Opponent's Review of Master Thesis

Author: Pranava Swaroop Madhyastha
Title: Exploring Higher Order Dependency Parsers
Supervisors: Prof. Michael Rosner & RNDr. Daniel Zeman

Review Content:

The thesis investigates how semantic and morphological features affect the accuracy of higher order (second and third) dependency parsing models. The author conducted unlabeled dependency parsing experiments by including various semantic and morphological features. The experimental results have been reported for English and Czech languages. The thesis contains 8 Chapters, an appendix and bibliography. Overall, the thesis is well organized.

The author clearly articulates his thesis objective in Chapter 1 and provides necessary theoretical background in Chapter 2. In Chapter 3, the author briefly surveys the current research in adding various features in dependency parsing. The Chapter 4 adequately describes about the data sets, data format and tools used in the experiments. However, the treatment of feature extraction is very brief. In Chapter 5, the author mainly conducted three experiments by incorporating: (1) fine grained semantic features (2) coarse grained semantic features and (3) morphological features with the second/third order dependency parsing models. The results for these experiments have been reported in Tables. The experimental results corroborates with the author's claims. However, this Chapter is somewhat less detailed and the author should have given more attention to detailed explanation of preprocessing steps involved in the experiments, for ex: listing of semantic files. Moreover, the examples provided ("I enjoy watching …", "Path of …") for experiments 1 & 2 are not convincing and should have been more concrete. The Chapter 6 compares the experimental results with the existing research work. The Chapters 7 & 8 concluding the thesis. The Appendix contains intermediate results obtained during the training of dependency models. The author conducted an extensive survey on his research topic and has included all relevant publications in the bibliography section.

The accompanying DVD of the thesis contains thesis source/pdf, author's scripts and third party parser/scripts used in the experiments. The DVD lacks adequate documentation on running programming codes. I recommend the author to add necessary information on how to run the scripts and how to reproduce the experimental results.

There are some minor mistakes in the thesis,

- Missing word "Figure" in figure references on page 8 and 32.
- The Chapter 2 abruptly ends.
- The definition of "argmax" is wrong for equation 2.3.
- Missing citations on page 21 (2nd para) and page 25.

- Spelling errors: On page 37, “tthese” → “these”, “witch” → “with”.
- On page 8: “the the …”, on page 23: “the the …”

Questions:

1. The inclusion of semantic word senses can be greatly helpful in determining dependency relations between word forms (i.e. labels), in view of that, how do you justify the word senses can be helpful in determining the structure? (i.e. heads and modifiers)
2. The baseline includes only the coarse grained POS tags. Is there any reason why fine-grained POS tags are not used in the baseline?
3. When looked at the sample data (from Experiment 2), the semantic files are actually mapped from fine-grained POS tags, in that case, (a) are the semantic files simply fine grained POS tags? and (b) how does the Experiment 2 differ from Experiment 3 for English?
4. The phrase “statistically significant” (for ex: on page 41) has a different meaning in statistics. Do you perform any statistical tests to determine whether the results are “statistically significant”?

Conclusion:

I recommend this thesis for defense.

Prague, 20.1.2012

Loganathan Ramasamy,
Institute of Formal and Applied Linguistics,
Charles University in Prague.