VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

AKTIVNÍ UČENÍ S NEURONOVÝMI SÍTĚMI

ACTIVE LEARNING WITH NEURAL NETWORKS

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE **TOMÁŠ BUREŠ**
AUTHOR

VEDOUCÍ PRÁCE **Ing. MICHAL HRADIŠ, Ph.D.**
SUPERVISOR

BRNO 2019

Vysoké učení technické v Brně
Fakulta informačních technologií

Ústav počítačové grafiky a multimédií (UPGM) Akademický rok 2018/2019

Zadání bakalářské práce

Student: **Bureš Tomáš**
Program: Informační technologie
Název: **Aktivní učení s neuronovými sítěmi**
Active Learning with Neural Networks
Kategorie: Zpracování obrazu

Zadání:

1. Prostudujte základy konvolučních neuronových sítí a aktivního učení.
2. Vytvořte si přehled o současných metodách využívajících aktivní učení a neuronové sítě.
3. Vyberte konkrétní metody a aplikace vhodné pro experimenty.
4. Implementujte navržené metody a proveďte experimenty nad vhodnou datovou sadou.
5. Porovnejte dosažené výsledky a diskutujte možnosti budoucího vývoje.
6. Vytvořte stručné video prezentující vaši práci, její cíle a výsledky.

Literatura:

- Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012
- Štěpán Beneš: Aktivní učení s neuronovými sítěmi, VUT, 2018.

Pro udělení zápočtu za první semestr je požadováno:

- Body 1 až 3.

Podrobné závazné pokyny pro vypracování práce viz http://www.fit.vutbr.cz/info/szz/

Vedoucí práce: **Hradiš Michal, Ing., Ph.D.**
Vedoucí ústavu: Černocký Jan, doc. Dr. Ing.
Datum zadání: 1. listopadu 2018
Datum odevzdání: 15. května 2019
Datum schválení: 1. listopadu 2018

Abstrakt

Práce se věnuje problematice aktivního učení a jeho spojení s neuronovými sítěmi. Nejprve obsahuje úvod do problematiky, nastínění metod prozkoumaných metod aktivního učení. Následuje praktická část s experimenty zkoumající jednotlivé strategie a jejich vyhodnocování.

Abstract

The topic of this thesis in active learning in conjunction with neural networks. First, it deals with theory of active learning and strategies used in real life scenarios. Followed by practical part, experimenting with active learning strategie and evaluating those experiments.

Klíčová slova

aktivní učení, neuronová síť, zpracování obrazu, strojové učení

Keywords

active learning, neural network, image processing, machine learning

Citace

BUREŠ, Tomáš. *Aktivní učení s neuronovými sítěmi.* Brno, 2019. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Michal Hradiš, Ph.D.

Aktivní učení s neuronovými sítěmi

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Michala Hradiše, Ph.D. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Tomáš Bureš
15. května 2019

Poděkování

Děkuji vedoucímu práce Ing. Michalu Hradišovi, Ph.D. za jeho odbornou pomoc.

Obsah

Kapitola 1

Úvod

Tato práce se zabývá metodami aktivního učení a jejich spojení s neuronovými sítěmi. Cílem spojení těchto dvou oblastí je redukce finanční náročnosti na získávání velkých datových sad, potřebných pro trénování neuronových sítí.

Cílem mé práce je pomocí experimentů spojujících tyto dvě oblasti pozorovat dopad aktivního učení na efektivitu modelu a jeho srovnání s použitím náhodného vzorkování a trénování na celé datové sadě.

Kapitola 2

Aktivní učení a jejich spojení s neuronovými sítěmi

2.1 Aktivní učení

Aktivní učení v informatice označuje speciální případ strojového učení, ve kterém učící se algoritmus aktivně dotazuje uživatele data na nichž se bude trénovat. Tento dotazující se algoritmus je pak schopen podávat lepší výsledky i při učení na menším množství dat. Tato vlastnost ještě dále nabývá na významu při spojení s neuronovými sítěmi, které pro podání co nejlepších výsledků většinou potřebují obrovské množství trénovacích dat. V mnoha oblastech jsou označené vzorky na nichž by se algoritmus mohl učit snadno získatelné, například ve spojení se sociálními sítěmi, kdy uživatelé sami svojí aktivitou poskytují dostatečné množství trénovacích dat. Ovšem v různých scénářích využívajících strojové učení, jsou kvalitně označené vzorky obtížně nebo velmi draze získatelné [23]. Například:

- *Rozpoznání řeči.* Přesné označování mluveného slova je časově velmi náročné a vyžaduje zkušeného lingvistu. Anotace nahrávky na úrovni slova může trvat desetkrát délku nahrávky a na úrovni fonémů až 400 násobek délky audia [36].
- *Extrakce informace.* Dobré systémy pro extrakci informací musí být trénovány velmi detailně anotovaných dokumentech. Uživatel označuje entity a vztahy, jenž nás v textu zajímají, jako třeba osoby a názvy organizací, nebo jestli dotyčná osoba pracuje pro konkrétní organizaci. Lokalizace entit může zabrat půl hodiny i pro stručný novinový článek [24].
- *Klasifikace a filtrování.* Učení klasifikace dokumentů (např.: článků nebo webových stránek) nebo jiných typů médií (např.: obraz, zvuk, video) vyžaduje aby uživatel v každém dokumentu nebo souboru označoval značkami, jako "důležité"a "nedůležité". Nutnost označení tisíců vzorků může být únavné a nekdy i redundantní [23].

Aktivní učení tento problém řeší optimalizací učícího procesu pomocí dotazování většinou lidského anotátora na neoznačené vzorky. Model si tedy sám vybírá nejvhodnější vzorky jejichž anotaci požaduje. Tímto se aktivně učící model snaží dosáhnout co nejvyšší efektivity s využitím co nejmenšího množství dat. Snižuje se tedy časová a finanční náročnost samotného označování. Aktivní učení tedy nabývá na významu hlavně v oblastech, kde není problém získat trénovací data ale je velmi náročné jejich označování.

Modely strojového učení ve spojení s aktivním učením mohou pracovat různými způsoby, většinou však následují podobný princip. Učící se model začíná s malou označenou sadou,

přičemž dle algoritmů aktivního učení vybírá vhodné vzorky z mnohem větší neoznačené sady. U těchto vybraných vzorků požaduje anotaci, učí se na nich a následně je přidá do označené sady [23].

2.2 Pokládání dotazů

Existuje několik způsobů jakými může model pokládat dotazy. Třemi hlavními zmiňovanými v literatuře jsou *Syntéza dotazu* (Membership query syntehis), *Selektivní vzorkování* (Stream-based selective sampling) a *Aktivní učení na fondu* (Pool-based active learning) [23].

2.2.1 Syntéza dotazu

Jedná se o jeden z prvních způsobů dotazování zkoumaných ve spojitosti s aktivním učením [23] [1]. V tomto případě učící se model může dotazovat anotaci jakéhokoliv neoznačeného vzorku vstupního prostoru, včetně vzorku uměle vytvořeného modelem, což se v tomto případě především předpokládá. Efektivní syntéza dotazů se je často velmi účinná pro učení modelu v oblasti konečných problémů [2] [23]. Syntéza dotazu byla také rozšířena pro regresní úkoly strojového učení [7].

Ačkoliv syntéza dotazu může být použita pro mnoho druhů praktických problémů, naráží na zásadní problém, značení těchto nově vytvořených vzorků může být pro lidského anotátora poměrně obtížné. Například Baum a Lang ve svojí práci [4] využili syntézu dotazů s lidskými anotátory k natrénování neuronové sítě k rozpoznávání ručně psaných znaků. Narazili však na nečekaný problém: velká část obrazů generovaných neuronovou sítí sestávala z nerozpoznatelných symbolů bez jakéhokoliv sémantického významu [23] [4]. Podobně si lze představit jak by vypadalo zapojení syntézy dotazu pro učení ve spojitosti s psaným či mluveným slovem.

