

Oponentní posudek na disertační práci (Doctoral thesis)

Název práce: Imperfect Learning of Multi-Classifier

Autor: Mgr. Ing. Radek Hřebík

Školitel: prof. Ing. Josef Jablonský, CSc. Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, Katedra ekonometrie

Školitel specialista doc. Ing. Jaromír Kukal, Ph.D., České vysoké učení technické v Praze
Fakulta jaderná a fyzikálně inženýrská, Katedra softwarového inženýrství

Oponent: prof. RNDr. PhDr. Antonín Slabý, CSc., Univerzita Hradec Králové, Fakulta informatiky a managementu, Katedra informatiky a kvantitativních metod

Struktura a obsah práce

Práce napsaná v anglickém jazyce má 160 stran textu. Základní text, který má rozsah 90 stran, je tvořen 9 částmi včetně nečíslovaného Úvodu a Závěru. Úvod (Introduction) - je stručným úvodem do tematiky práce, Kapitola 1 Klasifikační metody ve vektorovém prostoru (Classification Methods in Vector Space) je věnována vybraným základním faktům a pojmům, Kapitola 2 - Cíle práce (Thesis Goals) formuluje cíl práce a související dílčí cíle práce. Následují kapitoly dokumentující hlavní přínosy disertační práce a autorovy přístupy k řešení: Kapitola 3 - Diffusion Based Learning (Učení založené na difúzi), Kapitola 4 - Theory of Imperfect Parallel Learning. (Teorie neperfektního paralelního učení), Kapitola 5 - Imperfect Classifier (Neperfektní klasifikátor), Kapitola 6 - Experimental Part (Experimentální část), Kapitola 7 - Implementation in Matlab (Implementace v Matlabu). Conclusions (Závěry) pak představuje krátké shrnutí. Text doplňují Seznam symbolů (List of Symbols) a Seznam zkratk (List of Abbreviations) v úvodu textu a Seznam použité literatury (Bibliography), Ukázky zdrojového kódu (Source Code Examples) a fulltexty tří k tématu disertační práce navázaných publikací v impaktovaných časopisech (Selected Publications in the Impact Journals) v závěru textu.

Cíl práce

Tématikou práce jsou pojmový aparát a vybrané výsledky novátorské teorie neperfektního učení.

Smyslem práce je přispět podstatným věrohodným příspěvkem zejména k tematice klasifikace (zařazení do tříd), k oblastem klasifikátorů a metrik posuzujících kvalitu (= přesnost a citlivost) klasifikátorů, k metodám zvýšení přesnosti klasifikace s využitím novátorského konceptu kritické citlivosti umožňujícího zvýšit vliv tříd s malým počtem vzorů.

Autor formuluje v této souvislosti cíl a skupinu souvisejících dílčích cílů disertační práce.

Hlavním cílem práce je navrhnout konstrukci nového klasifikátoru založeného na novém principu a současně navrhnout k jeho hodnocení novou metriku tzv. kritickou citlivost představující nejnižší dosaženou citlivost v jednotlivých třídách, která hraje klíčovou roli při vytváření i vyhodnocení výkonu klasifikátoru. Zkonstruovaný klasifikátor zajišťuje zohlednění každé třídy a nezanedbává kvalitu nedostatečně zastoupených tříd.

Dále je v práci uvedena skupina následujících pěti souvisejících dílčích cílů:

Vytvoření konceptu skrytých tříd (počet skrytých tříd je obecně větší, než počet výstupních tříd.), jejichž účelem je přispět ke zkvalitnění klasifikátoru (= zkvalitnění rozhodovacího procesu spojeného s klasifikací - tj. zařazením do tříd a využitím tříd).

Návrh původní neperfektní strategie učení, založené na principu dílčího vzorkování (= sub-sampling), tradiční klasifikace a použití konceptu skrytých tříd. Toto dílčí vzorkování vede ke zvýšení kvality klasifikátoru využívajícího konceptu sjednocení tříd.

Vytvoření nové techniky (technik) samoorganizace dat, sloužící jako alternativa k tradičnímu Kohonenovu učení a založené na adaptaci (úpravě) učebních vah.

Porovnání efektivity jednotlivých heuristik s tradičními přístupy pomocí standardní křížové validace. Tento přístup poskytuje objektivní srovnání kvality klasifikátorů. Pro tento účel je použito několik tradičních a běžně dostupných datových sad.

Použití nové klasifikační techniky na reálný (netradiční) datový soubor včetně prezentace výsledků použití nového klasifikátoru včetně srovnání s tradičními řešeními.

Metody a postupy použité v práci

Autor vychází z rozsáhlé rešerše literárních zdrojů zkoumané problematiky. Definuje a využívá pojmový aparát, výsledky, algoritmy a postupy specifické aplikační oblasti, které patří věcně zejména do oblasti informatiky (umělé inteligence, strojového učení, algoritmizace apod.), a matematiky (matematické statistiky, operačního výzkumu - lin programování apod.). K realizaci algoritmů a experimentům s nimi autor využil systém Matlab.

Metody použité v práci lze souhrnně hodnotit, jako novátorské (invenční), adekvátní řešení problematiky a ke splnění cílů disertační práce vedoucí. Postupy použité v práci je možno hodnotit jako zajímavé a účinné.

V této souvislosti je třeba zdůraznit a velmi pozitivně hodnotit zejména skutečnost, že jsou v práci dokumentovány autorovy vlastní podstatné vytvořené metody, úpravy a postupy.

Aktuálnost tématu, disertabilita, splnění cílů práce

Tématika práce je vysoce aktuální, velmi obtížná, dosti široce pěstovaná a též široce využitelná. Práce svojí tematikou, obsahem, postupy a výsledky jednoznačně patří do oboru doktorského studia (Studijní program: Aplikace přírodních věd, Studijní obor: Matematické inženýrství).

Autor volil tematiku, která je jeho delším výzkumným zájmem a ve které již dosáhl kvalitních výsledků, na které může dobře navázat. Vytvořený pojmový aparát, elementy doprovodné teorie a výsledky jsou původní a originální a rozšiřující poznání. Práce navazuje i na tradici výzkumů, metodických postupů, zkušeností, výsledků a vybavení v dané oblasti na FJFI i expertní znalosti školitele a zejména školitele specialisty.

Cíl práce a související dílčí cíle formulované v práci byly splněny. Práce rozšiřuje pěkným, netradičním, překvapivým a rozsáhlým příspěvkem poznání v oblasti tématu.

Autor má 3 publikace v kvalitních časopisech s IF úzce související s tematikou disertační práce a tady překračuje publikační požadavky stanovené k obhajobě závěrečné obhajobě práce.

Přesnost práce, formální stránka práce

Práce je napsána v angličtině, jasným a přesným jazykem. Volba anglického jazyka má výhodu v možnosti používat anglické terminologie v oblasti, která není v českém jazyce plně stabilizována. Definice, postupy, algoritmy a výsledky jsou přesně formulovány. Práce demonstruje též velmi dobře autorovy přínosy. Formální stránka práce i přesnost vyjadřování, úprava vzorců, diagramů a grafů mají velmi dobrou úroveň. Práce má jasnou strukturu, proto je možno se v ní dobře orientovat.

Výsledky práce a přínosy práce

Práce přináší pestré, originální, dosti početné a vzájemně související výsledky v oblasti teoretické, metodologické a je dobře využitelná i v oblasti praktické.

Za hlavní teoreticko-metodologické výsledky lze považovat zejména následující: Je představen autorův nový neperfektní přístup (metodický postup) ke klasifikačním úlohám, a shlukování dat a je porovnán s existujícími (klasickými metodami a postupy). V provedeném porovnání se jeví autorův přístup často jako efektivnější. Je představena metrika kritické citlivosti jako hlavní míra pro hodnocení kvality klasifikátorů. Je formulována tzv. neperfektní učební strategie s použitím modelu lineárního programování. Jsou navrženy 2 nové metody sebeorganizace odvozené z tradiční Kohonenovy učební strategie spočívající v začlenění difuze do procesu učení a vedoucí k dynamičtějšímu a flexibilnějšímu přístupu k sebeorganizaci. Dva nově vyvinuté způsoby učení - strategie založené na difuzi – tradiční difuze a anomální difuze – ve většině případů vykazují zlepšení nebo alespoň srovnatelnost s tradiční metodou Kohonenovou. Je představena novátorská teorie skrytých tříd (které lze vytvořit několika způsoby) a použít v učení s učitelem i bez učitele. Je vytvořen nový obecný multi-klasifikátor s pětivrstvou strukturou. Teorie optimálního sjednocení založená na teorii lineárního programování je stěžejní částí konečné klasifikace. Navržený klasifikátor je flexibilní a uživatel může používat individuální nastavení.

Praktické výsledky a možnosti využití: Přístupy uvedené v práci jsou formulovány často pomocí skriptů Matlabu a mohou být opakovaně použity v řadě situací, aplikací a dále modifikovány a rozvinuty.

Poznámky a otázky do diskuse při obhajobě

Práce poskytuje řadu možností a oblastí k diskusi.

Navrhují diskusi věnovat například rozdílům ve filosofii a podstatě rozdílů mezi klasickými klasifikačními metodami a přístupy k učení a autorovým novátorským přístupem založeným na početnějších skrytých třídách, nové klasifikaci a neperfektním učení, souladu autorových přístupů s posledními trendy AI, přednostem autorových přístupů a problémům, které musel překonat. Existují nějaké limity, pokud jde o typ řešených problémů a typ aplikační oblasti?

Závěr:

Práce splňuje v nároky na disertační práce kladené. Jde o kvalitní, přesný a hluboký a přínosný text poskytující velmi pěkný vhled do problematiky řešené v práci a dobře demonstruje elementy autorova přístupu i rozsáhlé autorovy výsledky v rámci předložené novátorské a zajímavé teorie. Rozsah, originalitu, systematickosti a náročnosti aktivit studenta a kvalitu výsledků, tak jak je dokumentuje předložená disertační práce, je možno velmi pozitivně hodnotit. Disertační práci doporučuji k obhajobě a též, aby Mgr. Ing. Radku Hřebíkovi byl po úspěšné obhajobě udělen titul Ph.D.

Hradec Králové 15. 5. 2023

Antonín Slabý