

Web 评论文本的细粒度意见挖掘技术研究

黄胜

2014 年 06 月

中图分类号：TP391.1

UDC 分类号：681.37

Web 评论文本的细粒度意见挖掘技术研究

作者姓名	<u>黄 胜</u>
学院名称	<u>计算机学院</u>
指导教师	<u>牛振东教授</u>
答辩委员会主席	<u>林守勋研究员</u>
申请学位	<u>工学博士</u>
学科专业	<u>计算机软件与理论</u>
学位授予单位	<u>北京理工大学</u>
论文答辩日期	<u>2014 年 06 月</u>

Research on Fine-grained Opinion Mining Technologies of Web Review Texts

Candidate Name:	<u>Sheng Huang</u>
School or Department:	<u>Computer Science & Technology</u>
Faculty Mentor:	<u>Prof. Zhendong Niu</u>
Chair, Thesis Committee:	<u>Prof. Shouxun Lin</u>
Degree Applied:	<u>Doctor of Philosophy</u>
Major:	<u>Computer Software and Theory</u>
Degree by:	<u>Beijing Institute of Technology</u>
The Date of Defence:	<u>June, 2014</u>

Web 评论文本的细粒度意见挖掘技术研究

北京理工大学

研究成果声明

本人郑重声明：所提交的学位论文是我本人在指导教师的指导下进行的研究工作获得的研究成果。尽我所知，文中除特别标注和致谢的地方外，学位论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京理工大学或其它教育机构的学位或证书所使用过的材料。与我一同工作的合作者对此研究工作所做的任何贡献均已在学位论文中作了明确的说明并表示了谢意。

特此申明。

签 名： 日期：

关于学位论文使用权的说明

本人完全了解北京理工大学有关保管、使用学位论文的规定，其中包括：①学校有权保管、并向有关部门送交学位论文的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存学位论文；③学校可允许学位论文被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换学位论文；⑤学校可以公布学位论文的全部或部分内容（保密学位论文在解密后遵守此规定）。

签 名： 日期：

导师签名： 日期：

摘要

近年来,随着电子商务和社交媒体的蓬勃发展,网络用户可以针对包括在线商品、新闻时政、公众人物和个人经历等在内的各种主题和对象发表自己的主观评论意见。这类主观评论信息包含了用户个人的主观意见、情感、态度和观点等。随着用户的参与度不断提高,以 Web 为媒介的主观评论文本信息日益激增。面对海量的 Web 评论文本,如何针对特定信息需求,进行有效的用户主观意见挖掘和分析,即 Web 评论文本意见挖掘,日益成为当前智能信息处理、数据挖掘和计算语言学等领域的一个研究热点。Web 评论文本意见挖掘技术具有重要的理论研究和实际应用价值,可以广泛地应用于信息检索、商业智能、社会舆情分析等方面。

虽然粗粒度的主客观分类和情感分类技术已经相对比较成熟,但针对评论文本的细粒度意见挖掘研究仍面临着特征空间较大、数据稀疏、有效特征不足、自动化程度不高和领域依赖性问题。本文针对以上关键问题,从细粒度意见元素抽取、自适应评价目标聚类、领域情感词典自动构建、评价目标和情感词联合聚类四个方面开展研究工作,具体研究内容和创新性成果包括:

(1) 针对细粒度意见挖掘面临的特征缺乏以及多级别特征的有效融合问题,提出了基于序列标注学习和句法语义结构特征的意见元素抽取方法。由于用户评论数据通常具有不规范性的特点,细粒度意见挖掘研究比传统的信息抽取任务更加困难,需要解决特征空间较大、数据稀疏和缺乏有效特征的难题。本文将评价目标和情感词的抽取任务转化为序列标注学习过程,采用条件随机场 CRFs 模型构建有效融合多级别特征的统一抽取框架,并提出一种朴素的图剪枝算法对评价目标进行自动分类。通过引入句法语义结构特征,有效利用长距离的依存句法关系,以解决有效标注特征缺乏问题。实验结果验证了基于序列标注学习和句法语义结构特征的意见元素抽取方法的有效性。

(2) 针对评价目标聚类的领域依赖性和语义关联相似度计算问题,提出了基于约束的评价目标谱聚类方法。评价目标聚类是用户评论意见挖掘研究中的核心任务,已成为基于特征的意见摘要和推荐的基础。现有的研究工作通常忽略了计算评价目标之间语义关联的领域依赖性,且存在着有效关联信息不足的问题。因此,本文研究提出基于约束的评价目标谱聚类方法以解决上述问题。该方法通过挖掘评价目标之间的

词法约束和上下文约束信息，增强评价目标之间的领域关联性；采用约束谱聚类算法，在融合先验约束知识的同时，有效地降低聚类空间的高维性和稀疏性。实验结果表明，基于约束的谱聚类方法有效提高了评价目标聚类的效果。

(3) 针对情感词典构建存在的算法领域适应性差、种子词依赖性和准确率不高问题，提出了基于约束标签传播的领域情感词典自动构建方法。情感词典是自动化情感分析的重要基础，然而由于评论文本的领域性，情感词的倾向性并不是固定不变的，而是依赖于出现的领域和上下文语境。传统的构建方法通常都面临着领域依赖、自动构建程度不高、准确率低等问题。因此，本文研究提出基于约束标签传播的领域情感词典自动构建方法以解决上述问题。该方法针对情感词的领域性，利用组块依存知识和先验通用情感词典抽取领域候选情感词和短语；通过定义和抽取情感词之间的上下文约束和词法约束关系，增强情感词之间领域依赖的情感关联相似度；最后利用约束标签传播算法计算候选词的情感倾向，构建领域情感词典。实验结果表明，约束标签传播方法有效提高了领域情感词典构建的准确率，受种子词覆盖的影响较小。

(4) 针对细粒度意见挖掘中评价目标和情感词抽取以及匹配修饰关系计算问题，提出了评价目标和情感词的联合聚类方法。目前大多数的研究工作仅仅考虑了局部上下文中评价目标与情感词之间的直接共现关系，而忽略了全局领域上下文中评价目标与情感词之间的隐含修饰关系，且通常面临着标注训练语料缺乏、特征稀疏和领域依赖引起的准确率低问题。因此，本文采用联合聚类方法，将评价目标和情感词抽取以及它们之间的匹配修饰关系计算问题转化为基于先验约束的半监督学习过程。在评价目标具有相互约束信息的基础上，根据情感词与评价目标的关联关系引入情感词之间的约束关系，从而为联合聚类过程提供先验的指导知识。利用约束联合聚类算法，同步地对评价目标按照语义方面聚类，对情感词按照与评价目标类的关联聚类，从而获取情感词类与评价目标类之间的匹配修饰关系。实验结果验证了基于约束的评价目标和情感词联合聚类方法的有效性。

关键词：Web 评论文本；意见挖掘；谱聚类；情感词典；联合聚类

Abstract

With the flourishing development of e-commerce and social media in recent years, the web users can publish their reviews to various topics and objects including online products, news events, public figures and personal experiences, etc. These subjective reviews contain valuable opinions, sentiments, attitudes and feels from the users. As the public engagement increases, the subjective review texts have grown dramatically on the web. With such massive web reviews texts, how to effectively mine and analyze user opinions according to specific information need, also known as web review texts oriented opinion mining, has become a hot research topic in the fields of intelligent information processing, data mining and computational linguistics and so on. Meanwhile, web review texts oriented opinion mining has great research and application values, and can be widely used in information retrieval, business intelligence and public opinion analysis, etc.

Although the coarse-grained subjectivity and sentiment classifications have been relatively mature, the web review texts oriented fine-grained opinion mining still faces the problems like large feature space, data sparseness, lacking of effective features, little automation and domain dependency, etc. Regarding above critical problems, this paper mainly conducts research work on four aspects: fine-grained opinion elements extraction, adaptive opinion targets categorization, domain-specific sentiment lexicon construction and co-clustering of opinion targets and opinion words. The main research works and contributions are listed as follows:

(1) Regarding problems of insufficient features and effective combination of multi-level features, this paper proposes the opinion elements extraction approach by leveraging sequence labeling learning and syntactic semantic structure features. Since user review texts are usually nonstandard, fine-grained opinion mining is more difficult than traditional information extraction task, and needs to solve the challenges of large feature space, sparse data and lacking of effective features. This paper transforms the task of opinion targets and opinion words extraction into the process of sequence labeling learning by employing CRFs model to construct unified extraction framework which effectively combines multi-level features, and also proposes a naïve graph-pruning strategy to classify opinion targets into semantic categories. In order to solve the problem of insufficient labeling features, the syntactic semantic structure features are introduced to exploit the syntactic dependency

relations between long-distance words. The experiment results have verified the effectiveness of the opinion elements extraction approach.

(2) Regarding problems about domain dependency and semantic association calculation in opinion targets categorization, this paper proposes the constraints based opinion targets spectral clustering approach. Opinion targets categorization is a core task in review texts opinion mining, and has become the foundation of feature-level opinion summarization and recommendation. In current research works, the domain-dependent characteristics of semantic association calculation intra opinion targets are usually ignored, and there also exists the problem of lacking of effective association information. Therefore, this paper studies the constraints based opinion targets spectral clustering algorithm to solve above problems. The lexical and contextual constraint relations between opinion targets are mined to enhance their domain-specific associations. The constrained spectral clustering algorithm can not only incorporate prior constraints information, but also effectively reduce high dimensionality and sparseness of the clustering space. The experiment results show that the constraints based spectral clustering approach has effectively improved the opinion targets categorization results.

(3) Regarding problems of algorithmic domain dependency, sentiment seeds dependency and low accuracy, this paper proposes the constrained label propagation approach for automatic construction of domain-specific sentiment lexicon. Sentiment lexicon is the solid foundation for automatic sentiment analysis. However, due to the domain-dependent characteristics of review texts, the polarities of sentiment words are not fixed, but vary depending on the domain and context. The traditional construction methods usually face the problems of domain dependency, little automation and low accuracy. Therefore, this paper studies the constrained label propagation algorithm to solve these problems. The candidate sentiment terms are extracted from domain corpus by exploiting chunk dependency knowledge and prior generic sentiment lexicon. The pair-wise contextual and morphological constraint relations are defined and extracted between sentiment terms to enhance their domain-specific sentiment associations. At last, the constrained label propagation algorithm is employed to calculate the polarities of candidate sentiment terms, and construct the domain-specific sentiment lexicon. The experiment results show that the constrained label propagation approach has effectively improved the precision of domain sentiment lexicon construction, and is less affected by the sentiment seeds coverage problem.

(4) Regarding problems about opinion targets and opinion words extraction and their matching relationship calculation in fine-grained opinion mining, this paper proposes the opinion targets and opinion words co-clustering approach. Most current research works only considered the explicit co-occurrence relations in local context, while ignored the hidden matching relationship in global domain context, and there also exists the problems of lacking of annotated training data, and low accuracy caused by feature sparseness and domain dependency. Therefore, this paper studies the co-clustering approach to transform opinion targets and opinion words extraction and their matching relationship calculation into prior constraints based semi-supervised learning process. Besides the constraint relations intra opinion targets, the constraint relations intra opinion words are also introduced to serve as prior knowledge to supervise the co-clustering process. The constrained co-clustering algorithm simultaneously clusters opinion targets into semantic aspects and clusters opinion words into sentiment groups associated with the target aspects respectively, and consequently obtains the global matching relationship between semantic aspects and aspect-specific sentiments. The experimental results have verified the effectiveness of the constraints based co-clustering approach.

Key Words: Web review texts; opinion mining; spectral clustering; sentiment lexicon; co-clustering

图索引

图 1.1 Epinions 网站上关于数码相机的用户评论文本示例.....	2
图 1.2 电子商务和服务站点示例	2
图 1.3 评论文本意见挖掘涉及的主要领域和技术路线	5
图 2.1 线性条件随机场模型结构	23
图 2.2 训练和测试文档以及特征模板的格式示例	26
图 2.3 基于序列标注学习的细粒度意见挖掘框架	27
图 2.4 依存句法树示例	30
图 2.5 基于依存句法的分布上下文抽取算法流程	35
图 2.6 评价目标与分布上下文特征之间的二部关系图	35
图 2.7 基于图剪枝的评价目标分类算法流程	36
图 2.8 基于图剪枝的评价目标分类示例	37
图 2.9 评价目标抽取的实验结果对比	39
图 2.10 情感词抽取的实验结果对比	39
图 3.1 评价目标聚类示意图	44
图 3.2 评价目标空间与词法和上下文约束空间的对应关系示意图	48
图 4.1 两个领域情感词典之间重叠示例	59
图 4.2 基于词语知识库的情感词典构造方法	60
图 4.3 连词方法构造情感词典	61
图 4.4 基于依存句法树和组块分析构造组块依存树示例	65
图 4.5 TBOD 评论数据集中半结构化评论文本格式示例.....	66
图 4.6 参数 α 取值对约束传播算法收敛迭代次数的影响	79
图 4.7 Cars 领域中参数 λ 取值对微平均 F 值的影响分析.....	80
图 4.8 Hotels 领域中参数 λ 取值对微平均 F 值的影响分析.....	80
图 5.1 评价目标类与情感词类之间的匹配修饰关系示例	83
图 5.2 联合聚类示意图	86
图 5.3 评价目标和情感词联合聚类系统框架	87
图 5.4 评价目标与情感词之间的关联空间及约束关系空间示例	91
图 5.5 α 值对评价目标聚类准确率的影响	96

表索引

表 2.1 语义角色标签符号统计	33
表 2.2 关于数码相机的用户评论数据集的统计信息	37
表 2.3 CRFs+SSSI 方法在四种数据集上的评价目标抽取结果分析.....	39
表 2.4 评价目标分类结果示例	42
表 2.5 不同相似度计算方法的分类结果比较	42
表 3.1 验证数据集和评价目标类别统计	51
表 3.2 评价目标之间约束关系的抽取结果统计	52
表 3.3 不同聚类算法的 Entropy 值比较	53
表 3.4 不同聚类算法的 Purity 值比较	54
表 3.5 不同约束关系类型对 CSC 算法 Entropy 值的影响	55
表 3.6 不同约束关系类型对 CSC 算法 Purity 值的影响	55
表 3.7 不同上下文类型对 CSC 算法 Entropy 值的影响	56
表 3.8 不同上下文类型对 CSC 算法 Purity 值的影响	56
表 4.1 TBOD 和 OpinRankDataset 数据集的统计信息.....	72
表 4.2 情感词 c 在评论数据集中的列联表	72
表 4.3 候选情感词抽取结果统计	75
表 4.4 Cars 和 Hotels 领域中前 20 个正面和负面种子情感词示例.....	75
表 4.5 Cars 领域中宏平均和微平均准确率、召回率、F 值的结果对比	76
表 4.6 Hotels 领域中宏平均和微平均准确率、召回率、F 值的结果对比	76
表 4.7 Cars 领域中种子情感词数量影响分析.....	77
表 4.8 Hotels 领域中种子情感词数量影响分析.....	78
表 5.1 验证数据集统计	93
表 5.2 评价目标和情感词抽取结果统计	94
表 5.3 不同过滤策略的候选评价目标抽取准确率统计	94
表 5.4 评价目标约束关系的抽取结果统计	94
表 5.5 情感词约束关系的抽取结果统计	95
表 5.6 不同聚类算法的准确率比较	96
表 5.7 评价目标与情感词匹配修饰关系抽取结果和准确率统计	97

目录

摘要.....	I
ABSTRACT.....	III
图索引.....	VI
表索引.....	VII
目录.....	VIII
第 1 章 绪论.....	1
1.1 研究背景和意义.....	1
1.2 国内外研究现状及发展趋势.....	4
1.2.1 情感分类.....	5
1.2.2 细粒度意见挖掘.....	7
1.2.3 情感词典构建.....	9
1.2.4 跨领域意见挖掘.....	11
1.2.5 评论文本意见挖掘系统应用.....	12
1.2.6 存在的困难和不足.....	13
1.3 论文主要研究内容和创新点.....	14
1.3.1 论文主要研究内容.....	14
1.3.2 论文创新点.....	16
1.4 论文组织结构.....	17
第 2 章 基于序列标注学习的意见元素抽取.....	19
2.1 引言.....	19
2.2 研究现状分析.....	21
2.3 条件随机场理论.....	22
2.3.1 特征函数的选择.....	24
2.3.2 参数估计.....	24
2.3.3 模型推断.....	25
2.3.4 CRF++介绍.....	26

2.4 基于序列标注学习的细粒度意见挖掘框架.....	27
2.5 基于 CRFs 模型的评价目标和情感词抽取.....	28
2.5.1 特征选择.....	28
2.5.2 句法语义结构特征抽取.....	30
2.5.3 语义角色标注.....	32
2.6 基于图剪枝的评价目标分类.....	33
2.6.1 评价目标相似度计算.....	33
2.6.2 图剪枝分类算法.....	36
2.7 实验准备和结果分析.....	37
2.7.1 数据集准备.....	37
2.7.2 评价目标和情感词抽取结果分析.....	38
2.7.3 评价目标分类结果分析.....	40
2.8 本章小结.....	43
第 3 章 基于约束的评价目标谱聚类.....	44
3.1 引言.....	44
3.2 相关研究.....	45
3.2.1 基于无监督学习的评价目标聚类.....	45
3.2.2 基于主题模型的评价目标聚类.....	45
3.2.3 基于半监督学习的评价目标聚类.....	46
3.3 基于约束关系的谱聚类模型.....	46
3.3.1 评价目标约束关系的定义和抽取.....	46
3.3.2 约束传播算法.....	48
3.3.3 约束谱聚类算法.....	49
3.4 上下文特征抽取和表示.....	50
3.5 实验验证和结果讨论.....	51
3.5.1 数据准备和评价标准.....	51
3.5.2 基线方法.....	52
3.5.3 实验结果.....	52

3.6 本章小结.....	56
第4章 基于约束标签传播的领域情感词典自动构建.....	58
4.1 引言.....	58
4.2 相关研究.....	60
4.2.1 基于词汇知识库的情感词典构建.....	60
4.2.2 基于领域语料的情感词典构建.....	61
4.2.3 目前存在的主要问题.....	63
4.3 候选情感词和种子词抽取与关联度计算.....	64
4.3.1 领域候选情感词抽取.....	64
4.3.2 种子情感词抽取.....	66
4.3.3 语义关联图构建.....	67
4.4 基于约束关系的标签传播算法.....	68
4.4.1 情感词约束关系的定义和抽取.....	68
4.4.2 约束关系传播算法.....	69
4.4.3 约束标签传播算法.....	70
4.5 实验验证和结果分析.....	71
4.5.1 验证数据集准备.....	72
4.5.2 基线方法和评价标准.....	72
4.5.3 结果分析.....	74
4.6 本章小结.....	81
第5章 评价目标和情感词联合聚类.....	82
5.1 引言.....	82
5.2 相关研究.....	84
5.3 联合聚类.....	85
5.4 评价目标和情感词抽取与关联度计算.....	86
5.4.1 评价目标和情感词联合聚类系统框架.....	87
5.4.2 候选评价目标和情感词抽取.....	87
5.4.3 关联矩阵构建.....	89

5.5 基于约束的联合聚类算法.....	89
5.5.1 基于约束的联合聚类建模.....	89
5.5.2 约束关系的定义和抽取.....	91
5.5.3 约束联合聚类的求解算法.....	92
5.6 实验验证和结果讨论.....	93
5.6.1 数据准备和评价标准.....	93
5.6.2 实验结果分析.....	93
5.7 本章小结.....	97
第6章 总结和展望.....	99
参考文献.....	102
攻读学位期间发表论文与研究成果清单.....	118
致谢.....	120

第1章 绪论

1.1 研究背景和意义

随着 Internet 和 Web 技术的发展,互联网已经成为发布信息、交流意见和传播思想的重要途径。近年来,随着电子商务和社交媒体的蓬勃发展,网络用户可以自由地针对包括在线商品、新闻时政、公众人物和个人经历等在内的各种主题和对象发表自己的观点和言论。Web 上产生了大量的包含用户主观意见、态度、情感和观点的评论信息。这些评论信息通常以文本的形式出现,表达了人们的各种情感色彩和情感倾向,如喜、怒、哀、乐、批评和赞扬等。潜在的用户可以通过浏览这些具有主观色彩的评论来了解大众舆论对于某一事件或产品的看法。然而,面对日益激增的 Web 评论文本,仅依靠人工的方法难以进行海量数据的采集和处理。因此,如何针对特定信息需求,进行有效的用户主观意见分析和意见挖掘,是当前互联网智能信息处理、数据挖掘、计算语言学等领域的一个热点研究问题。评论文本意见挖掘具有重要的理论研究和实际应用价值,可以广泛地应用于产品评论挖掘、商业智能、社会舆情分析等方面。

意见性信息在人类社会生活中发挥着重要的作用。人们在做出选择和决策时,都倾向于参考他人的意见^[1]。他人的意见反映了社会群体对于某个事件、人物、产品等对象的情感倾向性,这种群体智慧可以用于对未来进行准确有力地预测^[2]。以电子商务领域为例,越来越多的消费者和经营者都意识到了用户评论信息的重要性。一方面,消费者在购买和使用一个产品之前,通常倾向于了解其他用户的评价和意见信息,从而做出更好的购买决策。美国商业周刊在 2008 年的调查发现,70%的顾客在购买前会查看产品的相关评论信息。华尔街日报在 2009 年的调查显示,92%的网民认为产品评论的作用远远大于推销者或者其它来源。Harris Poll 调查机构在 2010 年关于互联网产品评论的研究表明,45%的调查者在网上购物时会受到评论信息的影响,46%的人对产品的评论信息深信不疑^[3]。另一方面,经营者和厂商也经常需要了解大众或者消费者对于他们的产品和服务的反馈,从而更好地进行商业决策。目前,几乎所有的电子商务网站都提供了用户评价功能,图 1.1 描述了 Epinions.com 网站上典型的用户评论文本格式。此外,在社会政治领域中,政府和管理者需要了解社会民众对于政策和法规的支持度,以及对于政府管理和服务的满意程度,因此用户针对热点事件和政治人

物发表的评论文本信息也为互联网舆情监控提供了反映网络社会群体舆论倾向性的原始数据。

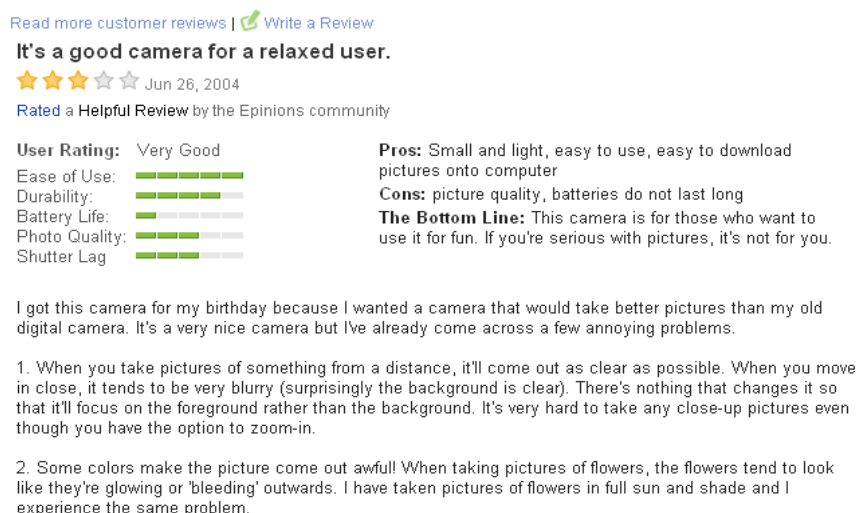


图 1.1 Epinions 网站上关于数码相机的用户评论文本示例

随着网络用户的参与度不断提高，Web 上的用户评论信息呈爆发式增长。这些用户评论信息涵盖了新闻时事、公众人物、在线商品和服务、个人经历和声誉等众多的讨论主题；同时这些信息也广泛地分布在新闻评论网站、电子商务站点、博客、论坛、微博和社交网络等各种互联网媒介。例如在商业智能领域，商品和服务的各种评论出现在各大电子商务网站以及论坛上（如图 1.2 所示）。面对海量的评论文本资源，对于消费者而言，阅读所有的评论文本来制定决策是不现实的，而采用人工方式进行统计、分析和挖掘，显然是低效和不切合实际的。因此，Web 评论文本数据并不是简单容易处理的，这给人们在短时间内获得有价值的意见性信息带来很大的困难。



图 1.2 电子商务和服务站点示例

面对 Web 上海量的用户评论文本数据，如何寻求一种有效的方法和手段从中挖掘出对用户有用的信息，已成为数据挖掘、自然语言处理、计算语言学等领域共同追寻的目标。评论文本意见挖掘技术就在这种需求背景下应运而生。

评论文本意见挖掘,通常也称为情感分析(Sentiment Analysis),是指针对用户产生的主观性评论文本数据,采用自动化/半自动化的手段进行分析、处理、归纳和推导,从中挖掘出用户(个人或群体,组织等)针对主题、人物、事件等表达的评论、观点和意见的过程。意见挖掘技术一方面起源于传统的数据挖掘、自然语言处理和计算语言学领域,目前的很多研究方法都是对传统数据挖掘和机器学习技术的应用和拓展;另一方面却又具有相当的文本理解能力,能够处理传统方法无法解决的问题。与以往的信息抽取、文本分类和文本摘要等任务不同,意见挖掘技术不仅要分析显式的语言表达结构,还需要对文本成分之间的关联语义进行深层次理解。可以说,意见挖掘技术的出现弥补了传统数据挖掘技术和文本挖掘技术的不足,是一项更具有实际应用价值的新技术。从涉及的文本语言层面来看,意见挖掘涉及了文本语言的各个粒度,从词汇层(分词和词性标注)、句法层(命名实体和语法分析)和语义层到篇章层面。

从二十世纪末开始,研究者针对文本意见挖掘展开了相应的研究工作,经过十多年的蓬勃发展,相关技术和应用取得了长足的进步。近年来文本意见挖掘日益成为一个热门的研究领域,吸引了众多的研究人员和学者。1997年,Hatzivassiloglou和McKeown^[4]最早开始了词语的情感倾向判别研究。在2000年前后,Wiebe等人^[5-7]对文本主观性分析展开了系统的研究。到2002年,Turney^[8]和Pang^[9]分别将无监督规则和监督学习的方法运用于解决评论文本的情感分类问题。

目前,评论文本意见挖掘已经涵盖了广泛的研究内容和技术领域。从挖掘对象的数据粒度来分,可以分为篇章级别情感分析、句子级别情感分析、词汇级别及词义级别情感分析。从挖掘的数据类型来分,可以分为新闻评论意见挖掘、博客和论坛数据意见挖掘、商业产品评论数据意见挖掘以及最新的微博数据意见挖掘等任务。从挖掘任务的不同来分,可以分为意见抽取和倾向判定、意见主体抽取、主体和情感表达关系挖掘、意见搜索以及情感摘要等任务。按照挖掘的领域和语言来分,可以分为相同领域意见挖掘、跨领域意见挖掘以及跨语言意见挖掘等。

评论文本意见挖掘具有重要的研究意义和广阔的应用前景,可以应用到现实生活中的许多方面。评论文本意见挖掘广泛运用于电子商务和产品评论分析。随着网络购物的兴起,用户在购买商品时倾向于参考其他顾客对该商品的评价意见,网络公司也需要用户和市场的反馈信息,评论文本意见挖掘能够有效地分析用户对产品的意见信息。评论文本意见挖掘广泛地运用于商业智能和信息检测。通过对网络数据的采集、

整理和分析,有效地挖掘出用户对商业产品和企业的评论信息,从而为产品开发和商业决策提供辅助支撑。评论文本意见挖掘技术还可以广泛运用于推荐和搜索系统,根据用户的情感需求,对搜索结果进行情感抽取和情感分类,推荐包含正面情感的结果给用户。除此之外,评论文本意见挖掘技术还可以应用于网络舆情监控、图书推荐、新闻报道评论、电子学习等诸多方面。

综上所述,Web 评论文本意见挖掘是自然语言理解、数据挖掘领域的一个重要的研究课题,已经成为学术界和工业界所关注的焦点。现有的研究已经取得了一些进展,但仍存在着许多尚未解决的难题和挑战。同时,评论文本意见挖掘也是一种新颖的、很有前途的技术,不仅具有重要的学术研究意义,而且具有很大的应用价值。

1.2 国内外研究现状及发展趋势

近十年来,针对 Web 评论文本意见挖掘的研究工作得到了很大的重视和发展,在英文和中文领域的相关研究都取得了重要进展。有关自然语言处理、人工智能、信息检索、数据挖掘以及 Web 应用等领域的多个国际顶级会议(AAAI, ACL, COLING, SIGIR, WWW, SIGKDD 等)都收录了意见挖掘的相关论文。美国国家自然科学基金委员会和 DAPRA 等投入了大量的研究经费^[10, 11], CMU、UIUC 等单位均开展了相关研究工作^[12-14]。在评测方面,文本信息检索国际会议和亚洲语言信息检索评测会议专门设立了意见挖掘相关的评测任务,包括意见检索、篇章情感分类、观点提取等。在中文文本挖掘领域,中国中文信息学会近年来陆续开展了中文倾向性分析评测(COAE 2008-2013),从词汇、句子和篇章等多粒度层面对中文文本进行意见挖掘。中国计算机学会自然语言处理与中文计算会议(NLP&CC 2012, 2013)也引入了面向中文微博数据的情感分析任务和中文词汇语义关系抽取任务。

图 1.3 描述了评论文本意见挖掘涉及的主要领域和技术路线。

在下文中,首先介绍粗粒度的情感分类研究,主要包括文档级别和句子级别情感分类;其次重点对细粒度的意见挖掘、情感词典构建和跨领域意见挖掘等任务及其国内外的研究现状和进展进行概括;接下来,对目前研究和开发的实际评论文本意见挖掘系统进行介绍;最后对目前研究中存在的困难和不足进行分析和总结。