2.2.2 Selektivní vzorkování

Alternativou k syntéze dotazu je *selektivní vzorkování* [6]. Klíčovým předpokladem je, že neoznačená trénovací data lze získat snadno a levně, takže data ze vstupní distribuce mohou být vzorkována a učící se model se rozhoduje bude-li chtít znát jejich označení nebo je zahodí. Tento přístup je občas nazýván *sekvenční aktivní učení*, jelikož vzorky jsou vytahovány z neoznačené sady jeden za druhým a model se rozhoduje zda-li se na ně bude dotazovat nebo je zahodí. Pokud je vstupní distribuce uniformní selektivní vzorkování se může podobat syntéze dotazu. Pokud ovšem vstupní distribuce není uniformní nebo je neznámá, selektivní vzorkování zaručuje smysluplnost dotazů, jelikož pocházejí ze vstupní distribuce [23].

Rozhodnutí zda-li se dotazovat na vzorek lze učinit několika způsoby. Jedním z nich je ohodnotit vzorky "měrou infomativnosti"nebo "dotazovací strategií"a učinit vážený náhodný výběr s tím, že u informativnějších vzorků bude větší pravděpodobnost, že budou vybrány [9]. Dalším způsobem je výpočet explicitního *regionu nejistoty* [6], nebo-li části vstupních dat jimiž si model není jistý a dotazovat se pouze na vzorky spadajícího do tohoto regionu. Nejjednodušším způsobem jak tohoto dosáhnout, je stanovení minimální hranice míry informativnosti definující tento region. Vzorky vyhodnocené jako nad touto hranicí jsou pak dotazovány. Jiný přístup k výpočtu tohoto regionu může být stanovení neoznačených vzorků jen jsou pro třídu modelu stále neznámé. Jinými slovy pokud se dva modely stejné

třídy, ale s jinými parametry, shodují na všech označených datech, ale rozchází u nějakého neoznačeného vzorku, pak tento vzorek spadá do *regionu nejistoty*. Výpočet tohoto regionu je však velmi náročný a musí dojít k jeho vypočítání znovu po každém dotazu. V praxi jsou tedy používány pouze aproximace [6] [27] [10].

Selektivní vzorkování bylo použito v praxi například časování senzorů [15] nebo získávání informací [34].

2.2.3 Aktivní učení na fondu

V mnoha reálných případech mohou být najednou sesbírány velké kolekce dat. S touto myšlenkou vzniklo *aktivní učení na fondu* [18], jenž předpokládá existenci malé sady $\mathcal{L}$ označených dat a velké sady $\mathcal{U}$ dat neoznačených. Vzorky jsou pak vybírány "hladově"(greedy) podle míry jejich informativnosti použité k ohodnocení celé sady $\mathcal{U}$. Aktivní učení na fondu bylo studováno pro mnoho realných problémů strojového učení zahrnujících klasifikaci textu [18] [20] [31] [13], extrakce informace [29] [25], klasifikace obrazu [30] [35], klasifikace videa [33] [12], rozpoznání mluveného slova [32], diagnózu rakoviny [19] a mnoha dalších.

Hlavní rozdíl mezi *selektivním vzorkováním* a *aktivním učením na fondu* spočívá v tom, že první prochází datovou sadou sekvenčně a vybírá vzorky individuálně, zatímco druhý ohodnocuje celou sadu než si vybere nejlepší vzorek.

2.3 Strategie výběru

Všechny metody pokládání dotazů zahrnují vyhodnocování informativnosti neoznačených vzorků, jenž moho být generovány *de novo* nebo vybírány z datové sady. Proto bylo vyvinuto mnoho strategií dotazování (query strategies). Tato sekce se zabývá těmi nejznámějšími [23].

2.3.1 Nejistota modelu

Pravděpodobně nejjednodušší strategií je ta, pracující s nejistotou modelu (uncertainty sampling) [23] [18]. V tomto případě se aktivně učící model dotazuje na vzorky u nichž si je nejméně jistý jak by je sám označil. Pro probabilistické učící se modely je tento přístup jednoduchý. Například, při trénování probabilistického modelu pro binární klasifikaci model jednoduše vybere k dotazování ty vzorky u nichž si je svým označením jistý s pravděpodobností 0,5 [18] [17]. Obecnější strategie pracující s nejistotou modelu využívají jako míru nejistoty entropii [28]

$$x^*_{ENT} = \underset{x}{\operatorname{argmax}} - \sum_i P(y_i|x;\theta) \log P(y_i|x;\theta),$$

kde y_i nabývá hodnot všech možných označení. Entropie je míra v oblasti teorie informace reprezentující míru informace potřebnou k zakódování distribuce. Jako taková je často brána jako míra nejistoty, či nečistoty ve strojovém učení. Pro binární klasifikaci se chová identicky jako dříve zmíněné určení pravděpodobnosti s jakou si je model jistý. Avšak je možno ji jednoduše zobecnit pro probabilistické klasifikátory pro klasifikaci více tříd a pro práci s komplexnějšími strukturovanými vzorky jako jsou sekvence [25] a stromy [14]. Alternativou k entropii je takzvaná nejnižší jistota (least confident)

$$x^*_{LC} = \underset{x}{\operatorname{argmin}}\, P(y^*|x;\theta),$$

kde $y^* = \text{argmax}_y P(y|x;\theta)$ je nejpravděpodobněji označení třídy. Tato strategie se ukázala jako velmi účinná pro úkoly extrakce informace [25] [8]. Pro binární klasifikaci je tento přístup ekvivalentní k přístupu s entropií.

2.3.2 Výbor modelů

Další, spíše teoretickou strategií výběru vzorků je algoritmus *výbor modelů* (query-by-committee) [27]. Tento přístup zahrnuje udržování výboru $\mathcal{C} = \{\theta^{(1)}, \ldots, \theta^{C}\}$ modelů, jenž jsou všechny trénovány na současné označené sadě $\mathcal{L}$, ale reprezentují odlišné hypotézy. Každý člen výboru potom volí označení kandidátů pro dotaz. Jako nejinformativnější vzorek je pak vybrán ten, na němž se modely nejméně shodují. Základní premisou je minimalizace prostoru verzí, což je soubor hypotéz jenž jsou konzistentní se současnou označenou sadou $\mathcal{L}$. Literatura se neshoduje na ideální velikosti výboru, jenž se mění s třídou modelů nebo se způsobem použití. Avšak i malé výbory čítající dva až tři členy se v praxi ukázaly jako velmi účinné [27] [20].

Pro určování míry nesouhlasu existují dva hlavní přístupy. První je *entropie hlasů* [9] popsaná jako

$$x^*_{VE} = \underset{x}{\text{argmax}} - \sum_i \frac{V(y_i)}{C} \log \frac{V(y_i)}{C},$$

kde y_i opět nabývá hodnot všech možných označení a $V(y_i)$ je počet "hlasů"jenž toto označení obdrželo od členů výboru. Tato metoda lze chápat jako zobecnění *nejistoty modelu* založené na entropii. Další navrhovanou měrou je průměrná *Kullback-Lieblerova divergence* [9]

$$x^*_{KL} = \underset{x}{\text{argmax}} \frac{1}{C} \sum_{c=1}^{C} D(P_{\theta^{(c)}} || P_{\mathcal{C}}),$$

kde

$$D(P_{\theta^{(c)}} || P_{\mathcal{C}}) = \sum_i P(y_i|x;\theta^{(C)}) \log \frac{P(y_i|x;\theta^{(c)})}{P(y_i|x;\mathcal{C})}.$$

Zde $\theta^{(c)}$ reprezentuje jednotlivé modely výboru a $\mathcal{C}$ reprezentuje celý výbor. $P_{\mathcal{C}}(y_i|x) = \frac{1}{C}\sum_{c=1}^{C} P_{\theta^{(c)}}(y_i|x)$ je modely dohodnutá pravděpodobnost, že y_i je správné označení. *Kullback-Lieblerova divergence* [16] je v teorii informace míra odlišnosti mezi dvěma pravděpodobnostními rozděleními. Tato míra považuje za nejinformativnější vzorek ten, jenž má největší průměrný rozdíl jakéhokoliv člena výboru oproti usnesené distribuci.

