# Description: 东大校徽1111

学校代码： 10286

分 类 号： TP311

密 级： 公开

U D C : 004.4

学 号： 131508

**基于依存句法分析的中文评价对象抽取**

**和情感倾向性分析**

研究生姓名**：贾治中**

导师姓名**：高志强 教授**

|  |  |
| --- | --- |
| 申请学位类别 工学硕士 | 学位授予单位 东 南 大 学 |
| 一级学科名称 计算机科学与技术 | 论文答辩日期 2016 年 6 月 5 日 |
| 二级学科名称 计算机软件与理论 | 学位授予日期 2016 年 月 日 |
| 答辩委员会主席 | 评 阅 人 匿名评阅人 |
|  | 匿名评阅人 |

2016 年 5 月30日

**CHINESE OPINION TARGET EXTRACTION AND ORIENTATION ANALYSIS BASED ON SYNTACTIC DEPENDENCIES**

A Thesis Submitted to

Southeast University

For the Degree of Master of Engineering

BY

JIA Zhizhong

Supervised by

Professor Gao Zhiqiang

School of Computer Science and Engineering

Southeast University

June 2016

**东 南 大 学 学 位 论 文 独 创 性 声 明**

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得东南大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

论文作者签名： 日期：

**东 南 大 学 学 位 论 文 使 用 授 权 声 明**

东南大学、中国科学技术信息研究所、国家图书馆有权保留本人所送交学位论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权东南大学研究生院办理。

论文作者签名： 导师签名： 日期：

# 摘要

随着互联网的发展，包含有观点和评论的文本大量涌现。人们一方面在浏览别人发表的评论，一方面也在不停地分享自己对于某些人或物的观点和情感。群体的观点发掘能更好地为决策者提供更好的决策依据，所以情感分析无论对于经济发展、政治决策还是社会个体的行为都有着极其重要的指引作用。目前中文评价对象抽取和情感倾向性分析的效果还很不理想，本文在前人研究的基础上，进行了一些有益的探索

本文主要利用语句中各个成分之间的依存关系，提取出中文短文本的评价对象和评价词，并分析出评价对象的情感倾向性。论文的主要工作如下：

1. 在现有词典的基础上，构建用于情感分析的情感词典，主要包括：正面情绪词典、负面情绪词典、正面评价词典、负面评价词典、观点引述词典、虚拟语气词典、转折词典、名词性情感词典等。
2. 由于传统的基于依存句法分析的评价对象抽取和倾向性分析效果不理想，本文在依存句法分析的基础上，利用语义角色标注，并添加了一系列的语义规则进行情感分析。实验结果表明在NLPCC2013的微博评测语料上，添加语义规则的基于依存句法分析的方法，能够显著提高识别效果。
3. 结合词义和词语相似度计算算法，提出一种评价对象搜索方法，用于改善在只抽取出代词或句法关系中无评价对象的情况下，搜索上下文中真正的评价对象的精确度。

本文实验表明：通过完善情感词库和添加语义知识，能够较为显著地提高评价对象抽取和情感倾向性分析效果。

**关键词：**依存句法分析，评价对象抽取，倾向性分析，语义规则

Abstract

With the development of the Internet, a large number of text with opinions and comments appear on the Internet. People browse in other people's comments on the one hand, on the other hand also share their views on some people or things continually. Mining the group view can provide better decision-making basis for policymakers, so sentiment analysis plays an important guidance role for economic development, political decisions and behaviors of the individual or social behavior. However, in the domain of the Chinese, the performance of opinion target extraction and orientation analysis is not very ideal. Based on the research results in this area, some useful exploration is carried out in this thesis.

Syntactic dependencies of components of the sentence are used to extract opinion targets and opinions from the Chinese short text, then sentiment orientation is analyzed in this thesis. Our major work is as follows:

1. On the basis of the existing dictionaries, sentiment dictionaries are built for sentiment analysis such as Positive Emotion Dictionary, Negative Emotion Dictionary, Positive Opinion Dictionary, Negative Opinion Dictionary, Opinion Statement Dictionary, Subjunctive Mood Dictionary, Adversative Dictionary, Noun Sentiment Dictionary and so on.
2. As a result of non-ideal performance of traditional opinion target extraction and orientation analysis based on syntactic dependencies, semantic role annotation and some semantic rules are introduced into sentiment analysis. The experimental results show that in the NLPCC 2013 micro-blog evaluation corpus, the method with semantic rules improve the performance significantly.
3. With combination of lexical meaning and word similarity computation algorithm, a opinion target search method is proposed, which is used to improve the precision of real opinion target search when the system only extract pronouns or opinion target is not found in syntactic dependencies.

The experimental results show that through improving the sentiment word library and adding the semantic knowledge, the performance of opinion target extraction and orientation analysis can be improved significantly.

**Key Words:** Syntactic Dependency, Opinion Target Extraction, Orientation Analysis, Semantic Rule

# 目录

[摘要 I](#_Toc445987299)

[Abstract II](#_Toc445987300)

[目录 III](#_Toc445987301)

[缩略词表 V](#_Toc445987302)

[第一章 绪论 1](#_Toc445987303)

[1.1 研究背景 1](#_Toc445987304)

[1.2 情感分析的基本概念 1](#_Toc445987305)

[1.3 中文评价对象抽取和倾向性分析的难点 2](#_Toc445987306)

[1.4 本文的研究目的和主要内容 3](#_Toc445987307)

[1.5 论文结构 4](#_Toc445987308)

[第二章 相关研究 5](#_Toc445987309)

[2.1 相关评测和情感资源 5](#_Toc445987310)

[2.2 评价对象抽取的研究现状 6](#_Toc445987311)

[**2.2.1.** **基于寻找频繁出现的名词和名词词组的抽取方法** 7](#_Toc445987312)

[**2.2.2.** **利用观点和评价对象的关系进行抽取的方法** 7](#_Toc445987313)

[**2.2.3.** **使用有监督学习进行抽取的方法** 8](#_Toc445987314)

[**2.2.4.** **使用话题模型进行抽取的方法** 8](#_Toc445987315)

[2.3 情感倾向性分析的研究现状 9](#_Toc445987316)

[**2.3.1.** **有监督的学习方法** 10](#_Toc445987317)

[**2.3.2.** **基于词典的方法** 10](#_Toc445987318)

[2.4 评价对象和情感词双向传播抽取方法 10](#_Toc445987319)

[2.5 本章小结 12](#_Toc445987320)

[第三章 依存句法分析 13](#_Toc445987321)

[3.1 概述 13](#_Toc445987322)

[3.2 分析方法 14](#_Toc445987323)

[**3.2.1.** **生成式依存分析方法** 15](#_Toc445987324)

[**3.2.2.** **判别式依存分析方法** 16](#_Toc445987325)

[**3.2.3.** **确定性依存分析方法** 17](#_Toc445987326)

[**3.2.4.** **基于序列标注的分层式依存分析方法** 18](#_Toc445987327)

[3.3 本章小结 19](#_Toc445987328)

[第四章 中文评价对象抽取和倾向性分析系统设计与实现 20](#_Toc445987329)

[4.1 情感资源的构建 20](#_Toc445987330)

[**4.1.1.** **情绪词典** 20](#_Toc445987331)

[**4.1.2.** **评价词典** 20](#_Toc445987332)

[**4.1.3.** **观点引述词典** 21](#_Toc445987333)

[**4.1.4.** **虚拟语气词典** 22](#_Toc445987334)

[**4.1.5.** **转折词典** 22](#_Toc445987335)

[**4.1.6.** **名词性情感词典** 22](#_Toc445987336)

[**4.1.7.** **否定词典** 22](#_Toc445987337)

[**4.1.8.** **伪否定词典** 22](#_Toc445987338)

[4.2 系统设计与实现 23](#_Toc445987339)

[4.3 预处理模块 24](#_Toc445987340)

[**4.3.1.** **无关符号过滤** 24](#_Toc445987341)

[**4.3.2.** **非核心成分过滤** 25](#_Toc445987342)

[**4.3.3.** **语料生成** 25](#_Toc445987343)

[4.4 评价对象和评价词抽取模块 25](#_Toc445987344)

[**4.4.1.** **SR规则** 26](#_Toc445987345)

[**4.4.2.** **ATT规则** 28](#_Toc445987346)

[**4.4.3.** **SN规则** 29](#_Toc445987347)

[**4.4.4.** **SO规则** 29](#_Toc445987348)

[**4.4.5.** **评价对象和评价词修正规则** 29](#_Toc445987349)

[**4.4.6.** **评价对象搜索算法** 30](#_Toc445987350)

[4.5 倾向性分析模块 31](#_Toc445987351)

[4.6 本章小结 32](#_Toc445987352)

[第五章 实验与分析 33](#_Toc445987353)

[第六章 总结和展望 34](#_Toc445987354)

[6.1 工作总结 34](#_Toc445987355)

[6.2 未来展望 34](#_Toc445987356)

[致谢 35](#_Toc445987357)

[参考文献 36](#_Toc445987358)

# 缩略词表

|  |  |  |
| --- | --- | --- |
| **缩略词** | **英文全称** | **中文全称** |
| NLPCC | The conference on Natural Language Processing and Chinese Computing | 自然语言处理与中文计算会议 |
| TREC | Text Retrieval Evaluation Conference | 文本检索评测会议 |
| NIST | National Institute of Standards and Technology | 美国国家标准与技术研究院 |
| DOD | Department of Defense | 美国国防部 |
| NTCIR | NACSIS Test Collection for IR | NACSIS信息检索评测 |
| JSPS | Japan Society for the Promotion of Science | 日本学术振兴会 |
| NACSIS | National Center for Science Information Systems | 日本国家科学咨询系统中心 |
| COAE | Chinese Opinion Analysis Evaluation | 中文倾向性分析评测 |
| CRF | Conditional Random Fields | 条件随机场 |
| HMMs | Hidden Markov Models | 隐马尔可夫模型 |
| PMI | Point-wise Mutual Information | 点式互信息 |
| pLSA | Probabilistic Latent Semantic Analysis | 概率潜语义分析 |
| LDA | Latent Dirichlet allocation | 隐式狄利克雷分布 |
| BN | Bayesian Networks | 贝叶斯网络 |
| MCMC | Markov chain Monte Carlo sampling | 马尔可夫链—蒙特卡洛采样 |
| NTU | National Taiwan University | 国立台湾大学 |
| MST | maximum spanning trees | 最大生成树 |
| LTP | Language Technology Platform | 哈工大的语言技术平台 |

# 绪论

本章主要阐述情感分析的研究背景、基本概念、中文评价对象抽取和倾向性分析的难点，最后介绍本文的研究目的和主要研究内容。

## 研究背景

观点或情感对于人类的一切活动和行为起着决定性的作用。我们的信仰、对现实的认知以及我们所作出的任何决定，从某种程度上，都取决于自身以及他人如何看待和评价所在的世界。正因为如此，每当我们需要做决定的时候，都会习惯性地先参考别人的观点。这对于个体不仅如此，对于大多数群体和组织而言更是如此。在现实生活中，个体消费者在购买某产品前往往想知道该产品使用者的评价；在选举中，选民也希望在投票前知道其他人对于候选者的观点；商业组织必须清楚地了解顾客或公众对于他们的产品或服务的看法，才能更好地做出商业计划和营销决策。

随着互联网的飞速发展，人们一方面在浏览别人发表的评论，一方面也在不停地分享自己对于某些人或物的观点和情感。尤其是微博、论坛、购物平台、新闻网站等公众媒体的快速发展，评论性信息迅速膨胀，仅靠人工的方法难以处理海量信息的收集、分类和分析。于此同时，仅仅获取个体的评价信息是没有意义的，群体的观点发掘才能更好地为决策者提供更好的决策依据。所以情感分析无论对于经济发展、政治决策还是社会个体的行为都有着极其重要的指引作用。因此，迫切需要计算机帮助用户快速获取和整理这些相关评价信息，情感分析（Sentiment Analysis）技术应运而生。观点以及与之相关联的概念包括情感、评价、态度和情绪等，都是情感分析和观点挖掘的主要研究对象。自从2000年开始，情感分析已经逐步成为自然语言处理领域中最为活跃的研究领域之一[1]。在数据挖掘、网络挖掘、文本挖掘等领域，它同样被广泛研究。由于情感分析对于商业和社会的重要性，它已经从计算机科学领域扩展到管理科学和社会科学领域。

