

11-712: NLP Lab Report

William Yang Wang
ww@cmu.edu

January 19, 2014

Abstract

Dependency parsing is a core task in NLP, and it is widely used by many applications such as information extraction, question answering, and machine translation. In general, the resources for Chinese dependency parsing are less accessible than English, and publicly available Chinese dependency parsers are still very limited. In this project, the goal is to build a Chinese dependency parser that can be used by others.

1 Basic Information about Chinese Dependency Parsing

Chinese dependency parsing has attracted many interests in the past decade. Bikel and Chiang (2000) and Chiang and Bikel (2002) are among the first to use Penn Chinese Tree Bank for dependency parsing, where they adapted Xia (1999)’s head rules. A few years later, the CoNLL shared task opened a track for multilingual dependency parsing, which also included Chinese (Buchholz and Marsi, 2006; Nilsson et al., 2007). These shared tasks soon popularized Chinese dependency parsing by making datasets available, and there has been growing amount of literature since then (Carreras, 2007; Che et al., 2010; Duan et al., 2007; Nivre et al., 2007; Sagae and Tsujii, 2007; Zhang and Clark, 2008). In this work, we aim at building a new publicly available Chinese dependency parsing, using new technologies that aim at improving the accuracy of the state-of-the-art.

- 2 Past Work on the Syntax of Chinese
- 3 Available Resources
- 4 Survey of Phenomena in Chinese Dependency Parsing
- 5 Initial Design
- 6 System Analysis on Corpus A
- 7 Lessons Learned and Revised Design
- 8 System Analysis on Corpus B
- 9 Final Revisions
- 10 Future Work

References

- Daniel M Bikel and David Chiang. Two statistical parsing models applied to the chinese treebank. In *Proceedings of the second workshop on Chinese language processing: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 12*, pages 1–6. Association for Computational Linguistics, 2000.
- Sabine Buchholz and Erwin Marsi. Conll-x shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, pages 149–164. Association for Computational Linguistics, 2006.
- Xavier Carreras. Experiments with a higher-order projective dependency parser. In *EMNLP-CoNLL*, pages 957–961, 2007.
- Wanxiang Che, Zhenghua Li, and Ting Liu. Ltp: A chinese language technology platform. In *Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations*, pages 13–16. Association for Computational Linguistics, 2010.
- David Chiang and Daniel M. Bikel. Recovering latent information in treebanks. In *Proceedings of the 19th International Conference on Computational Linguistics - Volume 1*, COLING ’02, pages 1–7, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1072228.1072354. URL <http://dx.doi.org/10.3115/1072228.1072354>.
- Xiangyu Duan, Jun Zhao, and Bo Xu. Probabilistic parsing action models for multi-lingual dependency parsing. In *EMNLP-CoNLL*, pages 940–946, 2007.
- Jens Nilsson, Sebastian Riedel, and Deniz Yuret. The conll 2007 shared task on dependency parsing. In *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL*, pages 915–932. sn, 2007.
- Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135, 2007.
- Kenji Sagae and Jun’ichi Tsujii. Dependency parsing and domain adaptation with lr models and parser ensembles. In *EMNLP-CoNLL*, volume 2007, pages 1044–1050, 2007.
- Fei Xia. Extracting tree adjoining grammars from bracketed corpora. In *Proceedings of the 5th Natural Language Processing Pacific Rim Symposium (NLPRS-99)*, pages 398–403, 1999.
- Yue Zhang and Stephen Clark. A tale of two parsers: investigating and combining graph-based and transition-based dependency parsing using beam-search. In *Proceedings of the Conference on Em-*

pirical Methods in Natural Language Processing, pages 562–571. Association for Computational Linguistics, 2008.