

Liang Huang, University of Pennsylvania, USA K-best Algorithms in Parsing and Machine Translation

Joint work with David Chiang (USC Information Sciences Institute, USA)

The general idea of k-best search is to find the top k solutions of a given problem, since in many cases the 1-best solution is not guaranteed to be optimal given more (global) information. Recently it has become a popular technique in natural language processing, especially in parsing and machine translation. However, fast and exact k-best algorithms are largely unknown to the NLP community. In this work, we develop a series of efficient algorithms for exact k-best search in the general framework of directed hypergraphs, and demonstrate their performance on state-of-the-art statistical parsers.

These algorithms have also been successfully applied to search problems in machine translation. In particular, I will talk about an adaptation called "forest rescoring", for decoding with integrated language models, which achieved more than ten fold speed-up against conventional beam search on both phrase-based and syntax-based translation systems.

References:

- Liang Huang and David Chiang (2005). Better k-best Parsing. In Proceedings of the 9th International Workshop on Parsing Technologies (IWPT). <http://www.cis.upenn.edu/~lhuang3/ Huang-iwpt-correct.pdf>
- Liang Huang and David Chiang (2007). Faster Algorithms for Decoding with Integrated Language Models. To appear in Proceedings of ACL 2007, Prague, Czech Rep. <http://www.cis.upenn.edu/~lhuang3/acl-cube.pdf>

Bio: Liang Huang is a fourth-year PhD student at the University of Pennsylvania, under Aravind Joshi. He is mainly interested in the theoretical aspects of computational linguistics, in particular, efficient algorithms in parsing and machine translation, and formal properties of synchronous grammars. He also works on applying computational linguistics to structural biology.