## 情感分析的基本概念

情感分析，又叫做观点挖掘，主要研究人们对于实体（如产品、服务、组织、个体、议题、事件、话题和它们的属性）的观点、情感、评价、评估、态度和情绪。根据任务的侧重，情感分析还有其他不同的名称，例如观点抽取、情感抽取、主观性分析、影响分析、情绪分析、回顾分析等。

按照处理文本的粒度不同，情感分析主要分为文档级、句子级、实体和特征级。按照处理文本的类别不同，可分为基于新闻评论的情感分析和基于产品评论的情感分析。

目前情感分析有六个比较关键的任务，分别是实体抽取和分类，方面（特征）抽取和分类，观点持有者抽取和分类，时间抽取和标准化，实体方面情感分类，观点元组生成。基于这个框架的情感分析常常被称为基于方面或者基于特征的情感分析[2][3]。

基于方面的情感分析又可分为两个子任务，分别是评价对象抽取和情感倾向性分析。文档级和句子级的情感分析在现实使用中往往不足以满足需求，因为它们无法识别出每个实体或属性及其对应的情感倾向性。即使我们假设每个文档评估一个单独的实体，一个含有关于某个实体正明评价的文档并不意味着作者对该实体的所有方面或属性都持有正面评价。对于更加完整的情感分析，我们需要抽取出所有的评价对象并判断对于该对象的评价是正明还是负面的。

评价对象抽取（Opinion Target Extraction）是文本情感分析中的重要研究任务之一，评价对象是指一段文本中所讨论或者描述的主题、实体或者实体的某些方面和属性等，具体表现为评论文本中修饰词语或评论词语所修饰的对象，如新闻评论中的某个事件/话题或者商品评论中的某个产品的组件（如“手机电池”）等。本文研究主要针对中文短文本中显式的评价对象进行抽取，而非隐藏在文本背后的评价对象。

倾向性分析（Orientation Analysis），又可以称作情感分类，指的是抽取出文本中描述评价对象的情感词、极性词，对其做情感倾向性分析（正面评价或负面评价），这里的情感词特指的是具有情感倾向性的词语。对于有些具体场景，比如购物网站上的商品评论，其评价对象可能就已经给定了，此时我们就不必再做评价对象的抽取工作，只需要判断该评价对象的情感倾向性。

## 中文评价对象抽取和倾向性分析的难点

由于中文词语语义的多样性，在不同语境下可能会有不同的情感表述效果，加上网络用语的随意性，每个人表述情感的方式个性化，中文评价对象抽取和倾向性分析在实际应用中的效果很不理想，尤其是微博。目前，中文评价对象抽取和倾向性分析的难点主要有以下几点：

1. 同一句中，评价对象与评价词距离较远，同时，二者之间存在其他干扰性的评 价对象和评价词，导致二者无法被正确地关联和抽取出来。
2. 情感词存在，但是句中无明显的评价对象或只有相应的代词存在时，即评价对 象缺失，如何根据上下文正确找到情感词关联的评价对象。
3. 句子含有情感词，但是并不包含情感倾向性，可能只是表达某种心理活动或非 评价性的想法、愿望等。如，我希望小明能取得好成绩！
4. 句子中无明显的情感词存在，但依旧表达情感倾向性。如，他的店铺很有人气！
5. 转折词、否定词、情态动词等表达方式及其位置给倾向性分析带来不确定性。
6. 一些与情感词关联的名词，干扰了真正评价对象的抽取，例如，查韦斯病逝， 又一个伟大的英雄离开了。该句中可能抽取出来的评价对象和评价词是[英雄， 伟大]。
7. 情感分析基础工作的错误带来的干扰，例如中文分词、句法分析、语义分析过 程中的标注错误可能给后续的分析带来影响，尤其是后两者的效果还不理想。
8. 英文情感词典已经有非常完备的SentiWordNet，而中文情感词典资源却质量不 高，不够细 致，前期的基础不牢固，要得到准确的效果就不容易。
9. 大多数语料库都是基于商品评论的，如电脑、酒店等，标注细致的新闻评论、 微博等语料库太少，而这些正是中文评论语境最为丰富的领域。
10. 新词不断产生，基本常识、上下文知识和各式各样的实体词，以及中文句式类 型的复杂性都对中文评价对象的抽取带来很大的影响。

## 本文的研究目的和主要内容

使用依存句法分析能够很好地揭示出评论语句中各成分之间的修饰关系或依赖关系，通常观点词都是和评价对象相互关联的，而多数情况下观点词可以根据词典或分词的标注结果而事先得知，所以利用依存关系分析可以通过观点词抽取出一些不频繁出现的评价对象。同时，通过依存句法关系分析，可以获得观点词的上下文特征（如转折关系、递进关系等），进而有利于分析出观点词的情感倾向。所以，基于句法依存分析的方法能够很好地对评价对象和观点词进行抽取和分析。然而，目前对于中文文本进行评价对象抽取和倾向性分析的精度还不是很好，因此，本文希望通过丰富情感词典、添加语义知识、改进评价对象搜索方法来改善目前基于依存句法分析的中文评价对象抽取和倾向性分析的性能。

由于中文本身具有的复杂性与特殊性，决定了中文评价对象的抽取和倾向性分析所面临的困难更大。本文将对此展开研究，主要内容如下：

1. 构建用于情感分析的情感词典，主要包括：正面情绪词典、负面情绪词典、正 面评价词典、负面评价词典、观点引述词典、虚拟语气词典、转折词典、名词 性情感词典等。
2. 针对实验所使用微博评测语料的语言成分的多样性，进行了大量的预处理工 作，包括去除无关字符、链接等，同时对非核心句子成分进行过滤，进而缩短 评价对象和评价词的关联距离。
3. 由于传统的基于依存句法分析的评价对象抽取和倾向性分析效果不理想，本文 在依存句法分析的基础上，利用语义角色标注，并添加了一系列的语义规则进 行情感分析。
4. 结合词义和词语相似度计算算法，提出一种评价对象搜索方法，用于改善在只 抽取出代词或句法关系中无评价对象的情况下，搜索上下文中真正的评价对象 的精确度。

## 论文结构

本文的后续章节组织如下：

第二章介绍中文评价对象抽取和倾向性分析的研究现状，并加以比较。

第三章介绍本文使用的依存句法分析的定义和主流算法。

第四章详细介绍基于依存句法分析的中文评价对象抽取和倾向性分析系统的设计与实现。首先介绍用于情感分析的情感词典的构建，介绍各个词典的用途。然后对本文使用的评价对象和情感词抽取规则及评价对象搜索算法进行详细介绍。

第五章在NLPCC 2013微博情感要素抽取评测语料上进行实验，给出实验结果和分析。

第六章对全文的工作进行总结和展望。

# 相关研究

尽管语言学和自认语言处理有着很长的历史，但在2000年以前，关于人们观点和情感的研究非常少。随着互联网的飞速发展，这个领域的研究开始变得活跃起来。出现这种情况有如下几个原因：第一，情感分析有着广阔的运用领域，无论是商业还是政治决策。第二，情感分析领域有很多具有挑战性的问题并且以前从未被研究过。第三，这是人类历史上第一次出现这样的情况，通过网络上的公众媒体，我们能够拥有如此海量的情感数据。正因为此，情感分析得到了飞速的发展。

## 相关评测和情感资源

情感分析最先引起了国际文本检索会议TREC1(Text Retrieval Evaluation Conference)的关注，该会议是由美国国家标准技术研究所（NIST）和美国国防部（DOD）联合主办的，自从1992年开始就作为TIPSTER Text Program下的附属计划，每年召开一次。该会议主要关注如下几个方向：问答系统、特定领域搜索、传统Web检索等。其目标是：1）鼓励基于大型语料库上的信息检索研究；2）通过创建一个开放的平台来促进工业界、学术界和政府间的交流与合作；3）通过验证优秀的检索方法和技术对于解决现实问题的效果，加速技术从研究领域转向商业产品；4）提供更为先进有效的评估技术供工业界和学术界使用。情感分析相关的任务从2006年开始加入到会议评测中，并且微博检索任务TREC Blog Track[4-5]是TREC最先关注的情感分析任务，主要针对英文文本中的情感词或观点信息进行检索。

NTCIR（NACSIS Test Collection for IR）2计划是由日本学术振兴会（JSPS）和日本国家科学咨询系统中心（NACSIS）自1997年开始联合主办的，每年举行一次，并拥有中、英、韩、日四种语言的标准语料库，主旨是为了加强信息索引、问答系统、文本摘要、信息抽取等信息获取技术的研究。其目标与TREC很相近，都是为了促进相应技术的发展，提供各个领域的交流平台，研究更好的评估方法以及构建和提供用于信息获取领域实验的大规模数据集。NICIR的情感分析评测同样出现在2006年，主要针对文本的情感分类以及观点持有者的抽取等[6-7]。

随着中文文本的情感分析得到了越来越多的学者和研究机构的关注，为了推动中文情感分析技术的发展，国内第一个情感分析方面的评测COAE（Chinese Opinion Analysis Evaluation）[8]于2008年举办第一届。COAE目的在于推动中文情感分析理论和技术的研究和应用，同时建立相关的分析语料库。COAE 共设置6个任务，可分为3个方面：一是

1. http://trec.nist.gov
2. http://ntcir.nii.ac.jp

中文评价词语的识别和分析，属于词语级的情感分析评测；二是中文文本倾向性相关要素的抽取，主要是抽取句子中的评价对象，以及对于其观点的倾向性判别，属于语句级的情感分析评测；三是中文文本主客观性及倾向性的判别，属于篇章级的情感分析评测。

除了COAE提供了产品类的评价语料库，中国科学院计算技术研究所的谭松波博士提供的较大规模的中文酒店评论语料，约有10000篇，并标注了褒贬类别，可以为中文的篇章级的情感分类提供一定的帮助。

另外自然语言处理与中文计算会议NLPCC（The conference on Natural Language Processing and Chinese Computing ）自2012年开始也设置了一些热门主题的研究任务，包括情感分析、信息关联和组织、问答系统等。其中对情感分析提供了中文微博评测语料和三个任务，分别是微博情感分类、情感句识别和分类、情感要素抽取。

中文的评价词词典资源有NTU评价词词典（繁体中文）和HowNet评价词词典。NTU 评价词词典由台湾大学收集，含有2812个褒义词与8276个贬义词。HowNet评价词词典包含9193个中文评价词语/短语，9142个英文评价词语或短语，并被分为褒贬两类。而且该词典提供了评价短语，为情感分析提供了更丰富的情感资源。

## 评价对象抽取的研究现状

情感分析问题的一些特征可以用来帮助评价对象的抽取[1]。其中一个关键特征就是一个观点往往都会有一个对应的评价对象，这个对象往往是句中的某一属性或话题。所以识别出句子中的每一个观点表达式和它的评价对象非常重要。同时我们也需要注意到，一些观点表达式可能扮演两个角色，一个是表示正面或负面情感的情感词，另一个是隐式的评价对象。例如，“这本书很便宜”。该句的评价词时“便宜”，对应的隐式评价对象是“价格”。处理该类型的文本时，需要构建一个评价词和隐式评价对象的对应词典。本文主要研究句中出现的显式评价对象的抽取工作。

目前针对显式评价对象的抽取主要有四种方法，分别是基于寻找频繁出现的名词和名词词组的抽取方法，利用观点和评价对象的关系进行抽取的方法，使用有监督学习进行抽取的方法以及使用话题模型进行抽取的方法。

现在既有的研究主要是针对网络评论，其评论格式主要有下面两种：

1. 赞成意见、反对意见和详细评论：评论者首先会分别描述简要的优缺点，然后 再写一个完整的评论内容。
2. 自由格式：评论者可以自由发表观点。

格式（1）主要来自于商品、酒店评论等，相对于格式（2）要容易处理。Liu等[9]在2005年提出一种基于序列学习的方法抽取评价对象，主要利用赞成和反对意见的文本特征，如它们通常比价简短，句块由标点符号、连词等隔开，每个句块最多含有一个评价对象等，这些特征都较为有效地提高了抽取精度。所以目前针对商品评论的情感分析效果要优于微博评论。本文主要研究自由格式文本的评价对象抽取。