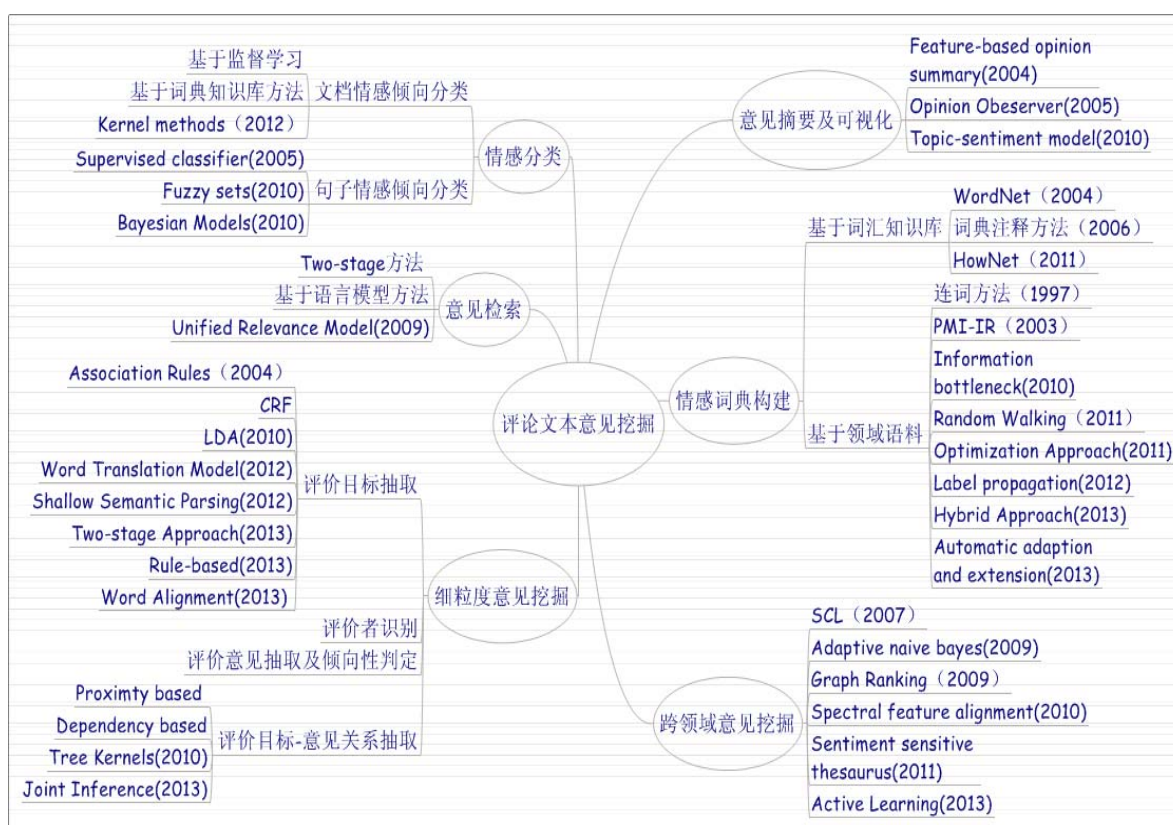


图 1.3 评论文本意见挖掘涉及的主要领域和技术路线

1.2.1 情感分类

1.2.1.1 文档级情感分类

文档级别情感分类是目前意见挖掘领域中研究最多的问题^[8, 15-18]。它以整篇文档作为基本的信息单位，判定文档中所表达的情感倾向（正面或负面），通常可以细分为两个子任务：主客观分类和情感倾向分类。主客观分类将一篇文档分类为描述事实性的客观文本和表述作者意见倾向的主观文本。情感倾向分类则是将文档按照作者所表达的情感倾向分为积极（正面）和消极（负面）两个倾向类别。目前的文档级别情感分类研究主要以主观性用户评论文本为研究对象。按照所采用的机器学习方法划分，目前的文档级别情感分类研究主要分为监督学习和无监督学习两类。

基于监督学习的文档情感分类。直观上，情感分类可以看作是二分类问题，可以采用监督学习的方法解决。基于监督学习的分类是利用人工标注文本情感极性，通过对语料、一元特征、二元特征、分类器的训练，得出文本的情感极性，实现对文本的情感分析。

Pang 等人^[9]最早将贝叶斯、最大熵和支持向量机等监督学习的方法运用于解决电

影评论的情感分类问题,实验结果显示支持向量机取得了最佳分类效果(82.9%)。文献^[19]中进一步运用了图的最小割方法来识别文档中的主观部分文本,这极大地方便了句子间上下文约束的引入。徐军等人^[20]研究了机器学习方法在新闻文本的情感分类中的应用,利用朴素贝叶斯和最大熵方法来判断新闻及评论语料的情感倾向性。

传统的文本分类研究表明特征选择通常对于性能的影响更为关键。对于基于情感的文本分类,选择具有语义倾向的词汇作为特征项,正确处理词性、否定词、句法结构等特征以及采用二值作为特征项权重往往能够提高分类的准确率^[20]。

基于词典知识库的文档情感分类。相对于基于监督机器学习方法的文档情感分类,基于词典知识库的文档情感分类属于无监督分类学习方法。基于无监督的分类方法基本是利用文本中词语的点互信息量(PMI)或者与基准词的比较来获得词语的情感极性,然后对整篇文章的情感词汇进行累加、平均,利用正负情感词汇的概率得出文本的情感极性。这种方式通常需要情感语义词典、词汇关系库等语义知识库的支撑,例如英文的 WordNet,中文的 HowNet、同义词词林等。

Turney^[8]在情感基准词的基础上,基于人工规则抽取蕴含情感的形容词和副词短语,利用点互信息量计算这些短语的情感倾向性,进而根据文本中包含的情感短语的平均情感倾向对文本整体情感倾向做出判定。Zagibalov 等人^[22]提出了一种对中文产品评论无监督情感分类的自动种子词选择方法。该方法不需要任何标注训练数据,只需要常用的否定词和状语词的信息,从而避免了与监督方法紧密相关的领域依存问题。Taboada 等人^[23]提出了一种基于情感词典对文本进行情感分析的方法,并考虑了情感增强和情感转义等影响因素。

1.2.1.2 句子级情感分类

评论文本通常并不是单一地陈述事实或表达情感,而是事实和情感相互交融。在阐述事实时会相应地表达个人情感,在表达情感时也需要引证对应的事实。目前的研究论文开始从篇章级别转移到以句子为粒度对作者的情感倾向进行分析,从整篇文档中抽取事实部分和对应的情感部分。由此,产生了句子/从句级别文本意见挖掘任务。句子级别意见挖掘的主要任务是对特定上下文的句子进行各种主观性信息的分析和抽取,包括句子的情感极性和情感极性的实施者、受体以及强度等各方面要素。

王根等人^[24]针对句子情感的分析问题,提出了一种基于多重冗余标记的 CRFs 模型。该方法不仅能有效地解决有序标记的分类问题,还能在保证情感分析中各子任务使用不同特征的前提下,将情感分析中的主客观分类、褒贬分类和褒贬强弱分类任务

统一在一个模型之中，从而在多个子任务上寻求联合最优，制约分布完成时误差的传播。姚天昉等人^[62]利用领域本体来抽取语句主题及其属性，然后基于句法分析结果识别主题与情感描述项之间的关系，并最终判断语句中每个主题的极性。

Fu 等人^[25]提出了一种基于模糊集（Fuzzy Sets）理论的中文句子情感分类方法。他们首先提出了一种由细粒度到粗粒度的句子情感强度估计方法；然后定义分别代表正面、中立和负面情感极性的三个情感模糊集，并建立了将中文句子分别归属于这三个情感模糊集的关系函数；最后，基于最大成员关系值确定句子的情感极性。Li 等人^[26]针对文本中情感表达方式自由多样的特点，提出了一种基于核的中文句子情感分类方法，分别设计了词核、N 元语法核和句法核三种函数来融合词语、N-gram 和句法级别的特征。

1.2.2 细粒度意见挖掘

粗粒度的情感分类技术只能从整体上判断文本的情感倾向性，而不能深入挖掘用户对于评论对象各个方面的评价。对于一项产品而言，整体上持有相同情感倾向性的用户对产品的局部细节可能持有不同的情感倾向性；不同的用户也往往关注不同的产品特征，并对产品不同方面的优缺点进行比较。一项热门的产品通常具有众多的评论数据，用户逐一阅读这些评论是一项耗费时间和精力事情。因此，近年来，细粒度的基于产品特征的意见挖掘技术逐渐成为意见挖掘研究领域的热点和核心问题。

细粒度意见挖掘（Fine-grained Opinion Mining），也称为基于特征的意见挖掘（Aspect-Oriented Opinion Mining），旨在深入到产品特征层面，提取评价信息中涉及的评论主体、评价特征、情感词以及对应的评价倾向等意见要素，从而为一些实际应用提供必要的细节信息^[33]。与粗粒度的情感分类不同，细粒度的产品评论挖掘涉及到产品特征抽取、意见表达抽取、产品特征的聚类 and 聚合、产品特征情感摘要、同类产品信誉度比较挖掘等研究问题。

Morinaga 等人^[53]最早从互联网上抓取产品的网页数据，通过制定句法和语言规则，判定用户的意见表达和正反面情感倾向，进而提出几种朴素的文本挖掘方法对情感按照产品进行归类，对产品的声誉度进行挖掘。Liu 和 Hu 等人^[15, 54, 55]最早对细粒度意见挖掘开展了系统的研究，提出了细粒度评论文本挖掘的基本框架，将在线产品评论挖掘的流程分为：频繁特征识别、情感词抽取、情感词倾向性判定、句子的情感倾向性计算和生成摘要信息五个步骤，并首先对隐式产品特征进行抽取^[2]。

目前，有监督和无监督的机器学习方法被广泛地应用于细粒度的评论文本意见挖

掘研究。IBM 的 Su 等人^[57]针对中文细粒度产品评论意见挖掘问题,提出了一种相互增强的方法。该方法基于内容信息和情感关联信息对产品特征和情感词集合进行迭代聚类,并根据产品特征和情感词构建相互间的关联集合。Jin 等人^[56]针对产品评论意见挖掘的问题,提出了一种基于 Lexicalized HMMs 的机器学习方法,从大量评论文本中抽取评价对象和评价意见。Jiang 等人^[58]提出一种基于树核的细粒度意见挖掘方法,通过树核的方法对评论语句的结构化特征建模。Qi 等人^[59]首次将基于条件随机场的判别式学习模型运用于产品评论的意见挖掘,从而避免了特征之间的条件独立性假设问题。Moghaddam 等人^[38]利用评论文本结构中包含的一部分已知的产品属性和用户评级信息,提出了一种无监督的 Opinion Digger 方法,该方法从评论文本中自动抽取重要的评价属性和对应的用户情感。施寒潇^[39]针对细粒度的属性分类和情感判定问题,采用有监督和半监督学习方法对属性分类问题进行研究,在情感词缺少对象属性的情况下,通过计算 PMI 值来确定评价对象属性类与情感词之间的关联关系。

近年来,主题模型也被越来越多地用于对评论文本中包含的主题和情感进行细粒度的建模。Brody 等人^[60]提出了一种无监督的产品特征 (Aspect) 和评论情感 (Sentiment) 抽取算法,该算法采用一种句子级别局部话题模型,通过设定一定数目的话题来抽取产品特征,并采用一种无监督方法生成正负面情感词的种子集,判定情感倾向。Lin 等人^[37]提出了一种基于 LDA 的联合情感-主题 (JST) 概率模型,该模型通过领域无关的情感词典引入弱监督信息,同时识别评论主题和用户情感;在此基础上,根据主题与情感的依赖关系定义了一种重新参数化的 Reverse-JST 模型,在建模过程中根据主题分布生成对应的情感。Yang 等人^[61]提出一种联合推理模型,该模型利用优化意见抽取子任务的预测知识寻求全局最优解,从评论文本中抽取意见相关的实体。

针对评论文本中的意见目标抽取问题,章剑锋等人^[63]将同一句子中共现的评价词与评价对象作为候选集合,应用最大熵模型并结合词、词性、语义和位置等特征进行关系抽取。该方法在一定程度上解决了指代消解以及评价对象遗漏问题。郝博一等人^[64]提出了一种基于泛化、繁殖和自举的意见目标抽取方法,该方法在泛化过程中提炼原子意见目标和意见目标模式,在繁殖过程中对复合意见目标进行扩展,并采用自举机制实现了意见目标的递增学习。李冰等人^[65]将浅层句法信息和启发式位置信息引入到条件随机场模型,在不增加领域情感词典的情况下,有效地提高了系统的准确率。张莉等人^[66]提出利用核心句进行学习的思想,继而确定了 10 种句法关系作为语言特

征,将原始句和核心句分别基于词、词性和句法关系利用条件随机场模型进行学习和比较。Jakob 等人^[49]尝试采用条件随机场 CRFs 模型解决单一领域和跨领域的评价目标抽取问题。

此外,研究者们也尝试了利用领域本体来解决细粒度意见挖掘问题^[33-36, 75]。郭冲等人^[33]针对细粒度的意见要素抽取和情感判定问题,定义了一种情感本体树结构,并通过评价搭配抽取算法、评价搭配倾向预测算法和特征聚合算法自动构建领域情感本体树。Wei 等人^[34]针对产品特征的层次关系识别和产品特征与情感词之间的对应关系识别问题,提出了一种基于情感本体树的层次学习(HL-SOT)方法,从产品评论中自动标注产品特征以及对应的情感词。Lau 等人^[35]提出了应用于意见挖掘的模糊领域本体及其自动构建方法,该方法重点关注领域特征、领域情感词及它们之间的对应关系的抽取,在人工标注的评论集上的文本情感倾向分类测试达到不错的效果^[33]。Cheng^[36]针对领域词汇的特点,首先采用半自动的方法基于一些现存的本体资源构建应用领域本体,然后将基于规则的命名实体识别技术与信息抽取引擎结合起来,识别被评价的汽车领域主题术语并给它们指派该应用领域的相关概念^[62]。姚天昉等人^[75]利用领域本体抽取汽车评论中的实体和特征,利用极性词词典识别用户评论意见并判断它们的褒贬性以及强度。

1.2.3 情感词典构建

情感词典是自动化情感分析研究的重要基础,包含了情感词/短语及其对应的先验情感倾向。词语是表达情感的基本单位,由词语组成情感表达短语,由短语组成情感从句,由从句组成表达复杂情感的句子。词语的情感倾向性是构成文本情感倾向性的基础。然而,文本的情感倾向性并不完全依赖于包含的情感词的倾向性,比如,正面情感词可能出现在负面评论文本中,负面情感词也可能出现在正面评论文本中。同时,情感词汇又分为领域无关和领域相关两类。领域无关情感词独立于具体的研究领域,具有固定的情感倾向(如美好、喜欢、讨厌等);领域相关情感词依附于具体的研究领域,其情感倾向会随着修饰对象的不同而改变(如大用来修饰显示器屏幕时多用于表达正面情感,而用来修饰存储器体积时常常表达负面情感)。如何针对特定领域自动抽取领域相关的情感词汇和短语,并判定其情感倾向是一个值得研究的问题。词语的情感倾向计算,是指对于词语的褒贬程度计算出一个度量值。为了方便比较和统计,目前比较常用的做法是将度量值规定为 $[-1, +1]$ 区间的实数,其中正负代表情感的褒贬倾向,数量的绝对值代表词语的极性强度^[30]。目前的研究方法通常分为基于大规模语

料统计的方法和基于词汇知识库的方法。

Hatzivassiloglou 和 McKeown^[4]最早开始了词语的情感倾向判别研究。他们尝试使用监督学习的方法对词语进行语义倾向判别,通过对训练语料的学习进行语义倾向判别,准确率约 82%,在加入篇章中形容词之间的接续信息后,准确率提升到 90%左右。

Turney^[8]首先通过词性抽取评论文本中的意见词和短语,然后分析词汇上下文信息研究其情感倾向,采用 PMI-IR 方法来计算所抽取短语的语义倾向。PMI 衡量了当我们发现了一个词时,能够找到另一个词的可能性。

$$PMI(term_1, term_2) = \log_2 \frac{P(term_1 \wedge term_2)}{P(term_1)P(term_2)} \quad (1.1)$$

一个词的语义倾向 (SO) 基于它和正面基准词 “excellent” 与负面基准词 “poor” 的关联程度进行计算:

$$SO(phrase) = PMI(phrase, "excellent") - PMI(phrase, "poor") \quad (1.2)$$

公式(1.1)中的概率值可以采用不同的方法来计算,通常可以采用向搜索引擎中输入查询后返回的匹配数目进行计算。通过计算评论中的所有短语的平均情感倾向,如果平均 SO 为正,那么将该评论归为正面评论,否则归为负面评论。

Wilson 等人^[27]针对如何自动区分词语的先验情感倾向性与上下文依赖倾向性问题进行研究,提出了一种结合主观性分类和倾向性分类的两阶段策略,系统地分析了不同的特征选取策略和不同的分类学习算法对实验效果的影响。Esuli 等人^[28, 29]提出了基于 WordNet 中的同义词集 (Synset) 构建语义情感词典 SENTIWORDNET: 基于每一个同义词集关联的注释 (Gloss) 的定量分析和向量化表示,通过八个不同的三元分类器组合,对同义词集进行半监督分类,计算同义词集的正面、中立和负面情感倾向性的评分。三种情感倾向性的评分之和等于 1。SENTIWORDNET 3.0 是对 SENTIWORDNET 1.0 的提升,其基于 WordNet 3.0 版本,并采用随机游走算法 (Random Walk) 优化同义词集三个方面的情感评分。

国内方面,关于汉语词汇、短语的情感倾向分析主要基于 HowNet 和同义词词林等语义词典展开。朱嫣岚等人^[30]基于 HowNet 提出了两种词汇语义倾向性计算的方法,即基于语义相似度的方法和基于语义相关场的方法。通过计算词语间的相似程度,对词语的褒贬倾向按照一定的计算法则进行赋值。根据所得的语义倾向度量值判别其褒贬倾向。其假设选择 k 对基准词,每对基准词包括一个褒义词和一个贬义词。与褒义

基准词联系越紧密，则词语的褒义倾向越强烈。与贬义基准词联系越紧密，则词语贬义倾向越明显。设定默认的阈值，最终倾向值大于阈值为褒义，小于阈值为贬义。单词 w 的语义倾向值计算公式如下：

$$Orientation(w) = \sum_{i=1}^k Similarity(key - p_i, w) - \sum_{j=1}^k Similarity(key - n_j, w) \quad (1.3)$$

其中， $Similarity$ 函数分别采用 HowNet 中提供的语义相似度和语义相关场两种计算方法。

王素格等人^[31]针对现有研究仅采用目标词与基准词的关联强度来确定目标词的情感倾向性和基准词选取问题，提出了一种基于类别区分能力与情感词词表相结合的方法。然后根据词汇与其同义词很大程度上具有相同的情感倾向，提出了基于同义词的词汇情感倾向判别方法，在一定程度上避免了数据稀疏问题。该方法不仅考虑了目标词与基准词的关联强度，而且也考虑了目标词的同义词与基准词的关联强度。实验结果表明，基于同义词的词汇情感倾向判别方法优于仅采用目标词与基准词的词汇情感倾向判别方法。

宋晓雷和王素格等人^[32]针对词汇情感倾向判定中的数据稀疏问题和对外部资源的依赖情况，利用概率潜在语义分析，提出了两种用于判别词汇情感倾向的方法。一是使用概率潜在语义分析获得目标词和基准词之间的相似度矩阵，再利用投票法决定其情感倾向；二是利用概率潜在语义分析获取目标词的语义聚类，然后借鉴基于同义词的词汇情感倾向判别方法对目标词的情感倾向做出判别。这两种方法的优点是均可在没有外部资源的条件下，实现词汇情感倾向的判别。

1.2.4 跨领域意见挖掘

领域适应性问题^[40-42]是机器学习和数据挖掘领域面临的重要难题之一，评论文本意见挖掘研究中同样存在算法和模型的领域自适应性问题。学术界对于单个领域的意见挖掘研究非常广泛，处理方法也相对比较成熟，而对于跨领域意见挖掘的研究工作还处在比较初始的阶段。

Blitzer 等人^[43]最早将结构对应学习（Structural Correspondence Learning, SCL）应用于跨领域情感分析，SCL 是一种应用广泛的跨领域文本分析算法，其目的是尽可能地建立训练集与测试集中特征之间的对应关系。Tan 等人^[44]将 SCL 应用于中文跨领域情感分析中，在文献^[21]中进一步提出了利用频繁共现熵从源领域和目标领域中抽取

泛化特征，并利用加权 EM 算法训练跨领域的朴素贝叶斯情感分类器。Wu 等人^[45]将图排序（Graph Ranking）思想应用于跨领域情感分析中，通过构造源领域与目标领域中的文档相似度矩阵，迭代计算目标领域中文档的情感得分。算法采用文档与源领域的情感中心向量之间距离作为文档情感初始值，图排序算法可以认为是一种迭代的 k-NN 算法^[46]。

Li 等人^[47]提出的基于约束非负矩阵三分解（Non-negative Matrix Tri-factorization）的方法，可以将源领域中文档的情感信息传递到目标领域。Liu 等人^[48]提出了一种两阶段的跨领域情感分类策略，该策略首先采用一种 Transfer-PLSA 的主题模型抽取领域间的公共主题，进而将源领域特征与目标领域特征对应起来；然后利用在源领域标注样本上训练的分类器，从目标领域中选择一些代表性样本，并基于选择的样本重新训练分类器。Glorot 等人^[50]针对大规模领域自适应情感分类问题提出了一种深度学习方法，该方法从领域评论文本中无监督学习潜在的特征表示，基于学习到的特征表示训练情感分类器，从而具有良好的可扩展性。Bollegala 等人^[51]提出了一种构建领域依赖情感词典的方法，在训练和测试分类器时利用构建的词典对特征向量进行扩展。Li 等人^[52]通过主动在目标领域中选择少量标注样本，将主动学习（Active Learning）的思想应用于解决领域之间具有显著差异的问题。

从以上研究可以看出，跨领域文本意见挖掘研究的主要问题在于如何寻找两个领域中情感特征之间的一种有效映射关系。目前很多研究利用半监督学习的方法，通过逐次迭代或组合的方式逐渐减少训练集和测试集之间的差异。

1.2.5 评论文本意见挖掘系统应用

评论文本意见挖掘技术广泛应用于电子商务、商业智能、社会政治分析等领域。近年来，国内外研究机构和公司开发了很多实用的意见挖掘系统和工具。

Dave 等人^[67]研究并开发的 ReviewSeer 是世界上第一个情感分析工具和第一个针对给定产品评论区别其褒贬性的系统。Gamon 等人^[68]研究并开发的 Pulse 系统可以自动挖掘网上用户所上载的文本中有关汽车评论中的褒贬信息和强弱程度。Liu 等人^[55]研究并开发的 Opinion Observer 系统可以处理网上在线顾客产品评论，对设计产品（如电子照相机）各种特征的优缺点进行统计，并采用可视化方法对若干种产品特征的综合质量进行比较。IBM Almaden 研究中心的 Yi 和 Niblack^[69]研究和开发了 WebFountain 系统中的意见挖掘器。WebFountain 系统是一个基于多类型数据、开放领域意见挖掘的研究和开发平台。Popescu 等人^[70]开发的 OPINE 是一个无监督的信息抽取系统，可

以自动从大量产品评论中抽取产品特征和用户评价意见。OPINE 考虑了对隐性特征的抽取, 并对意见强度进行排序。Wilson 等人^[71]研究并开发的 OpinionFinder 系统能够自动识别主观性句子以及句子中各种与主观性有关的成分。Scaffidi 等人^[72]研究和开发的 Red Opal 系统从用户评论数据中抽取产品特征, 并对产品的每一个特征进行评分, 为用户提供基于特征的产品查询。IBM 开发的 OpinionIt^[73]是一个多语言的意见挖掘系统, 该系统使用浅层语义分析技术将产品评论的方面进行聚类, 分别抽取这些特征的意见, 并给出正负意见的百分比。Zhang 等人^[74]研究并开发的 Weakness Finder 系统能够收集用户反馈数据, 抽取中文在线评论中的产品特征和用户满意程度, 进而对比不同产品之间的优缺点。

国内方面, 姚天昉等人^[75]开发了一个用于汉语汽车评论的意见挖掘系统, 该系统能够在各大论坛上挖掘顾客们对各种汽车品牌的不同性能指标的评论和意见, 并且判断这些意见的褒贬性以及强度, 根据统计信息给出可视化结果。测试结果表明该系统达到了 80% 的平均召回率和 60% 的平均准确率。郝博一等人^[76]研究并开发的 OPINAX 系统, 能够基于语言依存分析和语料库统计从原始评论语料中挖掘未登录 (OOV) 产品属性。

1.2.6 存在的困难和不足

从国内外的研究现状可以看出, 评论文本意见挖掘是一项复杂的技术, 而不单纯是文本挖掘或者文本分类的工作。虽然以上各个方面的研究都取得了不少进展, 但总体上仍处于发展阶段, 评论文本意见挖掘研究仍然面临着以下关键科学问题:

(1) 特征选择困难。与传统的信息抽取或者文本分类不同, 评论文本意见挖掘旨在获取文本中的主观性元素 (如意见、评论、观点等) 以及这些元素与事实性元素 (比如主题、意见持有者等) 之间的关系, 这些成分表达形式多样, 而且常常不是显式地、独立地表达。此外, 文本意见挖掘也涉及到自然语言文本从篇章结构到句法、词法以及语义等各个层面的处理。目前的解决方案更多地是利用传统方法进行常规特征的提取, 而忽略了主观性文本语言的多变性、省略性、情感隐蔽性等特点。因此, 评论文本意见挖掘面临着更加困难的特征选择问题。

(2) 数据稀疏和有效特征不足。评论文本意见挖掘针对的处理对象是用户发表的主观性评论信息, 然而由于发表评论的自由性和用户素质的良莠不齐, 这些评论文本常常在内容和形式上杂乱无章。随着语料规模的增大, 目前的挖掘模型和算法通常面临着特征空间巨大、数据稀疏和有效特征缺乏的问题, 对于隐式主题以及细粒度的

主题和情感关系的识别还存在困难。因此，如何挖掘和利用更深层次、更复杂的语义信息成为提高文本意见挖掘效果的关键。

(3) 领域依赖性问题。与大多数自然语言处理任务类似，评论文本意见挖掘也面临着强烈的领域依赖问题。当测试语料与训练语料属于不同的领域时，通常会导致性能明显地下降。并且，同样的算法和模型应用于不同的领域时，也会产生较大的性能差别。因此如何设计具有良好领域自适应性的模型和方法是评论文本意见挖掘技术面向大规模实际应用的基础问题。

1.3 论文主要研究内容和创新点

1.3.1 论文主要研究内容

Web 评论文本的细粒度意见挖掘涉及多项非常具有挑战性的任务。针对目前细粒度意见挖掘研究领域中的问题，如缺乏有效融合多级别特征的统一挖掘框架、数据稀疏、有效特征缺失、算法的领域自适应能力较差等，本文重点从细粒度意见元素抽取、自适应评价目标聚类方法、领域情感词典自动构建方法、评价目标和情感词联合聚类四个方面开展研究工作，并通过实验语料对模型和算法进行验证。具体研究内容论述如下：

(1) 细粒度的意见元素抽取。与传统的文档级别情感分类任务不同，细粒度意见挖掘旨在对主观性文本数据进行更深入的意见抽取和分类工作。目前针对文档级别的情感分类技术已经相对成熟，然而细粒度的意见抽取和分类技术的研究仍未能取得令人满意的效果。

细粒度意见挖掘任务通常包括评价目标抽取、意见持有者识别、意见表达抽取和意见表达情感倾向判定等子任务，其中评价目标和意见表达抽取以及情感倾向判定是评论文本细粒度意见挖掘研究的核心任务。由于用户评论数据通常具有不规范性的特点，细粒度意见挖掘研究比传统的信息抽取任务更加困难，需要解决多种特征的融合抽取问题。本文将重点研究基于评论领域的意见元素抽取方法。

(2) 自适应的领域评价目标聚类。评价目标聚类是评论文本意见挖掘中一类重要的研究任务，该任务针对评论数据中出现的同一评价特征通常具有不同表达方式的现象，对抽取的评价目标集合按照特征和属性进行聚类。在领域评论文本中，由于数据的不规范性和用户评论的随意性，针对同一评论特征，不同的用户通常会采用不同的表达方式。例如，在数码相机领域中，“image”、“photo”和“picture”等评价目标

都用于描述“照片”特征。这些用来描述同一特征的不同但相关的词和短语，通常称为领域同义词。对这些领域同义词按照特征聚类，对基于特征的意见挖掘和摘要具有重要的作用。

目前的研究工作主要分为基于词典或词汇知识库对评价目标归类的方法和基于领域语料库中分布相似度对评价目标聚类的方法。但这些研究工作通常忽略了评价目标的领域依赖性，很多表达方式与特定的领域关联以后才能表示同一评论特征；同时评价目标之间的语义关联也存在着有效信息不足问题。针对评论文本中同一评论特征的不同表达方式问题，我们重点研究领域自适应的评价目标聚类问题。

(3) 领域情感词典自动构建。领域情感词典包含了情感词/短语及其在对应领域中的先验情感倾向，是针对领域文本进行意见分类的基础。然而针对不同的应用领域，情感词的情感倾向并不是固定不变的，而是依赖于出现的语境。比如，在独立语境下具有情感倾向的词语可能在某一具体语境中表现为中性词；正面的情感词可能在某一具体语境中表达负面情感，负面的情感词也可能在某一具体语境中表达正面情感。因此领域情感词典自动构建算法研究是文本意见挖掘研究中重要的基础工作之一。

目前针对意见挖掘情感词典构建的研究中，多数方法利用词语在语料中的共现信息或者在词典中的语义关系获得词语相似度，然后通过与基准词的相似度比较判断词语的情感倾向性，构建情感词典。然而，这些方法通常都面临着领域依赖性、自动构建程度不高、准确率低下等问题。本文将针对上述领域情感词典自动构建问题开展研究。

(4) 评价目标和情感词的联合聚类。目前大多数的意见挖掘方法中，主要侧重于评价目标、情感词及它们之间的对应关系的判定，挖掘的结果通常以离散化的“意见目标—情感词”对的形式呈现给用户。随着评论文本规模的急剧增长和评论领域的不断多样化，亟需进一步挖掘和分析潜在的评价目标与情感词之间的依赖关系，这也是基于特征或主题的意见挖掘研究的根本任务。

评价目标和情感词联合聚类通过计算评价目标与情感词之间的语义关联度，同步迭代地对评价目标按照语义方面 (Aspect) 聚类，对情感词按照与评价目标类的关联聚类。目前的研究工作中，基于先验知识库的方法通常忽略了领域依赖性问题，同时也会产生覆盖率不足问题；基于领域语料的方法虽然能够在一定程度上缓解领域依赖性，却面临着特征空间的高维度和稀疏性问题。针对上述问题，本文将重点研究评价目标和情感词的联合聚类。通过采用联合聚类框架，将评价目标和情感词抽取以及匹

配修饰关系计算问题定义为基于先验约束知识的半监督学习问题。

1.3.2 论文创新点

本文围绕 Web 评论文本的细粒度意见挖掘技术开展研究,论文的主要工作和创新点包括:

(1) 提出一种基于序列标注学习和句法语义结构特征的意见元素抽取方法。针对现有的细粒度意见挖掘模型面临的特征空间巨大和有效特征缺乏等难题,本文将评价目标和情感词抽取任务转化为序列标注学习过程,采用条件随机场 CRFs 模型构建有效融合多级别标注特征的统一抽取框架,并提出一种朴素的图剪枝算法对评价目标进行自动分类。在综合考虑词汇特征和上下文特征的基础上,引入句法语义结构特征,能够有效地抽取长距离的依存句法结构信息,解决有效标注特征缺乏问题。实验结果表明,本文提出的意见元素抽取方法取得了良好的评价目标抽取和分类效果。

