Review of Master Thesis

Title: Assessing the impact of manual corrections in the Groningen Meaning Bank
Author: Benno Weck

Thesis

The goal of the thesis is exploring the possibilities of part-of-speech tagger re-training in order to improve annotations.

The submitted thesis is well and clearly structured. It contains seven chapters: in Introduction, the author explains the scope and objectives of his work. In Chapter 2, he provides some background of the Groningen Meaning Bank (GMB) and the C&C tools he experimented with. Also he describes the task of part-of-speech tagging. Then in the third chapter, he reviews related work on non-expert annotation, active learning, and error detection in annotated corpora. The core of the student's work is presented in next three chapters. In Chapter 4 he describes his framework for investigating whether a part-of-speech tagger can be re-trained using so called Bits of Wisdom to improve annotations of GMB. The investigation process concerns (i) designing of training and test sets, (ii) sampling methods for selecting additional training examples, (iii) the active learning approach, and (iv) an impact of part-of-speech tagging errors on syntactic parser performance. The results of experiments are presented in Chapter 5. The sixth chapter is devoted to building a part-of-speech tagger that can handle true labels (tags) of tokens if they are available. Finally, Chapter 7 contains a discussion and some ideas for future work.

The list of bibliographies consists of 70 items. No data and software packages are attached to the thesis. The thesis is written in understandable English.

Comments and Questions

1. The overall impression from the thesis is good. However, I would recommend to use tables or schemes instead of a number of paragraphs for description of some concepts. For example, setting the training and test data in Section 4.1.1 can be easily visualized using a scheme. In addition, presenting more examples would help a reader to understand the text a lot.
2. In Chapter 1, p. 5: "*There are many different layers of annotation with existing corrections in the GMB, but for this thesis we will concentrate on the POS level of annotation. The choice of this level … by two reasons: First, this level has the highest amount of corrections in comparison to the other levels ...*" I am curious why it happens.
3. In Section 4.1.2, there is a description of creating a silver-standard dataset. I am not sure whether the GMB training data set was created in the same way.
4. In Section 4.4, I have some doubt as to whether the definition of coverage is correct.
5. In Chapter 5, it would be useful to do more detailed analysis of the (un)matched BOWs, i.e. to see problematic tokens/tags/contexts.
6. In Section 5.3, what was the test data?
7. What was the reason for experimenting with the Maximum Entropy tagger? Have you thought of working with other algorithms?
8. What are your original contributions?

Conclusion

I recommend the submitted thesis for the defense and I suggest accepting it as a Master thesis.

Prague, January 28, 2016

Barbora Hladká

Institute of Formal and Applied Linguistic
Charles University in Prague