### **基于寻找频繁出现的名词和名词词组的抽取方法**

这种方法在给定的领域中从大规模的评论数据中寻找那些名词或名词性短语的显式评价对象表达式。Hu等[10]在2004年使用了一种数据挖掘算法，通过词性标注器识别出名词和名词性短语。计算这些潜在评价对象出现的频率，频繁出现达到阈值的对象才被保留下来。这种方法的依据是人们评论某个实体的不同属性时，所用词汇往往会有一定的聚集性。所以，那些频繁用到的名词往往是真正重要的评价对象。同样，那些非频繁的名词更可能是非评价对象或不重要的评价对象。尽管这种方法很简单，但是非常有效。有一些商业公司已经使用这种方法，效果也有所提高。

Hu的算法在Popescu等[11] 2005年的研究中得到了改进。他们的算法计算每个被发现的名词或名词短语与鉴别词（已经和某些实体类别关联的词）的点互信息（Pointwise Mutual Information, PMI）的评分来试图移除那些可能不是评价对象的名词或名词短语。以英文为例，对于汽车类的属性识别特征（或称为识别器）可能有：“of car”，“car has”，“car comes with”等，这些可以通过网络搜索来找到文本中的评价属性。PMI计算方式可以表示成下式：

（2.1）

其中，a表示使用频率搜索的方法识别出的潜在评价属性，d表示属性识别特征。网络搜索用于查找单独的评价属性和识别特征出现的次数和它们共现的次数。如果潜在评价对象的PMI值过低，则它不大可能是一个产品的某个组件或属性，因为a和d没有频繁共现。

Blair-Goldensohn等[12]在2008年通过只考虑含有情感词或某种句法模式发现情感倾向的语句，进一步改进了基于名词和名词性短语频率的抽取方法。Long等[13]通过基于频率和信息距离的方法抽取出评价对象。他们的方法首先使用基于频率的方法找到核心评价对象词，然后使用信息距离[14]来寻找其他与评价对象相关联的词语。

### **利用观点和评价对象的关系进行抽取的方法**

既然观点总是有对象的，所以他们总是相关联的。通常观点词都是很容易发现的，观点和评价对象之间的关系可以用来抽取评价对象。这种方法被Hu等[10]用来抽取非频繁出现的评价对象，他们没有使用任何分析器，当句子中不含有频繁出现的评价对象但含有一些情感词，那么离每个情感词最近的名词或名词短语被抽取出来。邻近评估函数的计算依赖于情感词和其所修饰的名词或名词短语的依存关系。Blair-Goldensohn等[12]基于情感模板的方法也是与此类似。这种基于关系的方法对于从含有观点的句子中发现重要的评价对象或主题都非常有效，因为无任何观点或情感词修饰的对象往往都不那么重要。

Zhuang等[15]在2006年使用句法依存关系分析器来抽取出依存关系，用于评价对象的抽取。Somasundaran[16]，Kobayashi[17]等人也相继使用类似的方法。Qiu等[18]在2011年进一步利用依存关系的思想融入双向传播的方法，同时抽取评价对象和情感词。Wu[19]等使用一个短语依存关系分析器从候选评价对象中抽取名词短语和动词短语，然后系统过滤掉那些非评价对象。目前利用依存关系进行抽取的方法已经被很多研究者运用到不同的研究任务中去[16]。另外，Zhou等[20]利用语义标注，结合评价对象和观点词之间的语义关系，使用含有大量语义知识的规则库来抽取评价对象，取得了不错的效果。

### **使用有监督学习进行抽取的方法**

评价对象抽取可以被看做一般信息抽取问题的特例。在过去的信息抽取领域，很多基于有监督学习的算法被提出。占主导地位的方法是基于序列标注的方法，因为这些都是有监督的学习方法，所以需要大量的标注数据进行训练。目前的序列标注学习方法有隐马尔科夫模型（Hidden Markov Models, HMM）[21]和条件随机场（Conditional Random Fields, CRF）[22]。Jin等[23]在2009年使用了词汇化隐马尔科夫模型来学习用于抽取评价对象和观点表达式的模板。Jakob等[24]在2010年利用词、词性标注、句法依存关系、词距离等特征，使用条件随机场在不同领域的评价语料上训练一个领域独立的评价对象抽取模型。Li等[25]融合了两种CRF的变种（Skip-CRF和Tree-CRF）来抽取评价对象和观点。不同于普通的链式CRF只能利用词序列进行学习，变种CRF可以进一步利用文本的结构特征。使用条件随机场方法的研究工作还有很多。

Kobayash等[26]在2007年提出的方法首先使用依存关系分析树寻找到候选的评价对象和观点词对，然后利用树结构分类算法学习和划分之前找到的评价对象和观点词对。Yu等[27]在2011年使用一个单类SVM[28]来标注某些是评价对象的正例。Ghani等[29]使用传统的有监督学习和半监督学习进行评价对象抽取。Kovelamudi等[30]利用维基百科上相关的信息进行有监督的学习和抽取。

### **使用话题模型进行抽取的方法**

近年来基于统计的话题模型开始成为从大量文本文档中抽取主题的一个主流方法。话题模型是一种无监督的学习方法，它假设每个文档是由多个话题混合组成，每个话题又是一个在词之上的概率分布。一个话题模型也就是一个文档的概率生成模型，其输出是一系列词的集合，每个集合构成了一个话题，同时也是文档集合中所有词的概率分布。目前有两种主要的基本模型：pLSA(Probabilistic Latent Semantic Analysis)[31] 和 LDA(Latent Dirichlet allocation)[32] 。

话题模型是一种基于贝叶斯网络（Bayesian Networks, BN）的图模型。话题模型尽管是用来对文本集合进行建模和话题抽取，但同时也可以对许多其他种类的信息进行建模。例如，在情感分析的上下文，基于观点词与评价对象存在对应关系，可以设计一个联合模型同时对观点词和话题进行建模。对于使用话题模型抽取评价对象仍然有一些不同之处，话题可以同时覆盖评价对象和评价词，对于情感分析而言，二者则需要分开。这种分开处理可以通过扩展基本模型（如LDA）来对评价对象和评价词联合建模。

Mei等[33]在2007年提出了一种联合模型用于情感分析。他们构建了一个评价对象和情感的混合模型（基于话题（评价对象）模型、正面情感模型和负面情感模型）。他们的方法主要是基于pLSA，其他大多数研究所使用的话题模型实施基于LDA的。

Titov等[34]在2008年提出像LDA的全局话题模型或许并不适合用于评价对象抽取。原因是LDA是利用文档间词共现和话题分布的不同来识别话题和话题间的词概率分布。然而，特定产品的评论所涉及的评价实体是相同的，这意味着全局话题模型将失效。Titov提出了多颗粒的话题模型，全局模型找到评价实体的同时，局部的模型找到实体的属性或某些方面（评价对象）。在这里，每个找到的评价对象都是一个一元语言模型。不同的词若表述相同或关联的事物则会自动聚合在同一个对象之下。然而，这种技术并没有分开评价对象和评价词。

Brody等[35]在2010年提出的方法首先利用话题模型识别评价对象，然后识别和评价对象相关的观点词（只考虑形容词）。Li等[36]在2010年提出两种联合模型（Sentiment-LDA和Depen-LDA），用来抽取带有正面和负面情感的评价对象。模型不是单独地寻找评价对象，而是将评价对象和观点词一并抽取。Mukherjee等[37]在2012年提出一种半监督的联合模型，模型允许用户提供评价对象的种子用来推理生成符合用户需求的评价对象分布。

尽管话题模型是一种可以扩展处理多种信息的概率推理模型，但它有一些缺点限制了它在情感分析应用中的实际使用。一个主要问题就是为了实现较理想的效果需要大量的数据和相当多的参数调优。更为糟糕的是，大多数话题模型使用了吉布斯采样（Markov chain Monte Carlo sampling, MCMC的一个特例），它会在每次运行时产生不同的结果。话题模型寻找频繁出现的评价对象并不难，难的是寻找那些局部的或全局非频繁的评价对象。那些局部频繁的评价对象往往在实际应用是最为重要的，因为它们往往是用户感兴趣的实体的某个方面或属性。

总之，话题模型是一个灵活而强大的建模工具，但对于现实情感分析而言，仅仅使用现有的话题模型很难得到理想的结果。一个有希望的研究方向是更多地融合自然语言和领域知识到模型中，这方面的工作已经有所进展[38-41]。

## 情感倾向性分析的研究现状

确定句子中每个评价对象的感情倾向主要有两种方法：有监督的学习方法和基于词典的方法。

### **有监督的学习方法**

Wei等[42]在2010年提出了一种层次分类模型，然而问题的关键在于如何确定每个情感词的覆盖范围，其是否修饰我们所关注的评价对象。一个主流方法是利用依存句法分析其依存关系以及其他的一些关联信息。Jiang等[43]在2011年使用一个依存关系分析器生成一系列的依存关系特征用于感情分类。Boiy等[44]在2009年提出的方法根据每个特征在分析树中相对于评价对象的位置来赋予每个特征相应的权重。对于比较句而言，“than”或其他相关的词可以用来分割句子[45-46]。

就目前而言，有监督的学习方法依赖于大量的有标注训练数据，某个领域训练好的模型在另一个领域往往表现的不是很理想。目前这个技术远远没有成熟，主要还是运用于文档级的情感分类，因为相对于句子，文档足够长且含有大量的用于分类的特征。

### **基于词典的方法**

基于词典的方法是一种无监督的方法，并且在很多领域里表现的都非常不错。很多研究主要使用一个包含了情感词、短语、成语的情感词典，利用复合表达式、观点规则以及可能需要的句法分析树来决定评价对象的情感倾向。当然基于词典的方法也要考虑情感转移词、转折从句和其他结构对于评价对象情感分析的影响。Ding等[47]在2008年提出的一个简单的方法主要使用了以下四个步骤确定一个情感词的情感倾向：

1. 标注出句子中包含的情感词和短语，并赋予正面评价和负面评价对应的评分。
2. 根据情感词附近的一些否定词等改变情感词的倾向。
3. 处理转折从句对情感分析的影响。
4. 合计句子中每个评价对象所有情感词倾向性的最后评分。

这个简单的算法在很多情况下都运行得相当不错。它可以相当好地处理类似于“Apple is doing very well in this bad economy”。除此外还有很多情感聚合的方法，Hu等[2]的方法是简单地对句中所有的情感词评分进行求和。Kim等[48]使用了词的情感评分的乘积。其他研究者也使用了类似的方法[49-50]。

事实上，上面简单的方法可以在很多方向得到加强。Blarir-Goldensohn等[51]在有监督方法的基础上融合了基于词典的方法。Kessler等[52]在2009年也提出了四种不同的决定情感词倾向性的策略，同时通过一些实验数据证明为什么在大量的标注数据上将情感词和评价对象关联起来是那么困难。Liu等[53]在2010年提出推导情感倾向性的一些基本规则。

## 评价对象和情感词双向传播抽取方法

评论词总有一个评价对象与之对应。基于这一点，Qiu等[18]通过利用情感词和潜在评价对象之间特定的句法关系以及一些种子情感词同时抽取出评价词和评价对象。基于观点词和评价对象之间的关系，情感词可以通过已知的评价对象识别出，同样评价对象也可以通过已知的评价词识别出。这种不断通过已识别出的评价对象或情感词识别出更多的情感词或评价对象的过程会一直进行直到没有更多的词被识别出。这种方法涉及到评价词和评价对象，因而被称为双向传播方法。其所使用的抽取规则是基于评价对象和评价词之间特定的依存关系。Qiu等使用依存文法[54]描述依赖关系，使用的依存关系分析器是minipar[55]。

这种方法对抽取的条件进行了限制，情感词是形容词，评价对象是名词或名词短语。同时，评价对象和评价词之间的依赖关系[54]（用{MR}集合表示）包括：mod，pnmod，subj，s，obj，obj2和desc。评价对象之间以及评价词之间的依赖关系（用{CONJ}集合表示）只有连词关系conj。具体的抽取规则见表2-1，其中OA-Rel表示情感词和潜在评价对象之间的关系，OO-Rel表示情感词之间的关系，AA-Rel表示潜在评价对象之间的关系。每一个关系都可以用一个三元组<POS(wi), R, POS(wj)>表示，POS(w)表示词w在句中的位置下标，R表示依存关系。

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observations | Output | Examples |
| R11  (OA-Rel) | *O→O-Dep→A*  ***s.t.*** *O∈{O}, O-Dep∈{MR},*  *POS(A)∈{NN}* | *a=A* | *The phone has a good “screen”.*  *good****→mod→****screen* |
| R12  (OA-Rel) | *O→O-Dep→H←A-Dep←A*  ***s.t.*** *O∈{O}, O/A-Dep∈{MR},*  *POS(A)∈{NN}* | *a=A* | *“iPod” is the best mp3 player.*  *best****→mod→****player****←subj←i****Pod* |
| R21  (OA-Rel) | *O→O-Dep→A*  ***s.t.*** *A∈{A}, O-Dep∈{MR},*  *POS(O)∈{JJ}* | *o=O* | *same as R11 with screen as the known word and good as the extracted word* |
| R22  (OA-Rel) | *O→O-Dep→H←A-Dep←A*  ***s.t.*** *A∈{A}, O/A-Dep∈{MR},*  *POS(O)∈{JJ}* | *o=O* | *same as R12 with iPod is the known word and best as the extract word.* |
| R31  (AA-Rel) | *Ai(j)→Ai(j)-Dep→Aj(i)*  ***s.t.*** *Aj(i) ∈{A}, Ai(j)-Dep∈{CONJ},*  *POS(Ai(j))∈{NN}* | *a=A i(j)* | *Does the player play dvd with audio and “video”?*  *video****→conj→****audio* |
| R32  (AA-Rel) | *Ai→Ai-Dep→H←Aj-Dep←Aj*  ***s.t.*** *Ai∈{A}, Ai-Dep=Aj-Dep OR*  *(Ai-Dep = subj AND Aj-Dep = obj),*  *POS(Aj)∈{NN}* | *a=Aj* | *Canon “G3” has a great len.*  *len****→obj→****has****←subj←****G3* |
| R41  (OO-Rel) | *Oi(j)→Oi(j)-Dep→Oj(i)*  ***s.t.*** *Oj(i)∈{O}, Oi(j)-Dep∈{CONJ},*  *POS(Oi(j))∈{JJ}* | *o=O i(j)* | *The camera is amazing and “easy” to use.*  *easy****→conj→****amazing* |
| R42  (OO-Rel) | *Oi→Oi-Dep→H←Oj-Dep←Oj*  ***s.t.*** *Oi∈{O}, Oi-Dep=Oj-Dep OR*  *(Oi /Oj-Dep ∈{pnmod, mod}),*  *POS(Oj)∈{JJ}* | *o=Oj* | *If you want to buy a sexy, “cool”, accessory-available mp3 player, you can choose iPod.*  *sexy****→mod→****player****←mod←****cool* |

表2-1 评价对象和评价词双向抽取规则

## 本章小结

本章首先介绍了情感分析的相关评测好和情感资源，然后介绍了国内外评价对象抽取和情感倾向性分析的主要研究方法。

# 依存句法分析

## 概述

“依存”指的是句子中词与词之间支配和被支配的关系，用这种关系描述语言结构的语法称为依存语法（dependency grammar），也可以称为从属关系语法。依存语法打破了传统语言学家一直强调的主谓关系，突出了句中各个词之间的直接或间接关系[57]。利用依存语法进行句法分析，称之为依存句法分析，又称依赖分析，其目标是找出句子中所有词之间的依赖关系。

依存关系可以用有向边（或称弧）表示，关系的支配者位于起始处，被支配者位于箭头处。也可以用依存树表示，其中父节点表示支配者。另外依存投射树利用高位置表示支配者，低位置表示被支配者的方式表示依存关系。这三种表示方式如图3-1，3-2，3-3所示。

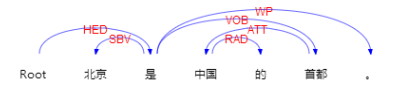


图 3-1有向图

图3-2依存树 图3-3依存投影树

在现实使用中，往往会给有向图或依存树的边、节点等添加更多的标记，以携带包括词法、句法、语义等更多的信息，如图3-4所示。

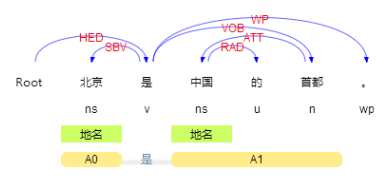


图3-4包含更多信息的依存关系图

20世纪70年代，语言学家J. Robinson[58]提出依存语法中关于依存关系的四条公理，在处理中文信息的研究中，中国学者冯志伟[59]提出了依存关系的第五条公理，总结如下：

1. 一个句子中只有一个成分是独立的。
2. 其它成分直接依存于某一成分。
3. 任何一个成分都不能依存与两个或两个以上的成分。
4. 如果A成分直接依存于B成分，而C成分在句中位于A和B之间，那么C或 者直接依存于B，或者直接依存于A和B之间的某一成分。
5. 中心成分左右两面的其它成分相互不发生关系。

这些公理的作用是为了对依存图或依存树进行条件约束，保证其具有唯一父节点、连通、无环和可投射以及句子的依赖关系构成一棵有根的树结构[57]。

## 分析方法

Hays[60]和Gaifman[61]是最早使用依存形式进行句法分析的，他们建立了一种类似于上下文无关文法（CFG）的依存文法[57]，Gaifman的理论体系包括三个规则：

1. LI：类似的规则表示范畴按照给定顺序依存 于范畴，位于位置。
2. LII：某一范畴所有词的规则均可列出，每个范畴至少有一个词，每个词至少属 于一个范畴，也可属于多个范畴。
3. LIII：支配一个句子的所有范畴规则均可列出。

由于和上下文无关文法差别不大，基于这种形式的句法分析推理算法也就和用于上下文无关句法分析的CYK算法[62][63]以及后来的Earley算法[64]基本相似，都是一种自底向上的动态规划算法[60]。

然而类似于CFG的算法对语言的投射性进行了限制，难以解决非投射现象。对此，Karlsson[65-66]，Maruyama[67]等人提出了约束语法和基于约束的依存句法分析方法。将依存句法分析看成约束满足问题，使用一系列形式化的约束性表达式来表达如何构造一棵约束满足的依存句法分析树，同时把不符合条件的进行剪裁。

2001年Covington等[68]人认为依存句法分析器的处理方式应当类似人的认知模式，也就是“依次读入”和“立即处理”。这种确定性句法分析器的基本原理是：处理一个句子时，每读入一个词，就对该词进行处理，判断其是否可以与之前的词建立某种依存关系（父节点还是子节点），在后面的分析中对于已确立依赖关系不再改变。

作为天然词汇化的语法理论，依存语法在纯粹的数据驱动框架下才充分发挥了它弱形式化的优势。随着自然语言处理的飞速发展，在统计机器学习的辅助下，出现了三种具有典型代表性的数据驱动的统计依存分析方法[57]：生成式依存分析方法、判别式依存分析方法和确定性依存分析方法。另外也有国内学者提出了基于序列标注的分层式依存分析方法，下文将逐个介绍。

### **生成式依存分析方法**

生成式依存分析方法采用了联合概率分布模型生成含有概率分值的依存分析树，然后采用相应的算法找到分值最高的打分树作为输出结果[57]。

（3-1）

式3-1中，表示句子，表示分析结果，表示模型参数，表示联合概率模型，为训练实例，为实例个数。Eisner在1996年提出了基于双词汇语法的生成式依存分析方法[69][70]，该方法为词间的依赖关系引入概率，同时给出了三个概率模型：

1. 二元亲和词汇模型（bigram lexical affinites）

（3-2）

式3-2中，表示词标记（）和词本身（）；取值0或1，表示位置和 处的词是否建立依存关系（）；是句子长度。该模型表示某位置标记的产生是由该位置的前两个标记决定的，该位置的词是由标记决定的，每一对词是否存在依赖是由他们的标记和词本身共同决定的。

1. 选择偏好模型（selectional preferences）

（3-3）

式3-3中的模型加入了选择偏好的信息，利用该信息限制模型为每个词偏好选择依赖关系的支配节点。

1. 递归生成模型（recursive generation）

（3-4） （3-5）

式3-4中，表示位置右边第个最近子节点（表示左子节点），和分别是左右两边的子节点个数。每个的左右子节点分别是由两个马尔可夫过程产生。每个子节点由其支配词和前一个子节点决定，是自顶向下的递归生成式模型，即双词汇语法[71]。

Eisner的模型可以采用动态规划算法，自底向上不断合并两个最优子树，最后生成一棵完整的最优依存树。由于该方法采用联合概率模型，其概率乘积分解需做近似估计，同时采用全局搜索，算法复杂度较高，效率较低，但是准确率有一定优势[57]。

### **判别式依存分析方法**

判别式模型不像生成式模型那样需要独立性假设，其训练是为了寻找使打分函数最大的。这方面最具代表性的研究成果就是R. McDonald等人开发的最大生成树（maximum spanning trees, MST）依存句法分析器（MSTParser）[72-74]。MSTPaerser是一种基于图的模型，其分析目标是寻找最高打分依存树，整棵树的打分由每条边的打分加权求和得到。

（3-6）

式3-6中，表示对边的打分，是句子的依存树，是权重向量，是节点与的高维二元特征函数向量，的取值0或1，表示节点与间的依赖关系。

判别式模型在依存句法分析方面需要解决三个问题：

1. 如何确定最优打分树。
2. 如何训练出最佳参数。
3. 如何选择特征函数。

对于如何确定打分树，若句子具有投射性，可以使用动态规划算法自顶向下找到最佳打分树[72]，反之，则使用Chu-Liu-Edmonds算法[73]。这个算法的基本思想是：一个收缩图的最大打分树等价于原始图的最大打分树，首先从每一个节点选择一个分值最大的入边，若破坏了无环条件则将这个环收缩到节点上重新选择入边。

在参数训练方面，MSTParser使用在线学习方法训练权重[74]，基本过程是：初始化，每读入一个实例，更新。更新采用的是边缘注入松弛算法：

（3-7）

MSTParser需要考虑的特征包括：父子节点对特征、父子节点间和父子节点两边其他节点的信息等。

生成式和判别式的分析方法都是用动态规划算法在全句范围内进行搜索，得到的结果也是全局最优的，但是算法的高复杂度限制了其实际使用。本文使用的依存句法分析模型就是基于图的分析模型。

### **确定性依存分析方法**

确定性依存分析方法的基本思路是对于当前分析的每一个词确定唯一的动作，成为前面已分析词的父节点或子节点，有时候可能需要对前面的动作进行回溯或修改。这种方法又称为基于转换的方法，用这种方法得到的依存句法树必然是唯一的。基于这种思路的有两个比较具有代表性的方法：Yamada算法[75]和Nivre算法[76]。

Yamada算法是基于“移进-归约”的方法，主要使用三个分析动作：移进、左归约和右归约。从句子开头遍历到结尾，多次使用线性判别分类器决定三个分析动作中的一个，循环遍历直到所有词都直接或间接依赖于唯一的根节点。基本思路如表4-1所示，其中，W表示分析窗口，一般宽度为2，A为当前已分析出的依赖关系集合，wi和wi+1是当前分析的一对词。

表3-1 Yamada算法的分析动作

|  |  |  |
| --- | --- | --- |
| 动作 | 当前窗口 | 此次结果 |
| 移进 | …W[wi, wi+1]… | …W[wi+1, wi+2]… |
| 左归约 | …W[wi, wi+1]… | …W[wi]…；A∪{wi→wi+1} |
| 右归约 | …W[wi, wi+1]… | …W[wi+1]…；A∪{wi+1→wi} |

Nivre算法是一种基于栈的算法，包含了两种分析动作，一种就是前面标准的“移进-归约”，包括移进、左归约、右归约；另一种取自Abney等人的Arc-eager算法[77]，此处不再做详细介绍，具体可参考文献[76]。

