

<<汉英机器翻译若干关键技术研究>>

图书基本信息

书名：<<汉英机器翻译若干关键技术研究>>

13位ISBN编号：9787302183587

10位ISBN编号：7302183589

出版时间：2008-10

出版时间：清华大学出版社

作者：刘群

页数：153

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<汉英机器翻译若干关键技术研究>>

前言

自然语言处理技术的产生可以追溯到20世纪50年代，它是一门集语言学、数学、计算机科学和认知科学等于一体的综合性交叉学科。

近几年来，随着计算机网络技术和通信技术的迅速发展和普及，自然语言处理技术的应用需求急剧增加，人们迫切需要实用的自然语言处理技术来帮助人们打破语言屏障，为人与人之间、人机之间的信息交流提供便捷、自然、有效的人性化服务。

但是，自然语言处理中的若干科学问题和技术难题尚未得到解决，有待于来自不同领域的学者深入研究和探索。

中文信息处理作为自然语言处理中的一个分支，近几年来备受关注。

一方面，随着中国经济的迅速发展和中国国力的不断增强，汉语正在成为一种新的强势语言而被世人瞩目，汉语理解所涉及的科学问题让国际计算语言学界无法回避；而另一方面，汉语使用者所拥有的巨大市场潜力令国际企业界不敢轻视。

因此，中文信息处理成为全球自然语言处理研究者们共同关注的问题已经是不争的事实。

目前国际上每年举行的颇具影响的几种技术评测，包括机器翻译评测、信息抽取评测和句法分析评测等，无不与汉语密切相关。

因此，作为炎黄子孙，我们没有理由不在这一领域的研究中做出应有的贡献。

中文信息处理所面临的困难既有其他任何一种自然语言处理都会遇到的共性问题，如生词识别问题、歧义消解问题等，也有中文处理本身所具有的个性问题，如汉语自动分词问题、词性定义规范问题等。

因此，从某种意义上讲，中文信息处理更具挑战性。

值得欣慰的是，中文信息处理在引起国际学术界和企业界关注的同时，得到了中国政府的重视和大力支持，它已经被列入国务院批准的“国家中长期科学技术发展规划纲要”。

因此，中文信息处理面临着前所未有的大好机遇。

近几年来，我国的中文信息处理技术得到了快速发展，无论是在基础理论研究方面，还是在技术开发和产业化发展方面，都取得了显著成绩，一大批青年学者投身到这一领域中。

为了使这一领域的广大学者，尤其是青年学生，全面了解中文信息处理的技术现状，进一步推动中文信息处理及其相关学科快速发展，我们组织编写并出版了这套中文信息处理丛书。

<<汉英机器翻译若干关键技术研究>>

内容概要

本书是作者所在的课题组近年来在汉英机器翻译研究方面所取得进展的一个阶段性总结。内容涉及汉英机器翻译的各个主要方面及关键技术，包括对目前国际上机器翻译研究进展的综述，汉语词法分析技术、汉语句法分析技术、汉语词汇语义相似度计算、汉英双语语料库的词语对齐、语料库的结构对齐、基于结构对齐语料库的翻译模板抽取、多引擎机器翻译方法等多方面的研究成果。

本书可供从事计算语言学、自然语言处理、中文信息处理、机器翻译等领域研究工作的人士参考，也可以作为大学相关专业高年级本科生和研究生课程的参考书。

<<汉英机器翻译若干关键技术研究>>

作者简介

刘群, 研究员, 教授, 博士生导师. 职务: 自然语言处理研究组组长 研究方向: 自然语言处理, 机器翻译, 信息提取 学术兼职: 中国计算机学会理事 中国计算机学会术语审定工作委员会主任 《中国计算机学会通讯》编委 中国中文信息学会机器翻译专委会副主任 中国中文信息学会语言资源建设和管理工作委员会 (ChineseLDC) 委员 《中文信息学报》编委 全国术语标准化技术委员会 (SAC/TC62) 委员 全国语言文字标准化技术委员会语法语篇分技术委员会副主任 中国科学院研究生院教授 SIGHAN2008程序委员会委员 ACL2008、ACL2009程序委员会委员 工作经历: 2003~现在中国科学院研究生院教授 (兼职) 2006~现在中国科学院计算技术研究所, 现任研究员、博士生导师 学历 1999.9~2004.5北京大学计算语言学研究所 (在职学习) 获理学博士学位 1989.9~1992.7中国科学院计算技术研究所获工学硕士学位 1984.9~1989.7中国科学技术大学计算机系获工学学士学位 主持的科研项目: 融合语言知识与统计模型的机器翻译方法研究 (国家自然科学基金重点项目, 2008-2011) 面向跨语言搜索的机器翻译关键技术研究 (863重点项目课题, 2007-2010) 基于短语结构转换模板的统计机器翻译方法研究 (国家自然科学基金, 2006-2008) 面向领域可定制的汉英机器翻译系统 (863重点项目, 1998 - 2000) 机器翻译新方法的研究 (863项目, 2005-2006) 通用机器翻译开发平台及汉英机器翻译系统 (863项目, 1996 - 1998)

<<汉英机器翻译若干关键技术研究>>

书籍目录

第1章 机器翻译方法综述 1.1 机器翻译的范式 1.2 基于平行语法的机器翻译方法 1.2.1 Alshawi的基于加权中心词转录机的统计机器翻译方法 1.2.2 吴德凯的反向转录语法 1.2.3 Takeda的基于模式的机器翻译上下文无关语法 1.3 基于实例的机器翻译方法 1.3.1 起源与发展 1.3.2 Sato和Nagao的方法 1.3.3 Kaii的方法 1.3.4 CMU的泛化的基于实例的机器翻译方法 1.3.5 基于实例的机器翻译方法的优缺点 1.4 基于信源信道模型的统计机器翻译方法 1.4.1 IBM的统计机器翻译方法 1.4.2 王野翊在卡内基·梅隆大学(CMU)的工作 1.4.3 约翰·霍普金斯大学(JHU)的统计机器翻译夏季研讨班 1.4.4 Yamada和Knight的工作——基于句法的统计翻译模型 1.4.5 Och等的工作 1.5 基于对数线性模型的统计机器翻译方法 1.5.1 对数线性模型 1.5.2 基于短语的统计翻译模型 1.5.3 基于句法的统计翻译模型 1.6 多引擎机器翻译方法 1.6.1 Pangloss系统 1.6.2 Verbmobil系统 1.7 机器翻译方法的分类 1.7.1 按翻译转换的层面进行分类 1.7.2 按语言知识的表示形式进行分类 1.8 小结第2章 基于层叠隐马尔可夫模型的汉语词法分析 2.1 汉语分析技术概述 2.1.1 汉语词法分析的难点 2.1.2 汉语词法分析的任务和前人的工作 2.2 汉语词法分析的层叠隐马尔可夫模型 2.2.1 隐马尔可夫模型简介 2.2.2 层叠隐马尔可夫模型的结构 2.2.3 层叠隐马尔可夫模型的核心数据结构——词图 2.2.4 层叠隐马尔可夫模型的参数训练 2.3 粗切分：基于一元语法的N最短路径方法 2.4 未定义词识别：基于角色的隐马尔可夫模型 2.4.1 模型的定义 2.4.2 角色的选取 2.4.3 角色的标注 2.4.4 未定义词的提取 2.4.5 参数训练 2.5 未定义词的概率估计：基于角色的词语生成模型 2.5.1 问题的由来 2.5.2 模型的定义 2.6 细切分：词汇化的隐马尔可夫模型 2.6.1 模型的定义 2.6.2 最短路径的求解 2.6.3 参数估计 2.7 词性标注：基于词性的隐马尔可夫模型 2.7.1 基于隐马尔可夫模型的词性标注 2.7.2 词性标记集的选择与转换 2.8 实验结果 2.8.1 各层隐马尔可夫模型的对比实验 2.8.2 在国家“973”计划评测中的测试结果 2.8.3 第一届国际分词大赛的评测结果 2.9 小结第3章 融合语义知识和词汇化上下文概率语法的汉语句法分析 3.1 前言 3.2 Baseline句法分析器 3.3 语义知识集成 3.3.1 语义类抽取 3.3.2 构建基于类的选择偏向模型 3.3.3 实验结果 3.3.4 性能改进分析 3.4 基于汉语宾州树库的句法分析相关工作 3.5 小结第4章 汉语词法分析与句法分析融合策略研究 4.1 引言 4.2 句法分析系统 4.2.1 融合语义知识的词汇化概率上下文无关语法模型 4.2.2 结构上下文模型 4.2.3 多子模型句法分析器 4.3 词法分析系统(ICTCLAS) 4.4 融合策略 4.4.1 切分转换：基于转换的错误驱动学习 4.4.2 标记转换：条件随机场 4.4.3 转换实验 4.5 实验与分析 4.6 比较 4.7 小结第5章 基于“知网”的词汇语义相似度计算 5.1 引言 5.2 词语相似度及其计算的方法 5.2.1 什么是词语相似度 5.2.2 词语相似度与词语距离 5.2.3 词语相似度与词语相关性 5.2.4 词语相似度的计算方法 5.3 “知网”简介 5.3.1 “知网”的结构 5.3.2 “知网”的知识描述语言 5.4 基于“知网”的语义相似度计算方法 5.4.1 词语相似度计算 5.4.2 义原相似度计算 5.4.3 虚词概念的相似度的计算 5.4.4 实词概念的相似度的计算 5.5 实验及结果 5.6 小结第6章 词语对齐的对数线性模型 6.1 引言 6.2 对数线性模型 6.3 特征函数 6.3.1 IBM翻译模型 6.3.2 词性标记转换模型 6.3.3 双语词典 6.4 训练 6.5 搜索 6.6 实验结果 6.7 小结第7章 一种双语短语结构对齐搜索算法 7.1 双语对齐技术概述 7.1.1 各种层次的语言单位上的对齐技术 7.1.2 短语结构对齐的定义 7.1.3 短语结构对齐的过程 7.1.4 短语结构对齐的问题和难点 7.1.5 现有的短语结构对齐技术 7.2 一种双语短语结构对齐的搜索算法 7.2.1 算法简介 7.2.2 局部对齐 7.2.3 短语结构对齐的柱形搜索(beam search)算法 7.2.4 局部对齐的归并 7.2.5 局部对齐的评分 7.2.6 搜索算法的时间复杂度分析 7.3 实验及结果分析 7.3.1 实验方案 7.3.2 实验语料来源及规模 7.3.3 短语结构对齐的实例分析 7.3.4 实验结果及分析 7.3.5 实验结果的进一步分析 7.4 小结第8章 短语结构转换模板的提取与应用 8.1 基于模板的机器翻译概述 8.2 短语结构转换模板定义 8.3 短语结构转换模板举例 8.4 短语结构转换模板的提取 8.5 短语结构转换模板的应用——基于模板的转换 8.6 实验结果 8.6.1 实验语料的来源及规模 8.6.2 实验结果分析 8.7 小结第9章 微引擎流水线机器翻译系统结构 9.1 微引擎流水线的基本思想 9.2 微引擎流水线的系统结构 9.3 微引擎流水线的公共数据结构 9.4 各种微引擎的程序接口和功能说明 9.5 微引擎调度算法 9.6 面向新闻领域的汉英机器翻译系统 9.6.1 研究背景 9.6.2 系统实现方案 9.7 实验结果及分析 9.8 小结第10章 总结及今后的工作附录 汉语词性标记集ICTPOS参考文献后

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>