

《多语自然语言处理》

图书基本信息

书名：《多语自然语言处理》

13位ISBN编号：9787111484916

出版时间：2015-2-1

作者：Daniel M. Bikel,Imed Zitouni

页数：448

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《多语自然语言处理》

内容概要

本书是第一本全面阐述如何构建健壮和准确的多语自然语言处理系统的图书，由两位资深专家编辑，集合了该领域众多尖端进展以及从广泛的研究和产业实践中总结出的实用解决方案。第一部分介绍现代自然语言处理的核心概念和理论基础，展示了如何理解单词和文档结构、分析语法、建模语言、识别蕴涵和检测冗余。第二部分彻底阐述与构建真实应用有关的实际考量，包括信息抽取、机器翻译、信息检索、文摘、问答、提炼、处理流水线等。

《多语自然语言处理》

作者简介

Daniel M. Bikel 现为Google公司高级研究科学家，正在开发用于自然语言处理和语音识别的新方法。在IBM工作期间，他为IBM的GALE多语种信息抽取和自动应答系统构架了拦截系统。在宾夕法尼亚大学攻读博士后期间，他建造了第一个可扩展的多语种语法分析引擎。

Imed Zitouni 现为微软公司高级研究员。2004~2012年，他是IBM公司高级研究科学家，领导IBM公司的阿拉伯语信息抽取和数据资源工作组。在此之前，他还曾领导DIALOCA的语音/自然语言处理组和Bell实验室/阿尔卡特朗讯的语言建模和呼叫路由工作。他的研究涉及机器翻译、自然语言处理和口语对话系统。

