Review of Master Thesis

Author: Benno Weck
Title: Assessing the Impact of Manual Corrections in the Groningen Meaning Bank
Supervisor: RNDr. Markéta Lopatková, Ph.D.

The goal of the thesis is to investigate a possibility of improving a performance of a POS tagger using so called "Bits of Wisdom" (BOWs, manually verified POS tags as well as other human corrections of an automatic annotation of the Groningen Meaning Bank, GMB). The author experiments with the tagger re-training – he adopts several techniques for data sampling and compares the obtained results. In addition, he introduces a pilot study on building a 'smarter' tagger that can process data with some correct tags.
The thesis consists of seven chapters; it includes a rich list of references, list of abbreviations and list of POS tags. No CD with software, data and the thesis text is attached.

After a short introductory chapter, the second chapter presents the research problem – it mentions the GMB corpus, NLP tools used for its annotation, and available corrections (BOWs); next chapter shortly reports on possible ways of correcting a corpus annotation and related work.
The core of the thesis is formed by chapters four and five. The fourth chapter starts with the problem of missing gold-standard data, i.e., manually corrected data which can serve for (re-) training of the tagger or for evaluation. As an inspiring solution, a testbed of so called silver-standard data is created – testing data from other corpora (WSJ and MASC) were enriched with sentences from selected GMB sections containing BOWs. Then a method used in the experiments is described: the method consists in re-training the original POS-tagger (i.e., the tagger used for the original annotation of the GMB corpus) using BOWs as additional training data. Stress is laid on different sampling strategies, i.e., how to select additional training data for most effective learning – the following approaches are tested: random data selection as a baseline, then longest-sentence-first selection and selection preferring sentences with high ratio of corrected tags on one side and active learning (voting by uncertainty of the tagger and voting by different systems) on the other side.
The fifth chapter focuses on results and their interpretation – it shows that data sampling preferring sentences with high ratio of BOWs and active learning based on uncertainty are most effective; both methods give good results using just part of available corrected data. These results suggest that sentences with high tagger uncertainty are best candidates for additional human annotation as their reliable annotation has the greatest impact on the improvement of the tagger accuracy. On the other hand, the experiments do not confirm an expectation that better POS tagging would imply better parsing.
The sixth chapter presents an additional (somehow inorganically appended) pilot study – it introduces and qualitatively evaluates a tagger that works on data where available corrected/verified tags are fixed.
The thesis is concluded with short summary, recommendation based on the experiments and possible topics for further investigation.
I can commend Benno especially for extensive bibliographic research (the bibliography contains 70 cited items; many of them are shortly presented in the thesis). Further, Benno properly and in detail discusses results of the experiments, he tries to interpret the findings and draw conclusions for effective improvement of the data.

The thesis has a clear and logical structure. It is written in decent English, I have only detected a small number of grammatical errors, which do not harm comprehension.

Question for discussion:
I do not understand why the active learning approaches give better results with reduced set of BOWs available (Figures 5.3 and 5.4). Can you explain the difference in sampling?

Conclusion
The reported thesis brings interesting results concerning possibilities of improving a performance of a POS tagger. It proves the author's ability to solve independently and creatively assigned tasks in the area of NLP and to clearly formulate his goals, used methods and the gained results.
The thesis complies with the requirements for Master Thesis at MFF. I recommend to accept the thesis for the defense.

Prague, January 27, 2016

doc. RNDr. Markéta Lopatková, Ph.D.
Institute of Formal and Applied Linguistics
Charles University in Prague

Excellent thesis suitable for student competitions NO