(2) 提出一种基于约束的评价目标谱聚类方法。针对计算评价目标之间相似度的领域依赖性和有效关联信息不足问题,对评价目标的词法结构和上下文信息进行建模,定义和抽取了评价目标之间的词法约束和上下文约束关系,增强评价目标之间的领域关联性。针对局部约束关系的稀疏性问题,采用约束传播算法将局部评价目标之间的约束关系传播到全局评价目标空间。采用约束谱聚类算法,在融合先验约束知识的同时,有效降低了聚类空间的高维性和稀疏性。在三种产品评论领域上的实验结果表明,基于约束的谱聚类方法有效提高了评价目标聚类效果。

(3) 提出一种基于约束标签传播的领域情感词典自动构建方法。该方法首先针对情感词的领域性,利用组块依存知识和先验通用情感词典抽取领域候选情感词和短语;然后利用情感词在评论语料中的分布上下文,构建情感词之间的情感关联相似度计算模型,并通过定义和抽取情感词之间的情感倾向约束关系,增强情感词之间领域依赖的情感关联相似度,为情感词典构建提供先验领域知识;最后采用标签传播算法,在融合先验约束知识的同时,快速收敛地将种子词的情感信息传播到全局候选情感词空间。在公开评论语料上的实验结果表明,约束标签传播方法有效提高了领域情感词典构建的准确率,受种子词覆盖的影响较小。

(4) 提出一种基于约束的评价目标和用户意见联合聚类方法。本文采用约束联合聚类算法框架,将评价目标和情感词抽取以及匹配修饰关系计算问题转化为基于约束的半监督学习过程。针对评价目标相似度以及情感词相似度的稀疏性和领域依赖问题,在评价目标具有相互约束信息的基础上,根据情感词与评价目标的关联关系引入

情感词之间的约束关系，为联合聚类过程提供先验的指导知识。利用约束联合聚类算法，在融合先验约束知识的同时，同步地对用户评价目标按照语义方面进行聚类，对情感词按照与评价目标类的关联进行聚类。实验结果表明，基于约束的联合聚类方法能够有效地解决评价目标和情感词抽取以及它们之间的匹配修饰关系计算问题，取得了比基线方法更高的聚类准确率。

1.4 论文组织结构

本文共分为 6 章，内容安排如下：

第 1 章，绪论。首先阐述了论文的研究背景和意义，介绍了国内外在 Web 评论文本意见挖掘各个方面的研究现状和发展趋势，并指出了论文的主要研究内容和创新点以及论文的组织结构。

第 2 章，基于序列标注学习的意见元素抽取。本章首先介绍了 Web 评论文本细粒度意见挖掘的基本内容以及研究现状，接着介绍了条件随机场理论和 CRF++ 开源工具。针对目前研究方法中存在的有效特征稀疏和缺乏融合多级别特征的统一框架问题，提出了基于序列标注学习和句法语义结构特征的意见元素抽取方法，并采用朴素的图剪枝算法对评价目标集合进行自动分类。最后，通过在实际数据集上的实验验证了本章提出的细粒度意见元素抽取方法的有效性。

第 3 章，基于约束的评价目标谱聚类。本章首先描述了评价目标聚类的基本内容和目前的研究方法，即基于无监督学习的方法、基于主题建模的方法以及基于半监督学习的方法。接下来，针对目前研究中存在的领域自适应性和语义关联信息不足问题，提出了基于约束谱聚类的评价目标聚类方法。通过定义和抽取评价目标之间的先验约束关系，增强了评价目标之间的领域关联性。最后，通过在三种实际产品评论领域上的实验验证了约束谱聚类算法在解决评价目标聚类问题上的有效性。

第 4 章，基于约束标签传播的领域情感词典自动构建。本章首先介绍了领域情感词典构建的研究背景和基本内容，然后介绍了目前的相关研究方法，即基于词汇知识库的构建方法和基于领域语料库的构建方法。针对目前研究中存在的领域依赖性、情感关联信息缺乏和种子词依赖问题，提出了基于约束标签传播的领域情感词典自动构建算法。通过定义情感词之间的词法约束和上下文约束关系，为领域情感词典构建提供先验的领域知识。最后，通过在公开评论语料上的实验验证了基于约束标签传播的领域情感词典自动构建方法的有效性。

第 5 章，评价目标和情感词联合聚类。本章首先介绍了评价目标和情感词联合聚类问题的产生背景和基本内容，然后介绍了目前的相关研究工作，对其中存在的缺陷和问题进行分析。接下来描述了联合聚类的理论和框架，以及目前重点研究的半监督联合聚类。针对目前研究中面临的标注训练语料缺乏、特征稀疏和领域依赖引起的准确率低下问题，提出了基于约束的联合聚类方法，将评价目标和情感词联合抽取以及匹配修饰关系计算问题转化为基于约束的半监督学习过程。最后，通过在公开评论语料上的实验验证了基于约束的评价目标和情感词联合聚类的有效性。

第 6 章对本文的整体工作进行了总结，分析了整个研究过程中的贡献以及存在的不足，并展望了下一步可能的研究方向。

第2章 基于序列标注学习的意见元素抽取

2.1 引言

按照挖掘粒度区分,文本意见挖掘可以分为粗粒度的主客观和情感倾向挖掘与细粒度的意见元素和关系挖掘。在早期,文本意见挖掘研究主要关注于粗粒度的主客观和情感倾向分类任务^[5, 8, 9, 19, 26, 71, 77, 78]。主客观挖掘通常针对给定的一篇文档、一个段落或者句子,分析其是否具有主观立场,即主观或者客观。情感倾向挖掘主要针对给定的具有主观性信息的文本,判断其情感倾向性,即正面或者反面^[2]。通常,挖掘的对象也涉及到了篇章、段落、句子以及词汇和短语等各种文本粒度。

近年来,随着互联网的广泛应用,Web 上涌现出大量的针对商品和服务的用户评论信息。这些评论信息包含了用户对商品或主题的各个方面的详细分析和评价,用户对不同的方面可能会做出截然不同的评价。整体上正面的评论和句子并不代表用户对评论对象的所有特征都持肯定态度;同理,负面的评论和句子中也可能包含用户对某个方面的肯定态度^[2]。单纯地对这些评论文本进行粗粒度的主客观性和情感极性分类已经不能满足人们的信息需求。越来越多的电子商务企业鼓励消费者对商品和服务的各个方面进行细致的评价,消费者也希望能够详细地了解 and 对比商品各个方面的优缺点,从而做出更为合理的购买决策。因此,针对评论文本的细粒度意见挖掘逐渐成为目前文本意见挖掘研究的热点问题。

与传统的粗粒度文档级别情感分类任务不同,细粒度意见挖掘旨在从主观性文本数据中抽取有价值的情感信息单元,主要涉及到意见元素和匹配修饰关系识别,以及情感倾向性计算工作。目前针对文档和句子级别的粗粒度情感分类技术已经相对成熟,然而细粒度的意见抽取和分类技术的研究仍未能取得令人满意的效果。在过去的几年间,许多研究工作^[14, 54-56, 59, 60, 79-83]都尝试明确定义和解决这一难题。然而,用户发表评论时通常具有很大的自由度和随意性,并且不同用户的表达习惯也并不相同,例如在数码相机领域中,不同用户可能分别采用“picture quality”和“image clarity”表示“照片”特征。因此,用户评论文本呈现出高度的随意性和不确定性特点。与传统的信息抽取和分析任务相比,针对 Web 评论文本的细粒度意见挖掘研究具有更大的困难。

根据文献^[90]中 Kim 和 Hovy 对文本意见的定义, 评价主题 (Topic)、意见持有者 (Holder)、意见陈述 (Claim) 和情感倾向性 (Polarity) 构成了意见的四个组成要素^[91]。对应地, 细粒度意见挖掘定义为评价目标抽取、意见持有者识别、情感词抽取、情感倾向性判定以及评价目标与情感词之间匹配修饰关系计算等子任务。通常将细粒度文本意见挖掘的结果表示为一个四元组 <Opinion Target, Opinion Holder, Opinion Word, Polarity(+/-)>, 分别对应评价目标、意见持有者、情感词和情感倾向性。

评价目标 (Opinion Target): 也称为意见目标, 是指用户讨论的目标, 具体表现为评论文本中评价词语所修饰的对象。例如在商品评论领域中, 评价目标包含商品本身、商品的特征和属性、以及属性的特征等, 通常以商品特征和属性作为基本粒度。

意见持有者 (Opinion Holder): 观点/评论的隶属者, 通常是指发表评论的用户或机构。如在商品评论领域中, 持有者通常被默认为评论文本的作者。在针对新闻评论的意见挖掘中, 意见持有者的识别显得尤为重要。

情感词 (Opinion Word): 也称为意见词, 是指能够直接表达用户正面或负面意见、态度、评价的词汇和短语。例如, “good”、“enjoy” 等词语为正面情感词, “poor” 和 “dislike” 等词语为负面情感词, 还有一些词如 “fast”、“big” 和 “long” 等在独立语境下表示中立情感, 其倾向性需要结合具体上下文语境进行判断。

情感倾向性 (Opinion Polarity): 用户对于评价对象表达的正面、负面或中立的态度, 通常也称为意见倾向性、情感极性等。情感倾向性包括最常见的褒贬二元类别和包含中立的三元类别, 以及更细致的多元类别等^[92, 93]。

在相关文献^[1, 15, 85, 94-96]中, 评价目标也称为评价特征 (Feature)、评价主题 (Topic)、意见方面 (Aspect) 等, 情感词也称为意见表达 (Expression)、意见陈述 (Claim) 等。本文中为了论述简便, 统一使用评价目标和情感词表述。

目前的意见元素抽取研究可以主要划分为基于统计规则的方法^[54, 55, 70, 80, 84-86]和基于机器学习的方法^[56, 58-60, 81, 87-89]两类。基于统计规则的方法通常利用手工设定的抽取规则抽取高频的名词短语及附近的形容词和动词作为候选评价目标和意见词。然而这类方法会引入大量的噪声数据和遗漏低频的评价目标, 导致较低的算法准确率, 同时也没有考虑到隐式特征的情形。基于机器学习的方法试图将细粒度的意见元素抽取看作概率学习过程, 但却面临着特征选择困难的问题。

在本章中, 我们将评价目标和情感词抽取任务转化为序列标注学习问题, 采用条件随机场 (Conditional Random Fields, CRFs) 来构建有效融合多级别特征的统一抽取

框架。针对标注特征选择和有效特征稀疏等问题,在传统词汇级别和上下文级别标注特征的基础上,引入句法语义结构特征,解决了相关词汇距离较远时缺乏充分约束关系的问题。在抽取得到用户评价目标集合之后,进一步地对评价目标集合基于主题/属性分类。通过引入基于语义知识和基于分布上下文的相似度计算方法,计算评价目标之间的语义相似度并构建语义关联图,我们提出一种朴素的基于图剪枝的评价目标分类方法。在实际数据集上的实验结果验证了本章提出的基于序列标注学习和句法语义结构特征的意见元素抽取方法的有效性。

2.2 研究现状分析

现有的意见元素抽取研究主要分为基于统计规则的方法和基于机器学习的方法。基于统计规则的方法^[15, 54, 70, 80]由领域专家手工设定抽取规则,从评论语料中选择高频的名词短语作为候选评价目标。这类方法抽取得到的特征词相对准确,但面临着遗漏低频特征词并引入许多高频的噪声名词问题。基于机器学习的方法^[56, 60, 81, 87]将细粒度意见挖掘看作在用户评论语料上的概率学习过程,通常面临着如何进行特征选择和有效标注特征稀疏问题。

Hu 和 Liu^[15, 54]最早开始研究从产品评论文本中挖掘产品特征并生成基于特征的意见摘要。他们利用关联规则选择高频的产品特征词,再利用 WordNet 中形容词之间的同义词和反义词关系,通过 bootstrapping 方法判定情感词的倾向性,最后根据句子中情感词的主导倾向性判断句子极性,生成基于特征的意见摘要。Popescu 和 Etzioni 等人在文献^[70, 80]中对 Hu 和 Liu 的早期工作进行了改进,通过清除高频的噪声名词及短语,在几乎无损召回率的情况下显著地提高了准确率。Wu 等人^[84]选择名词和动词短语作为候选产品特征,并基于产品特征词在产品评论文本中具有更高出现频率的假设,采用语言模型对候选产品特征进行过滤。Qiu 等人^[85]也是抽取名词和名词短语作为候选评价目标。然而,这些方法的局限性在于虽然能够取得较好的召回率,但准确率较差,并且不能有效地识别低频的产品特征。

近年来,主题模型逐渐被广泛应用于在线评论的特征和情感分析中。Brody 和 Elhadad^[60]提出了一种无监督的从评论文本中抽取用户评价特征并判定情感倾向的两步方法,该方法首先从句子级别的局部 LDA 模型中抽取产品特征,然后通过极性传播识别产品特征关联的情感词。但是,他们只考虑了评论句子仅包含单个产品特征的情况,而忽略了同时包含多个产品特征的情况。Zhao 等人^[81]将有监督的最大熵判别

式模型融入到无监督的主题产生式模型中，提出了一种最大熵-LDA 混合模型，该模型能够同时从评论文本中抽取评价特征和对应的情感词。Zhai 等人^[87]扩展了一种基于约束的 LDA 模型，使之具有处理大规模约束的能力，并提出两种自动抽取产品特征之间先验约束关系的方法。然而，这些方法主要基于对用户评价语料的统计学习，生成的主题并不能很好地对应细粒度的产品特征和情感词。

针对产品特征分类问题，Guo 等人^[97]提出了一种基于多层潜在语义关联(mLaSA)模型的无监督方法，第一层 LaSA 模型生成产品特征的潜在语义结构，第二层 LaSA 模型根据产品特征的潜在语义结构和在评论文本中的上下文片段对产品特征集合进行分类。Zhai 等人^[118]将无监督产品特征聚类问题转换为半监督的学习任务，提出了一种基于朴素贝叶斯分类的期望最大化(Expectation Maximum, EM)算法。Jin 等人^[56]提出了一种基于词汇化隐马尔科夫(Lexicalized HMMs)模型的 OpinionMiner 系统，该系统尽管被验证了比之前的方法更加有效，却依然面临着无法对输入数据序列上特征之间的相互依赖关系进行建模的问题。

Li 等人^[98, 99]将矩阵分解算法应用于许多意见挖掘任务中。文献^[98]提出了对词项-文档矩阵构建一种约束非负三分解(Non-negative Tri-factorization)模型，同时融入了领域无关的先验情感词法知识以及领域特定的部分文档情感标记信息。文献^[99]将非负矩阵三分解模型应用于解决跨领域情感分类任务，首先将源领域中已标注文档级别的情感转移到词汇级别，其次将词汇级别的情感转移到目标领域中的未标注文档。其假设源领域和目标领域都采用共同的词汇空间。Cambria 等人^[100]采用机器翻译技术将英语中的常识知识的分类体系和语义网络结构对应地翻译到汉语中，开发了用于中文情感分析的汉语常见常识知识库(Chinese Common and Common Sense Knowledge Base)。然而，这些方法需要利用额外的语义词典或领域知识库，通常也会面临着数据稀疏和先验领域知识缺乏问题。

2.3 条件随机场理论

条件随机场(Conditional Random Fields, CRFs)是一种用于标注和切分序列数据的条件概率模型，由 Lafferty 于 2001 年在最大熵模型和隐马尔科夫模型的基础上，提出的一种判别式概率无向图学习模型^[101, 102]。CRFs 模型是一种典型的判别式模型(Discriminative Model)，重点解决有序数据的序列化标注问题，目前已广泛应用于自然语言理解、生物信息学、机器视觉以及网络智能等领域。

在给定实际观测序列的条件下，条件随机场定义了标注序列空间上的对数线性分布。定义 X 表示未标注数据序列的随机变量， Y 表示对应的标记序列的随机变量，Lafferty 和 MaCallum^[101]定义了给定 X 和 Y 上的条件随机场如下：设 $G = (V, E)$ 是一个无向图，其中 V 表示 G 中结点的集合， E 表示 G 中边的集合。设 $Y = (Y_v)_{v \in V}$ 是以 G 中结点 v 为索引的随机变量 Y_v 构成的集合。在给定 X 的条件下，如果每个随机变量 Y_v 服从马尔科夫属性，即：

$$p(Y_v | X, Y_w, w \neq v) = p(Y_v | X, Y_w, w \sim v) \quad (2.1)$$

则 (X, Y) 就构成一个条件随机场。其中， $w \sim v$ 表示结点 w 和 v 是图 G 上的相邻结点。

最简单常用的是一阶链式结构，即线性链结构（Linear-chain）CRFs 模型。其结构是一个线性链，输出变量表示成一个序列。图 2.1 给出了其对应的图模型。

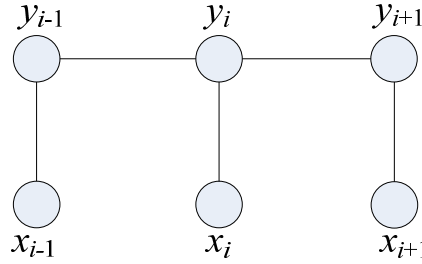


图 2.1 线性条件随机场模型结构

随机变量之间通过表示依赖关系的无向边进行连接。在线性链 CRFs 模型中，假设在各个输出结点之间存在一阶马尔科夫独立性，其输出结点被边连接成一条线性链。在给定观测序列 $X = \{x_1, x_2, \dots, x_n\}$ 的条件下，标注序列 $Y = \{y_1, y_2, \dots, y_n\}$ 上的联合分布概率为^[101]：

$$p_\theta(y | x) \propto \exp \left(\sum_{e \in E, k} \varphi_k h_k(e, y |_e, x) + \sum_{v \in V, k} \mu_k g_k(v, y |_v, x) \right) \quad (2.2)$$

其中， h_k 定义为边上的转移特征函数，依赖于当前位置 y_t 和前一位置 y_{t-1} ， g_k 定义为结点上的状态特征函数，只依赖于当前位置。 φ_k 和 μ_k 为每个特征函数的权值。我们可以将转移特征 h_k 和状态特征 g_k 用统一的特征函数符号 $f_k(y_{t-1}, y_t, x, t)$ 表示，则有：

$$p_\theta(y | x) = \frac{1}{Z_\theta(x)} \exp \left(\sum_{t=1}^T \sum_{k=1}^K \lambda_k f_k(y_{t-1}, y_t, x, t) \right) \quad (2.3)$$

其中, $Z_\theta(x)$ 为归一化因子, 确保所有可能的状态序列的条件概率之和为 1, 即它是所有可能的状态序列的“得分”之和^[46]:

$$Z_\theta(x) = \sum_y p_\theta(y|x) = \sum_y \exp\left(\sum_{t=1}^T \sum_{k=1}^K \lambda_k f_k(y_{t-1}, y_t, x, t)\right) \quad (2.4)$$

条件随机场模型在观测序列的基础上对目标序列进行建模, 既具有判别式模型的优点, 又具有产生式模型考虑到上下文标记间的转移概率、以序列化形式进行全局参数优化和解码的特点, 解决了其他判别式模型难以避免的标记偏置问题^[101]。

特征函数选择、参数估计和模型推断是条件随机场模型中需要解决的三个关键问题。分别描述如下。

2.3.1 特征函数的选择

特征函数用来对训练数据的经验分布特征进行建模。在条件随机场理论中, 通过定义转移特征函数和状态特征函数可以自由地从观测和状态组合中选取特征, 避免了 HMM, Naïve Bayes 的观测独立性假设的局限性。特征函数的选取直接关系模型的性能。在线性链 CRFs 模型中, 转移特征函数定义了观测序列的标记位置 $t-1$ 和 t 之间的依赖关系, 状态特征函数定义了当前位置观测变量和标记变量之间的依赖关系。

转移特征函数和状态特征函数通常都设为二值函数, 表示当观测变量和标记变量满足某一事实条件时其特征函数值为 1。例如^[103]:

$$h_k(y_{t-1}, y_t, x, t) = \begin{cases} 1, & y_{t-1} = B \text{ 且 } y_t = O \\ 0, & \text{否则} \end{cases} \quad (2.5)$$

该转移特征函数表示如果标记序列中相邻的两个标记, 前一标记是 B , 当前标记为 O , 则函数值就为 1, 否则为 0。

$$g_k(y_t, x, t) = \begin{cases} 1, & x_t = 'C' \text{ 且 } y_t = B \\ 0, & \text{否则} \end{cases} \quad (2.6)$$

该状态特征函数表示如果观测序列中当前位置 t 观测到的值为 ‘ C ’, 对应的标记为 B , 则函数值就为 1, 否则为 0。

2.3.2 参数估计

CRFs 模型的参数估计是从训练样本中学习参数向量 θ 的过程, 通常采用最大似然估计方法。已知训练样本集 $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$, 其中 $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_T^{(i)}\}$ 表示输入

数据序列, $y^{(i)} = \{y_1^{(i)}, y_2^{(i)}, \dots, y_T^{(i)}\}$ 表示对应的输出标签序列。在假设 D 中的每个 $\{x^{(i)}, y^{(i)}\}$ 对之间独立同分布的条件下, CRFs 模型在 D 上的对数似然概率公式为^[46, 104]:

$$L(\theta) = \sum_{i=1}^N \log P_{\theta}(y^{(i)} | x^{(i)}) \propto \sum_{x, y} \log p_{\theta}(y | x) \quad (2.7)$$

将条件概率公式(2.3)带入公式(2.7)中。在模型的训练过程中, 为避免参数过多产生过度拟合 (Overfit) 问题, 通常需要采用先验高斯分布对模型参数进行平滑^[46, 105], 调整后的对数似然概率公式为:

$$L(\theta) = \sum_{i=1}^N \sum_{t=1}^T \sum_{k=1}^K \lambda_k f_k(y_{t-1}, y_t, x, t) - \sum_{i=1}^N \log Z(x^{(i)}) - \sum_{k=1}^K \frac{\lambda_k^2}{2\sigma^2} \quad (2.8)$$

公式(2.8)即为 CRFs 模型中最大似然参数估计的目标函数。其中, 右边最后一项是用于缓解模型参数过度拟合的高斯先验值。

在实际的参数估计求解过程中, 通常采用一种迭代缩放或梯度的方法, 通过更新规则迭代地更新模型中的参数, 使得在更新参数条件下的模型具有更高的对数似然概率值。常用的迭代缩放实现有 Darroch 和 Ratcliff^[106]提出的 GIS (Generalised Iterative Scaling) 算法和 Della Pietra 等人^[107]在 GIS 基础上提出的 IIS (Improved Iterative Scaling) 算法。文献^[101]中 Lafferty 采用两种基于 IIS 的算法求解条件随机场的极大似然参数。IIS 算法虽然原理简单, 能够保证求解的收敛性, 但收敛速度非常缓慢。为了进一步地提高求解效率, Sha 和 Pereira^[108]采用了一种有限内存条件下的拟牛顿方法 (limited memory BFGS, L-BFGS)。L-BFGS 是一种基于历史的梯度和更新值来估计曲率值的二阶优化方法, 在大规模参数估计问题上表现显著。

2.3.3 模型推断

CRFs 理论中的模型推断主要针对在已知模型参数的条件下, 给定未标记的输入数据序列, 求解最可能的输出标记序列问题。设 $x = \{x_1, x_2, \dots, x_T\}$ 表示观察到的输入数据序列, $y = \{y_1, y_2, \dots, y_T\}$ 表示被预测的输出标记序列。已知模型参数 $\theta = \{\theta_1, \theta_2, \dots, \theta_K\}$ 的线性 CRF, 给定 x , 最可能的标记序列 \hat{y} 的条件概率为:

$$\hat{y} = \arg \max_y p_{\theta}(y | x) \quad (2.9)$$

将后验条件概率公式(2.3)带入到(2.9)中, 得到:

$$\hat{y} = \arg \max_y \frac{1}{Z_\theta(x)} \exp\left(\sum_{t=1}^T \sum_{k=1}^K \lambda_k f_k(y_{t-1}, y_t, x, t)\right) \quad (2.10)$$

对公式(2.10)求解可以采用类似 HMM 中的 Viterbi 算法。其采用动态规划的基本思想，对少数可能的路径进行计算，从而代替对全部路径的计算，通常只计算每个序列中最可能的一条路径。为了找到最优路径，需要在每一时刻都保存能够回溯前一时刻最优状态的数据^[46]。

2.3.4 CRF++介绍

CRF++¹是一个可用于分割和标注序列数据的简单、可定制的条件随机场开源工具，也是目前综合性能最佳的 CRF 工具。CRF++是为了通用目的而设计，可以应用于自然语言处理领域的一系列任务，如自动分词、命名实体识别、信息抽取和语块分析等。

CRF++工具包采用 C++语言编写，包括 Linux 环境下的源代码和 Windows 平台下的可执行程序。在本文中，我们使用的是 Windows 平台下的相关工具和说明文档。CRF++规定了训练和测试文本必须遵从一定的格式，通常包含多组符号，每组符号包含多个列，每一列都代表了这组的某个特征^[103]。在训练 CRF 模型之前，用户还需要预先定义好特征模板，模板文件描述了在训练和测试中用到的特征。图 2.2^[103]描述了训练和测试文档以及特征模板的格式示例。

I	PRP	B
love	VB	O
this	DT	B
vacuum	NN	F
cleaner	NN	F
模板:		
# Unigram		
U00:%x[0,0]		
U01:%x[0,1]		
U02:%x[0,0]/%x[0,1]		
# Bigram		
B		

图 2.2 训练和测试文档以及特征模板的格式示例

¹ <http://crfpp.googlecode.com/svn/trunk/doc/index.html>

2.4 基于序列标注学习的细粒度意见挖掘框架

图 2.3 描述了基于序列标注学习的细粒度意见挖掘的整体结构，包括用户评论采集、评论文本分析处理、条件随机场模型训练和测试、评价目标和情感词抽取与分类四个处理模块。各个模块的功能论述如下：

用户评论采集模块主要包括：针对用户指定的产品类别，从用户感兴趣的评论网站和论坛上抓取原始的用户评论网页，抽取其中的产品描述信息、用户评级、评论文本正文以及一些半结构化的用户标签等数据，然后将它们存入评论语料库。

评论文本分析处理模块主要包括：对评论文本进行段落和句子切分，对句子进行分词和词性标注，修改拼写错误和清理无意义的符号和标签；对单词取词干操作并对词性进行简化归并，标注单词的词干和词性特征；抽取评论句子中词汇之间的依存句法关系，依据句法语义结构特征的编码规则，标注单词的句法语义结构特征；定义词序列的语义角色标记，手工标注 CRFs 模型的训练语料和测试语料；提供基于 WordNet 和基于分布上下文计算评价目标之间相似度的调用接口。

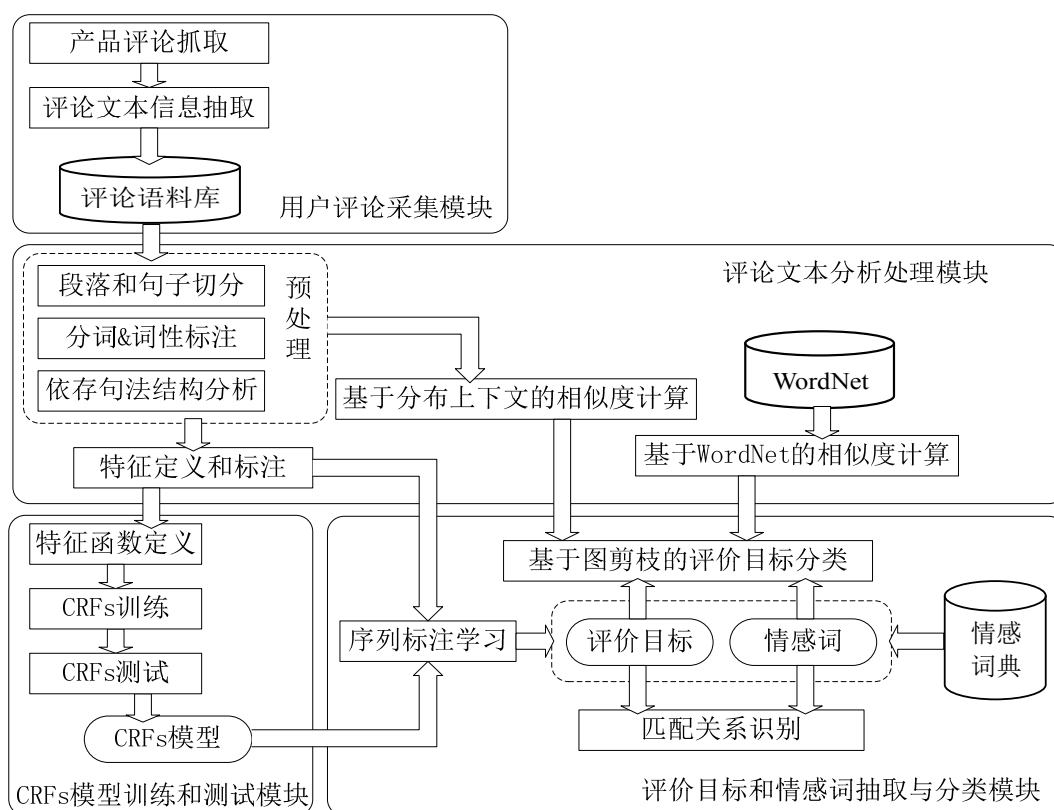


图 2.3 基于序列标注学习的细粒度意见挖掘框架

条件随机场模型训练和测试模块主要包括：设定 CRFs 模型的特征函数模板；使用训练语料进行迭代训练得到 CRFs 模型；使用测试语料评价 CRFs 模型的标注效果；使用训练得到的 CRFs 模型对新的语料进行序列标注。

评价目标和情感词抽取与分类模块主要包括：从 CRFs 模型对新的语料的序列标注结果中，抽取用户评价目标和情感词；基于句法依赖关系和近邻距离计算评价目标和情感词之间的匹配修饰关系；根据情感词典判定情感词的语义倾向；计算评价目标之间基于 WordNet 和基于分布上下文的相似度，对抽取得到的评价目标集合按照产品主题/属性进行分类。

2.5 基于 CRFs 模型的评价目标和情感词抽取

针对用户评论文本的意见挖掘研究中，意见持有者通常默认为评论文本的作者^[91, 93]。因此，我们主要抽取评论文本中的评价目标和情感词元素。我们定义评价目标的范围包括评论对象的组件和部件、评论对象的功能和属性、组件和部件的功能和属性以及其他相关对象等；定义情感词为用户的情感表达，通常为动词、形容词及其短语等。例如在数码相机领域中，“picture quality”、“lens cap”和“shutter delay”等都属于评价目标；“good”和“nice”、“terrible”和“hate”等分别表示正面和负面的情感词。基于 CRFs 模型的评价目标和情感词抽取通过选择标注特征和设定语义角色标记，训练 CRFs 模型并对新的评论语料进行自动标注。

