
The Charles University in Prague has invited me to act as an opponent in the Doctoral Thesis defense of Mgr. Aleš Tamchy. In this letter, I state my opinion on his Thesis “Lexical and Morphological Choices in Machine Translation”.

The Thesis consists of nine sections. The first four sections are introductory. Section 1 sketches up the structure of the Thesis, section 2 introduces into phrase-based machine translation, section 3 describes the English-Czech state-of-the-art system, section 4 introduces into machine learning background. Section 5 is the original contribution of the work and it describes the discriminative translation model utilizing source and target context. Section 6 contains integration into the phrase-based translation system. Section 7 describes experimental evaluation. Related approaches, their similarities and differences are described in section 8. Section 9 concludes the work presented in the Thesis and suggests future research directions. The work is well structured and clearly written, it has sufficient extent of 111 pages.

Lexical and morphological choices are highly relevant tasks. The first one is crucial for MT in general, the second one for languages relying on morphology. The authors propose a discriminative model operating on the level of phrase pairs, utilizing the current context. The model uses linguistic features extracted from the full source sentence and several preceding target words. The discriminative model is integrated as one additional feature into Moses log-linear model. The method is properly tested using automatic metrics on three language pairs and on different training data sizes.

I am attaching a list of remarks which occurred to me.

1) This approach relies on left-to-right decoding. What about extension to bottom-up (chart) decoding techniques used by Hiero (Chiang, 2005) or T2S, S2T, T2T?

2) Taking the source segmentation and reordering as given simplifies the training. On the other hand, many phrase pairs are not considered due to word alignment and rule extraction errors. Can we generalize this method to improve also word alignments?

3) Section 5.1: A simplified version of Figure 5.1 showing an example of a sentence, shared and translation features, and their interaction with translation scoring would improve readability.

4) Experiments were carried out in different setups and training data sizes. Integration of all results into one comprehensive table, adding results from multiple test sets, statistical significance (BLEU scores are often very close) would improve readability.
5) Baseline speed is reported as 0.8 sec/sec on page 67 vs. 0.28 s/sec in Table 7.4.

6) Table 7.4. suggests that context size 1 or 2 is the best, but context size 3 is even worse than context size 0. This is very disturbing. It suggests that the target features do not work. Tuning might be the culprit (as author suggests). Is that true for other language pairs? Are the small changes in BLEU statistically significant? This deserves thorough investigation.

7) It is important to state that the thesis proposal defense took place (the opponent was present) a couple of months before the first NMT works were published. Recent advances in the field of NMT have most likely outperformed the techniques presented in this work (as the author also suggests on several occasions). It does not diminish the relevance of this work, but it would be fair to compare to these results during this defense as well.

I hope that some of the comments above will be addressed during the defense. I believe that the exploration described in this work is sound, the work has enough substance, the proposed method is properly implemented and evaluated. In addition, I am familiar with author’s academic work, and aware of his repeated successes in the Czech-English WMT evaluation track.

I believe that the Thesis is a novel contribution to the field of statistical machine translation and that it clearly demonstrates the author’s ability to conduct research independently and to present its results.

In Prague, May 28th, 2017

Martin Čmejrek
IBM T. J. Watson Research, Yorktown Heights, NY and IBM Watson, Prague