Posudek oponenta diplomové práce

Autor/autorka práce: **Ondřej Pražák**

Název práce: **Vícejazyčné značkování sémantických rolí**

Obsah práce

Zvolené téma – vícejazyčné značkování sémantických rolí – je rozhodně náročné a hlavní cíl práce – vytvořit systém, který bude alespoň částečně přenositelný mezi různými jazyky bez nutnosti ručních úprav – je velmi ambiciózní. Z rozsáhlého seznamu prostudované literatury a celkové struktury práce je zřejmé, že se autor nad souvislostmi řešené úlohy důkladně předem zamyslel, bohužel ne vždy byla navržená „kostra" práce odpovídajícím způsobem doplněna potřebnými detaily. Moje největší výtka k obsahové stránce práce je následující:

Přestože je různým metodám strojového učení (využitelným pro zpracovávanou úlohu) věnován značný prostor v teoretickém úvodu – podstatná část kapitol 3 a 5 a v zásadě celá kapitola 6 – a důraz je kladen spíše na metody učení bez učitele (s logickým argumentem, že příprava trénovacích dat pro učení s učitelem v „nových" jazycích je drahá a náročná), v kapitole 8 věnované vlastnímu implementovanému systému je celá záležitost strojového učení odbyta jedním krátkým odstavcem s tím, že se ani nedozvíme, jaký algoritmus byl vlastně použit (je zmíněno pouze to, že jde o binární klasifikátory, trénované metodou učení s učitelem).

Kvalita řešení a dosažených výsledků

Autor práce pro svoje experimenty využil zbrusu nový formalismus pro syntaktické anotace, tzv. *Universal Dependencies* (UD), jehož specifikace byla dokončena teprve v roce 2015. I v jiných ohledech - algoritmy pro zpracování textových dat a (pravděpodobně, viz má výhrada v přechozím bodě) strojové učení – jsou použité metody plně „state-of-the-art", což nebývá v případě diplomových prací tak úplně pravidlem. Velmi pozitivně hodnotím, že diplomant srovnává svoje výsledky se světovými špičkami v daném oboru a nespokojí se pouze s tím, že by vyprodukoval „nějaké" hodnoty zvolených metrik a o těch pak bez uvedení kontextu prohlásil, že jsou dobré nebo špatné.

K úrovni zdrojových kódů nejsem schopen se vyjádřit, nejsa odborníkem na Javu. Softwarový návrh podle vysokoúrovňového popisu uvedeného v podkapitolách 8.1 až 8.4 ale považuji za velmi zdařilý.

Formální úroveň

V tomto ohledu s prací již tak spokojen nejsem. Text obsahuje časté překlepy – typicky takové, které neodhalí automatický korektor a přitom často změní význam sdělení – např. „nekonfigurovat" místo „nakonfigurovat" (str. 36), „netrénovaný" místo „natrénovaný" (str. 38) nebo „beyesovské" místo „bayesovské" (str. 21 či 22). Často je též nějaký termín použit a poté vysvětlen až o několik stran později („purity" a „collocation" použito na str. 22, vysvětleno na str. 24), případně pouze v seznamu zkratek (LDA a LSA použito na str. 10, „vysvětleno" v seznamu zkratek – kdyby byl na str. 10 odkaz na relevantní literaturu, stačilo by to). Vadil my i autorem často používaný styl číslovaného seznamu, kdy první položka je velkým písmenem a ty ostatní malým, ale to už je spíše typografický detail.

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

Práce s literaturou

Jak jsem již uvedl, seznam uvedené literatury je velmi rozsáhlý a snese srovnání s mnoha pracemi dizertačními. Namátkou prohlédnuté položky napovídají, že jde o zdroje relevantní ke zpracovávanému tématu. Je pouze škoda, že zrovna u své vlastní publikované práce [KP15] autor zapomněl uvést důležité bibliografické údaje, konkrétně název sborníku. V době Googlu (a uvedeného DOI) jde ale opět pouze o typografický prohřešek.

Splnění zadání

Zadání bylo splněno bez výhrad.

Doplňující informace k práci

Vzhledem k výše uvedeným výhradám a čistě na základě textu práce navrhuji mírně snížené hodnocení (viz níže). Při úspěšném průběhu obhajoby se ale nebráním posunu k hodnocení ještě lepšímu.

Dotazy k práci

Zopakuji zde svůj dotaz z první části hodnocení – jaký klasifikátor/klasifikátory jste ve své implementaci použil?

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni dne 6.6.2017

doc. Ing. Pavel Ircing, Ph.D.

SOUHLASÍ
S ORIGINÁLEM

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