

Statistical Translation Model Based On Source Syntax Structure*

Qun Liu, Yang Liu and Haitao Mi

Key Laboratory of Intelligent Information Processing
Institute of Computing Technology, Chinese Academy of Sciences
Beijing 100190, P.R.China
{liuqun, yliu, htimi}@ict.ac.cn

Abstract. Syntax-based statistical translation model is proved to be better than phrase-based model, especially for language pairs with very different syntax structures, such as Chinese and English. In this talk I will introduce a serial of statistical translation models based on source syntax structure. The tree-based model uses the one best syntax tree for translation. The forest-based model uses a compact forest which encodes exponential number of syntax trees in a polynomial spaces and lead to better performance. The joint parsing and translation model produces source parse trees, using the source side of the translation rules instead of separate parsing rules, and generate translations on the target side simultaneously, which outperforms the forest-based model. Some extensions of these models are introduced also.

Keywords: statistical machine translation, translation model, syntax-based model.

References

- Zhongjun He, Qun Liu, Shouxun Lin. Improving Statistical Machine Translation using Lexicalized Rule Selection. Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008), pages 321-328. Manchester, UK. Aug, 2008
- Qun Liu, Zhongjun He, Yang Liu, Shouxun Lin. Maximum Entropy based Rule Selection Model for Syntax-based Statistical Machine Translation. Proceedings of EMNLP 2008, pp 89-97, Honolulu, Hawaii
- Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-String Alignment Template for Statistical Machine Translation. Proceedings of COLING-ACL 2006, pp 609 – 616, Sydney, Australia, July 17-21.
- Yang Liu, Yun Huang, Qun Liu and Shouxun Lin, Forest-to-String Statistical Translation Rules, ACL2007, pp 704-711, Prague, Czech, June 2007
- Yang Liu, Yajuan Lü and Qun Liu, Improving Tree-to-Tree Translation with Packed Forests. Proceedings of ACL-IJCNLP2009, pp 558-566, Singapore, August 2-7, 2009
- Yang Liu, and Qun Liu. Joint Parsing and Translation. Proceedings of COLING 2010, pp 707-715, Beijing, China, August.
- Haitao Mi, Liang Huang and Qun Liu. Forest-Based Translation. Proceedings of ACL-08:HLT, pp192-199, Columbus, Ohio, USA. 2008
- Haitao Mi and Liang Huang, Forest-based Translation Rule Extraction. Proceedings of EMNLP 2008, pp 206-214, Honolulu, Hawaii

* This research is supported by the National Science Foundations of China (No. 60736014) and the High Technology Research and Development Program of China (No. 2006AA010108).