2.3.3 Očekávaná změna modelu

Další strategií výběru je dotazování vzorku jenž by současnému modelu přinesl největší změnu, pokud bychom znali jeho označení. Příkladem strategie tohoto typu je *očekávaná délka gradientu* (expected gradient length) pro diskriminativní probabilistické modely [26]. Jelikož diskriminativní probabilistické modely jsou většinou trénovány pomocí optimalizace gradientu, změna v modelu může být měřena jako délka trénovacího gradientu. Jinými slovy, trénovaný model dotazuje takový vzorek x, jenž by po označení a přidání do $\mathcal{L}$ vyústil v nerosáhlejší trénovací gradient. Nechť $\nabla\ell(\mathcal{L};\theta)$ je gradientem funkce ℓ vzhledem k parametrům modelu θ. Nechť $\nabla\ell(\mathcal{L} \cup \langle x, y\rangle;\theta)$ je nový gradient jenž by byl získán přidáním trénovací dvojice $\langle x, y\rangle$ do $\mathcal{L}$. Jelikož dotazovací algoritmus nezná opravdové označení y

předem, je třeba vypočítat délku jako předpoklad přes všechna možná označení

$$x^*_{EGL} = \underset{x}{\operatorname{argmax}} \sum_i P(y_i|x;\theta) \left\| \nabla \ell(\mathcal{L} \cup \langle x, y_i \rangle; \theta) \right\|,$$

kde $\|\cdot\|$ je Euklidovská norma každého výsledného gradientu vektoru. Metoda preferuje vzorky jenž pravděpodobně nejvíce ovlivní model, bez ohledu na výsledné označení vzorku. Tento přístup se ukázal být velmi dobře fungující v empirických studiích, ovšem může být velmi výpočetně náročný pokud prostor rysů a množina možných označení jsou velmi rozsáhlé [23].

2.3.4 Očekávaná redukce chyby

Strategie měřící potenciální redukci chybovosti modelu. Můžeme odhadnout budoucí chybovost modelu jenž by nastala, pokud by vzorek x byl označen a přidán do trénovací sady $\mathcal{L}$ a potom vybrat vzorek jenž minimalizuje toto očekávání. Strategie podobná předchozí, avšak nyní cílíme na nejnižší chybu u dotazovaného vzorku, místo největší změny v modelu. Metoda se zdá být téměř optimální a zároveň není závislá na třídě modelu. Vše co vyžaduje je správná *loss* funkce a způsob jak odhadnout pozdější pravděpodobnosti označení. Tato strategie byla s úspěchem použita s různými modely včetně naivního Bayesova klasifikátoru [22], Gaussovskými náhodnými poli [37], podpůrnými vektorovými stroji [21] a logistické regrese [11]. Naneštěstí také ve většině případů tato strategie bude výpočetně nejnáročnější mezi všemi zmíněnými. Nejen že vyžaduje odhadování budoucí chyby pro každý dotaz v celé $\mathcal{U}$, ale nový model musí být znovu trénován pro každé označení vzorku. Toto vede k drastickému nárůstu ceny výpočtů. Pro některé třídy modelů, jako jsou Gaussovská náhodná pole [37], je inkrementální učící procedura účinná a přesná, což tuto strategii dělá velmi parktickou. Ovšem pro většinu ostatních toto neplatí. Například binární logisticky regresní model by vyžadoval $O(ULG)$ časovou komplexitu jen pro výběr dalšího dotazu, kde U je velikost neoznačené sady $\mathcal{U}$, L je velikost současné trénovací sady $\mathcal{L}$ a G je počet výpočtů gradientů potřebných k optimalizační proceduře až do konvergence. Klasifikační úloha se třemi a více třídami označení využívající MaxEnt model [5] vyžaduje časovou komplexitu $O(M^2ULG)$, kde M je počet označovacích tříd. Sekvenční značkovací úlohy využívající *podmínková náhodná pole* (conditional random fields) časová komplexita dosahuje $O(TM^{T+2}ULG)$, kde T je délka vstupní sekvence. Z těchto důvodů se strategie s očekávanou redukcí chyby používají pouze u úloh s binární klasifikací [23].

2.3.5 Metoda vážené hustoty rozložení vzorků

Jak bylo zmíněno strategie pracující s nejistotou modelu a strategie s výborem modelů jsou náchylné k dotazování anomálních vzorků, což byl hlavní motivační faktor pro vznik strategie s očekávanou redukcí chyby [22] [37]. Hlavní myšlenkou strategie s váženou hustotou rozložení vzorků je, že informativní vzorky nejsou jen ty u nichž si je model nejistý, ale hlavně ty, jenž jsou reprezentativní pro vstupní distribuci. Proto v této strategii dotazujeme vzorky podle

$$x^*_{ID} = \underset{x}{\operatorname{argmax}} \varnothing_A(x) \times \left(\frac{1}{U} \sum_{u=1}^{U} \sim (x, x^{(u)}) \right)^{\beta}.$$

Zde $\varnothing_A(x)$ reprezentuje informativnost x podle nějaké bázové dotazovací strategie A, jako je *nejistota modelu* nebo *výbor modelů*. Druhá polovina rovnice váží informativnost vzorku

x podle jeho průměrné podobnosti k ostatním vzorkům vstupní distribuce. Parametr β řídí důležitost tohoto dílčího výsledku [23].

2.4 Efektivita aktivního učení

Zásadní otázkou je, zda-li aktivní učení v praxi opravdu funguje a jestli skutečně přináší nějaké výhody. Většina literatury a analýz se shoduje, že odpověď na první část otázky je kladná [6] [29] [26]. Stejně tak tato literatura poukazuje na nesporné výhody aktivního učení, jako je menší označená datová sada. Ovšem jsou zde i nevýhody. Trénovací sada je pevně spjatá s modelem jenž byl využit v konkrétním případě. Pokud bychom chtěli tento model změnit (což se v oblasti strojového učení děje velmi často), je třeba změnit také trénovací sadu [23]. Krom toho v některých případech se ukázalo, že aktivně učící model požadoval dokonce větší množtví označených vzorků, než pasivně se učící model [3].

Kapitola 3

Implementace a experimenty

K implementaci jsem zvolil Python knihovnu PyTorch. Postupně jsem implementoval metody pracující s nejistotou modelu a zkoumal jejich efektivitu v porovnání s náhodným vzorkováním a učením na celé datové sadě.

Obecně všechny zvolené metody dokázaly předčít náhodné vzorkování, dle očekávání, trénování na menší datové sadě nedosahuje kvalit trénování na sadě větší.

Kapitola 4

Závěr

Při experimentování se strategiemi aktivního učení jsem pozoroval jejich lepší efektivitu oproti náhodnému vzorkování. Ovšem i přes tyto úspěchy aktivní učení stále zaostává za pasivním učení na velkých datových sadách což jej pro praktické využití činí nevhodným. Trénování s využitím těchto metod je také mnohem výpočtě i časově náročnější než v případě pasivního učení. Ve výsledku tedy použití těchto metod nepřináší příliš pozitiv.

Literatura

[1] Angluin, D.: Queries and Concept Learning. *Machine Learning*, ročník 2, č. 4, Apr 1988: s. 319–342, ISSN 1573-0565, doi:10.1023/A:1022821128753.
URL https://doi.org/10.1023/A:1022821128753

[2] Angluin, D.: Queries Revisited. In *Algorithmic Learning Theory*, editace N. Abe; R. Khardon; T. Zeugmann, Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, ISBN 978-3-540-45583-7, s. 12–31.

[3] Baldridge, J.; Osborne, M.: Active Learning and the Total Cost of Annotation. 01 2004, s. 9–16.

