

Exploring Higher Order Dependency Parsers

Pranava Swaroop Madhyastha
European Masters Program in
Language and Communication Technologies.
2009-2011

October 31, 2011

Abstract

Parsing is one of the most important steps in understanding of natural languages. In this thesis, we focus on the dependency grammar formalism since, the core concepts of dependency grammar, namely the relational view of head and modifier asymmetry, has proven useful for diverse set of languages, especially accounting for the explanation of word order and relation between surface structure and meaning. Most of the recent efficient algorithms for dependency parsing work by factoring the dependency trees. In most of these approaches, the parser loses much of the contextual information during the process of factorization. In this thesis we investigate how features (syntacto-semantic) affect the higher order discriminative learning methods for dependency parsing. We will show that linguistic features in most cases provide a significant improvement in the parsing accuracy.

We start by presenting a survey on several discriminative learning methods for graph based statistical dependency parsers and explain the concept of higher order that is the generalization of the work done by [Koo et al., 2008] and [McDonald et al., 2006]. This leads us to the core of the thesis - feature engineering in higher order dependency parsers. Here, we experiment with several syntacto-semantic features then try to explain the theoretical foundation of these features. The experiments are done on two diverse languages - English and Czech, we have compared the several results obtained with different parsing algorithms.

References

- [Koo et al., 2008] Koo, T., Carreras, X., and Collins, M. (2008). Simple semi-supervised dependency parsing. In *Proceedings of 46th Annual Meeting of The Association for Computational Linguistics*.
- [McDonald et al., 2006] McDonald, R., Lerman, K., and Pereira, F. (2006). Multilingual dependency analysis with a two-stage discriminative parser. In *Proceedings of the Tenth Conference on Computational Natural Language Learning, CoNLL-X '06*, pages 216–220, Stroudsburg, PA, USA. Association for Computational Linguistics.