

数据库、信号与信息处理

从双语语料中获取翻译模板

张春祥¹, 梁颖红², 于林森³

- 1. 哈尔滨理工大学 软件学院, 哈尔滨 150080
- 2. 苏州市职业大学 计算机系, 江苏 苏州 215104
- 3. 哈尔滨理工大学 计算机科学与技术学院, 哈尔滨 150080

收稿日期 2008-12-30 修回日期 2009-2-2 网络版发布日期 2010-1-7 接受日期

摘要 翻译模板自动获取是提高MT译文输出质量和领域适应能力的关键性因素。利用Tree-to-String方法抽取等价对, 使用错误驱动的学习方法从中获取翻译模板并进行优化。将优化后的翻译模板用于一个基于转换的机器翻译系统中, 同时使用“863”对话语料对其进行评测。实验结果表明: 当使用自动获取并经优化的模板进行翻译时, 开放测试语料的译文评测分数有一定程度的提高。

关键词 [翻译模板](#) [等价对](#) [错误驱动](#)

分类号 [TP391.2](#)

Acquisition of translation template from bilingual corpus

ZHANG Chun-xiang¹, LIANG Ying-hong², YU Lin-sen³

- 1. School of Software, Harbin University of Science and Technology, Harbin 150080, China
- 2. School of Computer Engineering, Vocational University of Suzhou City, Suzhou, Jiangsu 215104, China
- 3. College of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China

Abstract

Automatic acquisition of translation templates is very important for MT system to improve its translation quality and its ability of adapting to new domain. In this paper, tree-to-string method is applied to extract translation equivalences. Error-driven learning method is used to acquire translation templates. A knowledge optimization tool is used to filter translation templates. Then these templates are applied to a transfer-based MT system, and “863” dialog corpus is used as open test corpus. The experiment shows that when new acquired and optimized templates are used, evaluation score for translation of open test corpus is improved.

Key words [translation template](#) [translation equivalence](#) [error-driven](#)

DOI: 10.3778/j.issn.1002-8331.2010.01.033

通讯作者 张春祥 z6c6x6@yahoo.com.cn

扩展功能	
本文信息	
▶	Supporting info
▶	PDF(527KB)
▶	[HTML全文](0KB)
▶	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
▶	复制索引
▶	Email Alert
▶	文章反馈
▶	浏览反馈信息
相关信息	
▶	本刊中 包含“翻译模板” 的相关文章
▶	本文作者相关文章
·	张春祥
·	梁颖红
·	于林森