2.5.1 特征选择

为了有效地抽取训练数据中包含的词法、上下文和语义结构特征信息，我们为 CRFs 模型的训练过程定义了词汇、上下文和句法语义结构三种级别的标注特征，分别描述如下：

(1) 词汇级别标注特征。反映了词汇属于评价目标还是情感词的可能性，包括词干特征和词性特征。在特定的评论领域中，一些词汇很明显就是评价目标词或者情感词，比如在数码相机领域中，“battery life”和“photo quality”、“nice”和“terrible”等分别表示评价目标和情感词。然而这些词具有不同的形式，如名词单复数、动词的不同形态以及形容词和副词的比较级形式等，这些词的不同形式在使用 CRFs 模型进行评价目标和情感词的抽取时会被识别为完全不同的词，因此不能充分利用词的不同形式的因素^[103]。为了消除由词形变化引起的特征空间稀疏问题，我们采用词干代替

词的不同形式。我们采用目前广泛应用的 Porter 算法提取单词的词干。Porter 算法是基于后缀剥离的词干提取算法，对文本中的单词递归地应用一系列的规则删除后缀，直到没有可用的规则。使用词干代替词本身进行特征函数的生成会放宽标记为评价目标和情感词的条件，使得标记为评价目标和情感词的词增多，从而提高召回率^[103]。

词性在一定程度上能够反映出候选词是评价目标还是情感词的可能性。目前的文献^[70, 80, 109, 110]通常认为，绝大部分的评价目标是以名词和名词短语的形式存在，绝大部分的情感词是以形容词和动词的形式存在。因此，词性特征在评价目标和情感词的识别中具有重要的作用。

我们利用开源的 Stanford Part-Of-Speech Tagger²工具^[111]获取句子中单词的词性。为了缩小词性标注特征空间，从而使词性相关的特征函数效果更加明显，需要对经常出现且对评价目标和情感词的抽取比较有意义的词性简化合并。我们将形容词相关的词性标注（如形容词 JJ、形容词比较级形式 JJR、形容词最高级形式 JJS 等）统一标记为形容词标注 JJ，将所有名词相关的词性标注（名词单数或物质名词 NN、名词复数形式 NNS、专有名词单数形式 NNP、专有名词复数形式 NNPS 等）统一标记为名词标注 NN，将所有副词相关的词性标注（副词 RB、副词比较级形式 RBR、副词最高级形式 RBS 等）统一标记为副词标注 RB，将所有动词相关的词性标注（动词 VB、动词过去时形式 VBD、动词现在分词形式 VBG、动词过去分词形式 VBN、动词第三人称单数形式 VBZ、动词非第三人称单数形式 VBP 等）统一标记为动词标注 VB。

词干特征和词性特征都是将原来不同形式的信息进行粗粒度化，使得具有相同词干或者相同词性的词都具有相同的特征函数，从而在 CRFs 模型训练和测试过程中具有统一的效果。在特征函数模板中，指定当前词的词干和词性作为词汇级别的标注特征。

（2）上下文级别标注特征。虽然词干和词性特征能够在一定程度上反映词汇属于评价目标还是情感词的可能性，但当前词的词干和词性都只是单个词的特征，无法用于对评价目标和情感词的上下文特征进行建模。基于 CRFs 模型的序列标注学习过程依然面临着缺乏有效特征的问题。因此，我们引入了上下文级别标注特征。

上下文级别标注特征反映了局部上下文范围内词汇之间的浅层语义结构关系，包含前后位置词汇的词汇级别特征和浅层语义分析特征。首先，指定 $[-1, +1]$ 范围为上下文窗口，包含当前词及其前一个词和后一个词的位置。在特征函数模板中，设定上下

² <http://nlp.stanford.edu/software/tagger.shtml>

文窗口内的词干和词性作为当前词的上下文级别标注特征；其次，对评论句子进行浅层语义分析，得到词汇的浅层语义标注。在特征函数模板中，设定浅层语义标注作为当前词的上下文级别标注特征。

(3) 句法语义结构级别标注特征。词汇和上下文级别标注特征只包含了当前和局部上下文范围内的有效特征，而忽略了词汇之间长距离的依存句法结构信息。如果两个相关词汇在句子中的距离较远，那么仅靠词汇和上下文级别特征就无法充分利用这两个词之间的关系。因此，我们通过对句子整体依存句法结构的分析，定义和抽取词汇之间的句法语义结构特征。通过引入词汇之间的句法语义结构特征，能够解决相关词汇距离较远时缺乏充分约束关系的问题。

2.5.2 句法语义结构特征抽取

2.5.2.1 依存句法树构建

以评论语句为单位，利用 Stanford 依存句法分析器³进行句法语义结构关系的解析，抽取词汇之间的依存句法关系。例如，给定评论句子“The video is good, but sounds are not very loud.”，抽取得到的词汇之间的依存句法关系如下：det(video-2, The-1)、nsubj(good-4, video-2)、cop(good-4, is-3)、root(ROOT-0, good-4)、nsubj(loud-11, sounds-7)、cop(loud-11, are-8)、neg(loud-11, not-9)、advmod(loud-11, very-10)、conj_but(good-4, loud-11)。以词汇为结点，以词汇之间的依存句法关系为边，构建得到的依存句法树结构如图 2.4 所示。

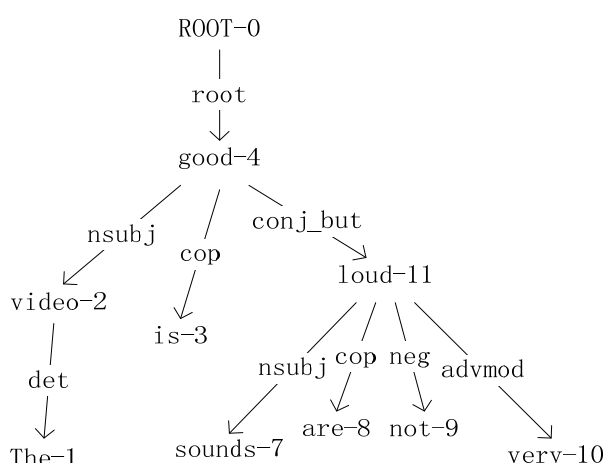


图 2.4 依存句法树示例

2.5.2.2 句法语义结构特征编码规则

³ <http://nlp.stanford.edu/software/lex-parser.shtml>

通过分析词汇之间的依存句法关系，本章定义了八种标注句法语义结构特征的编码规则，分别描述如下：

（1）形容词-名词短语结构

句子结构中出现形容词-名词短语的情况，其中形容词直接修饰名词。依赖的依存句法关系为 `amod`(名词, 形容词)。则将其中的名词的句法语义结构特征标记为 `NPNN`，将形容词的句法语义结构特征标记为 `NPADJ`。

（2）主系表结构

句子结构中出现主语-系动词-表语的情况，其中名词作为主语，形容词作为表语来修饰主语。依赖的依存句法关系为 `nsubj`(形容词, 名词)、`cop`(形容词, 系动词)。出现这种情况时则将其中的名词的句法语义结构特征标记为 `CNN`，将形容词的句法语义结构特征标记为 `CADJ`。

（3）形容词补语结构

句子结构中出现形容词充当补语修饰主语名词的情况。依赖的依存句法关系为 `nsubj`(动词, 名词)、`acomp`(动词, 形容词)。出现这种情况时则将其中的名词的句法语义结构特征标记为 `CNN`，将形容词的句法语义结构特征标记为 `CADJ`。

（4）名词词组结构

句子结构中出现由多个名词组成的名词词组，例如“photo quality”和“battery life”等都是词组形式存在的评价目标。依赖的依存句法关系为 `nn(governor, dependent)`，代表名词词组中单词间的关系，其中 `dependent` 用来修饰 `governor`。在这种情况下，则将其中 `governor` 的句法语义结构特征标记同时作为 `dependent` 的句法语义结构特征标记。

（5）并列名词或形容词结构

形容词-名词短语结构中出现并列形容词或并列名词，以及句子中出现并列主语的情况。依赖的依存句法关系为 `conj_and(governor, dependent)`，其中 `governor` 和 `dependent` 通常同时为名词或形容词。此时，将其中 `governor` 的句法语义结构特征标记同时作为 `dependent` 的句法语义结构特征标记。

（6）否定词修饰形容词结构

句子结构中出现否定词修饰形容词的情况。依赖的依存句法关系为 `neg`(形容词, 否定词)，表示形容词被否定词修饰。如果已知形容词被标记为 `NPADJ` 或 `CADJ`，且存在否定词修饰形容词，则重新将形容词的句法语义结构特征标记为 `NPADJNEG` 或

CADJNEG。

(7) 程度副词修饰形容词结构

句子结构中出现程度副词修饰形容词的情况。依赖的依存句法关系为 *advmod*(形容词, 程度副词)。如果已知形容词被标记为 CADJ 或 CADJNEG, 则标记程度副词的句法语义结构特征为 CADV; 如果已知形容词被标记为 NPADJ 或 NPADJNEG, 则标记程度副词的句法语义结构特征为 NPADV。

(8) 其他情况

除上述情况外的其他词都标记为 N。

综上所述, 句法语义结构特征共包含 9 种标记, 分别为: N、CNN、CADJ、CADJNEG、CA DV、NPADJ、NPADJNEG、NPADV、NPNN。

如图 2.4 所示的依存句法树示例中, 形容词 *good* 和名词 *video* 之间存在依存句法关系 *nsubj*(*good*-4, *video*-2)、*cop*(*good*-4, *is*-3), 符合规则 (2) 的条件, 因此将 *good* 和 *video* 的句法语义结构特征分别标记为 CADJ 和 CNN; 形容词 *loud* 和名词 *sounds* 之间存在依存句法关系 *nsubj*(*loud*-11, *sounds*-7)、*cop*(*loud*-11, *are*-8), 并且同时存在依存句法关系 *neg*(*loud*-11, *not*-9), 符合规则 (2) 和 (6) 的条件, 因此 *loud* 应该标记为 CADJNEG。此外, 由于存在依存句法关系 *advmod*(*loud*-11, *very*-10), 表示程度副词 *very* 修饰形容词 *loud*, 因此将 *very* 的句法语义结构特征标注为 CADV。

定义和引入句法语义结构级别的标注特征, 主要能够解决以下问题:

- (1) 相关词语在评论句子中距离较远时缺少关联约束的问题。
- (2) 处理出现并列名词或形容词的情况。
- (3) 通过定义 NEG 相关的标记, 处理出现否定词修饰情感词而改变情感极性的情况, 从而有助于对情感词的情感极性进行后期处理。

2.5.3 语义角色标注

我们定义了五种词序列的语义角色标记, 分别为 F、O、MOD、NOR 和 B。表 2.1 描述了这五种语义角色标记的详细意义。其中, F 用于标记属于评价目标的词, 如果是词组形式的评价目标, 对应的单词连续地标注为 F; O 用于标记用户的情感词, 如果是词组形式的情感词, 对应的单词连续地标注为 O; MOD 表示用于修饰情感词的程度词; NOR 表示否定词; B 用于表示句子中属于其他情况的背景词。

表 2.1 语义角色标签符号统计

语义角色标签符号	意义
F	评价目标
O	情感词
MOD	修饰情感词的程度词
NOR	否定词
B	其他情况

2.6 基于图剪枝的评价目标分类

2.6.1 评价目标相似度计算

对评价目标集合按照产品特征/属性进行分类的关键问题在于如何有效地计算评价目标之间的关联度或相似度。计算词汇之间相似度的方法主要分为两类：基于语义知识库的方法和基于语料分布的方法。基于语义知识库的方法主要利用按照概念间的层次结构关系组织的语义词典，如 WordNet、thesauri 或 encyclopedias 等，根据概念之间的上下位关系和同位关系来计算词语的相似度。因此，此类方法主要用于计算词汇之间领域无关的词法和语义相似度。例如，“picture”和“image”作名词时具有很高的 WordNet 相似度。基于语料分布的方法假设当两个词汇具有相同的上下文环境时，它们之间存在某种程度的语义相关性^[113]。此类方法主要用于计算词汇之间基于在大规模语料中分布上下文的领域相似度。

在本节中，为了有效地计算评价目标之间的相似度，我们分别采用基于 WordNet 的相似度计算方法和基于分布上下文的相似度计算方法。针对评价目标的特征表示问题，我们提出一种基于依存句法关系的分布上下文抽取策略，从领域评论语料中抽取评价目标的上下文特征。同时，我们提出一种简单有效的基于图剪枝的评价目标分类策略，对评价目标集合按照产品特征/属性进行分类。

（1）基于 WordNet 的评价目标相似度计算方法

WordNet^[114]是一个由 Princeton 大学认知科学实验室开发的著名英语词汇库，广泛地应用于自然语言处理研究领域。它基于词汇的语义概念将名词、动词、形容词和副词分别组织成为庞大的语义关系网络。目前 WordNet 已被广泛用于计算两个词语之间的相似度^[115-118]，我们也将 WordNet 应用于计算评价目标之间的相似度。由于 WordNet 是根据词性组织词语之间的语义关系，我们计算两个评价目标中具有相同词

性的成员词之间基于 WordNet 的相似度,进而通过计算所有成员词相似度的平均值得到评价目标之间基于 WordNet 的相似度。实验中采用常用的 Jiang & Conrath 算法计算成员词之间基于 WordNet 的相似度^[118],计算公式如下:

$$Jcn(word_i, word_j) = \frac{1}{IC(word_i) + IC(word_j) - 2 * IC(Ics)} \quad (2.11)$$

其中, Ics 表示 $word_i$ 和 $word_j$ 之间的最小公共包含 (Least Common Subsumer), IC 表示信息量 (Information Content)。

(2) 基于依存句法的分布上下文抽取

分布上下文特征抽取是基于语料分布计算评价目标之间相似度的基础。目前的相关研究工作^[118]通常采用基于上下文窗口的方法,抽取评价目标出现时前后邻接窗口内的词作为评价目标的上下文特征。本节中我们提出一种基于依存句法关系的分布上下文特征抽取策略。与传统的上下文窗口方法只能捕获短距离的邻接窗口上下文相比,依存句法方法能够获取长距离的依存句法上下文。对于每一个评价目标 f ,从用户评论语料库中抽取包含该评价目标的评论句子集合。我们利用 Stanford 依存句法分析器对每一条用户评论句子进行分词和词性标注,抽取词语之间的依存关系并构建依存句法树。对于评价目标 f ,从构建的依存树中抽取 f 的主导词 (governor) 集合 (包含父结点和祖先结点等) 以及 f 的依赖词 (dependent) 集合 (包含孩子结点和子孙结点等)。这些抽取的词语与 f 之间都具有直接或传递的依存关系,因此将它们作为 f 的上下文特征。为了降低评价目标的上下文特征的稀疏性,并不区分依存关系的类型。根据这些抽取得到的词集合,构建评价目标的上下文特征向量。为了降低上下文特征空间的高维性,去除其中的常见停用词。基于依存句法的评价目标分布上下文抽取算法流程的详细描述如图 2.5 所示。

(3) 基于分布上下文的评价目标相似度计算方法

在抽取得到评价目标的分布上下文特征以后,进一步计算评价目标之间基于语料分布的相似度。与传统的采用 TF-IDF 公式计算上下文特征的权值,并采用余弦公式的相似度计算方法不同,本节中我们提出一种基于二部图的分布上下文相似度计算方法,能够更加直观地对评价目标与分布上下文特征之间的关联进行建模。我们将评价目标的上下文特征空间表示为二部图模型 $\langle A, E, B \rangle$,其中结点集 A 表示评价目标集合,结点集 B 表示上下文特征词集合,边集 E 表示评价目标与上下文特征词之间的联接边集合。如果评价目标 f 的分布上下文特征中包含上下文特征词 cw ,则存在从评价

目标结点 f 到上下文特征词结点 cw 的联接边。评价目标与分布上下文特征词之间的二部图模型如图 2.6 所示。

算法 2.1: 基于依存句法的分布上下文抽取算法

输入: 评价目标 f

包含 f 的评论句子集合 S

对于集合 S 中的每一个评论句子 $s \in S$:

- (1) 利用句法分析器对 s 进行分析, 抽取词语之间的依存句法关系集合;
- (2) 根据依存句法关系构建一棵依存句法树, 并寻找到根节点;
- (3) 对于评价目标表达 f 中的每一个词 t , 收集其父结点和祖先结点;
- (4) 对于评价目标表达 f 中的每一个词 t , 收集其孩子结点和子孙结点;
- (5) 将收集到的结点词构成一个上下文特征向量。

输出: f 的依存上下文特征向量

图 2.5 基于依存句法的分布上下文抽取算法流程

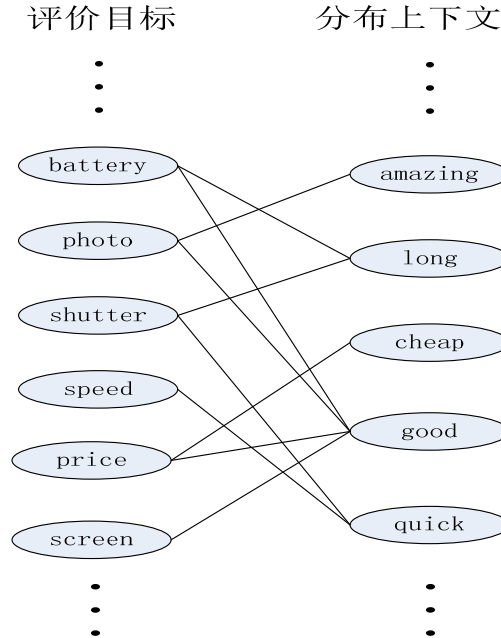


图 2.6 评价目标与分布上下文特征之间的二部关系图

如果集合 A 中有 m 个评价目标结点与 B 中的上下文特征词结点 cw 联接, 则赋予其中每条边的权值为 $1/m$ 。定义评价目标 f_i 和 f_j 之间基于分布上下文的相似度计算公

式如下：

$$Similarity_{context}(f_i, f_j) = \sum_{cw \in CooccurSet(f_i, f_j)} \frac{2}{OccurFreq(cw)} \quad (2.12)$$

其中， $CooccurSet(f_i, f_j)$ 表示同时在 f_i 和 f_j 的分布上下文特征向量中出现的特征词集合。 $OccurFreq(cw)$ 表示分布上下文特征向量中包含有特征词 cw 的评价目标的数量。

2.6.2 图剪枝分类算法

为了对评价目标集合按照用户评论的语义特征进行分类，在本节中我们提出一种简单而有效的基于图剪枝的评价目标分类策略。根据评价目标之间的相似度将评价目标集合表示为一个加权无向邻接图 $G = (V, E)$ ，其中 V 代表评价目标结点的集合，每一个结点表示一个评价目标； E 代表结点之间加权边的集合，每一个加权边表示评价目标之间的相似度。给定评价目标类别数 K ，我们重复查找和删除图 G 中权值最小的边，直到图 G 上连通子图的个数等于 K 。基于图剪枝的评价目标分类算法流程的详细描述如图 2.7 所示。

算法 2.2：基于图剪枝的评价目标分类算法

输入：评价目标集合 $Set_f = \{f_1, f_2, f_3, \dots, f_N\}$

评价目标类别数 K

- (1) 计算 Set_f 中评价目标元素之间的相似度，构建相似度矩阵，构建加权无向邻接图 G ，其中图中结点代表评价目标，图中边代表评价目标之间的相似度；
- (2) 计算图 G 中连通子图的个数 CB ，如果 CB 小于类别数 K ，转到步骤 (3)；否则转到步骤 (4)；
- (3) 寻找图 G 中权值最小的边并删除，转到步骤 (2)；
- (4) 将每一个连通子图对应为一个评价目标类别，由此得到评价目标表达集合 Set_f 的分类结果为 $AspectSet_f = \{Set_1, Set_2, \dots, Set_K\}$ ；
- (5) 返回 $AspectSet_f$ 并退出。

输出：评价目标分类结果 $AspectSet_f = \{Set_1, Set_2, \dots, Set_K\}$

图 2.7 基于图剪枝的评价目标分类算法流程

图 2.8 描述了在给定三种产品特征类别的情况下，采用图剪枝算法进行分类的示意图。如图 2.8 所示，其中 battery、photo/image、lens cap 三个评价目标结点之间的相

似度权值最小（子图 a），在删除权值最小的边之后，连通子图的个数等于产品特征类别数，每一个连通子图对应为一个产品特征类（子图 b）。

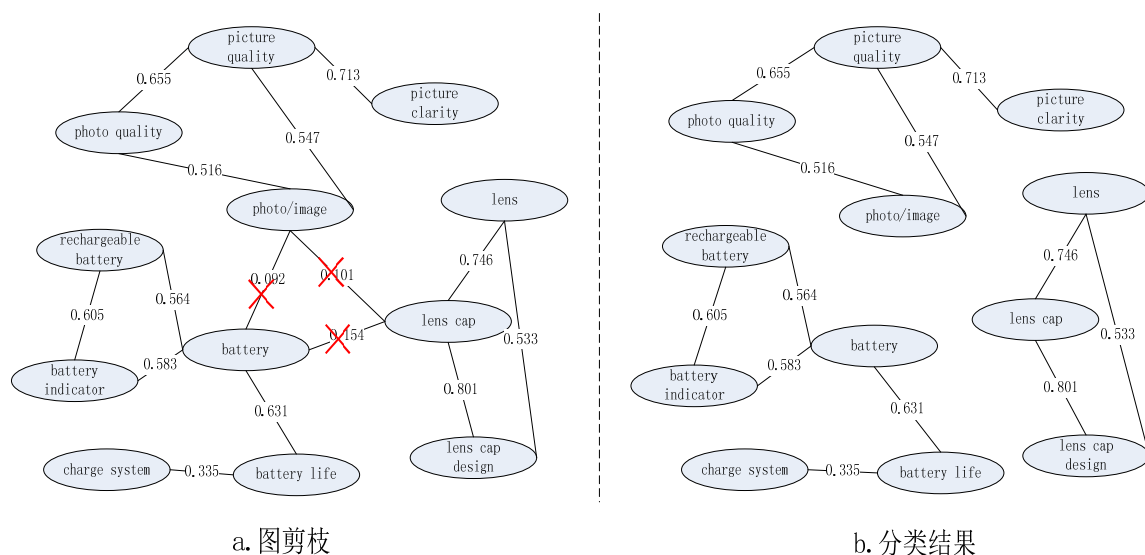


图 2.8 基于图剪枝的评价目标分类示例

2.7 实验准备和结果分析

2.7.1 数据集准备

实验部分采用数码相机领域的用户评论语料，主要包含 HL、YW1、YW2 和 YW3 四个数据集。其中，HL 是由 Hu 和 Liu 在文献^[54]中公开的产品评论数据集中关于数码相机 Canon G3 和 Nikon coolpix 4300 的评论文本构成，YW1、YW2 和 YW3 是从亚马逊网站⁴的数码相机评论页面抓取得到。我们对所有的数据集都预先进行句子切分、分词、词性标注和词干化处理，并对常见的编码和书写错误进行纠正，然后采用手工方式逐句地标注特征和语义角色。表 2.2 描述了实验数据集的详细统计。

表 2.2 关于数码相机的用户评论数据集的统计信息

数据集	用户评论数	评论句子数
HL	79	978
YW1	100	1098
YW2	163	974
YW3	182	1355

⁴ <http://www.amazon.com/>

2.7.2 评价目标和情感词抽取结果分析

2.7.2.1 抽取基线方法选择

为了评估评价目标和情感词的抽取效果，实验部分引入两种主流的细粒度意见元素挖掘算法作为基线方法：NVPs+LM 和 CRFs。本章提出的基于序列标注学习和句法语义结构特征的意见元素抽取方法标记为 CRFs+SSSI。

NVPs+LM: 抽取评价目标的一种朴素思想是以自然语言文本的词性标注为基础。文献^[84]通过对已标注的评论文本数据集进行分析，观察发现高达 98% 的产品特征都是以名词短语 (NPs) 和动词短语 (VPs) 的形式存在。因此，Wu 和 Zhang 等人^[84]首先选择所有的 NPs 和 VPs 作为候选产品特征，并基于领域评论语料训练语言模型去除候选集中的噪音词。我们引入该抽取算法作为一种基线对比方法。

CRFs: 为了验证本章引入的句法语义结构级别特征对提高评价目标和情感词抽取效果的贡献，我们引入文献^[59]中基于线性链 CRFs 模型的 Web 意见挖掘系统作为另一种基线对比方法。

实验采用准确率 (Precision)、召回率 (Recall) 和 F 值 (F-measure) 作为评价目标和情感词抽取结果的评价标准。准确率用于衡量抽取结果集中正确结果所占的比率，也称为查准率或正确率。召回率用于衡量抽取结果集中正确结果对标准结果的覆盖率，也称为查全率或覆盖率。F 值用于综合平衡准确率和召回率，计算一个综合反映整体性能的评价指标。准确率、召回率和 F 值的计算方法如下：

$\text{Precision} = \text{模型标注正确的实例数} / \text{模型标注的所有实例数}$

$\text{Recall} = \text{模型标注正确的实例数} / \text{人工标注的标准实例数}$

$\text{F-measure} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$

2.7.2.2 抽取结果对比分析

表 2.3 中给出了 CRFs+SSSI 模型在四个数据集上评价目标抽取的准确率、召回率和 F 值的实验结果。图 2.9 给出了在整体数据集上 CRFs+SSSI 方法与两种基线方法的评价目标抽取结果对比。图 2.10 给出了在整体数据集上 CRFs+SSSI 方法与基本 CRFs 模型的情感词抽取结果对比。

表 2.3 CRFs+SSSI 方法在四种数据集上的评价目标抽取结果分析

数据集	Precision	Recall	F-measure
HL	0.750	0.675	0.711
YW1	0.812	0.705	0.755
YW2	0.839	0.730	0.781
YW3	0.814	0.744	0.777
均值 (Average)	0.804	0.714	0.756

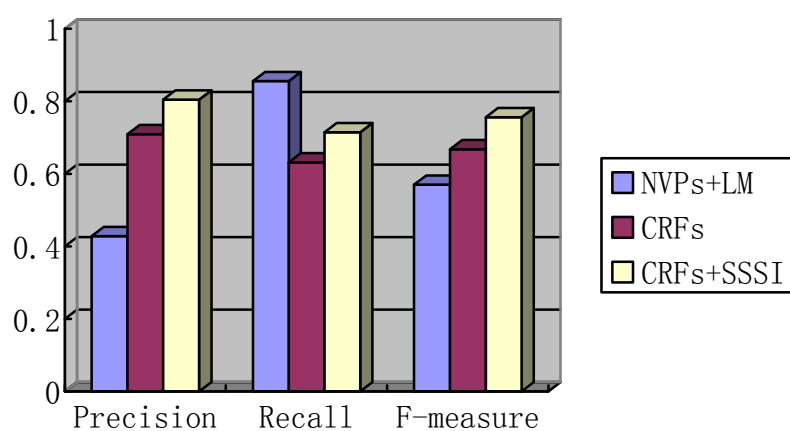


图 2.9 评价目标抽取的实验结果对比

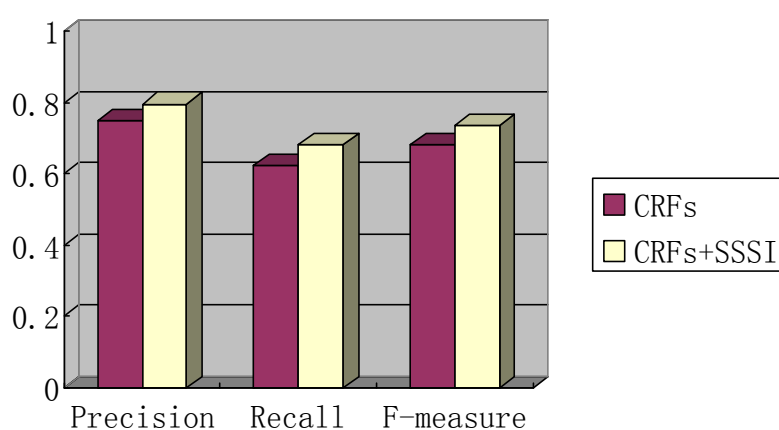


图 2.10 情感词抽取的实验结果对比

如表 2.3 所示, 我们可以分别比较在四个数据集上评价目标抽取结果的 Precision、Recall 和 F-measure。整体上本章提出的 CRFs+SSSI 方法在四个数据集上都取得了良

好的评价目标抽取结果。其中，在 YW1、YW2 和 YW3 三个数据集上取得了非常接近的抽取效果，相比在 HL 数据集上的实验效果要好一些。我们计算在四个数据集上的平均值，作为 CRFs+SSSI 方法在整体评论数据集上的实验效果。

如图 2.9 所示，与 NVPs+LM 和 CRFs 两种基线方法相比，本章提出的 CRFs+SSSI 方法取得了最优的评价目标抽取效果。从 Recall 来看，NVPs+LM 算法取得了最高的召回率值，明显优于 CRFs 和 CRFs+SSSI。然而从 Precision 和 F-measure 来看，与 CRFs 和 CRFs+SSSI 相比，NVPs+LM 算法的准确率和 F 值都达到最低。这也验证了大部分的评价目标都是以名词和动词短语形式存在的观察结论^[84, 119]。与基本 CRFs 模型相比，通过引入句法语义结构特征，能够有效地利用句子结构中长距离的依存关系，CRFs+SSSI 方法在 Precision、Recall 和 F-measure 上都取得了显著的提高。

如图 2.10 所示，与基本 CRFs 模型相比，本章提出的 CRFs+SSSI 方法在情感词抽取的 Precision、Recall 和 F-measure 三种评价标准上都取得了明显的提升。因此，本章提出的 CRFs+SSSI 方法能够取得良好的评价目标和情感词抽取效果。

2.7.3 评价目标分类结果分析

2.7.3.1 分类对比方法选择

通过对抽取得到的评价目标集合进行分析，可以发现许多评价目标都是包含有多个单词的词组形式，在这些评价目标之间通常存在一些共同词（Sharing Words）。例如，“shutter delay”与“shutter lag”之间存在共同的词 shutter，“picture quality”与“picture clarity”之间存在共同的词 picture，“price tag”与“price”之间存在共同的词 price 等。从语言学的角度来看，这些存在共同词的评价目标更有可能属于相同的产品特征类别^[87, 118]。因此，我们能够利用这种已知的共同词关系来增强评价目标之间的相似度计算。公式(2.13)用于计算评价目标之间基于共同词的相似度：

$$Similarity_{SharingWords}(f_i, f_j) = \frac{Num_{common}}{Num_{f_i} + Num_{f_j} - Num_{common}} \quad (2.13)$$

