Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Bohdan Ihnatchenko
Název práce	Multi-Target Machine Translation
Rok odevzdání	2020
Studijní program	Informatika **Studijní obor** Matematická lingvistika

Autor posudku	Ing. Tom Kocmi, Ph.D. **Role** Oponent
Pracoviště	Microsoft, Münich

Text posudku:

This master thesis investigates multi-target machine translation. The main problem with multi-target systems is that adding more languages decreases the performance of the system as a whole. The author proposes to group languages based on the language family in contrast to a random set of languages. He presents a nice summary of current works.
In the evaluation, he showed that grouping languages from the same language family preforms in most cases better than having the same number of languages across a random set of languages, even when using similar languages with different writing script (Polish and Russian). The discussion is sound, and I agree with the findings.

The idea to measure stalled steps is excellent! It is an interesting way to visualise how much is the model improving which I have not seen before despite its simplicity.

I am also excited to see that author uses an experiment manager (EM) to accumulate and evaluate results. Setting up an experiment manager is a time-consuming process. However, it dramatically improves researchers ability to evaluate a much larger number of experiments and notice outliers or non-standard behaviour than when done manually. Moreover, the author showed his skills also to use EM productively. For example, when he is evaluating behaviour in section 2.5.4 and discussing what each outlier could mean. Also, he can diagnose if there is an error in the training process. Moreover, this pipeline should not be lost after thesis defence, but it could be used in research.
It is refreshing to see the use of EM because it is not a standard in the machine learning community, where researchers often investigate runs manually or evaluate only the final BLEU performance. Therefore I would like to stress the significance of this part of the thesis.

As for formal criteria, despite occasional typos, grammatical errors, or unusual compound words the work is written with very good English. In general, the structure of the thesis could be improved. The author mixes general description, theory and implementation (e.g. section 2.4). Some concepts are used before they are defined (e.g. early stopping); some concepts are used without definition. Additionally, I would expect a longer description of the used Transformer model. There are a few missing/broken tables/figures (e.g. page 14), and some tables are not referenced in text (e.g. table 3.3). Formating of the thesis could be improved, when tables and figures would be closer to the page where they are referenced. Lastly, the author mixes a terminology, for example, in MT, we cannot say a language is high-resource, but a *language pair* can be. Moreover, some figures would benefit from a more detailed description (especially those in chapter 4).
Overall, the formal issues are minor and should not influence the final defence decision.

Questions for the defence (the first question is most important):
(page 39, 40) Can you explain why models with more languages have a larger vocabulary? You correctly describe the algorithm on page 5 that it takes vocabulary size as an initial parameter. Thus, it always generates a vocabulary of a specified size. Therefore I do not understand how do you analyse the increase in vocabulary size based on languages in the set.
(page 14) How did you check for overlapped sentences? Only an exact match or also small perturbances in characters? Furthermore, did you also checked for overlapped sentences in testset?
(sec 2.4.3) You are using the same setup as Popel, but I have not found out what optimiser and learning rate do you use? Both of them are important parameters in Popel's setup. Especially Adafactor optimiser.

I am strongly confident that this is an excellent master thesis, and I recommend it for defence.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhujete na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).
I would definitelly recommend it to be published as a scientific paper.

Datum 21. Srpna 2020 **Podpis**