[4] Baum, E. B.; Lang, K. J.: Query Learning Can Work Poorly when a Human Oracle is Used. Proceedings of the IEEE International Joint Conference on NeuralNetworks, 1992.

[5] Berger, A. L.; Pietra, V. J. D.; Pietra, S. A. D.: A Maximum Entropy Approach to Natural Language Processing. *Comput. Linguist.*, ročník 22, č. 1, Březen 1996: s. 39–71, ISSN 0891-2017.

[6] Cohn, D.; Atlas, L.; Ladner, R.: Improving generalization with active learning. *Machine Learning*, ročník 15, č. 2, May 1994: s. 201–221, ISSN 1573-0565, doi:10.1007/BF00993277.

[7] Cohn, D. A.; Ghahramani, Z.; Jordan, M. I.: Active Learning with Statistical Models. *CoRR*, ročník cs.AI/9603104, 1996.

[8] Culotta, A.; McCallum, A.: Reducing Labeling Effort for Structured Prediction Tasks. In *Proceedings of the 20th National Conference on Artificial Intelligence - Volume 2*, AAAI'05, AAAI Press, 2005, ISBN 1-57735-236-x, s. 746–751.

[9] Dagan, I.; Engelson, S. P.: Committee-Based Sampling For Training Probabilistic Classifiers. In *In Proceedings of the Twelfth International Conference on Machine Learning*, Morgan Kaufmann, 1995, s. 150–157.

[10] Dasgupta, S.; Hsu, D. J.; Monteleoni, C.: A general agnostic active learning algorithm. In *Advances in Neural Information Processing Systems 20*, editace J. C. Platt; D. Koller; Y. Singer; S. T. Roweis, Curran Associates, Inc., 2008, s. 353–360.

[11] Guo, Y.; Greiner, R.: Optimistic Active Learning Using Mutual Information. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007, s. 823–829.

[12] Hauptmann, A. G.; Lin, W.-H.; Yan, R.; aj.: Extreme Video Retrieval: Joint Maximization of Human and Computer Performance. In *Proceedings of the 14th ACM International Conference on Multimedia*, MM '06, New York, NY, USA: ACM, 2006, ISBN 1-59593-447-2, s. 385–394, doi:10.1145/1180639.1180721.

[13] Hoi, S. C. H.; Jin, R.; Lyu, M. R.: Large-scale Text Categorization by Batch Mode Active Learning. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, New York, NY, USA: ACM, 2006, ISBN 1-59593-323-9, s. 633–642, doi:10.1145/1135777.1135870.

[14] Hwa, R.: Sample Selection for Statistical Parsing. *Comput. Linguist.*, ročník 30, č. 3, Září 2004: s. 253–276, ISSN 0891-2017, doi:10.1162/0891201041850894.

[15] Krishnamurthy, V.: Algorithms for optimal scheduling and management of Hidden Markov model sensors. *Signal Processing, IEEE Transactions on*, ročník 50, 07 2002: s. 1382 – 1397, doi:10.1109/TSP.2002.1003062.

[16] Kullback, S.; Leibler, R. A.: On Information and Sufficiency. *Ann. Math. Statist.*, ročník 22, č. 1, 03 1951: s. 79–86, doi:10.1214/aoms/1177729694.

[17] Lewis, D. D.; Catlett, J.: Heterogeneous Uncertainty Sampling for Supervised Learning. In *In Proceedings of the Eleventh International Conference on Machine Learning*, Morgan Kaufmann, 1994, s. 148–156.

[18] Lewis, D. D.; Gale, W. A.: A Sequential Algorithm for Training Text Classifiers. *CoRR*, ročník abs/cmp-lg/9407020, 1994, cmp-lg/9407020.

[19] Liu, Y.: Active Learning with Support Vector Machine Applied to Gene Expression Data for Cancer Classification. *Journal of Chemical Information and Computer Sciences*, ročník 44, č. 6, 2004: s. 1936–1941, doi:10.1021/ci049810a, pMID: 15554662.

[20] McCallum, A.; Nigam, K.: Employing EM and Pool-Based Active Learning for Text Classification. In *Proceedings of the Fifteenth International Conference on Machine Learning*, ICML '98, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, ISBN 1-55860-556-8, s. 350–358.

[21] Moskovitch, R.; Nissim, N.; Stopel, D.; aj.: Improving the Detection of Unknown Computer Worms Activity Using Active Learning. In *KI 2007: Advances in Artificial Intelligence*, editace J. Hertzberg; M. Beetz; R. Englert, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, ISBN 978-3-540-74565-5, s. 489–493.

[22] Roy, N.; McCallum, A.: Toward Optimal Active Learning Through Sampling Estimation of Error Reduction. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, ISBN 1-55860-778-1, s. 441–448.

[23] Settles, B.: Active learning literature survey. Technická zpráva, 2010.

[24] Settles, B.; Craven, M.: Active Learning with Real Annotation Costs. 2008.

[25] Settles, B.; Craven, M.: An Analysis of Active Learning Strategies for Sequence Labeling Tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, s. 1070–1079.

[26] Settles, B.; Craven, M.; Ray, S.: Multiple-Instance Active Learning. In *Advances in Neural Information Processing Systems 20*, editace J. C. Platt; D. Koller; Y. Singer; S. T. Roweis, Curran Associates, Inc., 2008, s. 1289–1296.
URL
http://papers.nips.cc/paper/3252-multiple-instance-active-learning.pdf

[27] Seung, H. S.; Opper, M.; Sompolinsky, H.: Query by Committee. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, New York, NY, USA: ACM, 1992, ISBN 0-89791-497-X, s. 287–294, doi:10.1145/130385.130417.

[28] Shannon, C. E.: A Mathematical Theory of Communication. *Bell System Technical Journal*, ročník 27, č. 3, 1948: s. 379–423, doi:10.1002/j.1538-7305.1948.tb01338.x.

[29] Thompson, C. A.: Active learning for natural language parsing and information extraction. Morgan Kaufmann, 1999, s. 406–414.

[30] Tong, S.; Chang, E.: Support Vector Machine Active Learning for Image Retrieval. In *Proceedings of the Ninth ACM International Conference on Multimedia*, MULTIMEDIA '01, New York, NY, USA: ACM, 2001, ISBN 1-58113-394-4, s. 107–118, doi:10.1145/500141.500159.

[31] Tong, S.; Koller, D.: Support Vector Machine Active Learning with Applications to Text Classification. *J. Mach. Learn. Res.*, ročník 2, Březen 2002: s. 45–66, ISSN 1532-4435, doi:10.1162/153244302760185243.

[32] Tur, G.; Hakkani-Tur, D.; E. Schapire, R.: Combining active and semi-supervised learning for spoken language understanding. *Speech Communication*, ročník 45, 02 2005: s. 171–186, doi:10.1016/j.specom.2004.08.002.

[33] Yan, R.; Yang, J.; Hauptmann, A.: Automatically labeling video data using multi-class active learning. 11 2003, ISBN 0-7695-1950-4, s. 516–523 vol.1, doi:10.1109/ICCV.2003.1238391.

[34] Yu, H.: SVM Selective Sampling for Ranking with Application to Data Retrieval. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, KDD '05, New York, NY, USA: ACM, 2005, ISBN 1-59593-135-X, s. 354–363, doi:10.1145/1081870.1081911.

[35] Zhang, C.; Chen, T.: An active learning framework for content-based information retrieval. *IEEE Trans. Multimedia*, ročník 4, 2002: s. 260–268.

[36] Zhu, X.: Semi-Supervised Learning With Graphs. 01 2005.

[37] Zhu, X.; Lafferty, J.; Ghahramani, Z.: Combining Active Learning and Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. In *ICML 2003 workshop on The Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, 2003, s. 58–65.

Příloha A

Popis odevzdaného CD

Na přiloženém CD jsou zdrojové kódy, toto *pdf* a všechny soubory k jeho úspěšné kompilaci.