

Master's Thesis Review

Author: Miloš Stanojević
Title: Large-Scale Discriminative Training for Machine Translation into Morphologically Rich Languages
Supervisor: RNDr. Ondřej Bojar, Ph.D.

The thesis submitted by Miloš Stanojević examines machine learning techniques applied in machine translation (MT) that aim to improve translation quality by considering large sets of features. These features are then supposed help to discriminate better and worse candidate translations.

The thesis is well structured and contains only a small number of grammatical or typesetting errors. All the experiments are clearly described in appropriate detail and the text reads well.

The main contributions of the thesis are:

- a solid introduction to generative vs. discriminative models as applied in MT,
- a detailed description of state-of-the-art optimization algorithms,
- a thorough discussion of objective functions used, including a comparison of several variants of BLEU and ROUGE-S with human judgments,
- a method of projecting source-side dependency tree through translation in order to support rich features in model optimization,
- experiments with English-to-Czech and English-to-Serbian MT comparing:
 - MERT vs. MIRA with various (large) sets of simple features,
 - the spaces of hypotheses available for discriminative training,
 - various sets of features utilizing dependency analysis,
 - various objective functions (BLEU, sBLEU, ROUGE-S2).

Sadly, none of the tested setups performed significantly better than the baseline, but this can still be considered a good result because e.g. the largest model includes 44 thousand features and despite of that, the learning algorithms are able to reach the state-of-the-art score on a rather small tuning set. All the failures are also very well discussed in the text.

To conclude, the thesis documents that Miloš Stanojević was able to explore discriminative training methods in depth as well as breadth, bring in his own suggestions, empirically test them and clearly describe and discuss the results. I think that this is as excellent outcome and I suggest the thesis to be accepted as a M.Sc thesis at Charles University in Prague.

Prague, January 17, 2013.

RNDr. Ondřej Bojar, Ph.D.
Charles University in Prague, ÚFAL