由于确定性的分析方法得到的结果往往是局部最优的，其最大的问题就是错误传递，进而使得其准确率和生成式方法相比并无优势，但是它的时间复杂度是线性的，通过添加更多的特征，可以使其分析精度得到提高。

### **基于序列标注的分层式依存分析方法**

基于图的方法是以整个句子为依存关系搜索单元，确定性的方法是以邻近的词对作为分析单元，这两种方法的分析单位走了两个极端。于是，鉴萍等[78]人提出使用位于二者之间分析单元——依存层。依存层指的是依赖关系长度不大于1的词组成的结构。该方法使用CRF对分词后的句子进行序列标注，主要标注出相邻词间存在的依赖关系（依存层）；然后对依存层进行归约（只对当前已找到所有孩子节点的词进行归约），进入下一层序列标注和归约，直到所有词都归约到一个节点。基本流程以图3-5为例。





中间步骤略



图3-5 分层式依存句法分析示例

根据鉴萍的实验[78] ，这种方法分析的精度位于处于基于转换的方法和基于生成树的方法之间，但是利用序列标注模型显著地提高了分析效率，有很好的现实使用价值。

## 本章小结

本章主要介绍依存句法分析基本概念、当前的主流算法，下一章将详细阐述基于依存句法分析的中文评价对象抽取和倾向性分析系统的设计与实现。

# 中文评价对象抽取和倾向性分析系统设计与实现

## 情感资源的构建

情感资源的构建对于情感分析能否取到好的效果起到至关重要的作用，本文主要需要构建以下情感词典：正面情绪词典、负面情绪词典、正面评价词典、负面评价词典、观点引述词典、虚拟语气词典、转折词典、名词性情感词典、否定词典、伪否定词典。其中，情绪词典和评价词典是在NTU情感词典的基础上重新修改、整理出来的。下面将就各个词典的用途进行详细介绍。

### **情绪词典**

情绪词典主要分为正面情绪词典、负面情绪词典，主要反映的是评价主体内心对人或物（评价对象）的情感、喜恶、褒贬等主观感受。该类词一般以动词或形容词为主。

例1 我非常喜欢梁静茹的歌。

例1中“喜欢”一词是评价主体“我”对评价对象“梁静茹的歌”的主观情感，应当把“梁静茹的歌”作为评价对象抽取出来，同时“喜欢”是评价词。类似的词还有 “崇拜”、“拥护”、“力挺”、“讨厌”、“鄙视”等词，在Zhou等[20]人的研究中，这一类表达主观情感的动词称为后指动词。也就是说，无论该动词前面的成分是什么（被动句等个别句型除外），它的语义总是指向后面的宾语（即评价对象）。

与Zhou等人不同的是，本文的情绪词典不仅仅包括这类后指动词，还包括一些表达评价者主观情绪的词，如“高兴”、“难过”等，这一类词多以形容词为主，在句中可能不一定有所谓的宾语成分。

例2 我今天很高兴！

例3 梁静茹给我签名了，今天太高兴了！

例2是情绪句并非评价句，虽然含有情感词，但是仅仅表达内心的情绪，并没有对任何客体对象做评价。而例3中不仅仅有“高兴”这一情感词，还有与之关联的情感对象。从语义上来看，评价主体出现正面情感波动的原因是“梁静茹给我签名了”，表达的是评价主体对“梁静茹”的喜爱，“梁静茹”应当作为评价对象抽取出来。

### **评价词典**

评价词典主要分为正面评价词典和负面评价词典，与情绪词典中的词表达情绪、感受不同，该词典中的词更倾向于表达对人或物或事的观点或评论。这类词包括：“美丽”、“有趣”、“伤害”、“有利于”、“败类”等，多以形容词、名词、动词、副词等为主。一般情况下，当这类词为形容词或名词时，其修饰或指代的对象为评价对象；当这类词为动词或副词时，其动作的实施者或主语为评价对象。

例4 这是个有趣的故事！

例5 这个人是个十足的败类。

例4中评价词“有趣”直接修饰“故事”，例5中“败类”其实是主语“这个人”的宾语，既可以通过依赖关系将二者关联起来，也可以通过后文介绍的潜在评价对象搜索算法寻找名词“败类”指代的前文的潜在评价对象。

例6 不动产登记条例有利于楼市稳定！

例7 李敏成功地完成了任务。

例6中评价词“有利于”直接反映了主语“不动产登记条例”的成效，例7中副词“成功”修饰了主语“李敏”的动作，也是正面评价。

### **观点引述词典**

上文的情绪词典和评价词典主要是用来确定单个情感词的情感倾向性以及对应的评价对象位置，而观点引述词典主要是为了过滤句子中非核心成分，以减少不相关成分对评价对象抽取的干扰。本文定义的观点引述词分两种：一种是以介词开头，引述性词结尾的短语结构，如：“就…而言”，“据…判断”等；另一种是陈述性词语，如：“猜测”、“说明”、“感觉”等。

例8 查韦斯逝世了，就他和布什间的博弈而言，他是个了不起的总统。

例9不动产登记条例出台，潘石屹评论，这有利于经济软着陆。

对于介词开头的观点引述词，本文过滤掉介词和引述性词之间所有的成分（包括多余的标点符号）。如例8过滤后的结果是“查韦斯逝世了，他是个了不起的总统。”，这时候对情感词“了不起”寻找评价对象时，就少了“布什”等其他非评价对象的干扰。

对于陈述性词，如果其前面以名词短语或名词开头（如机构名、人名和一些专有名词）开头，那么名词短语和陈述性词之间的成分都可以过滤掉（包括多余的标点符号）。如例9过滤后的结果是“不动产登记条例出台，这有利于经济软着陆。”，这样就可以缩短评价词“有利于”和评价对象“不动产登记条例”之间的关联距离。这种过滤方法在Zhang等[56]人的研究中首次提及，他们是利用核心句和句法关系进行评价对象抽取。

### **虚拟语气词典**

虚拟语气词是用来过滤那些非真正表示情感倾向性或观点的情感词，这类词包括：“如果”、“希望”、“为了”等。这类词往往是表达作者某种期望、想法或未知猜测、条件假设等。

例10 我喜欢梁静茹，希望她的新作能取得成功！

例10中“希望”表达了作者对“她的新作”的期望，并非真正使用“成功”评价“她的新作”。但是需要注意的是，并不是含有虚拟语气词的句子就不含有评价成分。本文所做的仅仅是过滤掉位于虚拟语气词支配的句子成分中的评价词，这样例10中依旧可以抽取出评价对象“梁静茹”和评价词“喜欢”。

### **转折词典**

转折词主要是用来处理情感倾向性发生反转的情况。这类词主要包括：“但是”、“可是”、“然而”等。有时候在句中对于同一个评价对象，可能同时存在正面评价和负面评价，如果这个时候含有转折词，那么转折词后的评价词才是作者真正想要表达的对评价对象的倾向性。

例11 老罗的锤子手机，虽然不怎么好看，但是很用心。

例11中，“但是”之后的“用心”才是对“老罗的锤子手机”的真正评价，此时就可以忽略“但是”之前的“不怎么好看”。

### **名词性情感词典**

汉语中还有一些名词单独出现不表示任何情感，但是和“有”、“无”、“没”等词组合起来后就可以表示一定的情感倾向性。这类词包括“意思”、“意义”、“能力”、“地位”、“气质”、“性价比”等。这些词和“有”组合起来就可以表示正面评价，和“没有”、“没”、“无”等词组合起来就可以表示负面评价。

### **否定词典**

该词典主要包含表示否定意义的词，如“不”、“无”、“否”、“没有”等，用于情感倾向性反转。

### **伪否定词典**

该词典主要包含那些不表示否定意味但是含单个否定词的词，如“不得不”、“不可否认”、“非常”等，用于防止错误的情感倾向性反转。

## 系统设计与实现

原始语料

ATT规则

DO规则

主题语料

特有名词

语料

评估语料

对齐语料

预处理

非核心成分过滤

无关符号过滤

评测语料

句法分析

分词

语义角色标注

依存句法分析

命名实体识别

语言分析

词法分析

词性标注

语义分析

评价对象和评价词抽取

SR规则

情感

词库

评价对象和评价词修正规则

NE规则

倾向性分析

对齐和评估

图4-1 评价对象抽取和倾向性分析系统框架图

基于依存句法分析的中文评价对象抽取和倾向性分析系统主要分为六个部分：情感词库、预处理模块、语言分析模块、评价对象和情感词抽取模块、倾向性分析模块和评估模块。系统框架如图4-1所示，各个模块作用简介如下：

1. 情感词库：提供非核心成分过滤、情感词抽取和倾向性分析的知识库支持。
2. 预处理模块：处理与情感分析不相关的符号、链接、非核心成分等过滤，生成 出用于后面分析使用的主题语料、特有名词语料、评测语料、评估语料、对齐 语料。
3. 语言分析模块：对评测语料进行分词、词性标注、命名实体识别、依存句法分 析、语义标注处理。本文主要使用的是哈工大的语言技术平台LTP1进行语 言分析处理。
4. 评价对象和评价词抽取模块：使用语言分析模块生成的句子的依存关系和语义 语义标注，依据SR规则、ATT规则、SN规则、SO规则和情感词修正规则抽 取出评价对象和评价词。
5. 倾向性分析模块：对评价对象关联的所有评价词进行分析，计算出评价对象的 最终倾向性，并标注。
6. 评估模块：利用对齐语料得到评价对象在评估语料中的位置，评估系统性能。

## 预处理模块

### **无关符号过滤**

本文处理的语料主要取自中文微博，微博中含有大量不规范的语言成分，如网址链接、表情符号、用户自定义符号、主题分隔符（#）等等，这些对分词以及句法依存分析产生很大影响。因此，需要预先对其进行处理：

1. 删除微博中的所有http链接，以及和链接相关联的上下文，如：“来自新浪视 频http:// t.cn/zTzZAAN”，整个这段都应该被过滤掉。
2. 微博中很多表情符号和用户自定义的符号有时候并不真正表达其表面上的情 感，同时对句法分析和分词都带来很大的干扰，因此本文对其也进行了过滤。
3. 对主题分隔符‘#’进行进行替换处理，原因同上。
4. 对被括号括起来的含有@的字符串进行过滤，因为这些一般是微博中新闻来源 等标注，如：“（来自@新浪新闻）”。
5. http://www.ltp-cloud.com/

### **非核心成分过滤**

本文过滤的非核心成分主要是指句子中如“据我所知”、“根据有关专家分析”、“王石认为”等引述性成分。这些成分后面往往会包含评论性内容，然而这些成分中代词、名词或名词短语可能会因为依存分析或语义角色标注的结果干扰评论性内容中评价词和真正评价对象的关联。例如：“查韦斯逝世，王石认为他是一个伟大的反美斗士！”。这句中评价词“伟大”与“他”关联，在对“他”进行指代搜索时，可能会错误地关联到“王石”（当然也可以在指代搜索时进行非核心成分排除，但是会增加系统复杂性，本文选择在预处理时过滤掉这些成分）。

本文利用前文的创建的观点引述词典，对去除了无关符号的预处理语料进行非核心成分过滤。上例经过处理后变成“查韦斯逝世，他是一个伟大的反美斗士！”，此时在进行指代搜索时，“他”很容易对应到“查韦斯”，进而抽取出“[查韦斯，伟大]”。

### **语料生成**

经过前两部处理后，抽取生成出用于后面分析使用的主题语料、特有名词语料、评测语料、评估语料、对齐语料：

1. 对于微博开头或末尾以#或书名号（其中书名号与正文有空格隔开）包裹的内容，视为主题并抽取出来生成主题语料，同时过滤掉多余的#。主题语料将用于后面潜在评价对象搜索算法使用。
2. 特有名词语料主要指微博语料中备书名号包裹的内容，主要是书名、音乐名、电影名等该句中特有的名词。该语料可以用来区分句子中哪些未被书名号包裹的内容是特有名词。
3. 评测语料就是指经过上面过滤、抽取处理后剩下的微博语料，用于后面的语言分析和情感分析。
4. 评估预料中包含了每个微博语句的评价对象在句中的位置下标和相应的倾向性标注，用于后面的性能评估。
5. 对齐语料用于计算抽取出的评价对象在原句中的位置，用于评估计算。

## 评价对象和评价词抽取模块

在本模块中主要使用五个规则对前面语言分析的结果进行处理，抽取出评价对象和评价词。我们首先使用SR规则、ATT规则、SN规则和SO抽取抽取出评价对象和评价词，然后使用评价对象和评价词修正规则对抽取出来的结果进行修正，同时过滤掉不符合要求的结果。利用规则抽取时，可能需要使用到评价对象搜索算法，该算法主要是用于抽取并返回指定文本中包含的评价对象或代词等指代的评价对象，同时也返回搜索过程中找到的与评价对象对应的评价词（如果有，若没有则返回空），具体内容将在后面介绍。

### **SR规则**

SR（Semantic Role）规则是本文系统中唯一利用语义角色标注进行评价对象和观点词抽取的规则。系统重点关注的语义角色标注类型包括：谓词（PRED）、副词（ADV）、转折词（DIS）、动作的施事者（A0）、动作的受事者或对象（A1）、动作的间接受事者（A2，若前面出现过A1默认用A2对应的文本取代A1）。该规则融入了评价词的前指语义和情绪词的后指语义（见情感资源构建一节），本处不再介绍与之关联的细节部分。

**算法4.1**（**SR规则抽取算法**）

输入：以谓词为中心的语义角色标注结果（用PRED、ADV、DIS、A0、A1表示各个 标注对应的文本内容）；

输出：临时评价对象和评价词对集合S，已分析范围集合R.

