

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Michal Auersperger
Název práce English grammar checker and corrector: the determiners
Rok odevzdání 2017
Studijní program Informatika **Studijní obor** Matematická lingvistika

Autor posudku RNDr. Pavel Pecina Ph.D. **Role** vedoucí
Pracoviště Ústav formální a aplikované lingvistiky

Text posudku:

Michal Auersperger se v předkládané práci zabývá úlohou automatické korektury gramatických chyb v anglických textech se zaměřením na správné používání určitých a neurčitých členů. Podobně jako v dřívějších přístupech, je tato úloha řešena jako predikce členu (určitý/ neurčitý/žádný) pro každou jmennou frázi v textu, ekvivalentně jako klasifikace jmenných frází do tříd, které určují člen (určitý/neurčitý/žádný), který jí má v textu předcházet. V tomto smyslu je také úspěšnost řešení úlohy posuzována jako podíl správně určených členů.

Práce je psaná anglicky, celkem na 74 stranách, rozdělená do 4 očíslovaných kapitol, které tvoří hlavní část práce. Těmto kapitolám předchází jedna kapitola úvodní a po nich následuje kapitola závěrečná, seznam použité literatury, seznam obrázků a tabulek a také příloha, které obsahuje vzorek testovacího textu (jednak v originální podobě, jednak ve 5 variantách, ve kterých byly původní členy nejdříve odstraněny a poté doplněny nezávisle 5 různými lidmi – rodilými mluvčími), tato data byla použita pro srovnání výsledků automatických metod s výsledky člověka. Kapitola 1 je teoretická a diskutuje problematiku používání členů v angličtině. Kapitola 2 popisuje 2 hlavní metody strojového učení použité v práci (statistickou klasifikaci pomocí logistické regrese a *gradient boosted trees*. Kapitola 3 popisuje provedené experimenty (včetně detailů použitých metod) a kapitola 4 potom jejich výsledky (včetně nezbytné optimalizace parametrů), které srovnává s výsledky metody založené na použití jazykového modelu a výsledky lidskými.

Práci lze charakterizovat jako výzkumně-experimentální s tím, že v experimentech byly použity existující nástroje (např. pro lingvistickou analýzu textu nebo knihovny pro trénování modelů strojového učení). Celkově je práce psaná velmi čtivě, v podstatě bez jazykových chyb a překlepů. Text je dobře strukturovaný, srozumitelný a obsahuje všechny části, které by odborná práce tohoto typu měla mít (popis metod, experimentů, výsledků a jejich diskuse).

Návrh experimentů, jejich provedení a analýza výsledků je zcela v pořádku. Autor velmi vhodně zvolil posloupnost experimentů, které vedly k výběru optimální metody, jejich parametrů, pou-

žitých rysů apod. Práce obsahuje netriviální vědecký příspěvek, ten spočívá především v nově navržených rysech a použitím metodě klasifikace, které vedly ke zlepšení současného *state of the art* z 87% na 92% (accuracy), což je snížení chyby o třetinu. Vyzdvihl bych také srovnání s lidským výkonem, které ukázalo, že navržená metoda pracuje víceméně stejně jako nejlepší anotátor. Chválím také velmi dobře zpracovanou část teoretickou, které jasně vysvětluje obtížnost řešeného problému. Jedinou výtkou je to, že k práci nebyly přiloženy data, kódy a výsledky experimentů v elektronické podobě.

Předkládaná práce splňuje požadavky kladené na diplomovou práci na MFF UK. Její cíle byly splněny.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 31. 5. 2017

Podpis: