

本期目录 | 下期目录 | 过刊浏览 | 高级检索

[打印本页] [关闭]

人工智能及识别技术

基于句法调序的汉维统计机器翻译

陈丽娟^{1,2}, 张恒^{1,2}, 董兴华^{1,2}, 吐尔洪·吾司曼¹, 周俊林³

(1. 中国科学院新疆理化技术研究所, 乌鲁木齐 830011; 2. 中国科学院研究生院, 北京100049; 3. 中国科学院新疆分院, 乌鲁木齐 830011)

摘要: 在汉语到维语的统计机器翻译中, 2种语言在形态学及语序上差异较大, 导致未知词较多, 且产生的维语译文语序混乱。针对上述问题, 在对汉语和维语的语序进行研究的基础上, 提出一种汉语句法调序方法, 进而对维语进行形态学分析, 采用基于因素的统计机器翻译系统进行验证。实验结果证明, 该方法在性能上较基线系统有显著改进, BLEU评分由15.72提高到19.17。

关键词: 统计机器翻译 句法调序 形态学 因素模型 翻译模型

Chinese-Uyghur Statistical Machine Translation Based on Syntactical Reordering

CHEN Li-juan^{1,2}, ZHANG Heng^{1,2}, DONG Xing-hua^{1,2}, Turghun Osman¹, ZHOU Jun-lin³

(1. Xinjiang Technical Institute of Physics & Chemistry, Chinese Academy of Sciences, Urumqi 830011, China; 2. Graduate University of Chinese Academy of Science, Beijing 100049, China; 3. Xinjiang Branch of Chinese Academy of Sciences, Urumqi 830011, China)

Abstract: Chinese and Uyghur are very different in terms of morphological typology and word order, which leads to many unknown words and confusion word order in Uyghur when translate from Chinese to Uyghur using statistical method. On the basis of the word order of Chinese and Uyghur, a Chinese syntactic reordering method is proposed, and an analysis on Uyghur morphological information is made to resolve the difficulties. Experimental results on the factor-based SMT show that the approach achieves a substantial improvement in translation quality over the baseline phrase-based system, and the BLEU score is improved from 15.72 to 19.17.

Keywords: Statistical Machine Translation(SMT) syntactical reordering morphological factored model translation model

收稿日期 2011-07-25 修回日期 网络版发布日期 2012-02-05

DOI: 10.3969/j.issn.1000-3428.2012.03.057

基金项目:

中国科学院西部行动计划高新技术基金资助项目(KGC X2-YN-507)

通讯作者:

作者简介: 陈丽娟(1985—), 女, 硕士研究生, 主研方向: 自然语言处理; 张恒, 硕士研究生; 董兴华, 博士研究生; 吐尔洪·吾司曼, 助理研究员; 周俊林, 研究员

通讯作者E-mail: chenlijuanyx@sina.com

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF(254KB)
- ▶ [HTML] 下载
- ▶ 参考文献[PDF]
- ▶ 参考文献

服务与反馈

- ▶ 把本文推荐给朋友
- ▶ 加入我的书架
- ▶ 加入引用管理器
- ▶ 引用本文
- ▶ Email Alert
- ▶ 文章反馈
- ▶ 浏览反馈信息

本文关键词相关文章

- ▶ 统计机器翻译
- ▶ 句法调序
- ▶ 形态学
- ▶ 因素模型
- ▶ 翻译模型


本文作者相关文章

- ▶ 陈丽娟
- ▶ 张恒
- ▶ 董兴华
- ▶ 吐尔洪·吾司曼
- ▶ 周俊林

PubMed

- ▶ Article by Chen, L. J.
- ▶ Article by Zhang, H.
- ▶ Article by Dong, X. H.
- ▶ Article by Tu, E. H. ?W. C. M.
- ▶ Article by Zhou, D. L.

参考文献:

[2] Elming J. Syntactic Reordering Integrated with Phrase-based SMT[C]//Proc. of the 22nd International Conference on Computational Linguistics. Manchester.[J].UK: [s. n.].2008,;-


[4] 易坤琇, 高士杰. 维吾尔语语法.[M]. 北京: 中央民族大学出版社, 1998.

[5] 董兴华, 周俊林, 郭树盛, 等. 基于短语的汉维(维汉)统计机器翻译系统[J]. 计算机工程.2011, 37(9): 16-18 [浏览](#)

[6] Xue Nianwen, Xia Fei. The Bracketing Guidelines for the Penn Chinese Treebank(3.0) [EB/OL]. (2000-11-12). <http://www.cis.upenn.edu/~chinese>.

[7] Li Jinji, Kim Dong-II, Lee Jong-Hyeok. Annotation Guidelines for Chinese-Korean Word Alignment[EB/OL]. (2008-05-05). <http://www.mt-archive.info/LREC-2008-Li.pdf>.

本刊中的类似文章

1. 王彦, 谢晓方, 吴龙宝, 马钰, 朱宗健, 王丰.基于概率模型的裂纹识别算法[J]. 计算机工程, 2012,38(2): 189-191
2. 李敏花, 柏猛.基于数学形态学的复杂背景图像文字检测方法[J]. 计算机工程, 2012,38(04): 165-167
3. 王小强, 陈临强, 梁旭.实时全自动视频拼接方法[J]. 计算机工程, 2011,37(5): 291-封3
4. 尚赵伟, 国庆, 马尚君, 袁博, 杨建伟.基于二进小波变换的多车牌定位算法?[J]. 计算机工程, 2011,37(3): 16-18
5. 王奇文, 郑丽敏, 梅树立.基于形态学的小鼠舌头切片图像分割与实现[J]. 计算机工程, 2011,37(19): 5-8
6. 梁媛媛, 温佩芝, 黄文明, 任亚恒, 朱叶青, 罗桂海.一种指纹图像的快速分割方法[J]. 计算机工程, 2011,37(17): 208-210
7. 蒋国瑞, 吴琳.基于灰色聚类的谈判对手选择模型[J]. 计算机工程, 2011,37(13): 172-174,177
8. 潘巍, 杨娜菲, 安荣.基于梯度与形态学的低质量车牌定位[J]. 计算机工程, 2011,37(13): 144-146
9. 董兴华, 陈丽娟, 周喜, 周俊林, 吐尔洪?吾司曼.汉维统计机器翻译中的形态学处理[J]. 计算机工程, 2011,37(12): 150-152
10. 胡霍真, 田金文.基于合成孔径雷达图像的机场主跑道检测方法[J]. 计算机工程, 2011,37(12): 201-203

文章评论

反馈人	<input type="text"/>	邮箱地址	<input type="text"/>
反馈标题	<input type="text"/>	验证码	<input type="text" value="5194"/>
<input type="text"/>			