1. 若PRED或ADV或A0或A1含有情感词（即在情绪词典和评价词典中能找到 该词），则进入（2），反之进入（3）。
2. 依情感词出现的不同标注类型顺序进行评价对象和评价词抽取：
   1. 谓词是情感词（包括情绪词和评价词）。若A0、A1都不存在并且谓词是 评价词，则利用评价对象搜索算法寻找空的A0对应的潜在评价对象t，将 (t, PRED)加入S中；若A0 或A1存在一个：
3. 谓词是评价词，同时A0存在或A1在谓词之前（A0不存在），则利 用评价对象搜索算法寻找A0对应的潜在评价对象和评价词(t, o)；
4. 反之利用评价对象搜索算法寻找A1对应的潜在评价对象和评价词(t, o)。

将(t, o)和(t, PRED)加入S中。进入（4）。

* 1. 副词是情感词。若A0不存在，A1存在且谓词在A1的后面，利用评价对 象搜索算法寻找A1对应的潜在评价对象和评价词(t, o)；反之，利用评价 对象搜索算法寻找A0对应的潜在评价对象和评价词(t, o) 。o不为空则将 (t, o)加入S中，o为空则将(t, ADV)加入S中，进入（4）。
  2. A0或A1中含有情感词：

1. A1存在。当A1在谓词前面，则利用评价对象搜索算法寻找A1对应 的潜在评价对象和评价词(t, o)；反之优先利用评价对象搜索算法寻找 A0对应的潜在评价对象和评价词(t, o)，若未找到，则再利用评价对象 搜索算法寻找A1对应的潜在评价对象t和评价词o。若A1含情感词 o1则将(t, o1)加入S中，反之则将(t, o)加入S中。
2. A0存在。利用评价对象搜索算法寻找A0对应的潜在评价对象和评价 词(t\_0, o\_0)，若A1存在，则利用评价对象搜索算法寻找A1对应的潜 在评价对象和评价词(t\_1, o\_1)，检测A1中是否含有人称指代（见后文 的人称指代检测），若不含有或含有的人称指代对象恰好在A0中，则 将(t\_0, o\_1)加入S。若A0不是人称代词且谓词不是使A0成为潜在评 价对象的动词（如：“是”、“使得”、“作为”等词），则令潜在 情感要素集合{(t, o)} ={(t\_0, o\_0)}+{(t\_1, o\_1)}，否则{(t, o)} ={(t\_1, o\_1)}。遍历{(t, o)}，若A0含情感词o0则将(t, o0)加入S中，反之则 将(t, o)加入S中。

进入（4）。

1. 当A1、谓词都存在，谓词在A1之前且谓词含有“有”、“无”、“没”时，若A1在名词性情感词典中，则利用评价对象搜索算法寻找A0对应的潜在评价对象t。谓词对情感极性的影响（如“有意思”和“没意思”）将(t, 好)或(t, 坏)加入S中。
2. 将前面已经分析的文本内容范围加入到R中，返回。

**部分规则示例4.1**

1. 谓词是情感词：
2. “真好看啊！”这句从前文中寻找评价对象t，若找到则将(t, 好看)抽取出来。
3. “他很厉害！”这句寻找“他”指代的评价对象t，将(t, 厉害)抽取出来。
4. “喜欢上了查韦斯！”这句查韦斯是A1成分，将(查韦斯, 喜欢)抽取出来。
5. 副词是情感词：
6. “中国房地产 健康持续地发展”这句“中国房地产”是A1成分，作为“发展”的受事对象被标注出来，在谓词“发展”的前面，所以将(中国房地产, 健康)抽取出来。
7. “习近平 成功地完成了改革！”这句将(习近平, 成功)抽取出来。
8. A0或A1中含有情感词：
9. A0A1均有：“小明是个努力的孩子。”这句将(小明, 努力)抽取出来；“伟大的查韦斯离开了他的人民。” 这句将(查韦斯, 伟大)抽取出来。
10. 有A0无A1：“伟大的革命斗士走了！”这句需要对“伟大的革命斗士”进行评价对象搜索寻找真正的评价对象t，将(t, 伟大)抽取出来。
11. 有A1无A0：“不动产登记条例出台了，真是个好政策！”这句需要先进 行评价对象搜索找到潜在的A0对应的评价对象——“不动产登记条例”， 然后抽取出(不动产登记条例, 好)；若“不动产登记条例出台了”这个子句 不存在，则之抽取出(政策, 好)。
12. 谓词含“有”的例子：“这个故事非常有意思！”这句中“有意思”的分词成分虽然不具有情感，但是组合起来却表达正面倾向性，所以将“这个故事”作为评价对象抽取出来。

### **ATT规则**

ATT规则主要根据句法依存分析的结果，将所有ATT（Attribute，定中关系）标注范围内的且位置不在前面规则已分析范围中的词块抽取出来作为候选评价对象。这样是为了将未作为语义成分A0或A1但可能是评价对象的短语成分抽取出来，这些短语称为定中短语，其中心语一般是名词、代词、数词、体词性联合短语等，有时谓语性成分也可以充当中心语，例如“生活水平的改善”、“人工智能的发展”等。由此可见，这种做法能抽取出很多非名词性短语，比现有只抽取出名词性短语作为候选评价对象要好。需要注意的是，对于定中短语中可以进一步抽取出来作为评价对象的成分，如“伟大的毛主席”、“《三体》的成功”等，需要作进一步抽取。同时，我们将LTP词性标注集下的形容词、副词、动词、名词、惯用语等作为候选评价词。

本系统参照Qiu等[18]提出的双向传播规则对前文抽取出的候选评价对象和评价词进行筛选。通过分析，我们对提取规则进行了改进并扩展了新的间接抽取规则。实验证明扩展后的间接规则能进一步改善抽取效果。详细规则如下：

1. Head规则：







1. Direct规则：







1. Indirect规则：











### **SN规则**

SN（Single Noun）规则主要根据词性标注的结果，将除了表示方位、时间外的且不在前面规则已分析范围中的单个名词（包括一般名词、人名、地名、机构名、地理位置名、专有名词等）抽取出来作为候选评价对象。这样可以获取大多数以单个词形式出现的评价对象。同时，我们将LTP词性标注集下的形容词、副词、动词、名词、惯用语等作为候选评价词。接着利用类似上述ATT规则使用的改进的双向传播规则筛选出评价对象和评价词。

### **SO规则**

SO（Single Opinion）规则主要针对句中评价词无其他评价对象与之建立依赖关系，仅有根结点（Head）与之关联的情况。这时需要使用评价对象搜索算法寻找前文可能存在的与该评价词对应的评价对象。例如“查韦斯上台了。了不起！”，这段微博在处理下句时抽取出“了不起”，需要搜索到上文中的“查韦斯”作为其评价对象抽取出来。

### **评价对象和评价词修正规则**

评价对象和评价词修正规则主要针对前面规则已经抽取出来的评价对象和评价词进行检测和修正，主要包括以下几个子规则：

1. 有效性检测规则，符合以下任一条件的评价对象和评价词对都是无效结果，需要剔除：
2. 评价对象长度小于2且不是名词；
3. 评价对象是情感词且不是名词或评价对象包含有评价词；
4. 评价对象只含有副词、动词、惯用语、量词、数词或以方位词、时间词结尾等。
5. 评价对象是人称代词或以职位、称号（如“总统”、“斗士”）等结尾，则需要利用评价对象搜索算法在上下文中寻找人名。
6. 虚拟语气检测规则：对位置在“希望”、“猜测”、“不知道”等虚拟语气词支配的子句中的评价词及其对应的评价对象进行剔除。
7. 无效符号过滤规则：过滤掉评价对象首尾无效的标点符号，如“，”、“；”等，这些主要是语义标注时所引入的。
8. 情感词转折规则对SR规则抽取出来的评价对象和评价词对，若同一评价对象对应多个评价词，根据情感转折词典，优先选择位于转折词从句中的评价词。

### **评价对象搜索算法**

评价对象搜索算法主要处理两种种情况：当前分析成分中有评价词但无评价对象；有评价对象，但词性是代词或称呼类名词。这两种情况都需要从上下文中寻找真正的与评价词对应的评价对象。

**算法4.2**（**评价对象搜索算法**）

输入：当前分析上下文（包括当前的句子）的词性标注I，依存句法分析结果D，上 下文所有人名集合S，主题中的候选评价对象集合T，待处理评价对象t（可为 空）；

输出：评价对象target.

1. 若当前针对的是SR规则中空A0成分（即t为空）寻找评价词对应的评价对象， 优先根据D中的SBV（主谓关系）标注寻找当前分析成分的谓词对应的主语 作为t。
2. 若t是为人称代词或以称呼类名词结尾的词（利用词语相似度算法判断是否是 称呼类名词），则根据S寻找离当前t最近的人名作为target输出。
3. 若t为空或其他非人称的代词，则优先从当前位置（若是代词）的前一个位置 寻找连续名词短语（例如“锤子手机这款产品很不错！”）；若未找到，则从 t所在的子句开头至t所在位置部分寻找定中短语或可作为评价对象候选词的 名词或书名号包围的成分；若未找到，则从T中寻找第一个候选词；若未找到， 则选取同一上下文上一次分析中找到的评价对象；若都未找到，返回target为 空。

本文主要使用基于《知网》的词汇语义相似度计算算法[79]来判断当前评价对象是否是描述职位、称呼、称号的词或短语，进而决定是否在上下文寻找与之对应的人名评价对象。知网1（Hownet）使用某种知识描述语言将一系列的义原组织起来用以描述词语的语义（见图4-2）。

1. http://www.keenage.com/

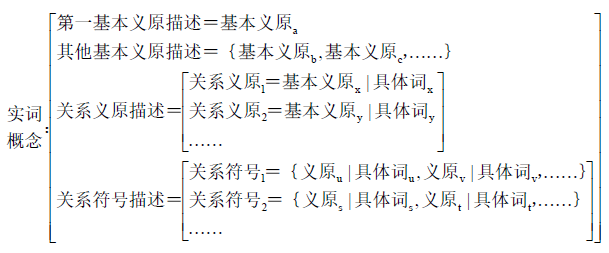
对于候选评价对象，若其最后一个词的第一基本义原是否是“人”且该词与“人”的相似度大于某一阈值（本文取0.7），则从该词开始优先从前面找最近的人名，若未找到则再往后寻找，若找到则作为其指代的评价对象抽出，若未找到，则将该候选评价对象作为真正的评价对象。

图4-2 实词概念描述结构

## 倾向性分析模块

倾向性分析主要根据每个评价对象对应的评价词，判断评价对象最终的情感倾向性。情感倾向性的判断主要使用情绪词典和评价词典，同时根据上下文中与之关联的副词决定是否对情感极性进行反转。

对于SR规则中抽取出来的评价对象，如果该评价对象和评价词取自同一个语义标注成分（例如“这不是一个有良知的政府！”），则从当前句子的依赖关系中寻找与当前谓词关联的所有副词作为当前评价词的分析参考词。这么做是考虑到评价对象和评价词已经在同一个语义成分中，仅仅需要考虑通过依存关系与之关联的副词，其他的副词很少会对此处的情感倾向性造成影响。如果评价对象和评价词不是取自同一个语义标注成分（例如“这届政府 不可否认是很给力！”），则选取当前以谓词为中心的语义标注范围内所有的副词作为当前评价词的分析参考词。

对于ATT规则、SN规则、SO规则等完全通过双向传播规则抽取出来的评价对象和评价词，我们需要将符合直接依赖（opinion-adv）和间接依赖（opinion-h-adv，h表示任一中间词）关系的副词抽取出来作为当前评价词的分析参考词。

本系统使用的倾向性计算算法（见算法4.3）主要考虑副词是否定词还是伪否定词对情感极性的影响。同时本文认为如果副词情感词和非副词情感词同时存在，副词情感词对情感极性的影响要稍微弱点，所以赋予副词情感词的基本评分要低一些。例如，“他这个人真心让我很反感！”这样的句子在分析时就会因为“真心”的副词评分低于“反感”，使该句的情感评分倾向于负面。

**算法4.3**（**倾向性计算算法**）

输入：当前评价对象对应的评价词集合S（包括评价词o及其对应的分析参考词集合 R），一般情感词基本分b1（本文设为1.0），副词情感词基本分b2（本文设 为0.7），绝对值最大评分maxScore（初始化为0）；

输出：评价对象情感倾向性标注：POS或NEG.