书籍目录

出版者的话

译者序

前言

关于作者

第一部分 理论

第1章 找出词的结构

1.1 词及其部件

1.1.1 词元

1.1.2 词形

1.1.3 词素

1.1.4 类型学

1.2 问题和挑战

1.2.1 不规则性

1.2.2 歧义性

1.2.3 能产性

1.3 形态模型

1.3.1 查词典

1.3.2 有限状态形态

1.3.3 基于合一的形态

1.3.4 函数式形态

1.3.5 形态归纳

1.4 总结

第2章 找出文档的结构

2.1 概述

2.1.1 句子边界检测

2.1.2 主题边界检测

2.2 方法

2.2.1 生成序列分类方法

2.2.2 判别性局部分类方法

2.2.3 判别性序列分类方法

2.2.4 混合方法

2.2.5 句子分割的全局建模扩展

2.3 方法的复杂度

2.4 方法的性能

2.5 特征

2.5.1 同时用于文本与语音的特征

2.5.2 只用于文本的特征

2.5.3 语音特征

2.6 处理阶段

2.7 讨论

2.8 总结

第3章 句法

3.1 自然语言分析

3.2 树库：句法分析的数据驱动方法

3.3 句法结构的表示

3.3.1 使用依存图的句法分析

3.3.2 使用短语结构树的句法分析

3.4 分析算法

3.4.1 移进归约分析

3.4.2 超图和线图分析

3.4.3 最小生成树和依存分析

3.5 分析中的歧义消解模型

3.5.1 概率上下文无关文法

3.5.2 句法分析的生成模型

3.5.3 句法分析的判别模型

3.6 多语言问题：什么是词元

3.6.1 词元切分、实例和编码

3.6.2 分词

3.6.3 形态学

3.7 总结

第4章 语义分析

4.1 概述

4.2 语义解释

4.2.1 结构歧义

4.2.2 词义

4.2.3 实体与事件消解

4.2.4 谓词 论元结构

4.2.5 意义表示

4.3 系统范式

4.4 词义

4.4.1 资源

4.4.2 系统

4.4.3 软件

4.5 谓词 论元结构

4.5.1 资源

4.5.2 系统

4.5.3 软件

4.6 意义表示

4.6.1 资源

4.6.2 系统

4.6.3 软件

4.7 总结

4.7.1 词义消歧

4.7.2 谓词 论元结构

4.7.3 意义表示

第5章 语言模型

5.1 概述

5.2 n元模型

5.3 语言模型评价

5.4 参数估计

5.4.1 最大似然估计和平滑

5.4.2 贝叶斯参数估计

5.4.3 大规模语言模型

5.5 语言模型适应

5.6 语言模型的类型

5.6.1 基于类的语言模型

- 5.6.2 变长语言模型
- 5.6.3 判别式语言模型
- 5.6.4 基于句法的语言模型
- 5.6.5 最大熵语言模型
- 5.6.6 因子化语言模型
- 5.6.7 其他基于树的语言模型
- 5.6.8 基于主题的贝叶斯语言模型
- 5.6.9 神经网络语言模型
- 5.7 特定语言建模问题
 - 5.7.1 形态丰富语言的建模
 - 5.7.2 亚词单元的选择
 - 5.7.3 形态类别建模
 - 5.7.4 无分词语言
 - 5.7.5 口语与书面语言
- 5.8 多语言和跨语言建模
 - 5.8.1 多语言建模
 - 5.8.2 跨语言建模
- 5.9 总结
- 第6章 文本蕴涵识别
 - 6.1 概述
 - 6.2 文本识别蕴涵任务
 - 6.2.1 问题定义
 - 6.2.2 RTE的挑战
 - 6.2.3 评估文本蕴涵系统性能
 - 6.2.4 文本蕴涵解决方案的应用
 - 6.2.5 其他语言中的RTE研究
 - 6.3 文本蕴涵识别的框架
 - 6.3.1 要求
 - 6.3.2 分析
 - 6.3.3 有用的组件
 - 6.3.4 通用模型
 - 6.3.5 实现
 - 6.3.6 对齐
 - 6.3.7 推理
 - 6.3.8 训练
 - 6.4 案例分析
 - 6.4.1 抽取语篇约束
 - 6.4.2 基于编辑距离的RTE
 - 6.4.3 基于转换的方法
 - 6.4.4 逻辑表示及推理
 - 6.4.5 独立于蕴涵学习对齐
 - 6.4.6 在RTE中利用多对齐
 - 6.4.7 自然逻辑
 - 6.4.8 句法树核
 - 6.4.9 使用有限依存上下文的全局相似度
 - 6.4.10 RTE的潜在对齐推理
 - 6.5 RTE的进一步研究
 - 6.5.1 改进分析器
 - 6.5.2 发明或解决新问题

- 6.5.3 开发知识库
- 6.5.4 更好的RTE评价
- 6.6 有用资源
 - 6.6.1 文献
 - 6.6.2 知识库
 - 6.6.3 自然语言处理包
- 6.7 总结
- 第7章 多语情感与主观性分析
 - 7.1 概述
 - 7.2 定义
 - 7.3 英语中的情感及主观性分析
 - 7.3.1 词典
 - 7.3.2 语料库
 - 7.3.3 工具
 - 7.4 词级和短语级标注
 - 7.4.1 基于字典的方法
 - 7.4.2 基于语料库的方法
 - 7.5 句子级标注
 - 7.5.1 基于字典
 - 7.5.2 基于语料库
 - 7.6 文档级标注
 - 7.6.1 基于字典
 - 7.6.2 基于语料库
 - 7.7 什么有效，什么无效
 - 7.7.1 最佳情况：已有人工标注的语料库
 - 7.7.2 次优情形：基于语料库的跨语言映射
 - 7.7.3 第三优情形：孳衍词典
 - 7.7.4 第四优情形：翻译词典
 - 7.7.5 各种可行方法的比较
 - 7.8 总结
- 第二部分 实践
- 第8章 实体检测和追踪
 - 8.1 概述
 - 8.2 提及检测
 - 8.2.1 数据驱动的分类
 - 8.2.2 搜索提及
 - 8.2.3 提及检测特征
 - 8.2.4 提及检测实验
 - 8.3 共指消解
 - 8.3.1 Bell树的构建
 - 8.3.2 共指模型：链接和引入模型
 - 8.3.3 最大熵链接模型
 - 8.3.4 共指消解实验
 - 8.4 总结
- 第9章 关系和事件
 - 9.1 概述
 - 9.2 关系与事件
 - 9.3 关系类别
 - 9.4 将关系抽取视为分类

- 9.4.1 算法
- 9.4.2 特征
- 9.4.3 分类器
- 9.5 关系抽取的其他方法
 - 9.5.1 无监督和半监督方法
 - 9.5.2 核方法
 - 9.5.3 实体和关系检测的联合方法
- 9.6 事件
- 9.7 事件抽取方法
- 9.8 超句
- 9.9 事件匹配
- 9.10 事件抽取的未来方向
- 9.11 总结
- 第10章 机器翻译
 - 10.1 机器翻译现状
 - 10.2 机器翻译评测
 - 10.2.1 人工评测
 - 10.2.2 自动评测
 - 10.2.3 WER、BLEU、METEOR等
 - 10.3 词对齐
 - 10.3.1 共现
 - 10.3.2 IBM模型1
 - 10.3.3 期望最大化
 - 10.3.4 对齐模型
 - 10.3.5 对称化
 - 10.3.6 作为机器学习问题的词对齐
 - 10.4 基于短语的翻译模型
 - 10.4.1 模型
 - 10.4.2 训练
 - 10.4.3 解码
 - 10.4.4 立方剪枝
 - 10.4.5 对数线性模型和参数调节
 - 10.4.6 控制模型的大小
 - 10.5 基于树的翻译模型
 - 10.5.1 层次短语翻译模型
 - 10.5.2 线图解码
 - 10.5.3 基于句法的模型
 - 10.6 语言学挑战
 - 10.6.1 译词选择
 - 10.6.2 形态学
 - 10.6.3 词序
 - 10.7 工具和数据资源
 - 10.7.1 基本工具
 - 10.7.2 机器翻译系统
 - 10.7.3 平行语料
 - 10.8 未来的方向
 - 10.9 总结
- 第11章 跨语言信息检索
 - 11.1 概述

- 11.2 文档预处理
 - 11.2.1 文档句法和编码
 - 11.2.2 词元化
 - 11.2.3 规范化
 - 11.2.4 预处理最佳实践
- 11.3 单语信息检索
 - 11.3.1 文档表示
 - 11.3.2 索引结构
 - 11.3.3 检索模型
 - 11.3.4 查询扩展
 - 11.3.5 文档先验模型
 - 11.3.6 模型选择的最佳实践
- 11.4 CLIR
 - 11.4.1 基于翻译的方法
 - 11.4.2 机器翻译
 - 11.4.3 中间语言文档表示
 - 11.4.4 最佳实践
- 11.5 多语言信息检索
 - 11.5.1 语言识别
 - 11.5.2 MLIR的索引建立
 - 11.5.3 翻译查询串
 - 11.5.4 聚合模型
 - 11.5.5 最佳实践
- 11.6 信息检索的评价
 - 11.6.1 建立实验环境
 - 11.6.2 相关性评估
 - 11.6.3 评价指标
 - 11.6.4 已有数据集
 - 11.6.5 最佳实践
- 11.7 工具、软件和资源
- 11.8 总结
- 第12章 多语自动文摘
 - 12.1 概述
 - 12.2 自动文摘方法
 - 12.2.1 传统方法
 - 12.2.2 基于图的方法
 - 12.2.3 学习如何做摘要
 - 12.2.4 多语自动摘要
 - 12.3 评测
 - 12.3.1 人工评价
 - 12.3.2 自动评价
 - 12.3.3 自动文摘评测系统的近期发展
 - 12.3.4 多语自动文摘的自动评测方法
 - 12.4 如何搭建自动文摘系统
 - 12.4.1 材料
 - 12.4.2 工具
 - 12.4.3 说明
 - 12.5 评测竞赛和数据集
 - 12.5.1 评测竞赛

- 12.5.2 数据集
- 12.6 总结
- 第13章 问答系统
 - 13.1 概述和历史
 - 13.2 架构
 - 13.3 源获取和预处理
 - 13.4 问题分析
 - 13.5 搜索及候选抽取
 - 13.5.1 非结构化资源搜索
 - 13.5.2 非结构化源文本的候选抽取
 - 13.5.3 结构化源文本的候选抽取
 - 13.6 回答评分
 - 13.6.1 方法概述
 - 13.6.2 证据结合
 - 13.6.3 扩展到列表型问题
 - 13.7 跨语言问答
 - 13.8 案例研究
 - 13.9 评测
 - 13.9.1 评测任务
 - 13.9.2 判断答案正确性
 - 13.9.3 性能度量
 - 13.10 当前和未来的挑战
 - 13.11 总结和进一步阅读
- 第14章 提炼
 - 14.1 概述
 - 14.2 示例
 - 14.3 相关性和冗余性
 - 14.4 Rosetta Consortium 提炼系统
 - 14.4.1 文档和语料库准备
 - 14.4.2 索引
 - 14.4.3 查询回答
 - 14.5 其他提炼方法
 - 14.5.1 系统架构
 - 14.5.2 相关度
 - 14.5.3 冗余
 - 14.5.4 多模态提炼
 - 14.5.5 跨语言提炼
 - 14.6 评测和指标
 - 14.7 总结
- 第15章 口语对话系统
 - 15.1 概述
 - 15.2 口语对话系统
 - 15.2.1 语音识别和理解
 - 15.2.2 语音生成
 - 15.2.3 对话管理器
 - 15.2.4 语音用户接口
 - 15.3 对话形式
 - 15.4 自然语言呼叫路由选择
 - 15.5 三代对话应用

- 15.6 持续的改进循环
- 15.7 口语句子的转录和标注
- 15.8 口语对话系统的本地化
 - 15.8.1 呼叫流程本地化
 - 15.8.2 提示本地化
 - 15.8.3 文法的本地化
 - 15.8.4 源端数据
 - 15.8.5 训练
 - 15.8.6 测试
- 15.9 总结
- 第16章 聚合自然语言处理引擎
 - 16.1 概述
 - 16.2 聚合语音和NLP引擎架构的期望属性
 - 16.2.1 灵活的分布式组件化
 - 16.2.2 计算效率
 - 16.2.3 数据操作功能
 - 16.2.4 鲁棒性处理
 - 16.3 聚合的架构
 - 16.3.1 UIMA
 - 16.3.2 GATE
 - 16.3.3 InfoSphere Streams
 - 16.4 案例研究
 - 16.4.1 GALE 互操作性演示系统
 - 16.4.2 跨语言自动语言开发系统
 - 16.4.3 实时翻译服务
 - 16.5 经验教训
 - 16.5.1 分割涉及延迟和精度之间的权衡
 - 16.5.2 联合优化与互操作性
 - 16.5.3 数据模型需要使用约定
 - 16.5.4 性能评估的挑战
 - 16.5.5 引擎的前向波训练
 - 16.6 总结
 - 16.7 UIMA样本代码
- 索引

精彩短评

1、厚重的翻译味，不过母语好处还是可以飞快地扫

《多语自然语言处理》

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:www.tushu000.com