

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Entity Linking and Disambiguation Using a Dialogue
Jméno autora:	Ta Van Duy
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	katedra kybernetiky/FEL
Oponent práce:	Ondřej Hubáček
Pracoviště oponenta práce:	katedra počítačů/FEL

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Úloha se skládá ze tří navazujících netriviálních částí: „Named Entity Recognition“ (NER), „Entity linking“ (EL) a „Named Entity Disambiguation“ (NED). Pro podúlohy NER a EL autor využívá existující řešení.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Splněno.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Autor experimentuje s vlatním modelem pro NER. Pro finální vyhodnocení však využívá existující řešení. Pro úlohu EL také využívá existující řešení. Pro NED autor navrhuje originální řešení pomocí dialogu. Vzhledem ke komplexnosti úlohy stálo za úvahu věnovat více pozornosti EL/NED na úkor NER.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Autor střídavě používá označení „Entity linking“ pro jednu část úlohy (EL) a pro celou úlohu (NER+EL+NED). Jinak prokazuje dobrý přehled v doméně. Pro samotné modelování využívá aktuální architektury neuronových sítí.	

Formální a jazyková úroveň, rozsah práce	C - dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práci by prospěl seznam použitých zkratk kvůli jejich nemalému množství.	
Sekce 2.2 „Named Entity Disambiguation Approaches“ obsahuje i přístupy k „Entity Linking“ (EL) ač to z nadpisu není patrné.	
Sekce 3.1 nazvaná „Problem Statement“ ve skutečnosti obsahuje návrh řešení problému.	
Termín RDF používán v 3.2.1-3.2.3 je zaveden a vysvětlen až v sekci 3.3 „RDF“.	
Kapitola 4 „Theoretical Analysis“, která tvoří třetinu textu, je pouze úvodem do neuronových sítí, o kterých se však autor hojně zmiňuje v předchozích kapitolách. Nevztahuje se přímo k použitým modelům a mohla by tak být uvedena daleko dříve a stručněji.	
Kapitola 5 na 7 stranách shrnuje použité modely, experimenty, výsledky a jejich diskuzi. Kapitola působí velmi zmateně. Například reportovaná accuracy v podsekcí NED (5.3.1) se vztahuje ke komponentě EL. Accuracy zde není vhodná metrika, nepostihuje totiž velikost množiny kandidátních entit. Dále například u výsledků jednoho z NER modelů (tabulka 5.6) autor hovoří o jisté nadřazenosti BLSTM nad BGRU, hned další výsledky za použití regularizace (tabulka 5.7) ukazují opak.	

Výběr zdrojů, korektnost citací

A - výborně

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Práce se zdroji je v pořádku. Autor cituje celkem 72 zdrojů. Citované jsou jak články týkající se přímo řešené úlohy, tak datasey, nástroje, databáze apt.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Největší část experimentů se věnuje nejprobádanější podúloze NER. Přes množství provedených experimentů zůstaly výsledky NER daleko za state-of-the-art. To však pro navazující části nepředstavovalo problém, jelikož autor pro NER a EL nakonec využil existující řešení. Řešení pro EL také zůstalo daleko za reportovaným state-of-the-art. Samotný podproblém NED by si zasloužil více prostoru.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Práce řeší velmi zajímavé téma. Nejzajímavějšímu podproblému se však text i experimenty věnují jen okrajově. První část práce (kap. 1-3) je velmi zdařilá, autor prokazuje dobrý přehled v doméně. Prostřední část (kap. 4) práce tvoří obecný úvod do neuronových sítí. Text k experimentální části (kap. 5) působí velmi zmateně a je nejslabší částí práce. Práce obsahuje nepřehledné množství formálních nedostatků.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **B** - velmi dobře.

Dotaz: Jak byste řešil problém velké množiny kandidátních entit (výstupu EL) pomocí dialogové NED?

Datum: 30/05/2018

Podpis: