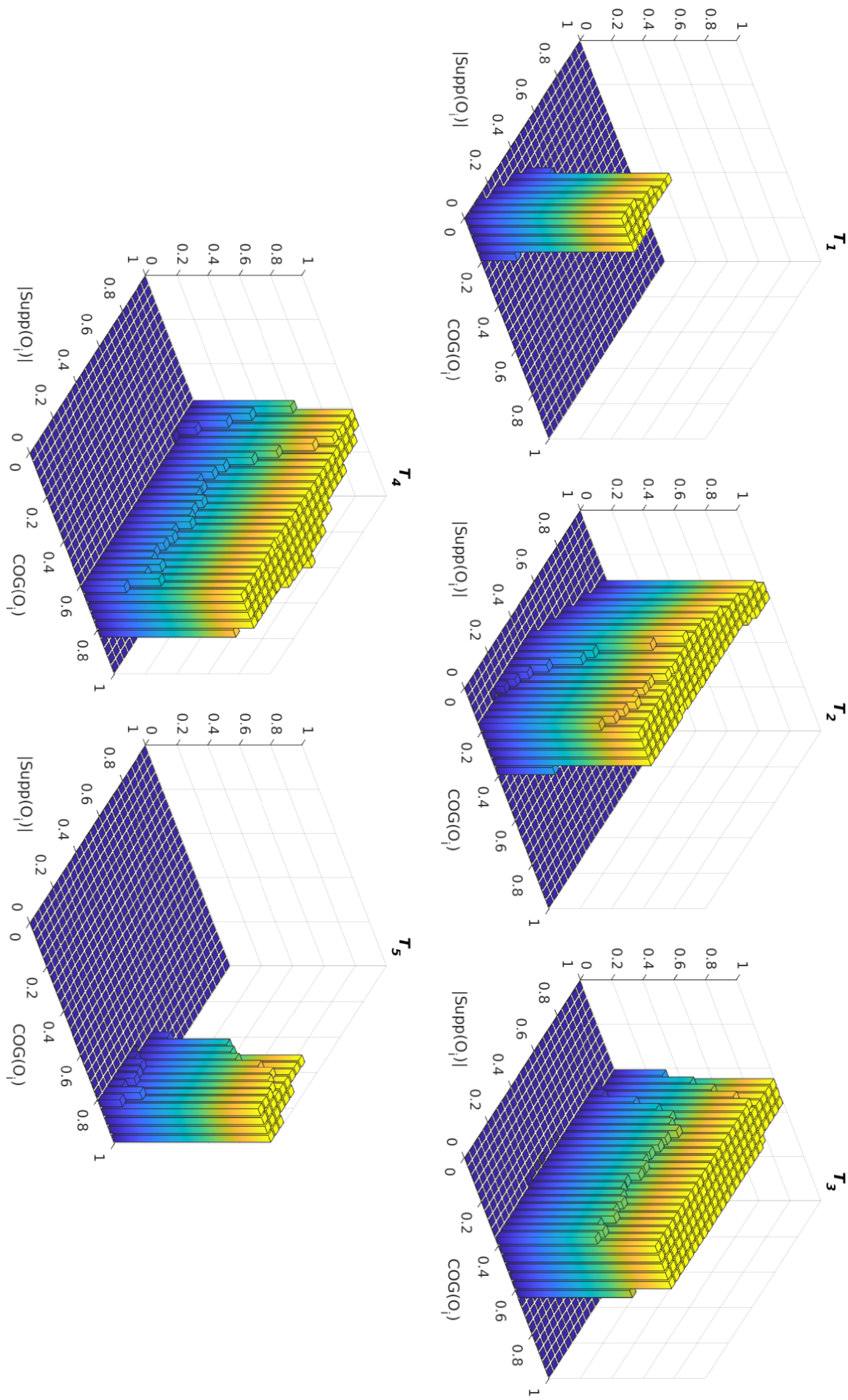


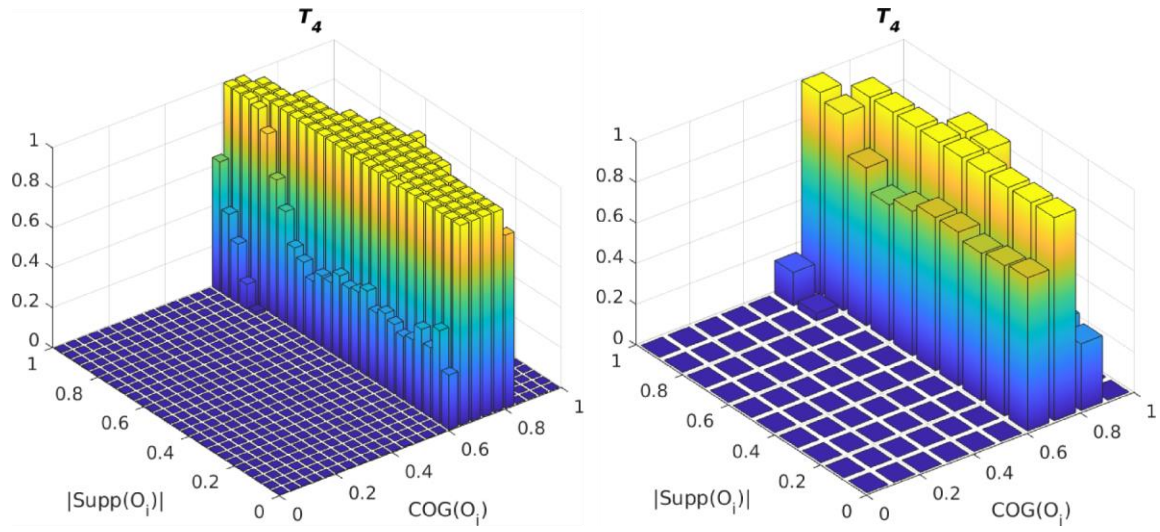
FIG. 2: Three-dimensional histogram representations of the results of the linguistic approximation of asymmetrical triangular fuzzy numbers. The bins are characterized by the COD of the fuzzy numbers and by the length of their supports. Each 3-D histogram summarizes the frequencies of suggesting the given linguistic term as the appropriate LA for the fuzzy numbers with characteristics belonging to the given bin.



Five linguistic terms $\mathcal{T}_1, \ldots, \mathcal{T}_5$ that form a linguistic scale were chosen with the respective meanings $T_1 \sim (0, 0, 0.25), T_2 \sim (0, 0.25, 0.5), T_3 \sim (0.25, 0.5, 0.75), T_4 \sim (0.5, 0.75, 1), T_5 \sim (0.75, 1, 1)$.

The results of the numerical example can be found in FIG. 2. Note, that FIG. 2 splits the graphical summary into five subfigures analogously to Talášek and Stoklasa (2016). Each subfigure of FIG. 2 represents the relative frequency of assigning the given linguistic term as LA for fuzzy numbers belonging to the respective bin. Although the use of histograms introduces a slight loss of information w.r.t. the approach proposed by Talášek and Stoklasa (2016), the limited amount of bins facilitates the comparison of the subfigures. It is now easy to see whether two corresponding bins in different subplots have nonzero height – in such case fuzzy numbers with similar (possibly even identical) characteristics can be assigned different linguistic approximations. Even though we have lost some information constructing the 3-D histograms (see FIG. 2), we can still obtain the same conclusions concerning the performance of Bhattacharyya distance as did Talášek and Stoklasa (2016). Note that even for a 10 times 10 bin 3-D histogram we can still clearly see that e.g. the result of the LA is less dependent on COG of the approximated fuzzy number when its length of support is high (see FIG. 3). In other words the reduction of information does not prevent us from identifying the important characteristics that were identified in the full information case (Talášek & Stoklasa, 2016). Moreover we can now obtain an estimate of the relative frequency of assigning specific linguistic terms for each bin. Note that such a piece of information cannot be obtained unless the bins are introduced.

FIG. 3: Three-dimensional histogram summarizes the frequencies of linguistic term $\mathcal{T}_4$ as the appropriate LA for the fuzzy numbers with characteristics belonging to the given bin. 25 times 25 bin representation (left), 10 times 10 bin representation (right). The dependence on the length of support is apparent for COG close to 0.5 in both subfigures.



Conclusion

The paper deals with the issue of the assessment of performance of linguistic approximation of asymmetrical fuzzy numbers. It proposes a 3-D histogram representation of the results of linguistic approximation. This representation allows for the estimation of relative frequencies of use of specific linguistic terms as linguistic approximation of fuzzy numbers with similar characteristics (COG and length of support define each bin). Even though the construction of histograms slightly distorts the information, important patterns are still recognisable. The proposed method relies on simulation to obtain the estimates and as such can be easily implemented with various distance/similarity measures and types of asymmetrical fuzzy numbers.

Literature:

Aherne, F., Thacker, N. & Rockett P. (1998). The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data. *Kybernetika,* 32(4), 363–368.

Dubois, D., & Prade, H. (1980). *Fuzzy sets and systems: theory and applications*. New York: Academic Press.

Klir, G. J., & Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Aplications*. New Jersey: Prentice Hall.

Ruspini, E. H. (1969). A new approach to clustering. *Information and Control,* 15(1), 22–32.

Talášek, T., & Stoklasa, J. (2016). The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers. In P. Slavíčková (Ed.), *Proceedings of the international scientific conference Knowledge for Market Use 2016*, 539–546. Olomouc: Societas Scientiarum Olomucensis II.

Talášek, T., & Stoklasa, J. (2017). A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation under Different Linguistic Scales. *Journal of Multiple-Valued Logic and Soft Computing,* 29(5), 485–503.

Yager, R. R. (2004). 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.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control,* 8(3), 338–353.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-I. Information Sciences, 8(3), 199–249, from http://doi.org/10.1016/0020-0255(75)90036-5.