在本节中，我们分别比较基于 WordNet 和基于分布上下文的评价目标相似度计算方法的有效性。基于共同词的相似度量属于先验的语言学知识，在计算评价目标之间基于 WordNet 和基于分布上下文的相似度时，也加入了基于共同词的相似度量。最后，我们给出综合利用三种相似度量计算评价目标之间相似度的分类实验效果。

基于共同词和 WordNet 的相似度计算方法：利用基于共同词和基于 WordNet 的

相似度量计算评价目标之间的相似度。

基于共同词和分布上下文的相似度计算方法：利用基于共同词和基于分布上下文的相似度量计算评价目标之间的相似度。

综合利用三种相似度量的计算方法：综合利用基于共同词、基于 WordNet 和基于分布上下文的相似度量计算评价目标之间的相似度。

2.7.3.2 评价标准

实验部分采用熵值 (Entropy) 和纯度 (Purity) 作为评价目标分类结果的评价指标^[118, 120]。已知抽取得到的评价目标集合 DS 和类别数 K ，对 DS 的标准分类记作 $G = \{g_1, \dots, g_j, \dots, g_K\}$ ；基于图剪枝的评价目标分类算法将 DS 划分为 K 个独立的子集 $\{DS_1, \dots, DS_i, \dots, DS_K\}$ 。熵值和纯度分别定义如下：

熵值 (Entropy)：衡量分类结果与标准划分相比的混乱程度，Entropy 值越小，分类结果的混乱程度越低。对于每一个评价目标分类子集 DS_i ，公式(2.14)计算 DS_i 的 Entropy 值，其中 $P_i(g_j)$ 表示结果类 DS_i 中包含标准类 g_j 中元素的比例。公式(2.15)计算评价目标分类结果的整体 Entropy 值。

$$Entropy(DS_i) = \sum_{j=1}^K P_i(g_j) \log_2 P_i(g_j) \quad (2.14)$$

$$Entropy_{total} = \sum_{i=1}^K \frac{|DS_i|}{|DS|} Entropy(DS_i) \quad (2.15)$$

纯度 (Purity)：衡量分类结果与标准划分的一致性程度，Purity 值越大，分类结果的纯度越高。对于每一个评价目标分类子集 DS_i ，公式(2.16)计算 DS_i 的 Purity 值。公式(2.17)计算评价目标分类结果的整体 Purity 值。

$$Purity(DS_i) = \max_j P_i(g_j) \quad (2.16)$$

$$Purity_{total} = \sum_{i=1}^K \frac{|DS_i|}{|DS|} Purity(DS_i) \quad (2.17)$$

2.7.3.3 分类结果对比分析

表 2.4 和表 2.5 验证了我们提出的基于图分割的评价目标分类方法的有效性。其中，表 2.4 中给出了一些评价目标分类结果的示例，表 2.5 中显示了不同评价目标相似度计算策略的分类结果比较。

表 2.4 评价目标分类结果示例

产品特征 类别	相关评价目标词	产品特征 类别	相关评价目标词
shutter delay	speed, shutter, shutter button, shutter delay, lag, shutter lag, shutter speed, lag time	auto focus	auto focus, focus, focus, system, light focus, auto focus delay, focus range, auto focus assist light
camera strap	strap, neck strap, wrist strap, camera strap	lens cap	lens, lens cover, lens cap design, lens barrel, lens cap
battery	battery, battery duration, standard battery, battery indicator, battery charger, rechargeable battery, backup battery, battery life, battery charge system, charge system	photo/ image	photo, picture, photo quality, picture quality, picture clarity, image, flash photo, image format, raw image, photograph
price	price, price range, price point, cost, charge, expense	display	display, display panel, lcd, screen, lcd screen
after-sale service	service, customer service, after-sale service, online service	resolution	resolution, resolution setting, high resolution, resolution limit

如表 2.4 所示，评价目标能够比较准确地分类到对应的产品特征类别。例如，关于快门延迟方面的评价目标词 speed、shutter、shutter button、shutter delay、lag、shutter lag、shutter speed、lag time 等都能够准确地划分到 shutter delay 类别；关于照片质量方面的评价目标词 photo、picture、photo quality、picture quality、picture clarity、image、flash photo、image format、raw image、photograph 等都能够准确地分类到 photo/image 类别。

表 2.5 不同相似度计算方法的分类结果比较

相似度方法	Entropy	Purity
基于共同词和 WordNet 的相似度计算方法	1.426	0.866
基于共同词和分布上下文的相似度计算方法	2.815	0.809
综合利用三种相似度量的计算方法	1.354	0.830

如表 2.5 所示，当综合利用基于共同词、基于 WordNet 和基于分布上下文的相似

度量计算评价目标之间的相似度时, 评价目标分类结果的 Entropy 值最低, 分类结果的混乱程度达到最低, 同时也取得了较好的 Purity 值, 分类结果的纯度较好。当仅仅利用基于共同词和基于 WordNet 的相似度量计算评价目标之间的相似度时, 也能够取得比较满意的评价目标分类效果。因此, 基于共同词和基于 WordNet 的相似度量在评价目标之间相似度计算中起到主要的作用。

2.8 本章小结

本章中, 针对 Web 评论文本中的细粒度意见元素抽取问题, 我们提出一种基于序列标注学习和句法语义结构特征的意见元素抽取方法, 采用条件随机场概率模型将评价目标和情感词抽取转化为对词汇序列的自动标注。通过定义词汇、上下文和句法语义结构三种级别的标注特征, 解决了特征选择和有效特征稀疏等关键问题; 采用基于 CRFs 模型的序列标注学习框架, 解决了多种标注特征的有效融合问题。同时我们提出一种无监督的基于图分割的评价目标分类算法, 分别利用词汇知识库和分布上下文计算评价目标之间的领域相似度。

实验结果表明, 与基线方法相比, 通过引入句法语义结构级别的标注特征, 能够有效地利用句子结构中长距离的依存关系, 本章提出方法的评价目标和情感词抽取效果取得了明显的提升。当综合利用共同词、词汇知识库和分布上下文计算评价目标之间的相似度时, 基于图剪枝的分类算法取得了良好的评价目标分类效果。以上结果验证了基于序列标注学习和句法语义结构特征的意见元素抽取方法在针对 Web 评论文本的细粒度意见元素抽取和分类中的有效性。

第3章 基于约束的评价目标谱聚类

3.1 引言

评价目标聚类是用户评论意见挖掘研究中的核心任务之一，旨在对用户发表主观意见的评价目标按照特征或属性进行归类。评价目标通常包括评价对象的特征、属性和部件，以及部件的特征和属性等^[75, 91, 121]。然而，由于用户评论的随意性和表达习惯的不同，针对同一对象特征或属性，不同的评论者通常会采用不同的描述方式。例如在数码相机评论领域中，“photo”、“photograph”、“image”和“picture”等都用于描述照片特征。这些不同但相关的描述方式统称为领域同义词，它们之间的关联相似度依赖于具体的领域。例如，在电影评论领域中，“movie”和“picture”用于描述影片属性；而在数码相机评论领域中，“picture”和“photo”、“movie”和“video”分别用于描述照片和视频特征。图 3.1 描述了在数码相机评论领域中评价目标聚类的示意图。评价目标聚类已成为基于特征的意见摘要和推荐的基础。目前，随着评论文本规模不断增大，抽取的评价目标数量显著增长，采用手工方式对其进行归类费时费力。此外，评论领域多样性的不断增强，也使得研究快速有效的评价目标聚类算法成为一个挑战性的问题。

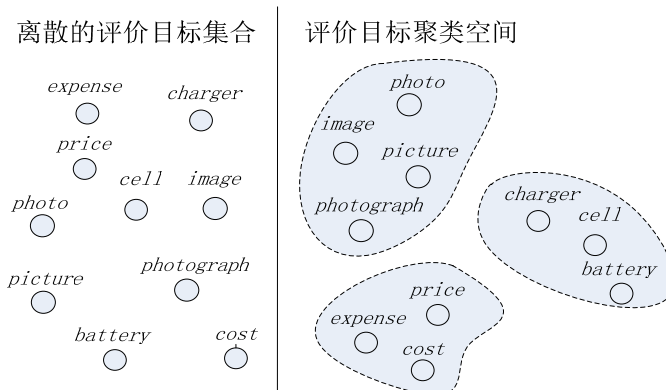


图 3.1 评价目标聚类示意图

目前的评价目标聚类研究主要分为基于无监督学习的方法和基于主题建模的方法。基于无监督学习的方法^[55, 57]通过计算评价目标之间的相似度，采用聚类或者分类算法对评价目标集合进行自动归类。其中如何有效计算评价目标之间基于领域的相似

度成为影响算法性能的关键。相似度计算主要分为基于词典或词汇知识库的方法^[55, 122]和基于领域语料库的方法^[112, 123]。基于词典的方法利用人工构建的词汇语义结构计算评价目标之间的相似度,但忽略了评价目标相似度的领域依赖性,同时也会产生覆盖率不足的问题。基于领域语料库的方法利用评价目标的上下文信息计算相似度,虽然在一定程度上缓解了评价目标相似度的领域依赖性问题,却面临着计算空间的高维度和稀疏性问题。基于主题建模的方法^[60, 124]通过在评论语料上的统计学习,训练主题生成模型,将生成的主题对应为评价目标类别。然而,这类方法生成的主题粒度较粗,并不能很好地对应细粒度的特征类。同时这些主题通常以单个词为单位,无法表示以短语形式存在的评价目标。

针对目前研究中存在的上述问题,本章提出一种基于约束的评价目标谱聚类方法。该方法通过定义和抽取评价目标之间的词法约束和上下文约束知识,增强了评价目标之间的领域关联性;利用约束传播算法,将局部评价目标之间的约束关系传播到全局评价目标空间;采用约束谱聚类算法,在融合先验约束知识的同时,有效降低了聚类空间的高维性和稀疏性。此外,本章还进一步论述了评价目标的上下文特征抽取策略,分析了邻接窗口上下文和依存句法上下文分别对评价目标聚类的影响。

3.2 相关研究

3.2.1 基于无监督学习的评价目标聚类

早期的评价目标聚类研究主要采用无监督学习方法。Liu 等人^[55]利用 WordNet 中词汇之间的语义关系,计算产品特征不同表达方式之间的近义词关系。Carenini 等人^[122]基于词之间的相似度量,将抽取的评价目标映射到已知的领域评价特征分类体系。Guo 等人^[97]提出了一种基于多层隐含语义关联的无监督算法,第一层隐含语义关联对评价目标的上下文环境建模,第二层隐含语义关联利用潜在语义结构和上下文片段对评价目标聚类。Su 等人^[57]提出了一种相互增强的聚类策略,该策略通过融合评价目标和情感词的内容相似度及两者之间的情感关联相似度,同步迭代地对评价目标和情感词进行聚类。Santosh 等人^[125]构建候选产品特征之间的关联图,利用小世界模型和图聚类算法对图结点按属性划分。

3.2.2 基于主题模型的评价目标聚类

近年来,主题模型也被广泛地应用于意见挖掘研究领域。Titov 等人^[124]提出了一

种多粒度主题模型，从评论文本中抽取产品特征并按照主题进行归类。Brody 等人^[60]基于 LDA 模型从评论句中抽取评价对象及对应的情感词，通过与种子词之间的情感传播判定用户表达的情感倾向，同时考虑了评价属性对情感倾向的影响。Zhao 等人^[81]提出了一种 MaxEn-LDA 混合模型，能够同时抽取评价特征和对应情感词。然而，这些方法主要基于对评论语料的统计学习，却忽略了局部上下文特征。

3.2.3 基于半监督学习的评价目标聚类

最近，一些研究者也尝试着在评价目标聚类过程中引入各种先验知识。Zhai 等人^[87]拓展了一种基于约束的 LDA 模型，并提出两种自动抽取产品特征之间先验约束关系的方法，能够处理大规模的产品特征之间的约束信息。Mukherjee 等人^[126]提出了一种半监督用户引导的产品特征抽取和聚类策略，引入了两种相关主题模型。Zhai 等人^[118, 127]将产品特征聚类转化为半监督学习问题，利用产品特征之间的词法特征自动标注部分训练数据。在本章中，我们针对评价目标聚类的领域依赖性和语义关联信息不足问题，通过构建评价目标之间的先验约束知识，采用基于约束的谱聚类策略来解决评价目标聚类问题。

3.3 基于约束关系的谱聚类模型

本节通过定义和抽取评价目标之间的先验约束关系，将评价目标聚类定义为一种基于约束的半监督学习过程。假设已知评价目标及对应情感词的抽取结果，基于约束的评价目标谱聚类算法包括三个步骤：约束关系定义和抽取、约束传播和约束谱聚类，分别论述如下。

3.3.1 评价目标约束关系的定义和抽取

定义和抽取评价目标之间的约束关系，能够为聚类过程提供先验的指导知识。通过对评价目标的词法结构和在评论语料中的上下文信息进行分析，我们定义评价目标之间的两类约束关系：词法约束关系和上下文约束关系。按照约束方向划分，词法约束和上下文约束又可以分为正向约束关系和反向约束关系。正向约束关系定义了评价目标之间属于相同特征类的先验关系及置信程度，反向约束关系定义了评价目标之间属于不同特征类的先验关系及置信程度。

词法约束关系对评价目标之间的词法关系进行建模。通过观察评价目标集合可以发现，大多数评价目标都是名词及名词短语。这些名词和名词短语之间通常存在着共

同词,例如“wide angle *lens*”与“*lens*”、“*picture* quality”与“*picture* clarity”、“*price* tag”与“*price*”等。从词法关系角度分析,这些包含共同词的评价目标更有可能属于相同的特征类^[87, 118]。因此,我们将评价目标之间包含共同词的词法关系定义为正向词法约束关系。

正向词法约束从词法关系角度衡量了评价目标之间属于相同特征类的置信程度。我们主要利用词法分析技术抽取评价目标之间的词法约束关系。在去除代词和常用停用词之后,如果两个评价目标之间存在共同的名词或者名词短语,则定义两者之间存在正向词法约束关系。

上下文约束关系对评价目标之间的上下文关联关系进行建模。我们主要通过通过对用户评论习惯和评论文本上下文结构的分析,抽取评价目标之间的反向上下文约束关系。反向上下文约束关系从上下文关联角度衡量了评价目标之间属于不同特征类的置信程度。通过观察评论文本可以发现,评价目标之间存在着句子级别的共现关系,且在同一评论文本单元中,描述同一特征类的评价目标之间具有相同的用户情感倾向。定义评价目标之间反向上下文约束关系的抽取规则如下:

(1) 在同一个评论语句中共现且没有“and”连接的各个评价目标,更倾向于隶属不同的特征类^[87]。例如在数码相机评论领域中,已知评论句子“*Optics* are top notch, *macro mode* is incredible, great *video recording* is possible as well.”,抽取得到的评价目标分别为“optics”、“macro mode”和“video recording”。这些评价目标分别属于不同的特征类,因为人们通常不会在同一个句子中针对同一个对象特征进行重复评价^[87]。因此我们利用这种评价目标之间存在的评论句子级别共现关系,抽取评价目标之间的反向上下文约束关系。连词“and”经常用于连接属于同一主题的评价目标,因此将出现“and”连接的情况排除在外。

(2) 在同一评论文本单元中,针对描述同一特征类的评价目标,用户保持相同的评价意见倾向。因此,在同一评论文本单元中,具有相反用户意见倾向的评价目标之间属于不同的特征类,因为评论者通常不会在同一评论文本单元中,针对同一个对象特征发表前后情感倾向相反的评价意见。因此我们利用这种属于同一特征类的评价目标之间在评论文本单元中存在的情感倾向一致性关系,抽取评价目标之间的反向上下文约束关系。

由于评论文本的不规范性,抽取的约束关系集合中可能存在着两个评价目标之间同时具有正向约束和反向约束关系的情况。当正向和反向约束关系发生冲突时,为了

不引入偏差数据，可以考虑同时丢弃这两种约束关系。

图 3.2 给出了评价目标空间与词法约束关系空间、上下文约束关系空间的对应关系示意图。词法约束空间和上下文约束空间分别包含评价目标之间的正向词法约束和反向上下文约束关系，评价目标空间中包含了所有词法和上下文约束关系。

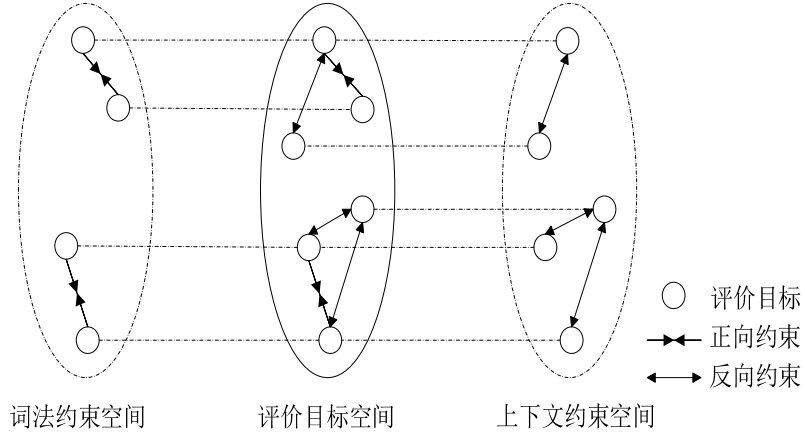


图 3.2 评价目标空间与词法和上下文约束空间的对应关系示意图

3.3.2 约束传播算法

已知评价目标集合 $X = \{x_1, \dots, x_N\}$ ，其中 N 表示评价目标数量。在完成评价目标之间约束关系抽取以后，定义正向约束关系集合 $D = \{(x_i, x_j) : z_i = z_j\}$ 和反向约束关系集合 $R = \{(x_i, x_j) : z_i \neq z_j\}$ ，其中 z_i 表示评价目标 x_i 的类别标签。为了直观地表达评价目标集合上存在的约束关系，定义约束矩阵 $Z_{N \times N}$ ，其中元素 z_{ij} 代表评价目标 x_i 和 x_j 之间的约束关系， $|z_{ij}|$ 表示约束关系的置信程度。初步地定义 z_{ij} 如下：

$$z_{ij} = \begin{cases} 1, & (x_i, x_j) \in D \\ -1, & (x_i, x_j) \in R \\ 0, & \text{其他情况} \end{cases} \quad (3.1)$$

约束矩阵 Z 包含了评价目标集合上的所有初始约束关系。尽管能够直接利用 Z 增强受约束评价目标之间的领域相似度，但 Z 中仅包含局部评价目标之间的约束知识。因此本章中，我们采用约束传播算法，将 Z 中局部评价目标之间的约束关系在评价目标空间 X 上进行全局传播。约束传播算法包含基于列的纵向传播和基于行的横向传播两阶段传播过程。

定义约束矩阵的赋值空间 $\tilde{E} = \{E = \{E_{ij}\}_{N \times N} : |E_{ij}| \leq 1\}$ ，其中 E 代表评价目标之间的约束矩阵状态， $|E|$ 代表约束关系的置信程度。 $E_{ij} \geq 0$ 表示评价目标 x_i 和 x_j 之间存在正向约

束关系 $(x_i, x_j) \in D$, $E_{ij} < 0$ 表示 x_i 和 x_j 之间存在反向约束 $(x_i, x_j) \in R$, $|E_{ij}|$ 代表 x_i 和 x_j 之间约束关系的置信程度。已知评价目标之间的相似度矩阵 $A_{N \times N}$, 其中 $A(x_i, x_j)$ 表示 x_i 和 x_j 之间的相似度值, $A(x_i, *)$ 表示 x_i 和相邻评价目标结点之间的相似度值之和。约束传播算法的详细描述如下:

(1) 基于相似度矩阵 A , 构建评价目标之间的权重矩阵 $W_{N \times N}$, 如公式(3.2)所示。令 $W = (W + W^T)/2$, 使得 W 成为对称矩阵:

$$W_{ij} = \begin{cases} \frac{A(x_i, x_j)}{\sqrt{A(x_i, *)} \sqrt{A(x_j, *)}}, & A(x_i, x_j) \geq 0 \\ 0, & \text{其他情况} \end{cases} \quad (3.2)$$

(2) 构建矩阵 $L = D^{-1/2} W D^{-1/2}$, 其中 $D_{N \times N}$ 为对角阵, 其第 (i, i) 元素等于 W 中第 i 行元素之和。

(3) 如公式(3.3)所示, 迭代执行纵向约束传播直至收敛, 其中 $E_v(t) \in \tilde{E}$, α 是约束传播速率的控制参数, 在 $(0, 1)$ 区间内取值。在实验过程中, α 的取值为 0.5。

$$E_v(t+1) = \alpha L E_v(t) + (1 - \alpha) Z \quad (3.3)$$

(4) 如公式(3.4)所示, 迭代执行横向约束传播直至收敛, 其中 $E_h(t) \in \tilde{E}$, E_v^* 是 $E_v(t)$ 的迭代收敛结果。

$$E_h(t+1) = \alpha E_h(t) L + (1 - \alpha) E_v^* \quad (3.4)$$

(5) 输出 $E^* = E_h^*$ 作为约束关系传播的最终收敛结果, 其中 E_h^* 是 $E_h(t)$ 的迭代收敛结果。

在步骤(1)中, 评价目标之间的权重矩阵 W 是从相似度矩阵 A 计算而来。在步骤(2)中, 构建矩阵 L 来对 W 进行对称归一化, 这是保证接下来纵向和横向约束传播能够收敛所必须的^[146]。步骤(3)和(4)分别执行纵向和横向约束传播过程, 分别利用矩阵 L 将局部约束关系沿着约束矩阵 Z 的列方向和行方向传播。文献^[128]证明了上述约束传播算法具有良好的计算效率和收敛特性。

3.3.3 约束谱聚类算法

通过约束传播算法, 将局部评价目标之间的约束关系传播到全局评价目标空间, 从而增强全局评价目标之间的领域相似度。在约束传播以后, 得到全体评价目标集合 X 上的约束矩阵 E^* , $|E^*|$ 代表约束关系的置信程度。基于约束的评价目标聚类的目标是得到满足 E^* 的聚类结果。基于约束矩阵 E^* 和初始权重矩阵 W , 定义约束权重矩阵

$\tilde{W}_{N \times N}$ 如下:

$$\tilde{W}_{ij} = \begin{cases} 1 - (1 - E_{ij}^*)(1 - W_{ij}), & E_{ij}^* \geq 0 \\ (1 + E_{ij}^*)W_{ij}, & E_{ij}^* < 0 \end{cases} \quad (3.5)$$

约束权重矩阵 \tilde{W} 融合了约束矩阵 E^* 和初始权重矩阵 W 。给定聚类的目标类别数 C ，约束谱聚类算法的详细描述如下:

(1) 构建矩阵 $Q = \tilde{D}^{-1/2} \tilde{W} \tilde{D}^{-1/2}$ ，其中 \tilde{D} 为对角阵，其 (i, i) 元素等于 \tilde{W} 中第 i 行元素之和。对 Q 进行特征值分解，取前 K 个最大的特征向量 v_1, \dots, v_K 。

(2) 构建矩阵 $F = [v_1, \dots, v_K]$ ，对行向量执行正则化操作，使其成为单位向量。其中第 i 行向量 $F_{i,*}$ 是评价目标 x_i 的低维表示向量。

(3) 对特征向量集合 $\{F_{i,*} : i = 1, \dots, N\}$ 执行 K-means 聚类，将评价目标集合 X 聚类为 C 个特征类别。

约束谱聚类算法在对评价目标集合进行聚类之前，通过特征值分解将评价目标映射到低维的特征向量空间， K 为特征向量空间的维数，从而降低了原始特征空间的高维性和数据稀疏性对评价目标聚类的影响。在实验过程中， K 的经验值取矩阵 Q 特征值分解后大于零的特征值数目。

3.4 上下文特征抽取和表示

针对评价目标之间相似度的领域依赖性问题，本章采用基于领域评论语料的评价目标相似度计算方法，从而不需要依赖于语义词典和词汇知识库等外部资源。基于评价目标在领域评论语料中的分布，本节介绍评价目标的上下文特征抽取策略、上下文特征空间表示和相似度计算函数。

针对评价目标的特征选择和表示问题，主要抽取两种类型的上下文特征：邻接窗口上下文和依存句法上下文。

邻接窗口上下文抽取左右邻接窗口内的词语作为上下文特征。设定邻接窗口大小为 T ，针对每一个包含评价目标 x_i 的评论语句，抽取 x_i 邻接的前 T 个词和后 T 个词作为邻接窗口上下文特征词，并对特征词的出现次数进行累加。在实验过程中， T 的取值为 5。

邻接窗口只能抽取评价目标前后有限窗口范围内的上下文，却忽略了长距离的依存句法特征。对于句法结构复杂的评论语句，邻接窗口上下文并不能有效地表示评价目标，而依存句法上下文能够抽取长距离的依存句法特征。因此，我们利用依存句法

分析器，对每一个包含评价目标的评论语句进行依存分析，生成依存句法树，抽取评价目标的上位词和下位词，作为依存句法上下文特征。依存句法上下文特征的抽取算法的详细描述如第 2 章 2.6.1 节中图 2.5 所示。

采用向量空间模型表示评价目标的上下文特征空间，每一个评价目标的上下文特征词集合都被转化为一个特征表示向量，使用 TF-IDF 公式计算其中每一维特征的权值，如公式(3.6)所示。为了降低上下文特征空间的维度，去除常见停用词。

$$TF-IDF = TF(d, t) * IDF(t) = \frac{d_t}{\sum d_i} * \log \frac{N}{N_t} \quad (3.6)$$

在计算得到评价目标对应的特征表示向量之后，采用余弦公式计算评价目标之间的相似度值，如公式(3.7)所示。

$$Cosine(v_i, v_j) = \frac{v_i \cdot v_j}{|v_i| \cdot |v_j|} \quad (3.7)$$

其中， v_i 和 v_j 分别表示评价目标 x_i 和 x_j 对应的特征表示向量。

3.5 实验验证和结果讨论

3.5.1 数据准备和评价标准

验证数据集包含三种产品领域的评论数据：数码相机(Digital camera)、手机(Cell phone)和吸尘器(Vacuum cleaner)。我们首先从 Amazon 网站上分别抓取三种产品的用户评论网页，抽取网页中的评论内容，每一条评论构成一个评论文本单元；然后利用现有的细粒度意见挖掘系统^[112]，抽取和标注用户关注的评价目标及对应的情感词，并采用人工方式对抽取结果进行检查和清理；最后，人工对得到的用户评价目标集合按照产品特征或属性进行归类，作为聚类任务的标准结果。表 3.1 给出了验证数据集及评价目标类别的统计情况。

表 3.1 验证数据集和评价目标类别统计

领域	Digital Camera	Cell Phone	Vacuum Cleaner
评论文本	524	204	856
评论语句	4406	3937	9878
评价目标	691	734	771
评价目标类别	35	40	38

实验部分采用 Entropy 和 Purity^[118, 120]作为聚类效果的评价标准。Entropy 衡量了实验结果与标准结果相比的混乱程度，Purity 衡量了实验结果与标准结果的一致性程度。Entropy 和 Purity 的详细定义和计算公式如第 2 章 2.7.3.2 小节所示。

3.5.2 基线方法

为了评估本章提出的约束谱聚类算法解决评价目标聚类问题的有效性，引入三种主流的聚类策略作为基线方法：

(1) K-均值算法 (K-means)。K-均值是传统聚类算法的典型代表，旨在将每一个数据点划分到均值距离最小的类别。为了验证约束谱聚类算法与传统聚类算法相比在评价目标聚类任务上的有效性，引入 K-均值算法作为一种基线方法。

(2) 谱聚类算法 (Spectral Clustering, SC)。谱聚类算法利用相似度矩阵上的谱分解理论，在执行聚类操作之前将数据表示映射到一个新的低维空间。已知谱聚类是一种快速有效的聚类算法，为了验证本章中定义和抽取的评价目标之间的先验约束关系对聚类效果的影响，引入朴素谱聚类算法作为一种基线方法。

(3) 基于自动标注的半监督 EM 分类算法 (L-EM)。L-EM^[118, 127]首先利用 WordNet 计算产品特征的词法相似度，自动构建少量的标注数据，然后训练 EM 分类算法对产品特征进行迭代分类。Zhai 等人^[118]证明了 L-EM 在产品特征聚类效果上优于传统聚类和主题建模的方法，因此引入 L-EM 作为一种基线方法。

3.5.3 实验结果

3.5.3.1 约束关系抽取结果分析

首先给出评价目标之间约束关系的抽取结果分析。本章中主要抽取了正向词法约束关系和反向上下文约束关系类型，作为评价目标聚类的先验知识。我们分别统计具有这两种约束关系的评价目标在全体评价目标集合中所占的比例。表 3.2 给出了在三种产品领域上的约束关系抽取结果统计。

表 3.2 评价目标之间约束关系的抽取结果统计

领域	Digital Camera	Cell Phone	Vacuum Cleaner
正向词法约束关系	9.87%	12.89%	11.35%
反向上下文约束关系	8.57%	6.08%	15.19%
两种约束关系之和	18.44%	18.97%	26.54%

分析表 3.2 中的结果可以发现，由于词法约束关系主要通过评价目标之间的词法

分析抽取，具有正向词法约束关系的评价目标比例并不直接受到领域评论数据集规模的影响；而上下文约束关系主要通过评价目标之间存在的句子级别和评论文本单元级别的上下文关系抽取，具有反向上下文约束关系的评价目标比例则直接依赖于领域评论数据集的规模。如表 3.2 所示，由于 Cell phone 领域的评论数据集规模较小，抽取的反向上下文约束关系的比例相对较低；而 Vacuum cleaner 领域的评论数据集规模较大，抽取的反向上下文约束关系也达到最高的比例。

3.5.3.2 评价目标聚类结果分析

为了验证本章提出的基于约束的评价目标谱聚类的有效性，我们在真实的验证数据集上与基线方法进行比较。表 3.3 和 3.4 分别给出了不同聚类算法在三种产品领域评论数据集上 Entropy 和 Purity 值的比较，其中 CSC 代表本章提出的基于约束的谱聚类方法。为了取得最好的聚类效果，我们同时利用评价目标之间的正向词法约束关系和反向上下文约束关系，以及邻接窗口上下文特征和依存句法上下文特征。

表 3.3 不同聚类算法的 Entropy 值比较

领域	Digital Camera	Cell Phone	Vacuum Cleaner
K-means	1.648	1.428	1.823
SC	1.875	1.406	1.834
L-EM	1.611	1.387	1.719
CSC	1.574	1.321	1.603

