

Posudek diplomové práce „Metody strojového učení pro analýzu sentimentu. Patočka, M. (2013)“

Ing. Ivan Habernal, Ph.D.
vedoucí práce

Katedra informatiky a výpočetní techniky, FAV, ZČU
Univerzitní 8, 306 14 Plzeň
habernal@kiv.zcu.cz

29. května 2013, Plzeň

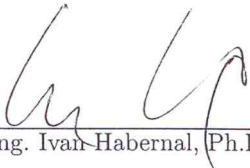
Bc. Michal Patočka se ve své práci zabývá analýzou sentimentu češtiny na doméně hodnocení produktů a možnostech využití metod strojového učení. Práce má rozsah 53 stran textu a obsahuje DVD s implementací všech metod a testů. Téma je velmi aktuální a odpovídá trendům ve výzkumu NLP (Natural Language Processing - zpracování přirozeného jazyka).

V teoretické části popisuje diplomant motivaci a současný stav výzkumu na angličtině, s tím že pro češtinu nebyly dosud publikovány žádné podobné práce. Jako zdrojů bylo využito množství vědeckých publikací. Dále se diplomant věnuje předzpracování a výběru příznaků pro algoritmy strojového učení pro daný úkol. Samotným metodám strojového učení je věnována celá sekce, stejně tak dalšímu podúkol, kterým je automatická selekce příznaků.

V experimentální části diplomant popisuje celý proces získání dat pro analýzu sentimentu z webového prostředí a jejich následného předzpracování včetně vhodných příznaků. Dále srovnává výsledky pro různé kombinace předzpracování, příznaků a tří metod strojového učení (Naïve Bayes, Maximum entropy klasifikátor a SVM), čímž získá tzv. "baseline". Velká část této sekce je věnována experimentům s různými nastaveními algoritmů selekce příznaků (celkem pět metod). Pro optimální nastavení metody Relevancy Score je dokonce překonán baseline výsledek. Všechny testované metody jsou vždy diskutovány, přičemž argumentace výsledků je důsledná a odhaluje vlastnosti daných metod.

Dva hlavní přínosy práce jsou následující. Zaprvé, diplomant prozkoumal oblast analýzy sentimentu v češtině na reálných datech a otestoval množství state-of-the-art method, jeho práce se dá označovat za pilotní. Zadruhé, jím implementované metody selekce příznaků budou zahrnuty do open-source knihovny pro strojové učení vyvíjené Ing. Konkolem (KIV).

Shrnutí. Diplomant vytvořil významnou práci vědeckého charakteru, na úloze pracoval samostatně, sestudoval existující anglickou literaturu (články) k tématu a přicházel s originálními nápady. Prozkoumal oblast analýzy sentimentu v češtině, naimplementoval a otestoval metody selekce příznaků a vše otestoval na reálných datech. Proto **doporučuji práci k obhajobě a hodnotím známkou výborně.**


Ing. Ivan Habernal, Ph.D.

Dodatečné otázky

1. (*K diskusi*) V čem spočívá výhoda přístupu strojového učení pro tuto doménu a úkol, např. v porovnání s metodami založenými na slovníku (tzv. sentiment lexicon/vocabulary)?

**SOUHLASÍ
S ORIGINÁLEM**



Západočeská univerzita v Plzni
Fakulta aplikovaných věd
katedra informatiky a výpočetní techniky