本刊检索　GO
高级检索　检索说明

| 首 页 | 关于我们 | 投稿指南 | 征订服务 | 诚邀合作 | 留 言 |

知识组织与知识管理

最新目录 | 下期目录 | 过刊浏览 | 高级检索　<< Previous Articles | Next Articles >>

树转录翻译模型解码优化

石崇德, 乔晓东, 王惠临

中国科学技术信息研究所 北京 100038

Shi Chongde, Qiao Xiaodong, Wang Huilin

Institute of Scientific & Technical Information of China, Beijing 100038, China

- 摘要
- 参考文献
- 相关文章

Download: PDF (487KB)　HTML (1KB)　Export: BibTeX or EndNote (RIS)　Supporting Info

摘要 针对树转录翻译模型中的规则二元化和解码算法进行深入研究,通过四分化的二元化转换方法减少词汇化同步转录规则的中间项目,通过实时判断中间项目有效性的RR-CKY算法来避免冗余项目生成。实验证明,这两种方法能有效减少解码过程中的中间项目,提高机器翻译解码效率,在一定程度上提高机器翻译效果。

关键词： 机器翻译　树转录翻译模型　句法分析　RR-CKY算法

Abstract： This paper proposes two methods to improve the efficiency of rule binarization and decoding in tree transducer based translation model. The authors convert synchronous transducer rules to four kinds of binary rules to reduce the temporary items, and propose RR-CKY decoding algorithm, which can avoid part of redundant items along with decoding. The experiments show that these two methods can reduce the number of temporary items and make decoding faster. They can also improve the quality of machine translation.

Keywords： Machine translation,　Tree transducer based translation model,　Parsing,　RR-CKY algorithm

Service
- 把本文推荐给朋友
- 加入我的书架
- 加入引用管理器
- Email Alert
- RSS

作者相关文章
- 石崇德
- 乔晓东
- 王惠临

[1]　Wu D. Toward Machine Translation with Statistics and Syntax and Semantics[C].In: *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU'09)*, Merano, Italy. 2009: 12-21.

[2]　Chiang D. Hierarchical Phrase-based Translation[J].*Computational Linguistics*,2007, 33(2):201-228.

[3]　Marcu D, Wang W, Echihabi A, et al. SPMT: Statistical Machine Translation with Syntactified Target Language Phrases[C].In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*,Sydney,Australia. 2006:44-52.

[4]　刘洋. 树到串统计翻译模型研究[D].北京:中国科学院计算技术研究所,2007.(Liu Yang. Research on Tree-to-String Statistical Translation Models[D]. Beijing: Institute of Computing Technology, Chinese Academy of Sciences,2007.)

[5]　蒋宏飞. 基于同步树替换文法的统计机器翻译方法研究[D]. 哈尔滨:哈尔滨工业大学,2010. (Jiang Hongfei. Research on Synchronous Tree Substitution Grammar Based Statistical Machine Translation Methods[D]. Harbin: Harbin Institute of Technology,2010.)

[6]　宗成庆. 统计自然语言处理[M].北京:清华大学出版社,2008.(Zong Chengqing. Statistical Natural Language Processing[M]. Beijing: Tsinghua University Press,2008.)

[7]　Zhang H, Huang L, Gildea D, et al. Synchronous Binarization for Machine Translation[C].In: *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics,2006:256-263.

[8] Wang W, Knight K, Marcu D. Binarizing Syntax Trees to Improve Syntax-based Machine Translation Accuracy[C].In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*,Prague, Czech Republic. Stroudsburg, PA, USA: Association for Computational Linguistics, 2007: 746-754.

[9] Fang L, Chung T, Gildea D. Terminal-aware Synchronous Binarization[C].In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oregon, USA. 2011: 401-406.

[10] The Penn Treebank Project[DB/OL]. [2013-06-15]. http://www.cis.upenn.edu/~treebank/.

[11] Collins M. Head-driven Statistical Models for Natural Language Parsing[D]. Philadelphia: University of Pennsylvania,1999.

[12] Charniak E. A Maximum-Entropy-Inspired Parser[C].In: *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference.* 2000:132-139.

[13] Klein D, Manning C D. Accurate Unlexicalized Parsing[C].In: *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics.*2003:423-430.

