

将语言计算的理论方法和最新成果 呈现给读者

—《统计自然语言处理(第 2 版)》评述

赵铁军¹

摘要 本文简要评述《统计自然语言处理》(第 2 版), 向读者推介一部自然语言处理领域的长篇力作。

关键词 自然语言处理, 统计方法, 专著

引用格式 赵铁军. 将语言计算的理论方法和最新成果呈现给读者—《统计自然语言处理(第 2 版)》评述, 2014, 40(5): 1024–1024

DOI 10.3724/SP.J.1004.2014.01024

Showing Readers the Fundamental Theories, Methods and New Achievements on Natural Language Computing

ZHAO Tie-Jun¹

Abstract This article briefly introduces the new edition of the book “Statistical Natural Language Processing”, which is an excellent long monograph in the area of computational linguistics and natural language processing. Some comments are given.

Key words Natural language processing, statistical method, monograph

Citation Zhao Tie-Jun. Showing readers the fundamental theories, methods and new achievements on natural language computing. *Acta Automatica Sinica*, 2014, 40(5): 1024–1024

近日喜获宗成庆老师所著的《统计自然语言处理》第 2 版(清华大学出版社, 2013 年 8 月), 谨向他表示诚挚的祝贺! 手里捧着沉甸甸的 87 万 5 千字的长篇巨著, 心中颇多感慨, 诚如该书作者在新版前言中所说: “对自己从日常繁忙的工作中挤出时间来从事本书写作的能力估计过高, 当然也对潮水般汹涌而来的各种事物所造成的巨大压力估计不足, 致使我不得不在生存和履行诺言之间苦苦地煎熬, 挣扎着一点一点地践行自己的承诺”。这一点对于很多在教学和科研工作一线的中国大陆同行来说, 无不有同感, 而宗老师最终“做到了我想做的事情”, 这就难能可贵了!

综观此书, 我觉得有两大特点: 一是全面, 二是独到。作者全面地介绍了自然语言处理领域近年来的最新研究成果, 对大量资料进行了详细的分析和综述, 工作量巨大。同时, 作者作为活跃在中国大陆自然语言处理领域前沿的知名学者, 在书中汇集了自己近年来指导博士生所取得的诸多创新成果。我相信, 这些成果必将带给广大读者宝贵的经验和启发。

作者在该书第一章给出了章节内容的关系图, 将全书分为基本概念与技术基础、关键技术和应用系统三个部分。这些内容广泛涉猎了自然语言处理所需要的基础知识、核心技术和各类应用。其中, 第 2 章到第 6 章介绍了自然语言处理

用到的基本概念和知识, 如概率论、信息论、 n 元模型、自动机方法和统计机器学习的常用方法, 并阐述了语料库和语言知识在自然语言处理中的基础地位。第二部分依照自然语言处理自底向上的处理过程, 分别介绍了词法分析、句法分析、语义分析和篇章分析这些关键技术, 每一章都有对各种理论和方法的细致描述, 点出了各个层面上所面临的难点问题。该书内容可谓全面但不庞杂, 系统且主次分明, 如句法分析以介绍短语结构句法分析和依存句法分析两大方法为主, 语义分析以词义消歧和语义角色标注为主。这些内容都曾经是或者仍然是自然语言处理研究的热点, 一直受到国内外研究者的关注。该书后 6 章对于自然语言处理的应用系统做了全面介绍, 其中, 语音翻译和人机对话系统两章在这一部分中颇具亮彩, 这与著者曾多年从事这一技术研究的切身经历密不可分。语音翻译与一般机器翻译相比有其不同之处, 只有抓住了语音翻译的特点和解决这些特殊问题的关键之处, 才能更好地实现高效、准确的口语翻译服务。该书第 12 章恰到好处地体现了这一理念。而自然语言处理在整个计算机应用中的另一个重要目的是架起人机交互的桥梁, 让人以自然语音的方式与机器自由交流, 这是人类久有的梦想, 而圆梦就得一步一步地从每一个技术环节做起。该书最后一章让人们看到了追逐这一梦想的路线图, 从口语现象分析、口语句子情感分析, 到口语解析、对话行为识别, 最终到口语生成和对话系统实现, 娓娓道来, 深入浅出, 颇有一种润物细无声的感觉。

说到该书的独到之处, 我不能不说这本巨著是作者多年来在科研一线辛勤耕耘的成果呈现。从这本书的每一章中我们都可以看到或感受到宗老师团队所开展的许多研究, 如第 8 章中介绍的基于序列标注的分层依存分析方法, 第 9 章中的双语联合语义角色标注方法, 第 11 章里的汉语机构名翻译方法等等。其中很多内容都在国际一流的学术期刊和学术会议上发表过。我相信这些内容定能为广大研究者和学生提供更多的启迪。

综上所述, 《统计自然语言处理》第 2 版的问世, 为我国自然语言处理领域的学子们馈赠了一件丰厚的礼品, 它既是一本很好的高年级本科生和研究生学习的教学参考书, 也可以成为自然语言处理技术研究和开发者们的参考资料。持书在手, 读者们自有体会, 我不必在此赘言, 只是表达我对著者好友的一种由衷钦佩!

赵铁军 哈尔滨工业大学计算机科学与技术学院教授。主要研究方向为自然语言处理、机器翻译。E-mail: tjzhao@hit.edu.cn
(ZHAO Tie-Jun Professor in the Department of Computer Science and Technology, Harbin Institute of Technology. His research interest covers natural language processing and machine translation.)

收稿日期 2014-03-24 录用日期 2014-04-02
Manuscript received March 24, 2014; accepted April 2, 2014

1. 哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001
1. Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001