如表 3.3 所示，与 K-means、SC 和 L-EM 算法相比，CSC 算法在三种产品领域上都取得了最小的 Entropy 值，聚类结果的混乱程度达到最低。其中，K-means 和 SC 算法在三种产品领域上分别取得了最大的 Entropy 值，聚类结果的混乱程度达到最高；相比较而言，L-EM 和 CSC 算法在三种产品领域上取得了较低的 Entropy 值。CSC 算法在 Entropy 值上明显地低于 K-means 和 SC。与 SC 算法相比，CSC 在 Digital Camera 和 Vacuum Cleaner 领域上分别降低了 0.301 和 0.231，而在 Cell Phone 领域上降低了 0.085；与 L-EM 算法相比，CSC 算法在 Digital Camera 和 Cell Phone 领域上的 Entropy 值分别降低了 0.037 和 0.066，而在 Vacuum Cleaner 领域中明显地降低了 0.116。

表 3.4 不同聚类算法的 Purity 值比较

领域	Digital Camera	Cell Phone	Vacuum Cleaner
K-means	0.394	0.497	0.390
SC	0.312	0.454	0.411
L-EM	0.403	0.471	0.435
CSC	0.457	0.528	0.514

如表 3.4 所示, 与 K-means、SC 和 L-EM 算法相比, CSC 算法在三种产品领域上都取得了最大的 Purity 值, 聚类结果的纯度达到最高。其中, SC 算法在 Digital Camera 和 Cell Phone 领域上取得了最小的 Purity 值, 而 K-means 算法在 Vacuum Cleaner 领域上取得了最小的 Purity 值。L-EM 与 K-means 算法在聚类结果的 Purity 值上表现相当, L-EM 在 Digital camera 和 Vacuum cleaner 领域上的 Purity 值大于 K-means, 而在 Cell phone 领域上的 Purity 值则小于 K-means。与 SC 算法相比, L-EM 算法在三种产品评论领域上的 Purity 值都有不同程度的提高。CSC 算法在 Purity 值上明显地高于 K-means、SC 和 L-EM 算法, 与 L-EM 相比, CSC 算法在三种产品领域上的 Purity 值都提升了 5%以上。

综上所述, 针对评价目标聚类任务, 基于约束的谱聚类算法取得了最优的实验结果, 混乱程度达到最低, 纯度达到最高。其中, 由于数据的稀疏性和缺乏先验的领域知识, 传统的 K-means 和 SC 算法的聚类效果较差。CSC 算法的聚类效果明显优于 K-means 和 SC, 验证了其与传统聚类算法相比的有效性, 以及引入评价目标之间的先验约束关系对于提高聚类效果的贡献; 与最新的 L-EM 算法相比, CSC 算法在三种产品领域上的聚类效果都取得了明显的提升, 验证了本章提出的基于约束的谱聚类方法解决评价目标聚类问题的有效性。

3.5.3.3 约束关系类型影响分析

如上所述, 引入评价目标之间的词法约束关系和上下文约束关系有效提高了评价目标聚类的效果, 我们进一步分析这两种约束关系类型分别对提高聚类效果的贡献。分别利用评价目标之间的正向词法约束关系、反向上下文约束关系以及两种类型约束关系之和试验 CSC 算法的聚类效果。表 3.5 和 3.6 分别给出了 CSC 算法在三种产品领域上 Entropy 和 Purity 值的比较。在实验过程中, 我们同时利用评价目标的邻接窗口和依存句法上下文特征。

表 3.5 不同约束关系类型对 CSC 算法 Entropy 值的影响

领域	Digital Camera	Cell Phone	Vacuum Cleaner
正向词法约束关系	1.658	1.362	1.621
反向上下文约束关系	1.672	1.549	1.614
两种类型约束关系之和	1.574	1.321	1.603

表 3.6 不同约束关系类型对 CSC 算法 Purity 值的影响

领域	Digital Camera	Cell Phone	Vacuum Cleaner
正向词法约束关系	0.429	0.497	0.486
反向上下文约束关系	0.403	0.482	0.507
两种类型约束关系之和	0.457	0.528	0.514

由表 3.5 和 3.6 所示，当同时利用评价目标之间的正向词法约束关系和反向上下文约束关系时，CSC 算法在三种产品领域上都取得了最优的聚类效果，混乱程度达到最低，纯度达到最高，明显优于仅仅采用一种约束关系类型时的聚类结果。在 Digital Camera 和 Cell Phone 领域中具有正向词法约束关系的评价目标比例大于具有反向上下文约束关系的评价目标比例，基于词法约束关系的聚类结果优于（Entropy 和 Purity 值分别低于和高于）基于上下文约束关系时的结果；而在 Vacuum Cleaner 领域中则刚好相反，基于上下文约束关系的聚类结果优于基于词法约束关系的结果。因此，CSC 算法对评价目标聚类的效果受到具有约束关系的评价目标比例的影响，比例越大，聚类的效果越好。当同时利用词法约束关系和上下文约束关系时，具有约束关系的评价目标所占的比例最大，聚类效果也达到最优。

3.5.3.4 上下文特征抽取影响分析

为了验证上下文特征抽取对评价目标聚类的影响，我们进一步分析 3.4 节中提出的邻接窗口上下文和依存句法上下文分别对 CSC 算法的影响。分别基于邻接窗口上下文、依存句法上下文以及两种类型上下文之和试验 CSC 算法的聚类效果。表 3.7 和 3.8 分别给出了 CSC 算法在三种产品领域上的 Entropy 和 Purity 值的比较。实验中，我们同时利用评价目标之间的正向词法约束关系和反向上下文约束关系。

表 3.7 不同上下文类型对 CSC 算法 Entropy 值的影响

领域	Digital Camera	Cell Phone	Vacuum Cleaner
邻接窗口上下文	1.637	1.378	1.703
依存句法上下文	1.694	1.421	1.774
两种类型上下文之和	1.574	1.321	1.603

表 3.8 不同上下文类型对 CSC 算法 Purity 值的影响

领域	Digital Camera	Cell Phone	Vacuum Cleaner
邻接窗口上下文	0.443	0.491	0.421
依存句法上下文	0.415	0.469	0.408
两种类型上下文之和	0.457	0.528	0.514

如表 3.7 和 3.8 所示，当同时利用评价目标的邻接窗口上下文和依存句法上下文特征时，CSC 算法在三种产品领域上都取得了最优的聚类效果，混乱程度达到最低，纯度达到最高，明显优于仅仅采用一种上下文类型时的聚类结果。其中基于邻接窗口上下文的聚类结果都要优于（Entropy 和 Purity 值分别大于和小于）基于依存句法上下文的聚类结果，这是因为依存句法上下文主要捕获长距离的依存特征，由此产生的特征空间较为稀疏。通过融合邻接窗口上下文和依存句法上下文，能够增强评价目标的特征向量，降低特征空间的数据稀疏性。

3.6 本章小结

本章提出了一种基于约束的谱聚类方法，将评价目标聚类问题转化为基于约束的半监督学习问题。通过对评价目标自身词法结构、用户评论习惯以及评论文本局部上下文进行分析，定义和抽取评价目标之间的先验约束关系，并采用约束传播将局部约束关系传播到全局评价目标空间，从而在相似度度量的基础上进一步增强了评价目标之间的领域关联性，能够有效地解决评价目标之间相似度的领域依赖性问题。采用约束谱聚类算法，在聚类操作之前将评价目标映射到低维特征向量空间，从而降低了聚类特征空间的高维性和稀疏性。同时，本章采用基于领域评论语料的相似度计算方法，减少了对 WordNet 等外部词典和词汇知识库的依赖；采用基于邻接窗口和依存句法相结合的上下文特征抽取策略，既能够抽取邻接窗口上下文特征，也能够有效地抽取长距离的依存句法上下文特征。

实验结果表明, 基于约束的谱聚类方法能够有效地解决评价目标聚类问题, 取得了比基线方法更好的实验效果。与 K-均值算法的比较, 验证了本章提出的方法优于传统聚类方法; 与谱聚类 and 半监督 EM 分类算法的比较, 验证了通过引入评价目标之间的先验约束知识, 基于约束的谱聚类方法有效地提高了评价目标聚类的效果。

第4章 基于约束标签传播的领域情感词典自动构建

4.1 引言

对文本进行情感挖掘，涉及到词汇、短语/结构、句子、段落和篇章等各个语言粒度。其中词汇和短语是构成文本的基本单位，文本中蕴含的主观情感也是以词语作为基本的情感表达单元。词汇和短语也是人类表达情感、态度、意见的基本元素。比如“wonderful”、“excellent”、“ugly”、“lazy”等，这类带有明显情感倾向性的词语称为情感词或者极性词。

情感词典是自动化情感分析研究的重要基础^[23, 79, 129-132]，其中包含了一系列的情感词/短语及其对应的先验情感倾向。然而，文本中的情感倾向并不完全依赖于包含的情感词的情感倾向。比如，正面情感词可以出现在负面评论文本中，负面情感词也可以出现在正面评论文本中。同时针对不同的应用领域，情感词的情感倾向并不是固定不变的，而是依赖于出现的具体领域和上下文语境。比如，在独立语境下具有情感倾向的词语可能在某一具体语境中表现为中立词，正面的情感词可能在某一具体语境中表达负面情感，负面的情感词可能在某一具体语境中表达正面情感。例如，“big”在用于形容汽车内部空间时，表达正面的情感倾向；而在用于表示手机电池体积时，则表达负面的情感倾向。总之，由于情感词的语境和领域依赖性，不可能构造一个适用于所有领域的完备情感词典。此外，随着网络用户的参与度不断提高，用户感兴趣的主题和领域日益增多，通过领域专家采用手工方式针对每一个评论领域构建特定的情感词典变得更加不切实际。因此针对目标领域，研究自动/半自动的情感词典构建方法已成为细粒度意见挖掘的基础任务之一。

根据定义，领域情感词典通常包含一系列领域特定的主观性词和短语，以及它们在目标领域中具有的情感倾向性。目前情感词典构建相关的工作主要划分为基于词汇知识库的规则方法和基于大规模语料库的统计方法。基于词汇知识库的规则方法^[28, 29, 54, 133]利用已知词汇知识库中存在的词条之间的语义关系及注释信息判定词语的情感极性；基于语料库的统计方法^[85, 134-136]基于表达相同倾向性的情感词之间彼此共现的假设，利用大规模领域评论语料中的统计共现信息计算词语的情感极性。通常，首先由用户指定少数的通用情感词作为种子词，分别赋予正面或负面的倾向性，进而

根据词语与种子词之间的语义关联强度计算候选词的情感倾向，构建情感词典。然而，目前的这些研究方法的不足之处在于：基于词汇知识库的方法通常完全依赖于先验的语义词典，无法适用于缺少这类先验词典的语言环境，同时忽略了情感词典具有的强烈领域依赖性。而基于语料库的统计方法虽然具有良好的领域依赖性，却忽略了在上下文和语义层面上存在的先验约束知识。

本章针对领域情感词典的自动构建问题开展研究工作，以如何定义和抽取领域依赖的先验知识，并将其融入统一的学习框架，实现领域情感词典构建为研究目标。本章的研究工作基于以下两个直观的假设：（1）按照情感词的领域依赖性划分，领域情感词典可以分为领域依赖情感词部分和领域独立情感词部分。领域独立情感词能够适用于大多数常见的评论领域，其情感极性和强度始终保持固定不变，如“good”、“great”、“bad”和“terrible”等；而领域依赖情感词的情感极性和强度依赖于具体的评论领域，当适用于其他评论领域时，需要根据上下文语境重新判定其情感极性和强度。如图 4.1 所示，对于不同的评论领域 A 和 B，重叠的阴影部分表示 A 和 B 共同的领域独立情感词部分，其他部分分别表示领域 A 和 B 特定的情感词部分。因此，我们可以选取少量的领域独立情感词作为种子词，通过构造具有领域适应性且高效的算法和模型，将种子情感词的情感信息传递到整个领域候选情感词空间，从而判定领域特定情感词的情感极性和强弱程度，自动构建领域情感词典。（2）从评论语料库中定义和抽取的情感词上下文和词法约束关系，在目标领域内始终保持有效性。例如，已知“helpful”与“helpless”之间构成词法反义词关系，在多种评论领域中它们始终具有相反的情感倾向性。

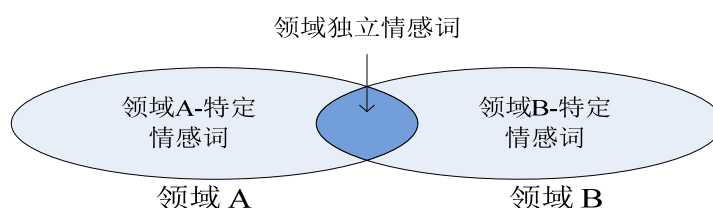


图 4.1 两个领域情感词典之间重叠示例

本章提出一种基于约束标签传播的领域情感词典自动构建策略。针对情感词典的领域依赖性问题，提出了利用组块依存分析和通用情感词典从领域评论语料中抽取候选情感词项。通过定义和抽取情感词之间的上下文约束和词法约束关系，为领域情感词典构建提供先验知识。针对约束关系的稀疏性问题，采用约束传播算法将局部约束

关系传播到全局情感词空间。最后，采用约束标签传播算法，在融合全局约束关系的同时，计算词语的情感倾向性，构建领域情感词典。

4.2 相关研究

近年来，学术界针对领域情感词典自动构建问题开展了一系列的研究工作。其中大部分方法的基本流程如下：首先手工指定一些已知情感极性的种子词，通过计算候选词与种子词的相似度关系判定候选词的倾向性，并不断地扩充种子词规模，从而构建情感词典。根据获取候选情感词与种子词之间相似度的方法不同，目前的研究方法可以主要划分为基于词汇知识库的构建方法和基于领域语料库的构建方法两类。

4.2.1 基于词汇知识库的情感词典构建

一些研究^[28, 29, 54, 133, 137]中提出利用已知的词汇知识库构建领域情感词典，例如英语中的 WordNet、General Inquirer 和 Open Office Thesaurus 等，汉语中的 HowNet、同义词词林等。这类方法主要依赖词汇知识库提供的词条之间的语义关系（如同义、反义、上下位关系等）以及词条解释来计算词语相似度，通过指定少数种子词逐步扩展情感词典（见图 4.2^[2, 54]）。基于词汇知识库的构建方法的基本思想是：具有同义词关系的词语之间具有相同的情感极性，而具有反义词关系的词语之间具有相反的情感极性。比如：褒义词“fast”的同义词“swift”、“prompt”和“quick”等都是褒义的，贬义词“slow”的同义词“dilatory”、“tardy”和“laggard”等都是贬义的；由于“fast”与“slow”是反义词，它们具有相反的情感极性。

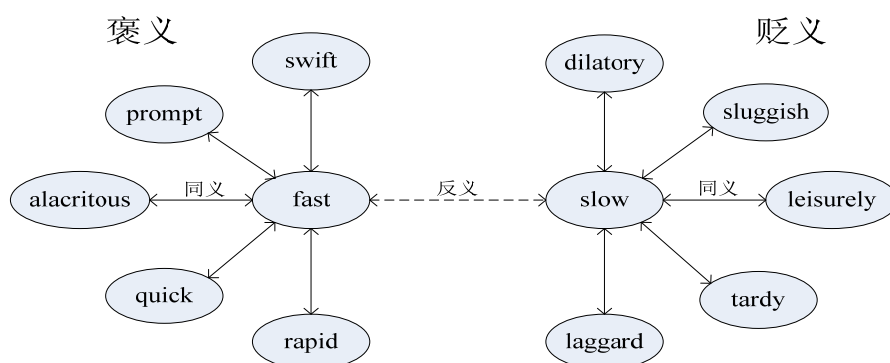


图 4.2 基于词语知识库的情感词典构造方法

Hu 和 Liu 在文献^[54]中利用 WordNet 中提供的形容词之间的同义/反义词集关系，通过给定足够规模的种子情感词集，判定候选形容词的情感倾向性，从而不断地扩充情感词典。但由于种子情感词的规模对情感词典的覆盖率影响较大，且忽略了同一情

感词的不同语义具有不同情感倾向问题，该方法具有一定的局限性。Kamps 等人^[137]通过连接 WordNet 中形容词之间的同义词关系构建词汇网络，根据在网络中与种子词（如“good”、“bad”）的距离计算候选情感词的情感倾向性。Esuli 等人^[28, 29, 133]利用 WordNet 中的同义词集（Synset）作为情感单位，利用关联的注释信息（Gloss）分别标注它们的积极、消极和中立倾向的情感得分。标注过程可以划分为半监督分类学习和随机游走算法两个阶段。Esuli 和 Sebastiani 在文献^[138]中进一步提出了随机游走算法的逆向和双向流模型，并验证了这两种新模型取得了优于 PageRank 模型^[133]的实验效果。

然而，基于词汇知识库的方法通常完全依赖于先验的词汇知识库资源，而忽略了情感词典的领域依赖性问题。因此，这类方法通常不能应用于缺乏语义词典资源的语言场景。此外，这类方法也面临着缺乏扩展性问题，难以处理词典未覆盖的词汇。

4.2.2 基于领域语料的情感词典构建

与基于词汇知识库的方法不同，基于领域语料的方法中情感词之间相似度关系的度量是以它们在领域评论语料中的统计分布信息为依据。这类方法通常都依赖于一个基本假设：在评论文本的局部上下文中，用户倾向于保持一致的情感倾向，因此在局部上下文范围内共现的情感词通常具有相同的情感倾向。

Hatzivassiloglou 和 McKeown 在文献^[4]中提出根据局部上下文中形容词之间的正向和反向连词关系构建词汇网络，并判断它们的情感极性。这种方法假设：通过“and”这类正向连词连接的形容词具有相似的极性，而通过“but”这类反向连词连接的形容词具有相反的极性，如图 4.3 所示^[2]。

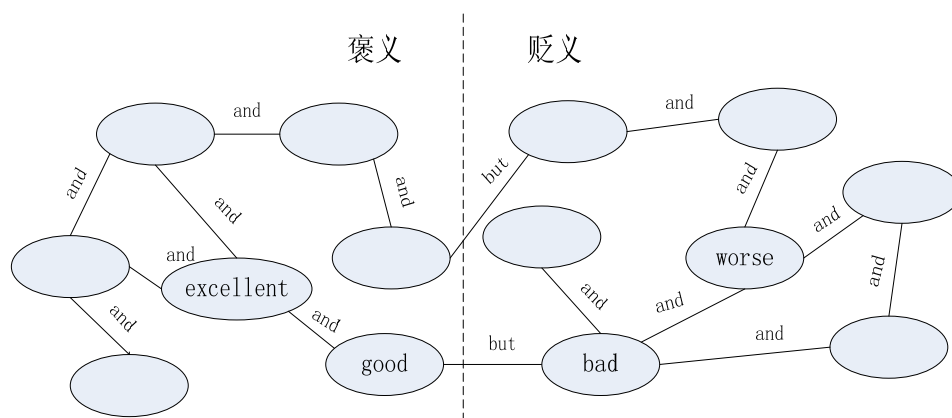


图 4.3 连词方法构造情感词典

Turney 在文献^[8]中首先基于固定模式抽取文本中的形容词和副词短语作为候选情感词和短语,然后采用点互信息(Pointwise Mutual Information, PMI)方法计算候选情感词与正负面种子词之间的 PMI 之差来判定情感词的语义倾向。Turney^[139]提出了一种无监督的 PMI-IR 方法,利用搜索引擎的返回结果计算词对之间的语义关联信息。Cruz 等人^[134]提出了一种半监督的领域情感词典构建算法,通过标注少量的领域种子情感词,并统计语料库中情感词之间的关联结构构建语义关联图,最后采用随机游走算法对关联图上各个结点的正负面情感得分排序,从而实现对领域词典的扩充。Popescu 和 Etzioni 在文献^[80]中采用松弛标记(Relaxation Labeling)的方法识别词语的情感极性。该方法通过利用包括句内共现和词典同义词在内的多种特征迭代地指派词语的极性。Wilson 等人^[140]提出一种监督学习的两阶段分类算法,对影响情感词和短语的上下文主客观性判定和情感极性判定的特征因素进行了详细分析。Kanayama 和 Nasukawa 在文献^[135]中提出了一种无监督的词典构建方法,该方法利用上下文一致性(Context Coherence)、句内和句间的共现关系来判定情感词和短语的极性。与其他研究工作^[4]相比,该方法涵盖了更广泛的词语共现关系类型。Qiu 等人^[85, 141]针对情感词典扩充和评价目标抽取的任务,分析和总结了八种情感词与评价目标之间的依存句法关系,提出了一种双向关联信息传播算法,并基于情感一致性假设对情感词进行正负面情感判定。该方法依赖于评论数据中的依存句法信息,对文本格式不规范的数据具有一定局限性。

Rao 和 Ravichandran 在文献^[142]中采用基于图的半监督方法识别词语的情感极性,其中待判定情感极性的词语表示为图上结点,词语之间的关联关系表示为结点之间的加权边,并通过实验证明了标签传播模型的效果优于最小割(Mincuts)模型和随机化最小割(Randomized Mincuts)模型。Weichselbraun 等人^[143]提出了一种半自动的情感词典构建方法,采用众包(Crowd-sourcing)策略为情感词赋值,并引入一种未标记领域文档上的自展(Bootstrapping)过程,对构建的词典按照具体用例进行扩展和定制。Lau 等人^[179]提出了一种两阶段伪标记方法,挖掘和利用了语料中隐含的词语之间以及词语与文档之间的关系。文献^[144]提出了一种基于函数优化的方法,分别利用最小割和词网络的聚团性两种目标函数来求解情感词典构建问题。

基于领域语料的方法通常基于共现信息计算词语之间的语义关联,或者直接利用局部上下文关系判定词语的情感极性。然而,它们并没有充分考虑情感词之间在上下文和词法层面存在的一致性或转折性约束关系。此外,这类方法也更容易受到种子词

覆盖问题的影响。

最近,一些研究还尝试利用源领域(Source Domain)中的先验情感知识,来辅助目标领域(Target Domain)中的情感词典构建。Tan 和 Wu^[136]针对计算情感词语义关联时出现的信息不足问题,通过融合源领域中的情感词典和文档情感标签信息,从源领域和目标领域数据中构建四种关联关系:情感词之间、情感文档之间、情感词和情感文档之间、情感文档和情感词之间四种关联图,进而利用随机游走算法融合这四种关联关系,计算情感词的正负面情感得分。但由于计算情感词和文档之间语义相似度时具有领域依赖性问题,且需要已知领域内的情感词典和情感文档等数据,该方法具有一定程度的局限性。Bollegala 等人^[145]通过利用多个源领域中的标记和未标记数据构建跨领域情感词库,从而发现在不同领域中表达相似情感的词语之间的关联关系。与这些工作相比,本章研究的情感词典构建策略不需要任何其他领域中的先验情感标签信息。当面对新的目标领域时,通常难以找到合适的具有情感标签信息的源领域。因此,本章提出的构建策略具有更好的适用性和实用性。

4.2.3 目前存在的主要问题

自动构建高质量的领域情感词典是领域评论文本意见挖掘的核心技术之一。目前的研究方法主要是利用词语在先验词汇知识库中的语义关系或者领域语料中的共现信息计算词语相似度,然后通过比较与正面和负面种子词的相似度大小来判定词语的情感极性,构建情感词典。然而,目前的研究中仍然存在着以下主要问题:

(1) 情感词典构建算法的领域自适应性问题。情感词典具有领域依赖性,很多词语往往与特定的描述对象关联起来才能表现出情感倾向性。目前在针对领域文本的意见分类中,仍然采用通用的领域无关情感词典,由此造成了准确率相对低下的问题。

(2) 情感词关联相似度计算模型。目前的大多数情感词典构建算法都是基于情感词与种子词之间的关联相似度,计算情感词的情感倾向性。现有的基于词典知识库和语料库的相似度计算方法难以有效反映情感词之间的情感关联相似度。

(3) 基准词依赖问题。由于目前多数情感词典构建算法使用的是半监督式的学习算法,影响该类算法性能的一个重要因素是基准词的选择和数量,如何减少基准词的依赖性是一个衡量算法有效性的关键问题。

针对以上三个主要问题,本章从情感词之间领域关联性的角度,认为解决上述问题的关键在于如何有效地利用领域依赖知识,即通过定义和抽取候选情感词之间的先验约束关系,更加准确地计算基于领域的情感词关联相似度,从而提高构建情感词典

的准确性。根据问题的关键，本章提出一种基于约束关系的标签传播算法，自动构建领域情感词典。通过对领域评论语料进行分析，利用组块依存关系分析和先验通用情感词典识别和抽取领域候选情感词项；利用半结构化的评论文本格式，从领域评论语料中选择若干有代表性的种子情感词；基于情感词在领域评论语料中的分布上下文，计算它们之间的语义关联度，构建语义关联图；通过定义和抽取情感词两两之间的上下文约束和词法约束关系，为领域情感词典构建提供先验知识；利用约束传播算法，将局部情感词之间的约束关系传播到全局领域情感词空间；最后使用半监督的约束标签传播算法，在融合先验约束关系的同时，计算候选词的情感倾向性。

4.3 候选情感词和种子词抽取与关联度计算

4.3.1 领域候选情感词抽取

识别和抽取领域候选情感词和短语是构建领域情感词典的首要任务。针对候选情感词的领域性，我们通过对领域评论语料进行句法和语义分析，利用组块依存知识和通用情感词典来识别和抽取领域特定的情感词和短语。

已知名词、形容词、动词、副词以及它们的短语是文本中最常见的具有情感倾向性的词汇类型^[1]。名词和名词短语通常用于表达客观的情感倾向状态，并不一定直接代表评论者的主观意见。形容词、动词及其短语通常用于直接表达评论人对于评论对象的主观情感和意见。副词和副词短语修饰形容词或动词，通常用于加强或减弱所修饰的形容词和动词的情感强度。例如，在一条汽车领域的评论文本片段如“This car has *great safety rating* and the *appearance* is *luxurious*. I would *strongly recommend* it!”中，形容词“great”和“luxurious”分别直接表达评论者对于“safety rating”和“appearance”的正面意见，副词“strongly”用于增强所修饰的动词“recommend”表达的正面情感强度。尽管名词“safety”能够暗示积极的情感状态，在这里仅作为用户评论的目标特征。在多个不同的用户评论领域中，名词、动词及其短语通常能够保持始终一致的情感倾向性，而形容词和副词则倾向于具有不同的情感倾向性。例如，形容词“luxurious”在以上的汽车评论文本片段中表达正面的情感倾向，但用于修饰生活方式时，则能够表达负面的情感倾向。形容词、动词和副词及其短语通常是构成领域情感词典的主要词汇类型^[135, 136]。

给定领域语料库中的一篇用户评论文本，首先将其划分为独立的评论句子集合，

对每一条评论语句执行分词、词性标注和组块分析处理。我们采用 Stanford Parser⁵分析每一条评论语句中的依存句法关系，并构建依存句法树。根据依存句法树和组块分析结果，进一步构建组块依存树^[84]。通过构建组块依存树，能够有效地识别和保留常用的情感短语、习语和俗语等不规则的文本表达形式。图 4.4 描述了给定评论语句“The car is beautiful and strong.”，基于依存句法树和组块分析构建组块依存树的示例。组块依存树的构建过程直观而简单：首先将词汇组块表示为组块依存树中的结点，然后将依存句法树中词语之间的依存关系添加到组块依存树中。如果两个词语属于相同的词汇组块，则将它们之间的依存关系添加到对应的组块结点之内；如果两个词语分别属于不同的词汇组块，则将它们之间的依存关系添加到对应的两个组块结点之间。

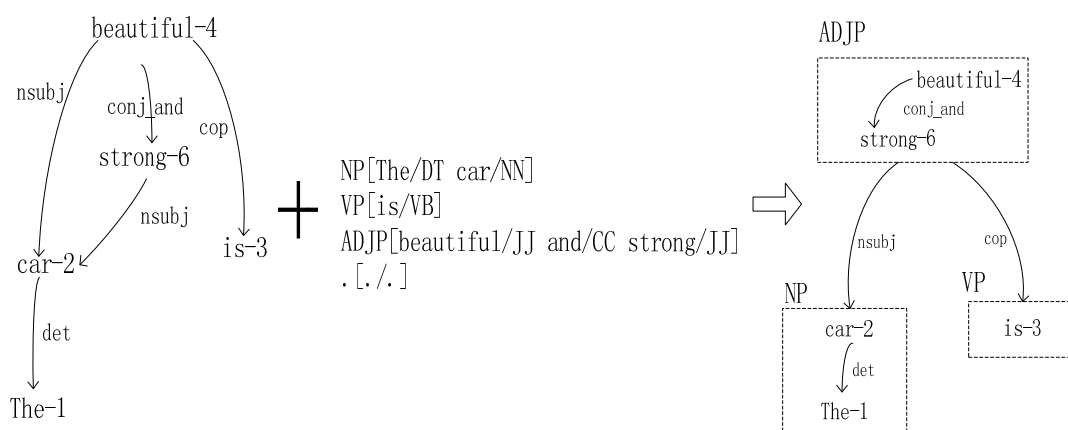


图 4.4 基于依存句法树和组块分析构造组块依存树示例

进一步地以已知的通用情感词典为基础，识别组块依存树中包含了通用情感词项的候选组块，这些候选组块通常直接包含了用户主观情感，因此可以从中抽取候选的领域情感词和短语。已知文献^[140]中给出了通用的 hltmnlp05clues 情感词典，其中涵盖了手工构建的八千多个主观性词项，并分别标注了情感倾向和强度。为了缩小领域情感词空间，提高构建领域情感词典的精确性，本章中我们仅选择了 hltmnlp05clues 中具有强烈主观性（“strongsubj”）的情感词项。在组块依存树中，我们抽取包含通用情感词项的组块作为候选情感组块。此外，组块之间的依存关系也能够用来过滤候选情感组块。如果候选组块与邻接组块之间存在形式主语（nominal subject, “nsbj”）依存关系，则该候选组块与目标领域具有更强的相关性。

为了去除无关修饰词和重复数据，进一步对候选情感组块执行一些修剪和清洗操作。为了保持情感短语和俗语的完整性，从候选组块的两端移除无关的介词、连词和

⁵ <http://nlp.stanford.edu/software/lex-parser.shtml>

代词等停用词。如果候选情感组块包含由连词连接的多个情感词项，则将每个情感词项划分为一个独立的情感组块。对于形容词名词结构的名词短语组块，其中形容词直接表达对于名词的情感，例如“a beautiful car”，我们也抽取这些形容词作为候选情感词。然而，这些名词短语组块中通常会包含着一些与特定领域无关的噪声。通过计算这些名词与一些领域特定的主题词之间的语义相似度，进而去除那些具有较低相似度的名词短语候选组块。

4.3.2 种子情感词抽取

作为一种半监督学习过程，我们提出的情感词典构建策略同样也需要一些情感种子词来启动传播学习过程。种子情感词主要由领域无关的词汇组成，能够在多个不同的评论领域中保持一致的情感倾向性。获取种子情感词的方法通常包括从领域语料库中自动抽取或者由人工指定。由于现实中的领域评论文本通常具有半结构化的格式，我们提出一种简单而有效的机制，通过利用半结构的评论文本格式，抽取领域种子情感词。用户在评论某一商品时，首先会简短地给出整篇评论的标题，分别列举正面和负面的评价方面，最后对个人的主观感受进行详细描述。图 4.5 给出了半结构化评论文本的格式示例。

Title: Great small SUV .
Pros: Great acceleration .
 Good fuel economy .