1. 根据情感词典分析当前评价词o是正面情感词还是负面情感词，相应赋予o基 本评分score=1或score=-1，根据情感词是否是副词更新score\*=b1或score\*=b2。
2. 遍历R中的词r，若r含否定词但是在伪否定词词典中，则不做处理；反之， score=-score。
3. 若|score|>|maxScore|，则maxScore=score。回到（1）直到遍历完S。
4. 若maxScore>0，输出POS，反之输出NEG。

## 本章小结

本章主要介绍本文所使用的情感资源的构建以及基于依存句法分析的评价对象抽取和情感倾向性分析系统的设计和实现，包括系统所使用的几个主要算法。

# 实验与分析

本文使用了自然语言处理与中文计算会议（NLPCC）2013年情感要素抽取和倾向性分析评测语料，共10个话题，约10000条微博。情感要素抽取和倾向性分析评测任务的要求[80]是：找出微博中每条观点句（观点句已给出，不需要自动识别，同时非观点句保留）作者的评价对象，即情感对象（包括句中位置）。每条观点句可能含有若干情感对象，同时判断针对情感对象的观点极性。

## 评价标准

评测标准采用精确评价和宽松评价两种方式，使用准确率、召回率和F1值作为评价标准。

在精确评价中，要求提交的情感对象的offset 和答案完全相同并且情感对象极性也相同时才算正确。如式（5-1）、（5-2）、（5-3），其中#correct\_label是实验标注出的评价对象（包括位置）和情感倾向完全正确的数目，#all\_label表示实验标注出的所有的评价对象的数目，#all\_evaluation表示评测语料中所有评价对象的数目。

（5-1）



（5-2）



（5-3）

在宽松评价中，一个结果包含4个参与评测的元素：句子微博wId，句子sId，情感对象区间s（由起始位置和终止位置构成）和极性，即r=(wId,sId,s,p)。首先定义两个结果r和r’之间的覆盖率c：



（5-4）



（5-5）

其中|\*|表示计算区间的长度，两个结果集合R和R’之间覆盖率C定义为：

（5-6）



假设实验标注结果集合为R’，评估语料中标注结果集合为R，则精度、召回率和F值为：



（5-7）



（5-8）



（5-9）

## 实验结果

本文实验结果如表5-1所示。就结果而言，与所有参加NLPCC2013情感要素抽取任务的评测结果相比，处于前列，但是离实际应用还有很长的一段路要走。

表5-1 整体情感分析结果

|  |  |  |  |
| --- | --- | --- | --- |
| 精确标准 | 准确率 | 召回率 | F1值 |
| 微平均 | 0.375 | 0.3155 | 0.3427 |
| 宏平均 | 0.3852 | 0.3246 | 0.3505 |
| 宽松标准 | 准确率 | 召回率 | F1值 |
| 微平均 | 0.5006 | 0.4203 | 0.457 |
| 宏平均 | 0.5056 | 0.4218 | 0.4576 |

### **语义规则对情感分析的影响**

图5-1 语义规则对整体情感分析性能的影响

图5-2 语义规则对评价对象抽取性能的影响

从图5-1可以看出，通过添加语义规则，情感分析系统的准确率得到显著提高，由40.21%提升到50.06%，从而明显提高F1值。然而，召回率仅由38.59%提升到42.03%。进一步分析图5-2可以发现，语义规则的添加并没有非常明显地使基于依存句法分析的方法抽取出更多的评价对象，但是却使得评价对象的抽取精确性得到很大的提高。通过语义角色标注可以较为清晰的标出以谓词为中心的动作参与成分，缩小以该谓词为中心的子句中评价对象的搜索范围。同时，利用包含语义知识的规则可以在一定程度上根据情感词的前指倾向或后指倾向决定抽取哪部分成分作为评价对象。这样可以避免完全依赖依存关系这种粗粒度的方法进行评价对象抽取，尤其是在评价词与多个评价对象具有依存关系的情况下。根据双向传播规则，与评价词相关的评价对象候选词都会被抽取出来。而语义规则会利用语义知识对参与情感表达的成分进行判断和选择，只要包含更加丰富的词库，这种细粒度的方法甚至还能识别出更多表面无情感但表达倾向性的评价词，包括一些惯用语、句式等。

### **依存关系间接规则对情感分析的影响**

图5-3 依存关系间接规则对情感分析性能的影响

从图5-3可以看出，在当前语料下，无论是否有语义规则，添加间接依存规则确实能提高系统召回率，而准确率都有所下降。综合起来，F1值均有所提高。

进一步观察可以发现，在没有语义规则的情况下，添加间接依存关系规则能显著提高系统召回率（约8%），同时准确率却由46.29%下降到40.21%，F1值提升约2%；而在有语义规则的情况下，添加间接依存关系规则系统召回率仅仅提高了约2%，准确率仅仅降低了约1%，F1值提升了约0.7%。由此可见，在无语义规则的时候，间接依存规则对于系统识别的准确率和召回率的影响非常明显，它在抽取出更多的评价对象的同时，也识别出很多无关成分，因此带来了精确度的损失。而有了语义规则之后，间接依存规则对系统性能的影响明显被削弱了。通过对实验标注结果进行分析，语义规则中使用的评价对象搜索算法能帮助识别出离评价词距离稍远的评价对象，在某种程度上类似于间接依存规则抽取方法寻找远距离的评价对象和评价词，但是由于其是基于语义知识的，在正确性上要更高一些。

从整体上看，合理地对间接依存规则中使用的关系进行约束，在一定程度上还是提高了系统的整体性能。

### **ATT规则与名词短语方法的对比实验**

图5-4 ATT规则与名词短语方法对比实验图

名词短语规则使用ICTCLAS1（其词性标注类型更为丰富）对微博文件进行分词和词性标注之后，需要将部分名词和相关词按名词短语合并规则[42]进行合并以形成短语，便于抽取出潜在的评价对象。短语部分作为一个完整的分词单元参与到句法依存分析中，有利于分析出短语与其他成分的依赖关系，缩短依赖关系分析距离。

1. http://ictclas.nlpir.org/

名词短语合并规则如下:



其中，NP、AP、M、Q分别表示名词短语、形容词短语、数词、量词，a、n、vn、r、word分别表示形容词、名词、名动词、代词和任意词，udel表示“的”。

名词短语规则主要关注的是以名词为核心的短语结构，而ATT规则关注的定中短语的中心语一般是名词、代词、数词、体词性联合短语等，有时谓语性成分也可以充当中心语。名词短语规则是在分词和词性标注的基础上，以规则为驱动的出去方法；ATT规则是在依存句法分析的基础上，将能在句法结构层面上通过定中关系关联的成分作为评价对象候选词，这是以句子整体结构分析结果为驱动的。因为语义规则中也使用了定中短语抽取评价对象，为了保证语义规则不对验证实验造成影响，我们在去除语义规则的情况下做了ATT规则和名词短语规则的对比实验。从图5-4可以看出，使用ATT规则的识别准确率比名词短语规则要高近21%，召回率高出近5%，F1值高出约15%。实验结果也验证了使用ATT规则抽取出的短语作为评价对象候选词在准确率还是召回率都取得不错的结果，它能抽出更丰富的短语类型，以句法结构为基础，准确性也相对高一些。可以猜想，如果再融入语义知识，准确性会有进一步提升的可能，而评价对象搜索算法中使用词义相似度就是基于这一点。

### **评价对象搜索算法对情感分析的影响**

图5-4 评价对象搜索算法对情感分析的影响

评价对象搜索算法

表5-2人名实体检测对不同话题的情感分析性能对比表

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 话题 | 0 | 1 | 2 | 5 | 4 | 5 | 6 | 7 | 8 | 9 |
| 无人名  实体检测 | 0.5485 | 0.3633 | 0.4846 | 0.4376 | 0.3574 | 0.3328 | 0.6169 | 0.4748 | 0.3293 | 0.4569 |
| 有人名  实体检测 | 0.5538 | 0.5388 | 0.4883 | 0.4354 | 0.3494 | 0.3339 | 0.6116 | 0.4784 | 0.3295 | 0.4569 |

注：此处性能以宽松标准下的F1值为主。编号和话题对应如下：0-不动产登记条例，1-查韦斯，2-锤子ROM，3-厨子戏子痞子，4-毒玩具，5-曼联VS皇马，6-魔境仙踪，7-王语嫣，8-笑傲江湖，9-中国方言式英语。

## 实验结果

# 总结和展望

## 工作总结

## 未来展望

# 致谢

首先，衷心地感谢高志强老师，无论是在技术实现、问题解决还是论文的修改上，都给我细心的指点和极大的帮助。他一丝不苟的治学态度、严谨的学术思维都是值得我敬佩和学习的！

其次，非常感谢实验室的师兄师姐在实验进行和论文修改中提出的宝贵建议。特别是鲁廷明和杨沛在此方面的悉心帮助，感谢鲁师兄在项目繁忙时给了我极大的空间去认真完成毕业设计，并帮我审阅论文，感谢他对我的诸多指点。还要感谢中国传媒大学国家语言资源监测和研究有声媒体中心的侯敏教授和周红照师姐在系统语义特征方面给予的帮助。

同时，对评审论文的各位专家和教授表示衷心的感谢！他们给我提出的宝贵意见和真知灼见，在我日后的工作和学习上一定会受益匪浅！

最后，我要感谢我的家人和朋友，是他们在精神上了给我极大的支持和鼓励！

再一次对他们表达我深深的谢意和敬意！

# 参考文献

1. Bing Liu. *Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers*, May 2012.
2. Hu, Minqing and Bing Liu. *Mining and summarizing customer reviews.* in *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*. 2004.
3. Hu, Nan, Paul A Pavlou, and Jennifer Zhang. *Can online reviews reveal a product's true quality: empirical findings and analytical modeling of Online word-of-mouth communication. in Proceedings of Electronic Commerce (EC-2006)*. 2006.
4. Ellen M, Voorhees. *Overview of TREC 2010[C] // Proceedings of the Nineteenth Text Retrieval Conference Proceedings (TREC 2010)*, NIST, 2010.
5. Iadh Ounis, Craig Macdonald, Ian Soboroff. *Overview of the TREC 2010 Blog Track[C] // Proceedings of the Nineteenth Text Retrieval Conference Proceedings (TREC 2010)*, NIST, 2010.
6. Ku L W, Lo Y S, Chen H H. *Using Polarity Scores of Words for Sentence-level opinion extraction[C] // Proceedings of the NTCIR-6 Workshop Meeting, 2007:316-322*, 2007*.*
7. *Hideo Joho, Tetsuya Sakai. Overview of NTCIR-10[C] // Proceedings of the 10th NTCIR Conference, Tokyo*, Japan, 2013.
8. 赵军, 许洪波, 黄萱菁等．中文倾向性分析评测技术报告[R]．北京:中文信息学会,2008．
9. Liu, Bing, Mingqing Hu, and Junsheng Cheng. *Opinion observer: Analyzing and comparing opinions on the web.* in *Proceedings of International Conference on World Wide Web(WWW-2005)*. 2005.
10. Hu, Minqing and Bing Liu. *Mining and summarizing customer reviews*. in *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*. 2004.
11. Popescu, Ana-Maria and Oren Etzioni. *Extracting product features and opinions from reviews.* in *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2005).* 2005.
12. Blair-Goldensohn, Sasha, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A. Reis, and Jeff Reynar. *Building a sentiment summarizer for local service reviews.* in *Proceedings of WWW-2008 workshop on NLP in the Information Explosion Era.* 2008.
13. Long, Chong, Jie Zhang, and Xiaoyan Zhu. *A review selection approach for accurate feature rating estimation.* in *Proceedings of Coling 2010:Poster Volume*. 2010.
14. Cilibrasi, Rudi L. and Paul M. B. Vitanyi. *The google similarity distance. IEEE Transactions on Knowledge and Data Engineering, 2007. 19(3): p. 370-383.*
15. Zhuang, Li, Feng Jing, and Xiaoyan Zhu. *Movie review mining and summarization.* in *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2006).* 2006.
16. Somasundaran, S., J. Ruppenhofer, and J. Wiebe. *Discourse level opinion relations: An annotation study.* in *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue.* 2008.
17. Kobayashi, Nozomi, Kentaro Inui, and Yuji Matsumoto. *Extracting aspect evaluation and aspect-of relations in opinion mining. in Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. 2007.
18. Qiu, Guang, Bing Liu, Jiajun Bu, and Chun Chen. *Opinion Word Expansion and Target Extraction through Double Propagation. Computational Linguistics, Vol. 37, No. 1: 9.27,* 2011.
19. Wu, Yuanbin, Qi Zhang, Xuanjing Huang, and Lide Wu. *Phrase dependency parsing for opinion mining.* in *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2009).*2009.
20. 周红照，侯明午，颜彭莉等. 语义特征在评价对象抽取与极性判定中的作用. 北京大学学报, 2014.
21. Rabiner, Lawrence R. *A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 1989. 77(2): p. 257-286*.
22. Lafferty, John, Andrew McCallum, and Fernando Pereira. *Conditional random fields: Probabilistic models for segmenting and labeling sequence data.* in *Proceedings of International Conference on Machine Learning (ICML-2001)*. 2001.
23. Jin, Wei and Hung Hay Ho. *A novel lexicalized HMM-based learning framework for web opinion mining.* in *Proceedings of International Conference on Machine Learning (ICML-2009).* 2009.
24. Jakob, Niklas and Iryna Gurevych. *Extracting Opinion Targets in a Singleand Cross-Domain Setting with Conditional Random Fields.* in *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010).* 2010.
25. Li, Binyang, Lanjun Zhou, Shi Feng, and Kam-Fai Wong. *A Unified Graph Model for Sentence-Based Opinion Retrieval.* in *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010).* 2010.
26. Kobayashi, Nozomi, Kentaro Inui, and Yuji Matsumoto. *Extracting aspectevaluation and aspect-of relations in opinion mining.* in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning.* 2007.
27. Yu, Jianxing, Zheng-Jun Zha, Meng Wang, and Tat-Seng Chua. *Aspect ranking: identifying important product aspects from online consumer reviews.* in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics.* 2011.
28. Manevitz, Larry M. and Malik Yousef. *One-class SVMs for document classification. The Journal of Machine Learning Research, 2002. 2: p. 139-154*.
29. Ghani, Rayid, Katharina Probst, Yan Liu, Marko Krema, and Andrew Fano. *Text mining for product attribute extraction. ACM SIGKDD Explorations Newsletter, 2006. 8(1): p. 41-48.*
30. Kovelamudi, Sudheer, Sethu Ramalingam, Arpit Sood, and Vasudeva Varma. *Domain Independent Model for Product Attribute Extraction from User Reviews using Wikipedia.* in *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP-2010).* 2011.
31. Hofmann, Thomas. *Probabilistic latent semantic indexing.* in *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI-1999)*. 1999.
32. Blei, David M. and Jon D. McAuliffe. *Supervised topic models.* in *Proceedings of NIPS.* 2007.
33. Mei, Qiaozhu, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. *Topic sentiment mixture: modeling facets and opinions in weblogs.* in *Proceedings of International Conference on World Wide Web (WWW-2007)*. 2007.
34. Titov, Ivan and Ryan McDonald. *Modeling online reviews with multi-grain topic models.* in *Proceedings of International Conference on World Wide Web (WWW-2008)*. 2008.
35. Brody, Samuel and Noemie Elhadad. *An Unsupervised Aspect-Sentiment Model for Online Reviews.* in *Proceedings of The 2010 Annual Conference of the North American Chapter of the ACL.* 2010.
36. Li, Fangtao, Minlie Huang, Yi Yang, and Xiaoyan Zhu. *Learning to Identify Review Spam.* in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-2011)*. 2011.
37. Mukherjee, Arjun, Bing Liu, and Natalie Glance. *Spotting Fake Reviewer Groups in Consumer Reviews.* in *Proceedings of International World Web Conference (WWW-2012)*. 2012.
38. Andrzejewski, David and Xiaojin Zhu. *Latent Dirichlet Allocation with topic-in-set knowledge.* in *Proceedings of NAACL HLT.* 2009.
39. Andrzejewski, David, Xiaojin Zhu, and Mark Craven. *Incorporating domain knowledge into topic modeling via Dirichlet forest priors.* In *Proceedings of ICML.* 2009.
40. Mukherjee, Arjun, Bing Liu, Junhui Wang, Natalie Glance, and Nitin Jindal. *Detecting Group Review Spam.* in *Proceedings of International Conference on World Wide Web (WWW-2011, poster paper).* 2011.
41. Zhai, Zhongwu, Bing Liu, Hua Xu, and Peifa Jia. *Constrained LDA for Grouping Product Features in Opinion Mining.* in *Proceedings of PAKDD-2011.* 2011.
42. Wei, Wei and Jon Atle Gulla. *Sentiment learning on product reviews via sentiment ontology tree.* in *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010).* 2010.
43. Jiang, Long, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. *Target ependent twitter sentiment cassification.* in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011).* 2011.
44. Boiy, Erik and Marie-Francine Moens. *A machine learning approach to sentiment analysis in multilingual Web texts. Information retrieval, 2009. 12(5): p. 526-558.*
45. Ding, Xiaowen, Bing Liu, and Lei Zhang. *Entity discovery and assignment for opinion mining applications.* in *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2009).* 2009.
46. Ganapathibhotla, Murthy and Bing Liu. *Mining opinions in comparative sentences.* in *Proceedings of International Conference on Computational Linguistics (COLING-2008).* 2008.
47. Ding, Xiaowen, Bing Liu, and Philip S. Yu. *A holistic lexicon-based approach to opinion mining.* in *Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2008).* 2008.
48. Kim, Soo-Min and Eduard Hovy. *Determining the sentiment of opinions.* in *Proceedings of Interntional Conference on Computational Linguistics (COLING-2004).* 2004.
49. Wan, Xiaojun. *Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis.* in *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2008).* 2008.
50. Zhu, Jingbo, Huizhen Wang, Benjamin K. Tsou, and Muhua Zhu. *Multiaspect opinion polling from textual reviews.* in *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2009).* 2009.
51. Blair-Goldensohn, Sasha, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A. Reis, and Jeff Reynar. *Building a sentiment summarizer for local service reviews.* in *Proceedings of WWW-2008 workshop on NLP in the Information Explosion Era.* 2008.
52. Kessler, Jason S. and Nicolas Nicolov. *Targeting sentiment expressions through supervised ranking of linguistic configurations.* in *Proceedings of the Third International AAAI Conference on Weblogs and Social Media (ICWSM-2009).* 2009
53. Liu, Feifan, Dong Wang, Bin Li, and Yang Liu. *Improving blog polarity classification via topic analysis and adaptive methods.* in *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL (HLT-NAACL-2010).* 2010.
54. Tesniere, L. *Élements de syntaxe structurale: Préf. de Jean Fourquet*1959: C. Klincksieck.
55. Lin, Dekang. Minipar. http://webdocs.cs.ualberta.ca/lindek/minipar.htm. 2007.
56. 张莉，钱玲飞，许鑫. 基于核心句及句法关系的评价对象抽取. 中文信息学报. Vol.25, No. 3. May, 2011.
57. 宗成庆. 《统计自然语言处理》.清华大学出版社. p221. 2013
58. Robinson JJ. 1970. *Dependency Structures and Transformational Rules. Language, 46(2): 259-285.*
59. 冯志伟. 1998. 判断从属树合格性的五个条件. 第二届全国应用语言学讨论会论文集.
60. Heys DG. 1964. *Dependency Theory: A Formalism and Some Observations. Language, 40: 511-525*.
61. Gaifman H. 1965. *Dependency Systems and Phrase-structure Systems. Information and Control, 8: 304-337*.
62. Kasami T. 1965. *An Efficient Recognition and Syntax Analysis Algorithm for Context-free Languages. Technical Report AFCRL-65-758,* Air Force Cambridge Research Laboratory Bedford, MA.
63. Younger DH. 1967. *Recognition and Parsing of Context-free Language in Time n3. Information and Control, 10(2): 189-208*.
64. Earley J. 1970. *An Efficient Context-free Parsing Algorithm. Communications of the Association for Computing Machinery, 13(2): 94-102*.
65. Karlsson F. 1990. *Constraint Grammar as a Framework for Parsing Running Text. in Proceedings for COLING, pages 168-173*.
66. Karlsson F, Voutilainen A, Heikkila J, and Anttila A. 1995. *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text*. Mouton de Gruyter.
67. Maruyama H. 1990. *Structural Disambiguation with Constraint Propagation. in Proceedings of ACL, pages 31-38*.
68. Covington MA. 2001. *A Fundamental Algorithm for Dependency Parsing. in Proceedings of ACM Southeast Conference, pages 95-102*.
69. Eisner J. 1996. *Three New Probabilistic Models for Dependency Parsing: An Exploration. In Proceedings of COLING, pages 340-345*.
70. Eisner J. 1996. *An Empirical Comparison of Probability Models for Dependency Grammar. Technical Report IFCA-96-11*, Institute for Research in Cognitive Science, University of Pennsylvania.
71. Eisner J. 2000. *Bilexical Grammars and Their Oubic-time Parsing Algorithms. in Bunt H. Nijholt A(eds.) Advances in Probabilistic and Other Parsing Technologies*. Kluwer.
72. McDonald R, Crammer K, Pereira F. 2005. *Online Large-margin Training of Dependency Parsers. in Proceedings of ACL, pages 91-98*.
73. McDonald R, Pereira F, Ribaroy K, Hajic J. 2005. *Non-Projective Dependency Parsing Using Spanning Tree Algorithms. in Proceedings of HLT\_EMNLP, pages 523-530*.
74. McDonald R, Lerman K, and Pereira F. 2006. *Multilingual Dependency Analysis with a Two-stage Discriminative Parser. in: Proceedings of CoNLL, pages 216-220*.
75. Yamada H, Matsumoto Y. 2003. *Statistical Dependency Analysis with Support Vector Machines. in Proceedings of IWPT, pages 195-206*.
76. Nivre J and Nilsson J. 2003. *Three Algorithms for Deterministic Dependency Parsing. in Proceeings of NODALIDA.*
77. Abney S. 1991. *Parsing By Chunks. in Robert Berwick, Steven Abney, Carol Tenny(eds.) Principle-Based Parsing.* Kluwer Academic Publishers.
78. 鉴萍. 2010. 依存句法分析方法研究与系统实现. 中国科学院自动化研究所.
79. 刘群，李素建. 2002. 基于《知网》的词汇语义相似度计算. 《中文计算语言学》.
80. 中文微博观点要素抽取评测大纲. NLPCC 2013. http://tcci.ccf.org.cn/conference/2013.