Master Thesis Review

Faculty of Mathematics and Physics, Charles University

Thesis author	Bohdan Ihnatchenko
Thesis title	Multi-Target Machine Translation
Year of submission	2020
Study programme	Informatika **Study branch** Matematická lingvistika

Review author	doc. RNDr. Ondřej Bojar, Ph.D. **Role** vedoucí
Department	Ústav formální a aplikované lingvistiky

Review:

The master thesis submitted by Bohdan Ihnatchenko explores multilingual machine translation systems, specifically focussing on translating into multiple target languages at once. The motivation lies in supporting multilingual environments, including international conferences, where the same content needs to be quickly made available in many languages. While this goal can be achieved simply with a large collection of standard pairwise systems, savings in resources (disk space for models and GPUs at runtime) are more than desirable.

The thesis is structured into five chapters plus introduction and conclusion. After the necessary review of the background, the setup of the experiments is described in depth. Chapters 3 and 4 are devoted to the baselines and two linguistically-motivated language groupings, respectively. The results are discussed in Chapter 5.

Overall, the thesis text is shorter than I would have liked, but this is offset by the magnitude of experimentation that Bohdan had to carry out: one set of experiments relies on a dataset for translating from English into 36 languages, another set of experiments is English-to-5. Bohdan experimented with up to 5 target languages in a model, so the choice of languages alone allows for far too many possible configurations to run. Bohdan was thus sampling over this space, but still had to cope with too many models to keep them on disk. The experiment management alone was very complex, requiring automatic stopping and deletion of models as soon as theirs scores were recorded. Monitoring a large number of experiments running also required custom-made tools to quickly check if all training jobs are running smoothly. Added to this was the complication that one of the datasets we relied upon was split into training and test data in a way which invalidated multilingual experiments and Bohdan had to rerun a large number of his experiments after he discovered this.

Bohdan's experiments with a random selection of target languages in the model are in line with recent related work: the more languages in a fixed-size model, the worse the results. We kept

the model size unchanged, to avoid yet another experimental dimension. Later results by Google indicate that if the model size grows by up to three orders of magnitude, gains are observed. The truth is though that with this big models, no GPU memory saving is actually possible.

The results with targetting related languages (tested on two groupings: Germanic languages, and Slavic languages with Cyrillic) are more diverse. The common result is that a decrease in quality is observed with more languages in the mix, but exceptions exist where the loss is smaller or some gain is obtained, such as spontaneous speech domain, small-data baselines or shared target Cyrillic script.

Bohdan carried out a very large number of experiments and described all of the clearly so that the work can be reproduced (subject to the randomness in configuration selection and training). He also did a good job in finguring out how to best present the results in a concise way and provided some first observations. As said, further analysis of these results would be highly desirable, including a small manual evaluation and comparison of outputs, but I confirm that the amount of work Bohdan performed is definitely in line other master theses at our department. It was the extensive but necessary preparatory work that took so much more time that Bohdan did not get to further analysis.

Unfortunately, the text quality is affected by the relatively short time in which it was written. There are occasional typos (e.g. “foolowing”) or missing references to tables. These errors do not compromise the understandability of the text overall.

As documented by Bohdan’s thesis, Bohdan can carry out research on his own, manage a very large number of concurrent experiment runs, obtain comparable scores from them and interpret and present the results concisely. More details could have been provided in the text, on experiments and esp. in the final discussion, but already the submitted version is undoubtedly sufficient to meet the requirements for master theses at Charles University. I thus recommend the thesis to be accepted.

I recommend the thesis to be accepted.

I do not propose the thesis for special recognition.

In Prague, 6. 9. 2020

Signature: