

<<自然语言处理的形式模型>>

图书基本信息

书名：<<自然语言处理的形式模型>>

13位ISBN编号：9787312022531

10位ISBN编号：7312022537

出版时间：2010

出版时间：中国科学技术大学出版社

作者：冯志伟

页数：638

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>

## <<自然语言处理的形式模型>>

### 前言

大学最重要的功能是向社会输送人才，大学对于一个国家、民族乃至世界的重要性和贡献度，很大程度上是通过毕业生在社会各领域所取得的成就来体现的。

中国科学技术大学建校只有短短的五十年，之所以迅速成为享有较高国际声誉的著名大学之一，主要就是因为她培养出了一大批德才兼备的优秀毕业生。他们志向高远、基础扎实、综合素质高、创新能力强，在国内外科技、经济、教育等领域做出了杰出的贡献，为中国科大赢得了“科技英才的摇篮”的美誉。

2008年9月，胡锦涛总书记为中国科大建校五十周年发来贺信，信中称赞说：半个世纪以来，中国科学技术大学依托中国科学院，按照全院办校、所系结合的方针，弘扬红专并进、理实交融的校风，努力推进教学和科研工作的改革创新，为党和国家培养了一大批科技人才，取得了一系列具有世界先进水平的原创性科技成果，为推动我国科教事业发展和社会主义现代化建设做出了重要贡献。

据统计，中国科大迄今已毕业的5万人中，已有42人当选中国科学院和中国工程院院士，是同期（自1963年以来）毕业生中当选院士数最多的高校之一。

其中，本科毕业生中平均每1000人就产生1名院士和700多名硕士、博士，比例位居全国高校之首。

还有众多的中青年才俊成为我国科技、企业、教育等领域的领军人物和骨干。

在历年评选的“中国青年五四奖章”获得者中，作为科技界、科技创新型企业界青年才俊代表，科大毕业生已连续多年榜上有名，获奖总人数位居全国高校前列。

鲜为人知的是，有数千名优秀毕业生踏上国防战线，为科技强军做出了重要贡献，涌现出20多名科技将军和一大批国防科技中坚。

## <<自然语言处理的形式模型>>

### 内容概要

本书对自然语言处理中的各种形式模型进行了系统的梳理,分别讨论了基于短语结构语法的形式模型、基于合一运算的形式模型、基于依存和配价的形式模型、基于格语法的形式模型、基于词汇主义的形式模型、语义自动处理的形式模型、语用自动处理的形式模型、隐马尔可夫模型、统计机器翻译的形式模型。

《自然语言处理的形式模型》说理透彻、语言流畅、实例丰富、深入浅出,适合于从事自然语言处理教学和研究的科研人员、大学师生阅读,也可以作为人工智能、计算语言学等课程的教学参考。

## &lt;&lt;自然语言处理的形式模型&gt;&gt;

## 书籍目录

总序前言第1章 自然语言处理的学科定位 1.1 从自然语言处理的过程来考察其学科定位 1.2 从自然语言处理的范围来考察其学科定位 1.3 从自然语言处理的历史来考察其学科定位 1.4 当前自然语言处理发展的几个特点第2章 语言计算研究的先驱 2.1 Markov链 2.2 Zipf定律 2.3 Shannon关于“熵”的研究 2.4 Bar-Hillel的范畴语法 2.5 Harris的语言串分析法 2.6 o.c.KysmrHHa的语言集合论模型第3章 基于短语结构语法的形式模型 3.1 语法的Chomsky层级 3.2 有限状态语法和它的局限性 3.3 短语结构语法 3.4 递归转移网络和扩充转移网络 3.5 自底向上分析和自顶向下分析 3.6 通用句法生成器和线图分析法 3.7 Earley算法 3.8 左角分析法 3.9 CYK算法 3.10 Tomita算法 3.11 管辖-约束理论与最简方案 3.12 Joshi的树邻接语法 3.13 汉字结构的形式描述第4章 基于合一运算的形式模型 4.1 中文信息MMT模型 4.2 Kaplan的词汇功能语法 4.3 MartinKay的功能合一语法 4.4 Gazdar的广义短语结构语法 4.5 Shieber的PATR 4.6 Pollard的中心语驱动的短语结构语法 4.7 Pereira和Warren定子句语法第5章 基于依存和配价的形式模型 5.1 配价观念的起源 5.2 Tesnière的依存语法 5.3 依存语法在自然语言处理中的应用 5.4 配价语法 5.5 配价语法在自然语言处理中的应用第6章 基于格语法的形式模型 6.1 Fillmore的格语法 6.2 Fillmore的框架网络第7章 基于词汇主义的形式模型 7.1 Gross的词汇语法 7.2 链语法 7.3 词汇语义学 7.4 知识本体 7.5 词网WordNet 7.6 知网HowNet第8章 语义自动处理的形式模型 8.1 义素分析法 8.2 语义场 8.3 语义网络 8.4 Montague的蒙塔鸠语法 8.5 Wilks的优选语义学 8.6 Schank的概念依存理论 8.7 Mel'chuk的意义(≡)文本理论 8.8 词义排歧方法第9章 系统功能语法 9.1 系统功能语法的基本概念 9.2 系统功能语法在自然语言处理中的应用第10章 语用自动处理的形式模型 10.1 Mann和Thompson的修辞结构理论 10.2 文本连贯中的常识推理技术第11章 概率语法 11.1 概率上下文无关语法与句子的歧义 11.2 概率上下文无关语法的基本原理 11.3 概率上下文无关语法的三个假设 11.4 概率词汇化上下文无关语法第12章 Bayes公式与动态规划算法 12.1 拼写错误的检查与更正 12.2 Bayes公式与噪声信道模型 12.3 最小编辑距离算法 12.4 发音问题研究中的Bayes方法 12.5 发音变异的决策树模型 12.6 加权自动机 12.7 向前算法 12.8 Viterbi算法 本章附录第13章 N元语法和数据平滑 13.1 N元语法 13.2 数据平滑第14章 隐马尔可夫模型(HMM) 14.1 HMM模型概述 14.2 HMM模型在语音识别中的应用第15章 统计机器翻译中的形式模型 15.1 机器翻译与噪声信道模型 15.2 最大熵模型 15.3 基于平行概率语法的形式模型 15.4 基于短语的统计机器翻译 15.5 基于句法的统计机器翻译第16章 自然语言处理系统的评测 16.1 评测的一般原则和方法 16.2 语音合成和文语转换系统的评测 16.3 机器翻译系统的评测 16.4 语料库系统的评测 16.5 国外自然语言处理系统的评测第17章 自然语言处理中的理性主义与经验主义 17.1 哲学中的理性主义和经验主义 17.2 自然语言处理中理性主义和经验主义的消长 17.3 理性主义和经验主义的利弊得失 17.4 探索理性主义方法和经验主义方法结合的途径

## &lt;&lt;自然语言处理的形式模型&gt;&gt;

## 章节摘录

HAL也应该能够做信息检索 (information retrieval, 发现它所需要的文本资源在哪里) 和信息抽取 (information extraction, 从文本资源中抽取它所需要的信息), 并且进行知识推理 (reference, 根据已知的事实得出结论)。

尽管这些问题现在还远远没有完全解决, HAL需要的一些与语言相关的技术现在已经研制出来了, 有一部分技术已经商品化。

解决这样的问题以及其他类似的问题, 是自然语言处理、计算语言学、语音识别与语音合成的主要研究内容。

我们把它们统称为语音与语言的计算机处理 (speech and language processing), 或者简单地称为自然语言处理 (natural language processing), 因此, 自然语言处理也同时包括了语音处理的内容。

像HAL这样有复杂的语言能力的智能机器人将要求非常广泛和深刻的语言知识。

我们只要读一读前面在。

HAL和Dave之间进行的对话, 我们就可以了解到这样的更加复杂的应用所需要的语言知识的范围和种类。

为了确定Dave讲什么, HAL必须能够分析它所接收的声音信号, 并且把Dave的这些信号复原成词的系列。

与此相似, 为了生成回答, HAL必须把它的回答组织成词的系列, 并且生成Dave能够识别的声音信号。

要完成这两方面的任务, 需要语音学 (phonetics) 和音系学 (phonology) 的知识, 这样的知识可以帮助我们建立词如何在话语中发音的模型。

值得注意的是, HAL还能够说出如像Im和cant这样的缩约形式, HAL必须把它们分别还原为I am和can not, 才能在它的词库中找到这些单词的对应物, 从而明白这些缩约形式究竟代表什么样的语言成分。

HAL还要能够产生并且识别单词的这样或那样的变体 (例如, 识别doors是复数)。

这些都要求HAL具有形态学方面的知识, 这些知识能够反映关于上下文中词的形态和行为的有关信息。

除了处理一个一个的单词之外, HAL还应该知道怎样分析Dave所提出的请求的结构。

这样的分析能够使HAL确定, Dave说的话是关于要HAL采取某种行动的一个请求, 这样的请求不同于下面关于陈述客观世界的简单命题, 也不同于下面关于door的问话, 它们是Dave请求的不同变体:

HAL, the pod bay door is open. (HAL, 分离舱的门是开着的。

) HAL, is the pod bay door open?

(HAL, 分离舱的门是开着的吗?)

) 此外, HAL还必须使用类似的结构知识把一个个的单词组织成为符号串, 构成它的回答。

例如, HAL必须知道, 下面的单词序列对于Dave是没有意义的, 尽管这个单词系列所包含的单词与它原来的回答中所包含的单词完全一样: Im I do, Sorry that afraid Dave Im cant. 这里所说的关于组词成句的知识, 叫做句法 (syntax)。

显而易见, 如果只是知道Dave所说的话语的各个单词以及句法结构, 并不能使HAL了解Dave提出的请求的实质。

为了理解Dave的请求事实上是关于要求关闭pod bay door (分离舱门) 的一个命令, 而不是讲关于当天中饭的菜单的事情, 就要有复合词的语义的知识、词汇语义学 (lexical semantics) 的知识以及如何把这样的复合词组成更大的意义的知识, 即关于组合语义学 (compositional semantics) 的知识。

<<自然语言处理的形式模型>>

编辑推荐

当代科学技术基础理论与前沿问题研究丛书：中国科学技术大学校友文库

<<自然语言处理的形式模型>>

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>