[14] Song X, Ding S, Lin C Y. Better Binarization for the CKY Parsing[C].In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*,Honolulu,Hawaii,USA. 2008:167-176.

[15] Schmid H. Efficient Parsing of Highly Ambiguous Context-free Grammars with Bit Vectors[C].In: *Proceedings of the 20th International Conference on Computational Linguistics.*2004.

[16] Fox H J. Phrasal Cohesion and Statistical Machine Translation[C].In: *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing.* Stroudsburg,PA,USA: Association for Computational Linguistics,2002:304-311.

[17] Galley M, Hopkins M, Knight K, et al. What's in a Translation Rule?[C].In: *Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (HLT-NAACL 2004)*,Boston, Massachusetts, USA. 2004:273-280.

[18] Graehl J, Knight K, May J. Training Tree Transducers[J]. *Computational Linguistics*, 2008,34(3):391-427.

[19] Goodman J. Semiring Parsing[J]. *Computational Linguistics*,1999,25(4):573-605.

[20] Venugopal A,Zollmann A,Vogel S. An Efficient Two-Pass Approach to Synchronous-CFG Driven Statistical MT[C].In: *Proceedings of Human Language Technology and North American Association for Computational Linguistics Conference*,Rocherster, NY, USA. 2007:500-507.

[21] GIZA + +[CP/OL]. [2013-06-15]. http://code.google.com/p/giza-pp/.

[22] The Stanford Parser[CP/OL]. [2013-06-15]. http://nlp.stanford.edu/software/lex-parser.shtml.

[23] SRILM[CP/OL]. [2013-06-15]. http://www.speech.sri.com/projects/srilm/.

[24] NIST Open Machine Translation (OpenMT) Evaluation[DB/OL]. [2013-06-15]. http://www.itl.nist.gov/iad/mig//tests/mt/.

[25] Och F J. Minimum Error Rate Training in Statistical Machine Translation[C]. In: *Proceedings of the 41st Annual meeting on Association for Computational Linguistics,* Sapporo, Japan. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003:160-167.

[1] 袁冬, 熊晶, 刘永革.面向甲骨文的实例机器翻译技术研究[J]. 现代图书情报技术, 2012,28(5): 48-54

[2] 石崇德, 王惠临.统计机器翻译中文分词优化技术研究[J]. 现代图书情报技术, 2012,28(4): 29-34

[3] 王东波, 朱丹浩, 谢靖.面向汉语自动句法分析的语法知识库构建[J]. 现代图书情报技术, 2011,27(4): 42-47

[4] 孙镇 王惠临.命名实体识别研究进展综述[J]. 现代图书情报技术, 2010,26(6): 42-47

[5] 仲夏 张志平 王惠临.词汇化树邻接语法研究述评及中文应用初探*[J]. 现代图书情报技术, 2010,26(5): 35-42

[6] 丁志强 王惠临.基于Glue Semantics 和DRT的自然语言理解系统的设计与实现*[J]. 现代图书情报技术, 2010,26(4): 46-52

[7] 祝清松,王惠临.中英文句法分析系统及验证平台的设计与实现*[J]. 现代图书情报技术, 2010,26(2): 38-43

[8] 张亮,陈肇雄,黄河燕,马玉芝.面向应用的汉语句法分析辅助系统的设计与实现*[J]. 现代图书情报技术, 2006,22(1): 47-50

[9] 汤艳莉,赖茂生.Ontology在自然语言检索中的应用研究[J]. 现代图书情报技术, 2005,21(2): 33-36

[10] 王金铨.翻译记忆(TM)——计算机翻译技术的新发展[J]. 现代图书情报技术, 2004,20(5): 13-16

[11] 葛敬民.期待翻译软件的新突破[J]. 现代图书情报技术, 2002,18(1): 78-79

[12] 葛敬民.浅谈国内机器翻译软件[J]. 现代图书情报技术, 1998,14(3): 53-55

[13] 朱爱群.自动标引和自动文摘对机器翻译的影响[J]. 现代图书情报技术, 1997,13(1): 47-50

[14] 刘海涛.Internte 网上自然语言通讯问题初探[J]. 现代图书情报技术, 1996,12(4): 7-11

[15] 杜长征,李韧之,徐亭起.英汉图书题录微机自动翻译研究[J]. 现代图书情报技术, 1989,5(1): 31-34