

POSUDEK VEDOUCÍHO BAKALÁŘSKÉ PRÁCE

Název: SVM Classifiers and Heuristics for Feature Selection

Autor: Tomáš Krupka

Shrnutí obsahu práce a celkové hodnocení

Práce pojednává o heuristických metodách pro selekci příznaků v rámci strojového učení „s učitelem“, a konkrétně v souvislosti s SVM klasifikátory. V souladu se zadáním práce se autor zaměřuje speciálně na známou heuristiku pro výběr příznaků SVM-RFE (Recursive Feature Elimination) a zkoumá výkonnost jejích modifikací. Testování výkonnosti zkoumaných algoritmů je prováděno na klasifikační úloze definované pro sémantickou disambiguaci vybraných anglických sloves. Trénovací data student obdržel již hotová.

Autor práce samostatně navrhl nové rozšíření metody SVM-RFE, které nazval EB-RFE (Evaluation-Based RFE) a experimentálně vyhodnotil, že na daném typu úlohy jeho inovativní modifikace dosahuje lepších výsledků ve srovnání s originálním SVM-RFE algoritmem, a to jak co do úspěšnosti automatické klasifikace, tak co do míry redukce počtu příznaků. V tomto smyslu autor zcela úspěšně zpracoval zadané téma bakalářské práce a její zadaný cíl velmi dobře splnil. Originální návrh EB-RFE a experimentální otestování této heuristiky je vlastním příspěvkem autora. Za významné považuji, že hlavní myšlenka EB-RFE, jakkoliv je jednoduchá, přinesla až překvapivě dobré výsledky a prokazatelné zlepšení standardní SVM-RFE metody. Zda je přínos EB-RFE univerzální, mohou ovšem ukázat pouze rozsáhlejší testy na různých typech dat.

Práce obsahuje Úvod (1 strana), dvě kapitoly převážně rešeršní (18 stran) a třetí kapitolu, kde autor popisuje své vlastní experimenty (7 stran). Matematický popis studovaných metod včetně popisu algoritmů je až na několik výjimek přesný. Rešerše z literatury je pojata tvůrčím způsobem, s ohledem na zamýšlené cíle práce. Text končí jednostránkovým, pěkně napsaným závěrem, kde autor výstižně shrnuje obsah a výsledky své práce. Práce obsahuje jako přílohu také naprogramované procedury, konkrétně dohromady cca 450 řádků v jazyce R. Chybí však jakákoliv dokumentace, takže použitelnost tohoto kódu pro osoby odlišné od autora bude patrně velmi problematická.

Text práce je napsán slušnou, dobře čitelnou angličtinou bez většího množství gramatických chyb. Po formální stránce má předložená práce všechny náležitosti, grafická úprava je bezchybná. Citace zdrojů jsou dostatečné, vytýkáám však nepořádné zpracování seznamu citované literatury (str. 29-30), kde mnohé položky neobsahují všechny náležité a povinné údaje.

Připomínky a otázky

- Jaký je váš názor na míru závislosti úspěšnosti EB-RFE na daném typu úlohy a konkrétních datech?

Závěr

Práci považuji za velmi dobrou a rozhodně ji doporučuji uznat jako bakalářskou práci.

Vedoucí bakalářské práce

RNDr. Martin Holub, PhD
Ústav formální a aplikované lingvistiky
MFF UK v Praze

V Praze, 26. 8. 2016