

Shallow Discourse Parser Comparison and Combination

Paul Opuchlich, 794745
Malte Klingenberg, 794394

Introduction

The field of text discourse theory has been around for several decades, with first theoretical approaches such as Rhetorical Structure Theory (RST) developed in the 1980s [Mann and Thompson, 1988]. Advancements in machine learning have fueled the development of many different kinds of discourse parsers, but while significant improvements have been made in recent years, the existing systems are still far from perfect. Explicit discourse relations (those marked by a connective such as "and", "because", "after" etc.) can be identified by modern systems with high accuracy, but implicit and entity relations pose a far greater challenge. Because the argument spans of discourse relations can vary wildly (from single clauses to several paragraphs), identifying them is another difficult task.

Because the many available parsers are based on different machine learning algorithms, for example support vector machines (SVMs) and different types of neural networks such as convolutional (CNN) and recurrent (RNN), there might be systematical differences in their performances on different tasks. It is not unreasonable to assume that, for example, one machine learning algorithm might perform better at classifying implicit senses, while another algorithm might be best suited to identifying argument spans.

Our goal in this project was to compare the performances of several different parsers and check whether any systematic patterns can be found. We also tried to find a way to combine several parsers in a way that gives better results than the individual parsers to see whether a "crowd intelligence" approach of combining different systems can be a good way to improve performance.

Dataset

CoNLL challenge [Xue et al., 2016]

- incomplete data (sense-only), generate dummy data (see below)

- used parsers (differences, some only guess a limited number of senses)

Approach

Preprocessing and Randomising

- Mapping the Relations (using CoNLL methods)

- Calculate Parameters for each Parser + Relation, some evaluation

- Create combined model by different voting approaches (best-wins, weighted...)

Evaluation

general overview

- agreement between three parsers
- compare performances of all parsers
- relationship between architecture and results?
- combination (different approaches)

Conclusion

- parsers are(n't) good at different things
- combination (does not) improve results

References

- [Mann and Thompson, 1988] Mann, W. C. and Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.
- [Xue et al., 2016] Xue, N., Ng, H. T., Pradhan, S., Rutherford, A., Webber, B., Wang, C., and Wang, H. (2016). Conll 2016 shared task on multilingual shallow discourse parsing. *Proceedings of the CoNLL-16 shared task*, pages 1–19.