Cons: Big blind spot .
 Fairly small rear seats .

Text: I've had several Toyota trucks and all are impressive .

图 4.5 TBOD 评论数据集中半结构化评论文本格式示例

如图 4.5 所示，“Title”部分概括了用户对于商品的整体性评价；“Pros”和“Cons”部分分别包含了用户对于商品的积极和消极评价方面的简短论述；“Text”部分中对用户评论的每一个商品特征展开详细的描述。“Pros”和“Cons”通常由名词以及形容词名词结构的短语构成。这些包含的形容词通常都具有主观倾向性，分别表达了正面和负面的用户情感。因此，我们可以选择这些形容词作为候选种子情感词。我们抽取“Pros”中包含的形容词，例如 great、good 和 excellent 等，作为候选正面种子情感词，并将它们添加到一个候选正面种子词集合中；抽取“Cons”中包含的形容词，例

如 poor、small 和 expensive 等，作为候选负面种子情感词，并将它们添加到一个候选负面种子词集合中。由于当上下文中存在否定词修饰情感词时，否定词能够逆转情感词原本的极性，从而改变上下文的情感倾向性，我们在抽取种子情感词时也考虑到了否定词对情感极性的影响。本章中，我们收集了一个常用的否定词列表，包含 not、no、never、hardly 和 seldom 等，在抽取“Pros”和“Cons”中的形容词时，我们检查形容词的邻接上下文中是否存在否定词。邻接上下文的范围设定为形容词之前的五个词和之后的两个词。当邻接上下文中存在否定词修饰形容词时，我们去除否定词，并对邻接上下文的情感倾向性取反，作为余下形容词的情感极性。

然而，由于内容的简短性以及用户发表评论时的随意性，“Pros”和“Cons”部分中不可避免地存在噪声数据。因此，为了过滤情感极性判断错误的候选种子情感词，我们对候选种子情感词集合中的情感词按照出现频率进行排序。对于同时出现在“Pros”和“Cons”中的形容词，它们可能同时出现在正面和负面候选种子词集合中。对于同时出现在正面和负面候选种子词集合中的形容词，我们从出现频率较低的候选种子词集合中移除该候选形容词。最后从正面和负面候选种子词集合中分别挑选前 N 个频繁的形容词作为正面和负面种子情感词。在本章的实验部分中，我们对种子情感词的抽取结果进行验证，实验结果表明种子情感词的选择机制是有效的。

对于评论文本不具备如图 4.5 所示的半结构化格式的商品评论领域，种子情感词可以由人工方式指定，或者直接借用其他已知评论领域中的种子情感词。由于种子情感词通常都是领域无关的，采用手工指定或者借用的方法也是可行的。

4.3.3 语义关联图构建

计算情感词的关联相似度，构建情感词之间语义关联图是基于种子词和标签传递算法构建领域情感词典的关键基础。基于语义关联图，能够将种子词的情感极性传递到候选情感词上，从而判定候选情感词的情感极性。

定义情感词集合上的语义关联图为 $G=(X, A)$ ，其中 $X = \{x_1, x_2, \dots, x_{|X|}\}$ 表示情感词集合， $|X|$ 表示抽取得到的候选情感词的数量， A 表示情感词之间的语义关联集合，以关联相似度矩阵的形式表示。

关联相似度量是邻接情感词彼此之间传递情感信息的桥梁。目前用于计算词汇关联度的方法很多，例如基于上下文连接和转折关系的方法^[4]、基于句间和句内共现信息的方法^[135]、基于词典语义关系的方法^[54]等。然而，情感词的关联相似度量应该具有可传递性，并且能够反映情感词之间领域特定的关联性。

在本章中，采用点互信息（PMI）公式计算情感词之间的关联相似度。PMI 公式通常用在信息论和统计学中衡量对象之间的语义关联关系。PMI 值具有可传递性，并且能够有效地基于情感词在领域评论语料中的分布上下文计算它们之间领域依赖的关联相似度，进而可以很容易地转化为语义关联图表示。情感词 x_i 和 x_j 之间的 PMI 计算公式定义如下：

$$PMI(x_i, x_j) = \log_2 \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \quad (4.1)$$

其中， $p(x_i, x_j)$ 表示 x_i 和 x_j 在评论语料中的共现概率， $p(x_i)$ 和 $p(x_j)$ 分别表示 x_i 和 x_j 在评论语料中的独立发生概率^[8]。点互信息值表示已知某一个情感词时，能够得到另一个情感词的可能性。

4.4 基于约束关系的标签传播算法

4.4.1 情感词约束关系的定义和抽取

与直接利用情感词之间的上下文连接和转折关系构建语义关联图方法^[134]不同，我们通过定义和抽取在目标评论领域中情感词之间存在的一些约束关系，作为先验知识来提高领域情感词典构建的效果。在本章中，我们定义了情感词之间的两种约束关系类型：上下文约束关系和词法约束关系。

上下文约束关系（Contextual Constraints）定义了候选情感词之间在评论文本上下文中存在的情感极性一致性或转折性关系。按照约束方向不同，上下文约束关系可以划分为正向上下文约束关系和反向上下文约束关系。正向上下文约束定义了候选情感词之间具有相同情感倾向性的先验关系，反向上下文约束定义了候选情感词之间具有相反情感倾向性的先验关系。上下文约束关系通常由一些上下文线索表示。我们定义正向上下文线索为并列连词，如 *and*、*as well as* 等，以及并列结构；定义反向上下文线索为转折连词，如 *but*、*however* 等，以及否定结构。当两个情感词在评论文本中连续出现时，如果存在上下文线索连接它们，则定义它们之间存在上下文约束关系。由于上下文约束关系完全依赖于情感词在领域评论文本中出现时的上下文，因此具有领域依赖性，只能够适用于具体的目标领域。下面分别给出了两个正向和反向上下文约束关系示例。其中，情感词由粗斜体所示，上下文线索词由下划线所示。

正向上下文约束关系示例：

(1) This fine looking luxury car is ***well appointed*** and is ***very pleasurable*** to drive.

(2) It is a *strong* car, with really *beautiful* looks.

反向上下文约束关系示例：

(3) It is an *interesting* but *filthy* hotel.

(4) The hotel bed is *narrow* and *not soft*.

词法约束关系 (Morphological Constraints) 定义了候选情感词之间在词法级别存在的情感极性一致性或转折性关系。与上下文约束关系类似，词法约束关系按照约束方向不同也可以划分为正向词法约束关系和反向词法约束关系。由于反向词法约束关系在领域候选情感词之间更为常见，本章中我们主要利用和抽取反向类型的词法约束关系。与上下文约束关系具有领域依赖性不同，词法约束关系完全依赖于候选情感词之间的词法分析，因此具有领域无关性，能够适用于多个不同的评论领域。下面给出了三个反向词法约束关系示例。

反向词法约束关系示例：

(5) “*natural*”与 “*unnatural*”

(6) “*practical*” 与 “*impractical*”

(7) “*hopeful*” 与 “*hopeless*”

如上述例子所示，“*natural*”与 “*unnatural*”构成词法反义词关系，通过前缀“un-”使得两词之间具有相反的情感极性。同样地，“*practical*”与 “*impractical*”之间也是通过前缀“im-”构成词法反义词关系。在例(7)中，与“*hopeful*”的正向情感极性相比，通过改变后缀为“-less”，“*hopeless*”具有相反的情感极性。

对于约束关系抽取，当情感词在评论文本中连续出现时，我们首先通过识别上下文线索抽取情感词之间的上下文约束关系。在图 4.5 所示的半结构化的评论文本格式中，Pros 和 Cons 部分分别表达了相反的用户情感倾向，它们中包含的情感词之间通常具有反向上下文约束关系。此外，当候选情感组块中包含上下文线索连结多个情感词时，这些情感词之间也具有上下文约束关系。通过指定常见的词法前缀和后缀，词法约束关系主要通过情感词之间的词法分析抽取。当正向和反向约束类型冲突时，为了减少和过滤偶然错误和噪声，仅保留占多数的约束关系类型。

此外，约束关系的定义和抽取都是基于语言规则，符合真实的语言使用环境，且无需过多的人工干预，因此能够具有良好的领域自适应性。

4.4.2 约束关系传播算法

抽取候选情感词之间的约束关系以后，我们定义所有正向上下文和词法约束关系

(Must-link) 的集合为 M , 所有反向上下文和词法约束关系 (Cannot-link) 的集合为 R 。定义约束表示矩阵 $D = \{d_{ij}\}_{|X| \times |X|}$, 其中元素 d_{ij} 定义为情感词 x_i 和 x_j 之间的先验约束关系, $|d_{ij}|$ 表示约束关系的置信度。初步地定义 d_{ij} 值如下:

$$d_{ij} = \begin{cases} 1, & (x_i, x_j) \in M \\ -1, & (x_i, x_j) \in R \\ 0, & \text{其他情况} \end{cases} \quad (4.2)$$

针对 D 中局部情感词之间约束关系的稀疏性问题, 我们采用一种两阶段的约束传播算法, 将局部情感词之间的约束关系传播到全局候选情感词空间。约束传播算法包含基于约束表示矩阵 D 中列上的纵向约束传播阶段和行上的横向约束传播阶段^[128]。约束标签传播算法的详细描述如第 3 章 3.3.2 节所示, 在此不再赘述。其中, 参数 α 指定了约束传播算法中约束信息的传播速率, 取值范围为 $(0, 1)$ 区间。在本章的实验部分中, 我们将对 α 的取值进行分析。

4.4.3 约束标签传播算法

在约束传播以后, 得到一个在全局候选情感词之间的约束关系集合, 表示为全局约束矩阵 E^* 以及对应的置信度 $|E^*|$ 。进而, 我们融合全局约束矩阵和标签传播框架, 提出一种基于约束关系的标签传播算法, 基于少数种子情感词判定领域情感词典中候选词的情感倾向性。标签传播算法是一种基于图的半监督学习框架, 具有良好的计算效率和收敛性^[147]。以下给出基于半监督的情感词典构建问题的形式化定义, 并描述约束标签传播算法。

已知 $L = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ 表示已知情感标签的种子词结点集合, 其中 x_l 表示种子情感词, y_l 表示 x_l 对应的情感极性标签。 $Y = \{y_{pos}, y_{neg}\}$ 表示情感极性标签的取值范围, $y_i \in Y$, 其中 y_{pos} 表示正面情感极性, y_{neg} 表示负面情感极性。令 $U = \{x_{l+1}, x_{l+2}, \dots, x_{l+u}\}$ 表示未知情感标签的候选情感词结点集合。因此, 可以得到全体领域情感词集合 $X = L \cup U = \{x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_{l+u}\}$, $|X| = l + u$, 通常假设 $l \ll |X|$ 。令 $Y_L = \{y_1, y_2, \dots, y_l\}$ 表示 L 中种子词对应的情感标签集合, $Y_U = \{y_{l+1}, y_{l+2}, \dots, y_{l+u}\}$ 表示 U 中候选情感词对应的未知情感标签集合。因此, 领域情感词典自动构建需要解决的问题即是基于 X 和 Y_L 求解 Y_U 。

已知领域情感词的关联相似度矩阵 A , 为了在标签传播框架中融入全局约束关系矩阵 E^* , 计算情感词之间的约束关联相似度如下:

$$\tilde{A}_{ij} = \begin{cases} 1 - (1 - E_{ij}^*)(1 - A_{ij}), & E_{ij}^* \geq 0 \\ (1 + E_{ij}^*)A_{ij}, & E_{ij}^* < 0 \end{cases} \quad (4.3)$$

其中, \tilde{A} 定义为领域情感词集合 X 上的约束关联相似度矩阵, 基于矩阵 A 和 E^* 计算得到。当情感词之间具有正向约束关系时, 增大初始的关联相似度; 当情感词之间具有反向约束关系时, 减小初始的关联相似度。进而能够定义 X 上的概率转移矩阵 $T_{|X| \times |X|}$, 转移概率元素 T_{ij} 计算如下:

$$T_{ij} = p(i \rightarrow j) = \frac{\tilde{A}(x_i, x_j)}{\sum_{k=1}^{|X|} \tilde{A}(x_i, x_k)} \quad (4.4)$$

其中, $p(i \rightarrow j)$ 表示从情感词结点 x_i 转移到另一个情感词结点 x_j 的概率, $\tilde{A}(x_i, x_j)$ 表示 x_i 和 x_j 之间的约束关联相似度值。

假设 $f^t = \{f_1^t, f_2^t, \dots, f_{|X|}^t\}$ 表示第 t 次迭代时 X 中情感词的情感评分向量, 定义向量 f^0 用于记录种子词和候选情感词的初始情感评分值。对于正面和负面种子词, f^0 中对应的维度分别设置为+1 和-1; 对于其他未知情感极性的候选情感词, 统一设置为 0。

标签传播算法定义描述如下:

(1) 对于 X 中的所有情感词结点, 根据邻接结点执行一步标签传播:

$$f_j^{t+1} = \lambda \sum_{i=1}^{|X|} T_{ij} f_i^t + (1 - \lambda) f_j^0 \quad (4.5)$$

每一个情感词结点既能够接收从邻接结点传递的情感信息, 同时也具有保持自身初始情感标签的倾向性。其中 λ 是传播速率参数, 用于指定当前情感词结点从邻接结点传递的情感信息与保持初始情感标签的相对比率。 $0 < \lambda < 1$ 。

(2) 对于种子情感词, 重置它们的情感评分为初始值:

$$f_l^{t+1} = f_l^0 \quad (4.6)$$

(3) 重复步骤 (1) 和 (2), 直至收敛。

最后得到迭代传播以后情感评分向量 f^t 的收敛结果。对于任一候选情感词结点 x , 根据 x 在 f^t 中对应维度的情感评分值判定 x 的情感倾向性。如果情感评分为正值, 则判定候选情感词具有正面倾向性; 如果情感评分为负值, 则判定候选情感词具有负面倾向性。此外, 情感评分很少出现零值的情况, 当出现零值时, 则判定候选情感词为中立情感词。

4.5 实验验证和结果分析

4.5.1 验证数据集准备

实验部分采用 TBOD⁶和 OpinRankDataset⁷数据集作为领域评论语料, 对本章提出的基于约束标签传播的领域情感词典自动构建效果进行评估。TBOD 数据集^[134]中包含三种领域的用户评论数据, 用户评论文本从 Epinions.com 网站收集, 均采用如图 4.5 所示的半结构化格式。OpinRankDataset 数据集^[148]中包含从 Tripadvisor 和 Edmunds 网站上收集的完整用户评论文本。我们选择 Cars 和 Hotels 领域的评论语料作为验证数据集。TBOD 和 OpinRankDataset 数据集的详细统计信息如表 4.1 所示。

表 4.1 TBOD 和 OpinRankDataset 数据集的统计信息

数据集 领域	TBOD 数据集		OpinRankDataset 数据集	
	Cars	Hotels	Cars	Hotels
评论文档数	972	988	42,288	254,539
平均句子数	27	34	8	8
平均词汇数	506	639	124	189

4.5.2 基线方法和评价标准

为了衡量基于约束标签传播算法自动构建领域情感词典的性能, 引入三种当前主流的词语情感倾向判定基线方法: 基于卡方检验 (Chi-square) 的情感倾向性判定^[149]、基于点互信息 (PMI-IR) 的情感倾向性判定^[8]和基于基本标签传播 (Label Propagation) 模型的情感词典构建^[142]。

对于每一个候选情感词 c , 统计它在数据集中的列联表如下:

表 4.2 情感词 c 在评论数据集中的列联表

候选情感词	正面 (Pos)	负面 (Neg)
c	$f(c, pos)$	$f(c, neg)$
$\neg c$	$f(\neg c, pos)$	$f(\neg c, neg)$

其中, $f(c, pos)$ 表示情感词 c 在正面评论中出现的频率, $f(c, neg)$ 表示情感词 c 在负面评论中出现的频率, $f(\neg c, pos)$ 统计除 c 之外的其他情感词在正面评论中出现的频率之和, $f(\neg c, neg)$ 统计除 c 之外的其他情感词在负面评论中出现的频率之和^[149]。我们根据用户评分值判定整篇评论文本的情感倾向性, 评分值大于 3 时判定为正面评论,

⁶ <http://www.lsi.us.es/~fermin/index.php/Datasets>.

⁷ <http://archive.ics.uci.edu/ml/datasets/OpinRank+Review+Dataset>.

评分值小于 3 时判定为负面评论，评分值等于 3 时判定为中立评论。

Chi-square: 卡方值以一种朴素假设——候选情感词 c 在正面评论中出现的概率等于其在负面评论中出现的概率，作为前提条件。卡方值的计算公式如下：

$$\chi^2 = \sum_{x \in (c, -c)} \sum_{y \in (pos, neg)} \frac{\{f(x, y) - f^*(x, y)\}^2}{f^*(x, y)} \quad (4.7)$$

其中， $f^*(x, y)$ 表示 $f(x, y)$ 基于朴素假设条件下的期望值^[149]。当 c 在正面评论中出现较多时，判定 c 的情感倾向值大于零，否则判定 c 的情感倾向值小于零。最终的情感评分计算如下：

$$PV_{\chi^2}(c) = \begin{cases} \chi^2(c), & p(c|neg) < p(c|pos) \\ -\chi^2(c), & \text{其他情况} \end{cases} \quad (4.8)$$

其中， $p(c|pos)$ 为 c 在正面评论中出现的概率， $p(c|neg)$ 为 c 在负面评论中出现的概率，它们的值分别从表 4.2 中估算得出^[149]。

PMI-IR: PMI-IR 算法采用点互信息和信息检索方法来估计词汇或短语之间的语义关联度。给定一个候选情感词，通过比较它与正面参考词“excellent”和负面参考词“poor”之间的关联程度，判定它的语义倾向性（ SO ）。 SO 值的正负性判定为词语的正负面极性，量值判定为词语的情感强度。 SO 值的计算公式定义如下：

$$SO(phrase) = PMI(phrase, "excellent") - PMI(phrase, "poor") \quad (4.9)$$

为了减小候选情感词与特定参考词之间共现稀疏的影响，我们分别选取了五个正面参考词和五个负面参考词，分别计算采用不同正面和负面参考词时候选情感词的 SO 值。最大的 SO 值被用来判定候选情感词的情感倾向性。

Label Propagation (LP): 为了验证本章中引入的情感词之间的正向和反向约束关系对提高领域情感词典构建效果的贡献，我们引入 Rao 和 Ravichandran 在文献^[142]中提出的基于基本标签传播模型的情感词典构建算法作为一种基线方法。该方法没有考虑任何情感词之间的先验约束关系。

我们采用信息检索领域中标准的准确率（Precision）、召回率（Recall）和 F 值（F-measure）作为领域情感词典自动构建效果的评价标准。准确率衡量了正确分类的情感词数与全体情感词总数的比率，召回率衡量了正确分类的情感词数与人工标注的情感词总数的比率，F 值计算准确率和召回率的调和平均值。定义准确率、召回率和 F 值的计算公式如下所示：

$$P_i = \frac{TP_i}{TP_i + FP_i} \quad R_i = \frac{TP_i}{TP_i + FN_i} \quad F_i = \frac{2 * P_i * R_i}{P_i + R_i} \quad (4.10)$$

其中, P_i 、 R_i 和 F_i 分别表示第 i 个情感类别的准确率、召回率和 F 值。 TP_i 表示被正确分配到第 i 类的情感词数量, FP_i 表示被错误分配到第 i 类的情感词数量, FN_i 表示属于第 i 类但被错误分配到其他类的情感词数量。

在分别计算得到正面情感词类和负面情感词类的准确率、召回率和 F 值之后, 为了从整体角度衡量领域情感词典构建的效果, 进一步计算正面情感词类和负面情感词类的宏平均 (Macro-averaged) 和微平均 (Micro-averaged) 值。

宏平均值用来衡量算法在所有类别上的整体效果, 在所有情感类别上求平均值。由于赋予所有的情感词类相同的权重, 更容易受到稀有类的影响。定义宏平均准确率、召回率和 F 值的计算公式如下:

$$Macro - averaged Precision = \frac{\sum_{i=1}^M P_i}{M} \quad (4.11)$$

$$Macro - averaged Recall = \frac{\sum_{i=1}^M R_i}{M} \quad (4.12)$$

$$Macro - averaged F - measure = \frac{\sum_{i=1}^M F_i}{M} \quad (4.13)$$

其中, M 表示情感倾向性的类别数。

微平均值用来衡量算法在所有情感词上的平均效果。由于赋予所有的情感词相同的权重, 更容易受到常见情感词类的影响。定义微平均准确率、召回率和 F 值的计算公式如下:

$$Micro - averaged Precision (MiaP) = \frac{\sum_{i=1}^M TP_i}{\sum_{i=1}^M (TP_i + FP_i)} \quad (4.14)$$

$$Micro - averaged Recall (MiaR) = \frac{\sum_{i=1}^M TP_i}{\sum_{i=1}^M (TP_i + FN_i)} \quad (4.15)$$

$$Micro - averaged F - measure = \frac{2 * MiaP * MiaR}{MiaP + MiaR} \quad (4.16)$$

4.5.3 结果分析

4.5.3.1 候选情感词抽取结果分析

针对 Cars 和 Hotels 领域，分别从用户评论语料中抽取候选的领域情感词和短语，构建领域情感词典。为了对构建效果进行评估，我们手工对每一个候选情感词标注领域特定的倾向性。根据情感词对领域的依赖性和情感倾向性，可以将词典划分为四个部分：领域独立负面情感词（domain-independent negative terms, INT）、领域依赖负面情感词（domain-dependent negative terms, DNT）、领域独立正面情感词（domain-independent positive terms, IPT）和领域依赖正面情感词（domain-dependent positive terms, DPT）。手工标注的情感词典作为接下来自动算法的标准结果。表 4.3 给出了 Cars 和 Hotels 领域中候选情感词的抽取结果统计。

表 4.3 候选情感词抽取结果统计

评论领域	INT	DNT	IPT	DPT	合计
Cars	202	90	194	135	621
Hotels	190	108	200	177	675

4.5.3.2 种子情感词抽取结果分析

在领域情感词典自动构建过程中，为了分析不同种子词数量的影响，我们分别选择前 5、10、20 和 50 个正面和负面的种子词进行实验。表 4.4 中给出了 Cars 和 Hotels 领域中出现频率最高的前 20 个正面和负面种子情感词示例。

表 4.4 Cars 和 Hotels 领域中前 20 个正面和负面种子情感词示例

评论领域	正面种子情感词	负面种子情感词
Cars	good, great, nice, comfortable, smooth, reliable, excellent, powerful, easy, quiet, roomy, fast, sporty, decent, dependable, safe, affordable, quick, strong, stylish	poor, expensive, small, bad, cheap, little, terrible, uncomfortable, pricey, unreliable, hard, questionable, horrible, slow, plastic, limited, average, ugly, bumpy, wrong
Hotels	great, good, nice, friendly, clean, beautiful, comfortable, convenient, excellent, large, free, quiet, wonderful, helpful, close, affordable, spacious, easy, elegant, cheap	small, poor, expensive, dirty, pricey, older, bad, terrible, noisy, rude, overpriced, little, loud, high, slow, tiny, horrible, outdated, hard, unfriendly

通过人工对抽取结果进行评估可以发现，大部分的种子情感词都是领域无关的。

如表 4.4 所示, 前 20 个正面和负面种子词中, 大约 45% 最常见的词语部分同时出现在了 Cars 和 Hotels 领域中并且保持相同的情感倾向。因此, 采用手工指定最常见的情感词作为种子词或者直接借用其他已知评论领域中的种子情感词的方法是可行的。种子情感词抽取方法能够取得令人满意的效果。

4.5.3.3 情感词典构建结果分析

本小节中给出了领域情感词典自动构建结果的对比分析。本章提出的约束标签传播算法表示为 ConsLP。我们选取前 10 个正面和负面的种子词进行实验, 约束标签传播算法的迭代次数设置为 100, 参数 α 和 λ 取值为 0.1。表 4.5 和表 4.6 分别给出了在 Cars 和 Hotels 领域中情感词典构建效果的宏平均和微平均准确率、召回率、F 值的结果对比。

表 4.5 Cars 领域中宏平均和微平均准确率、召回率、F 值的结果对比

	宏平均 (Macro-averaged)			微平均 (Micro-averaged)		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Chi-square	0.679	0.626	0.650	0.676	0.631	0.653
PMI-IR	0.714	0.408	0.501	0.701	0.404	0.513
LP	0.764	0.738	0.741	0.756	0.732	0.744
ConsLP	0.822	0.782	0.789	0.815	0.778	0.796

表 4.6 Hotels 领域中宏平均和微平均准确率、召回率、F 值的结果对比

	宏平均 (Macro-averaged)			微平均 (Micro-averaged)		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Chi-square	0.732	0.715	0.722	0.729	0.696	0.712
PMI-IR	0.781	0.475	0.545	0.734	0.450	0.558
LP	0.812	0.781	0.797	0.817	0.785	0.801
ConsLP	0.848	0.805	0.826	0.854	0.811	0.832

如表 4.5 和表 4.6 所示, 与三种基线方法相比, ConsLP 算法在 Cars 和 Hotels 领域中都取得了最高的宏平均和微平均准确率、召回率和 F 值。PMI-IR 算法在 Cars 和 Hotels 领域中都取得了最低的宏平均和微平均 F 值, 且它的准确率值要明显大于召回率值。与在 Cars 领域中的表现相比, 由于 Hotels 领域中的评论语料规模明显增大,

PMI-IR 算法的性能表现取得了明显提高。与 PMI-IR 算法相比, Chi-square 算法提高了召回率和 F 值,但在准确率上有所降低。与 Chi-square 和 PMI-IR 算法相比, LP 和 ConsLP 算法在两个领域中都取得了明显更好的宏平均和微平均准确率、召回率和 F 值。与 LP 算法相比,我们提出的 ConsLP 算法取得了明显的提升。如表 4.5 所示,在 Cars 领域中与 LP 算法相比,ConsLP 算法在宏平均和微平均准确率上分别提高了 5.8% 和 5.9%,在宏平均和微平均召回率上分别提高了 4.4%和 4.6%,在宏平均和微平均 F 值上分别提高了 4.8%和 5.2%。如表 4.6 所示,在 Hotels 领域中,与 LP 算法相比,ConsLP 算法在宏平均和微平均准确率上分别提高了 3.6%和 3.7%,在宏平均和微平均召回率上分别提高了 2.4%和 2.6%,在宏平均和微平均 F 值上分别提高了 2.9%和 3.1%。

因此,本章提出的约束标签传播算法在 Cars 和 Hotels 领域中都取得了最佳的情感词典构建效果,与三种基线方法相比,取得了明显的改进。其中,LP 和 ConsLP 算法都是基于情感信息传播的方法,相比 Chi-square 和 PMI-IR 方法,能够取得更好的实验结果。与 LP 基线方法相比,我们提出的 ConsLP 算法取得了明显的性能改进,这也验证了通过引入情感词之间先验的上下文和词法约束关系,确实能够有效地提高领域情感词典构建的效果。

4.5.3.4 种子情感词数量影响分析

为了验证种子情感词数量对领域情感词典构建结果的影响,我们选取不同数量的种子情感词进行实验。在 Cars 和 Hotels 领域中,我们分别选取前 5、10、20 和 50 个正面和负面的种子情感词。实验中,设置约束标签传播算法的迭代次数为 100,参数 α 和 λ 值均设置为 0.1。表 4.7 和表 4.8 分别给出了在 Cars 和 Hotels 领域中选择不同数量的种子情感词对领域情感词典构建效果的影响。

表 4.7 Cars 领域中种子情感词数量影响分析

	宏平均 (Macro-averaged)			微平均 (Micro-averaged)		
	Precision	Recall	F-measure	Precision	Recall	F-measure
5	0.728	0.692	0.700	0.719	0.702	0.711
10	0.822	0.782	0.789	0.815	0.778	0.796
20	0.827	0.789	0.801	0.821	0.786	0.803
50	0.829	0.791	0.804	0.821	0.790	0.805

表 4.8 Hotels 领域中种子情感词数量影响分析

	宏平均 (Macro-averaged)			微平均 (Micro-averaged)		
	Precision	Recall	F-measure	Precision	Recall	F-measure
5	0.837	0.799	0.817	0.842	0.804	0.823
10	0.848	0.805	0.826	0.854	0.811	0.832
20	0.851	0.810	0.830	0.858	0.817	0.837
50	0.854	0.817	0.835	0.859	0.824	0.841

如表 4.7 和表 4.8 所示, 当仅仅选择前 5 个正面和负面的种子情感词时, 在 Cars 和 Hotels 领域的结果表现最差。随着种子情感词数量的增加, 领域情感词典构建的效果进一步提升。如表 4.7 所示, 在 Cars 领域中, 当种子情感词从前 5 增加到 10 时, 宏平均和微平均 F 值分别增加了 8.9%和 8.5%。如表 4.8 所示, 在 Hotels 领域中, 宏平均和微平均 F 值对应地分别增加了 0.9%和 0.9%。然而, 当种子情感词从前 10 增加到 20 时, 以及从前 20 增加到 50 时, 表 4.7 和表 4.8 中性能提高的幅度都有所下降。当种子情感词从前 10 增长到 20 时, 宏平均和微平均 F 值在 Cars 领域中仅仅分别增长了 1.2%和 0.7%, 对应地在 Hotels 领域中分别增长了 0.4%和 0.5%。当种子情感词从前 20 增长到 50 时, 宏平均和微平均 F 值在 Cars 领域中仅仅分别增长了 0.3%和 0.2%, 对应地在 Hotels 领域中仅仅分别增长了 0.5%和 0.4%。

在 Cars 和 Hotels 两个领域中, 与选择前 10 个种子情感词的实验结果相比, 当选择前 20 和 50 个种子情感词时并没有带来明显的性能提升。通过观察可以发现, 随着种子情感词数量的增加, 后加入的种子情感词的频率逐渐降低, 与其他候选情感词的关联程度也逐渐减弱。因此可以认为, 当分别选择前 10、20 和 50 个正面和负面种子情感词时, 情感词典构建的效果差别不大。这也表明了我们提出的算法能够较少地受到种子情感词覆盖问题的影响。当情感词典构建的效果近似时, 利用的种子情感词越少, 算法的性能越好。因此在本章实验中, 仅选择前 10 个正面和负面的种子情感词。

4.5.3.5 参数设置分析

在约束传播算法中, 参数 α 指定了情感词之间约束关系信息的传播速率。为了准确地设置 α 值, 我们在(0, 1)区间内测试 α 的取值, 以 0.1 为步长逐步增加。实验结果表明, α 参数的取值确实能够影响到约束传播算法的收敛速度。图 4.6 描述了在 Cars 和 Hotels 领域中 α 取值不同时, 约束传播算法达到收敛状态时需要的迭代次数。我们

