

Review of Master's Thesis

Student: Jon Josef, Bc.
Title: Exploring Contextual Information in Neural Machine Translation (id 21979)
Reviewer: Fajčík Martin, Ing., UPGM FIT VUT

- 1. Assignment complexity** **more demanding assignment**

Študent musel v rámci práce preštudovať súčasný stav poznania v strojovom preklade založenom na neurónových sieťach (NMT) a zároveň preštudovať, akým spôsobom je možné do takýchto systémov zaintegrovať kontextovú informáciu. V rámci práce taktiež študent musel vytvoriť testovaciu sadu vhodnú pre vyhodnocovanie NMT s kontextom.
- 2. Completeness of assignment requirements** **assignment fulfilled**

Zadanie bolo splnené vo všetkých jeho bodoch.
- 3. Length of technical report** **in usual extent**

Práca je v obvyklom rozsahu.
- 4. Presentation level of technical report** **93 p. (A)**
 - Logická štruktúra práce je kvalitná. Úvod do stavu poznania je popísaný najprv z pohľadu tradičnej lingvistiky a až potom aj z pohľadu počítačovej lingvistiky. Druhá kapitola poskytuje čitateľovi prehľad súčasných metód strojového prekladu založenom na neurónových sieťach (NMT) a aj nedávnych prácach, ktoré sa snažili tieto metódy rozšíriť o reprezentáciu kontextu. V nasledujúcich kapitolách autor pokračuje v tejto kvalite. V kapitole 6 autor poskytuje k výsledkom rozsiahlu diskusiu, v ktorej poukazuje na silné aj slabé miesta vybraných metód.
 - Veľmi vzácné sa dá v práci naraziť na termíny a skratky, ktoré neboli vysvetlené a nie sú bežne používané mimo doménu práce (Truecasing, CAT systém, bootstrap resampling, ...).
 - V rovnici 4.3 bola nájdená chyba.
- 5. Formal aspects of technical report** **78 p. (C)**

Pri typografickej analýze boli nájdené nasledovné nedostatky:

 - Autor sa vzácné referuje ku kapitolám ich názvom, nie číslom.
 - U niektorých popisov obrázkov sa symbol "." presunul do nového riadku (viď obrázok 2.1 a 2.2).
 - U obrázku 2.5 popis v texte pod obrázkom nesúhlasí s obsahom obrázku, po diskusii s autorom sme odhalili chybu v texte.
 - Str.16 obsahuje neplatnú referenciu v texte ([?]).
 - Str. 21 symbol hviezdica (*) namiesto symbolu kartézskeho súčinu (×).
 - Obrázok 5.1 (str. 36) je príliš posunutý doľava.

Pri analýze jazykovej stránky boli odhalené nasledovné nedostatky:

 - str.18 termín "layers weights", mal pravdepodobne byť layers' weights?
- 6. Literature usage** **100 p. (A)**

Študent veľmi extenzívne popisuje veľké množstvo vysoko aktuálnych zdrojov. Veľké množstvo popisovaných zdrojov je z ročného rozmedia 2017-2019. Všetky uvedené zdroje spĺňajú citačné normy a zvyklosti.
- 7. Implementation results** **100 p. (A)**

Kód je veľmi prehľadne napísaný a okomentovaný. Časti riešenia obsahujú README súbory, ktoré popisujú funkciu súborov v priečinkoch. Riešenie je v súlade s autorským právom. Kód je verejne dostupný aj v online git repozitári pod licenciou GNU GPL.
- 8. Utilizability of results**

Práca prináša nové výsledky, ktoré preukazujú vlastnosti existujúcich metód pri nastavení rôznych parametrov. Práca napríklad ukazuje, že mnoho metód prináša zlepšenie iba na menších dátových sadách. Po doplnení experimentov popísaných v závere je práca vhodná pre použitie k publikačnej činnosti.
- 9. Questions for defence**

-
- 10. Total assessment** **93 p. excellent (A)**

Diplomová práca obsahuje kvalitnú logickú štruktúru, do detailu popisuje netriviálne nedávno publikované techniky neurálneho strojového prekladu, experimentuje s nimi a v rozsiahlej diskusii poukazuje na ich výhody a nevýhody. Autor tiež v niektorých prípadoch aplikoval techniku štatistického testovania hypotéz a ukázal, v ktorých prípadoch sú dosiahnuté výsledky štatisticky významné. Napriek nedostatkom vo formálnej úprave považujem prácu vzhľadom na úroveň obtiažnosti zadania a dosiahnuté výsledky za vysoko nadpriemernú.

In Brno 29. May 2019

Fajčík Martin, Ing.
reviewer