将横向约束传播和纵向约束传播的迭代次数相加，设置迭代的收敛阈值为 10^{-14} 。

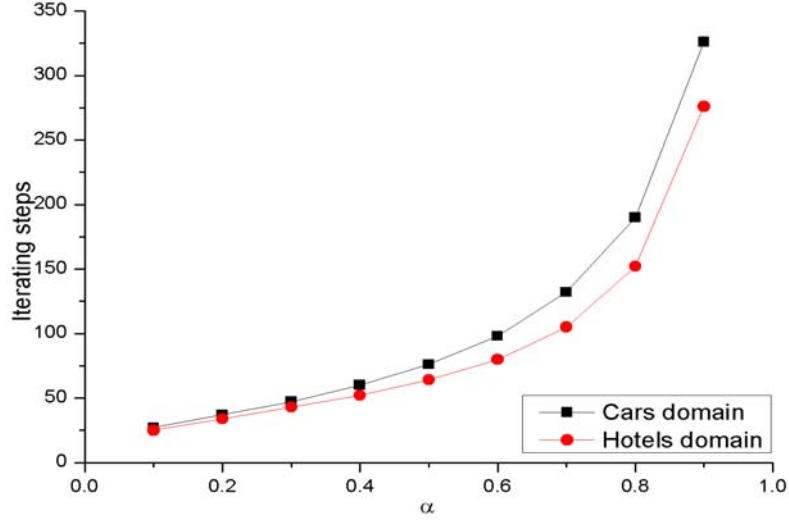
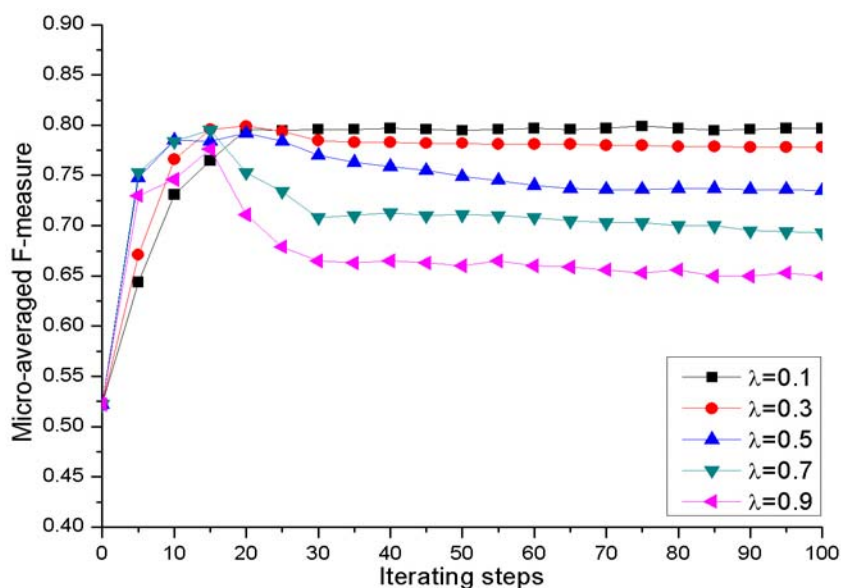
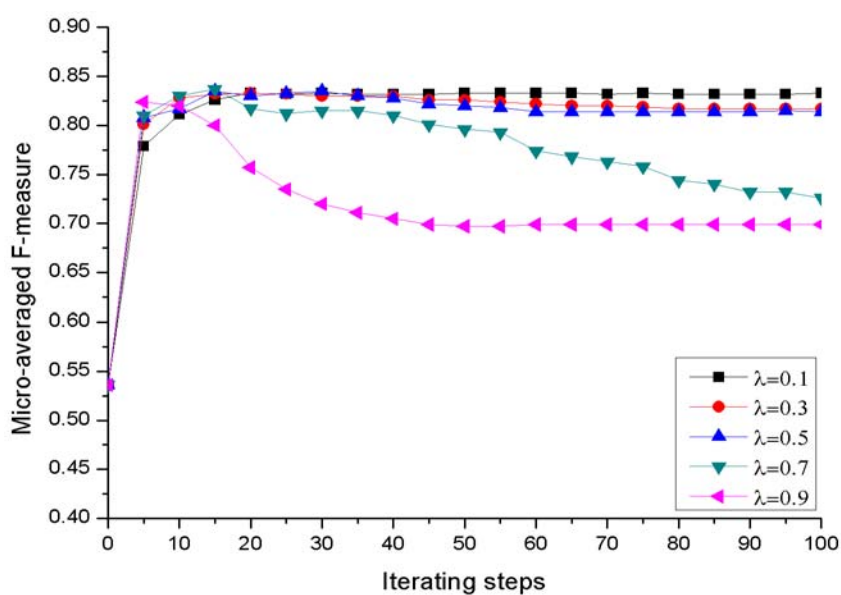


图 4.6 参数 α 取值对约束传播算法收敛迭代次数的影响

如图 4.6 所示，当 α 在 0.1 到 0.6 范围内取值时，在 Cars 和 Hotels 领域中收敛迭代次数保持平缓的增长速率。当 α 取值大于 0.6 时，约束传播算法的收敛过程逐渐开始振荡起来，导致在 Cars 和 Hotels 领域中，约束传播算法达到收敛所需要的迭代次数急剧增加。因此，我们取 $[0.1, 0.6]$ 为参数 α 的可行取值区间，约束传播算法的收敛速率令人满意。当 α 取值为 0.1 时，在 Cars 和 Hotels 领域中分别需要约 30 次迭代均可以达到收敛状态。已知领域情感词典中的候选情感词数量为 $|X|$ ，则每一步迭代过程的时间复杂度为 $O(|X|^3)$ 。因此在实验过程中，我们设置参数 α 值为 0.1。

参数 λ 指定了在约束标签传播算法中情感信息的传播速率。为了设置 λ 的合理取值，我们在 $(0, 1)$ 区间内测试 λ 的取值，以 0.1 为步长逐步增加。设置约束标签传播算法的迭代次数为 100，分析不同的 λ 取值对约束标签传播算法收敛性能的影响。实验中分别选取前 10 个正面和负面的种子情感词。图 4.7 和 4.8 分别描述了在 Cars 和 Hotels 领域中 λ 取值变化时，约束标签传播算法中微平均 F 值的收敛曲线。为了简明清晰，图 4.7 和图 4.8 中仅仅画出了 λ 取值为 0.1、0.3、0.5、0.7 和 0.9 时的 F 值收敛曲线。

图 4.7 Cars 领域中参数 λ 取值对微平均 F 值的影响分析图 4.8 Hotels 领域中参数 λ 取值对微平均 F 值的影响分析

从图 4.7 可以看出，在 Cars 领域中，当 λ 取值为 0.1 和 0.3 时，F 值的变化曲线在整个迭代过程中能够始终保持稳定的收敛状态，F 值收敛于最优值；当 λ 取值为 0.5 和 0.7 时，F 值曲线很快达到了最优值但却不能保持稳定的收敛过程，F 值随着迭代过程下降趋势明显；当 λ 取值为 0.9 时，F 值曲线在初始迭代时保持上升的趋势，但却无法达到最优值，随后呈现出剧烈的下降趋势，导致最终收敛于局部值。从图 4.8 可以看出，在 Hotels 领域中具有类似的现象，当 λ 取值为 0.1、0.3 和 0.5 时，F 值的

变化曲线在整个迭代过程中能够始终保持稳定的收敛状态，F 值收敛于最优值；当 λ 取值为 0.7 和 0.9 时，F 值曲线逐渐变得振荡起来，并最终收敛于局部值。我们进一步比较 λ 取值对 F 值曲线收敛速率的影响。如图 4.7 和图 4.8 所示，当 λ 取值为 0.3（在 Hotels 领域中甚至可以取值为 0.5）时，虽然 F 值曲线的收敛值稍有下降，但它与 λ 取值为 0.1 时的 F 值曲线相比，具有更快的收敛速率。因此，我们取 $[0.1, 0.3]$ 为参数 λ 的可行取值空间，在 Cars 和 Hotels 领域中都取得良好的收敛稳定性和较快的收敛速率。如图 4.7 和图 4.8 所示，当 λ 取值为 0.1 时，经过大约 20 次迭代传播之后，微平均 F 值曲线分别达到了最优值，并能够保持稳定的收敛状态。已知领域候选情感词的数量为 $|X|$ ，约束标签传播算法的每一次迭代过程的时间复杂度为 $O(|X|^2)$ 。因此在实验过程中，我们设置 λ 值为 0.1。基于约束标签传播算法的领域情感词典构建不仅具有良好的收敛稳定性，也具有较快的收敛速率。

4.6 本章小结

本章以提高算法的领域自适应性和情感词关联度计算模型的准确性为研究目标，提出了一种基于约束标签传播的领域情感词典自动构建方法。该方法通过分析领域评论文本中存在的局部上下文情感的一致性和转折性，以及情感词之间的词法关系，定义和抽取情感词之间的上下文和词法约束关系，能够额外地获取在特定领域上下文中情感词之间的依赖关系；采用约束传播算法能够有效地将局部约束关系传播到全局情感词空间，从而解决了局部约束关系的稀疏性问题；最后使用半监督的标签传播算法，在融合先验约束知识的同时，计算候选词的情感倾向性，构建领域情感词典。

在公开评论语料上的实验结果表明，约束标签传播方法有效地提高了领域情感词典自动构建的效果。与三种主流的基线方法相比，在准确率、召回率和 F 值上都取得了明显的提升。通过种子词数量影响分析表明，该方法较少地受到种子词覆盖问题的影响。通过参数设置分析表明，该方法能够保证良好的计算效率和收敛性能。此外，该方法无需任何的人工干预，能够轻易地适用于其他目标领域。因此，以上结果验证了本章提出的约束标签传播算法在领域情感词典自动构建中的有效性。

第5章 评价目标和情感词联合聚类

5.1 引言

在领域评论文本中，评价目标和情感词是构成用户意见表达的核心要素。评价目标主要包括领域对象的特征、属性、功能和部件等^[55, 112]。例如，在 Cars 领域中，评价目标包含“body design”、“style”、“inner space”和“fuel consumption”等；在 MP3 领域中，评价目标通常有“color screen”、“headphone”、“memory”和“price”等。情感词表达了用户对于目标实体的正面或者负面的评价。例如，“beautiful”、“good”、“thoughtful”、“clean”等表示正面情感词，而“bad”、“dirty”、“arrogant”、“disordered”等表示负面情感词。因此，评价目标和情感词抽取以及它们之间的匹配修饰关系计算是用户评论意见挖掘中的基本任务，已成为基于特征的意见摘要和推荐的基础。

目前大多数的研究工作^[54, 58, 70, 84, 152]只是简单地将细粒度意见挖掘划分为独立的子任务，主要侧重于评价目标和情感词的识别以及它们之间的对应关系的判定，挖掘的结果通常以离散化的“评价目标—情感词”对的形式呈现给用户。在判定评价目标与情感词之间的修饰关系时通常是以评论句子为单位，仅仅在局部上下文中抽取评价目标与情感词之间的直接共现关系，却忽略了在全局领域上下文中评价目标与情感词之间的隐含修饰关系。

由于用户评论的随意性和表达习惯不同，针对同一对象特征，不同的用户通常会采用不同的描述方式。此外，同一个表达短语在不同的领域中也可能描述不同的特征方面。因此，评价目标集合可以按照意见特征/属性划分为不同的语义方面（Aspect）^[91, 97, 118, 150]。同时，情感词集合按照与评价目标的匹配修饰关系可以划分为与不同语义方面匹配的情感词类^[57, 151]。例如在 Cars 领域中，“stylish”、“plain”、“attractive”等情感词通常用来匹配修饰“appearance”方面，而“expensive”、“cheap”和“affordable”更多地直接用来匹配修饰“price”方面。图 5.1 描述了评价目标类与情感词类之间的匹配修饰关系，其中实线表示评价目标与情感词之间的显式修饰关系，而虚线表示情感词与评价目标类的隐含修饰关系^[57]。随着用户评论领域不断多样化和评论文本规模急剧增长，抽取的评价目标和情感词数量显著增长。因此，采用手工方式对其归类费时费力，亟需进一步挖掘和分析潜在的评价目标与情感词之间的依赖关系，这也是基

于特征或主题的意见挖掘研究的根本任务。

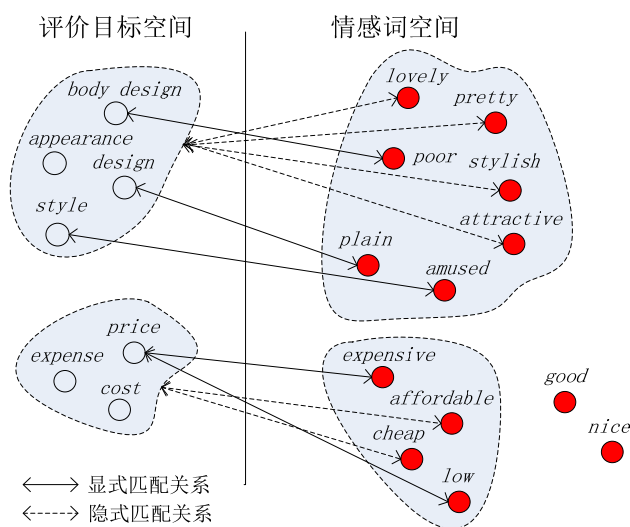


图 5.1 评价目标类与情感词类之间的匹配修饰关系示例

本章主要研究评价目标和情感词的抽取以及它们之间的匹配修饰关系计算问题。目前的研究工作主要利用无监督和有监督学习的方法以及主题建模的方法。无监督学习方法^[54, 70]主要利用人工制定的规则和词语的频率信息来识别评价目标和情感词，在局部上下文中根据距离窗口或浅层依存关系判定评价目标与情感词的匹配修饰关系。这类方法具有无监督的优点，不需要标注大量的领域训练语料，但通常性能不高，同时也会产生覆盖率不足的问题。而有监督学习方法^[58, 84]则是从训练语料中定义和抽取特征，利用分类和序列标注的方法来识别评价目标和情感词，并计算它们之间的匹配关系。相比较而言，这类方法能够取得较好的性能，但通常面临着训练语料不足和领域依赖性问题，在一个领域中训练的模型无法直接适用于其他领域。此外，特征空间的高维度和有效特征稀疏性也是重要的问题。基于主题建模的方法^[60, 124]通过在领域评论语料上的统计学习，训练主题生成模型，生成特征主题和对应的评价情感。然而，这类方法主要基于对评论语料的统计学习，生成的主题粒度较粗，同时也忽略了局部上下文特征。

针对上述问题，本章提出一种基于约束的评价目标和情感词联合聚类方法。通过采用联合聚类框架，将评价目标和情感词抽取以及它们之间的匹配修饰关系计算问题定义为基于先验约束知识的半监督学习问题。在评价目标之间相互关联的基础上，进一步引入情感词与评价目标之间的关联关系，从联合聚类的角度解决评价目标和情感词抽取问题。在评价目标之间具有相互约束信息的基础上，本章进一步根据情感词与

评价目标的关联关系引入情感词之间的约束关系,为联合聚类过程提供先验的指导知识。在满足评价目标和情感词的先验约束条件下,评价目标和情感词联合聚类通过计算评价目标与情感词之间的语义关联度,同步迭代地对评价目标按照语义方面进行聚类,对情感词按照与评价目标类的关联进行聚类。根据评价目标类与情感词类之间的关联强度抽取用户评论主题-情感联合类,从而获取评价目标与情感词之间在全局领域上下文中的匹配修饰关系。实验结果验证了本章提出的评价目标和情感词联合聚类方法的有效性。

5.2 相关研究

早期针对在线产品评论的意见挖掘研究^[54, 70, 152]主要采用无监督方法,通常直接根据产品特征和情感词之间的邻接距离和共现信息判定它们之间的匹配修饰关系。在文献^[79]中,Ding 等人针对多个具有不同倾向性的情感词同时在句子中出现的问题,提出了一种基于词典聚合的意见挖掘方法,该方法通过考虑情感词与评价目标之间的距离来衡量各情感词的权重^[2]。在文献^[85, 141]中,Qiu 等人通过人工分析和定义了八条评价目标与情感词之间的依存句法关系,利用基于句法分析的双向传播算法,迭代地同时扩充评价目标集合和情感词集合。

文献^[58, 84, 119]采用监督分类的方法判定评价目标与情感词之间的匹配修饰关系。Zhang 和 Wu 等人^[84, 119]利用浅层依存分析构造树核空间,基于规则和先验词典抽取候选评价目标和情感词,并通过训练 SVM 分类器判定它们之间的匹配对应关系。Jiang 等人^[58]针对意见挖掘的两阶段任务进一步地提出了四类树核空间,包括目标情感树、泛化的目标情感树、带边界标记的泛化目标情感树和带极性标记的泛化目标情感树。并通过将上述四类树核与多项式核进行组合构成复合核,进一步提高了系统的性能。然而,这些研究工作仅仅考虑了在局部上下文中评价目标与情感词之间的直接修饰关系,而忽略了在全局领域上下文中的隐含修饰关系。

近年来,主题模型逐渐被广泛地应用于文本意见挖掘研究领域。在文献^[60]中,Brody 和 Elhadad 提出了一种无监督的评价方面和用户情感抽取方法,该方法首先采用一种句子级别局部 LDA 主题模型,通过设定一定数目的主题来抽取评价方面;然后采用一种无监督方法生成正负面情感词的种子集,通过极性传播识别产品特征关联的情感词,判定情感倾向性。该方法能够有效地处理多领域和多语言数据集,并考虑了产品特征对情感倾向的影响。Zhao 等人^[81]提出了一种 MaxEn-LDA 混合模型,能够

同时抽取评价目标 and 对应情感词。然而，这些方法主要基于对评论语料的统计学习，生成的主题粒度较粗，并不能很好地对应细粒度的产品特征类；同时忽略了局部上下文特征，无法表示以短语形式存在的评价目标。

在文献^[57]中，Su 等人针对细粒度意见挖掘中产品特征与情感词之间的隐含情感关联抽取问题，提出一种基于相互增强（Mutual Reinforcement）的产品特征和情感词迭代聚类方法。该方法首先通过融合产品特征和情感词的内容相似度及两者之间的情感关联相似度，同步迭代地对产品特征和情感词进行聚类，在聚类过程中也加入了语义和文本结构知识；进而根据聚类结果，从产品特征类与对应的情感词类中识别前 n 个最强烈的情感联接，构建产品特征与情感词的情感关联集合。在文献^[151]中，Fu 等人采用基于信息理论（Information-theoretic）的联合聚类方法，该方法通过构建评价目标与情感词之间的关联矩阵，对评价目标和情感词进行联合聚类。然而，这些研究工作虽然考虑到了评价目标与情感词之间的隐含修饰关系，却只是简单地利用它们之间的直接关联信息，而忽略了在词法和上下文级别存在的先验约束知识。本章中，我们通过构建评价目标之间以及情感词之间的先验约束关系，增强它们之间的领域关联性，为联合聚类过程提供指导知识，从而提高它们之间匹配修饰关系计算的准确率。

5.3 联合聚类

联合聚类（Co-clustering），也称为二部聚类（Bi-clustering）、协同聚类，是指从多个维度对数据进行整体全局性的聚类。与传统的单向聚类相比，由于联合聚类充分利用了数据的整体信息，因而其具有全局最优解的优越性，更利于发现数据中隐藏的聚类结构。联合聚类最初用于基因表达、文本分析等领域^[153, 154]，可以同时为基因和所处表达环境或者文本和词汇进行聚类。联合聚类从数据的整体分布出发，聚类结果不仅能从任何单一维度学习到新的聚类知识，同时还能结合两种类型甚至多种类型，自动地挖掘出两种类型甚至多种类型上聚簇之间的对应关系及其相关知识^[155]。一般情况下，对于使用矩阵方式表达的训练数据，当行和列同时具有相关性时，应当考虑使用联合聚类，因为无论从哪一个维度进行单独聚类时，都会忽略另一维的相关信息^[156]。图 5.2 给出了联合聚类的结果示意图，其中子图 a 为行和列类型数据的聚簇划分，子图 b 为行和列聚簇的对应关系，称为联合类（Co-cluster）。

联合聚类的基本原理是通过行聚类和列聚类两个步骤进行循环迭代直至收敛。联合聚类不仅能对特征空间进行降维，在达到高效的同时又尽可能多地利用丰富的数据

信息来达到更好的效果。随着联合聚类广泛地应用于自然语言处理、图像语音及视频分析、协同过滤推荐等领域，研究者们提出了许多优秀的实现算法^[157-163]。Dhillon^[154]在 2001 年最早提出了关于文本挖掘的联合聚类算法，并于 2003 年提出了一种以 Kullback-Leibler (KL) 距离最小为标准的联合聚类方法^[164]。Banerjee 等人^[163]提出了一种同时考虑类别内部均值和行与列向量全局均值的联合聚类方法，采用最小 Bregman 信息原则进行一般化最大信息熵与标准平方，以得到最好的矩阵逼近。

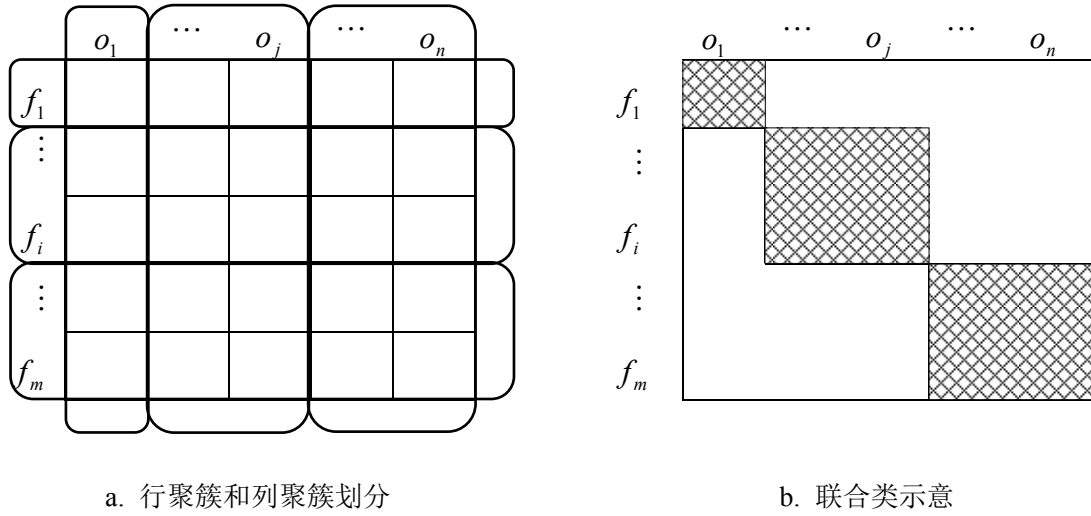


图 5.2 联合聚类示意图

近年来，融合领域经验知识的约束联合聚类研究逐渐引起了研究者的关注^[165-169]。约束联合聚类通过引入对象 (Object) 和属性 (Feature) 以及聚类参数的约束关系，为联合聚类过程提供先验指导知识，从而减少数据稀疏和噪音的影响，进一步提高聚类效果。Pensa 等人在文献^[166, 170]中提出了基于约束条件的联合聚类框架，通过引入对象之间以及属性之间的正向 (Must-link) 和反向 (Cannot-link) 约束关系，为联合聚类过程提供先验指导知识。针对基因表达数据分析中的时序依赖问题，他们进一步引入了对象之间以及属性之间的间序 (Interval) 约束关系。Shi 等人^[171]提出了一种基于谱约束模型的约束联合聚类求解算法，该算法将目标函数定义为矩阵迹 (Trace) 的最小化问题。Song 等人在文献^[167, 168]中采用两面隐马尔科夫随机场 (HMRF) 对文本之间以及词之间的约束关系建模，提出了基于监督和无监督约束条件的文本联合聚类模型，并利用 EM 算法优化求解约束联合聚类模型。

5.4 评价目标和情感词抽取与关联度计算

5.4.1 评价目标和情感词联合聚类系统框架

图 5.3 描述了基于约束联合聚类的评价目标和情感词抽取系统的整体框架。针对特定的产品主题，系统首先从 WWW 上抓取相关的用户评论网页，抽取其中的产品描述信息、用户评级、评论文本正文和一些半结构化的用户标签等数据，并将它们存入评论语料库。然后对评论文本进行解析处理，主要包括对评论文本进行段落和句子切分、对句子进行分词和词性标注、修改拼写错误和清理无意义的符号和标签、对单词取词干操作并对词性进行简化归并等。之后，从评论文本的解析结果中抽取候选评价目标词集合和情感词集合，并对其执行过滤操作，去除错误和噪声数据。我们进一步地根据评论文本上下文抽取评价目标之间以及情感词之间的约束关系，包含正向和反向约束关系。最后利用先验约束关系，我们对评价目标集合和情感词集合进行联合聚类，对评价目标按照语义主题或方面聚类，对情感词按照与评价目标类的关联聚类，进而根据评价目标类与情感词类之间的关联强度，抽取用户评论主题-情感联合类集合。利用得到的评价目标与情感词之间的匹配修饰关系，能够更加有效地对在线评论进行基于特征的细粒度意见挖掘。

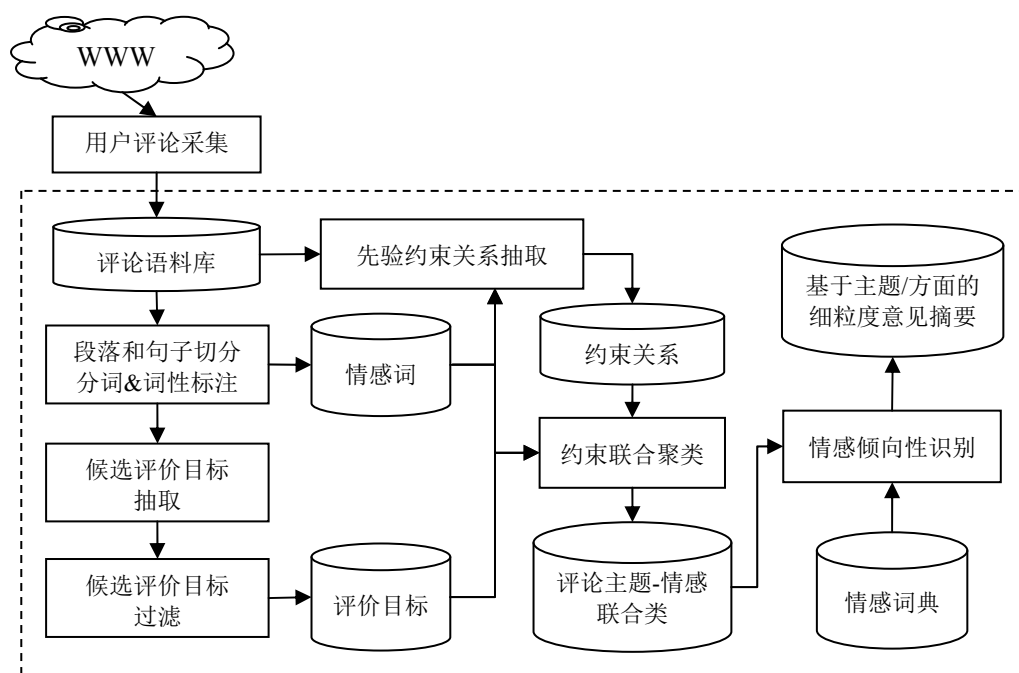


图 5.3 评价目标和情感词联合聚类系统框架

5.4.2 候选评价目标和情感词抽取

目前的研究工作^[15, 54, 172, 173]中，通常直接以词性为基础，选择名词和名词短语作

为评价目标。然而，并不是所有的名词和名词短语都是有效的评价目标，这种简单的处理策略会引入大量的噪音，即将一些普通名词错误识别为评价目标。因此，必须要对抽取的评价目标集合进行筛选和过滤。在本章中，我们首先抽取领域评论文本中的名词和名词短语作为候选评价目标词。为了降低候选评价目标空间的稀疏性和去除错误和噪音数据，采用以下策略对候选评价目标集合进行过滤：

(1) 停用词和不规则表达方式处理。首先去除候选评价目标中包含的常见停用词、连词和代词等。由于评论文本的随意性，经常会出现不规范的表达被分词工具错误地识别为名词的情况，如“gb”、“^_^”等，我们采用规则的方式，建立特殊符号表，并且限制评价目标词的长度范围来过滤候选名词。

(2) 命名实体识别。在产品评论中，一些与产品相关的品牌、机构等名词会大量出现，例如 mp3 领域中的“apple”、“ISO”、“ipod”等。这类名词与评价目标非常相似，修饰产品特征的评价词基本也可以修饰这些名词。Qiu 等人^[85]采用基于语法规则的方法，将比较句中的核心名词识别为品牌和机构名，这种方法准确率较低。本章中，我们采用 Stanford 命名实体识别工具，去除候选评价目标词集合中的人名、地点名、机构名和品牌名以及其它的通用名词。

(3) 基于出现频率过滤。为了去除大量偶尔出现的低频词，我们也根据出现频率对候选评价目标集合进行过滤。设定频率阈值为 T_F ，统计候选评价目标在领域评论文本中出现的频率，仅保留频率大于 T_F 的词汇。

情感词表达了用户对评价目标的正面或负面的主观情感倾向。在相关文献^[4, 6, 54, 85, 135, 137]中，通常直接选择形容词和动词作为情感词。因此在本章中，我们首先抽取领域评论文本中的形容词和动词作为候选情感词。为了降低领域情感词空间的高维性和稀疏性，提高抽取结果的精确性，采用以下策略对候选情感词集合进行过滤：

(1) 基于通用情感词典过滤。目前已知一些由人工整理的通用情感词典^[28, 29, 140]，其中包含了所有具有主观情感倾向的词语。文献^[140]中给出了通用的 hltmnlp05clues 情感词典，其中涵盖了手工构建的八千多个主观性词项，并分别标注了情感倾向和强度。因此，我们利用 hltmnlp05clues 对候选情感词集合进行过滤，仅保留具有强烈主观性（“strongsubj”）的词项。

(2) 基于出现频率过滤。同时我们也根据出现频率对候选情感词集合进行过滤。设定频率阈值为 T_O ，统计候选情感词在领域评论文本中出现的频率，仅保留频率大于 T_O 的候选情感词项。

5.4.3 关联矩阵构建

计算评价目标和情感词之间的关联矩阵 E 是进行联合聚类的基础。通过对用户评论文本进行分析，我们可以利用两种先验共现信息计算评价目标与情感词之间的关联度：评论句子级别共现和评论文档级别共现。

评价目标与情感词在评论句子级别的共现。评论句子是用户发表主观意见的基本单位，通常包含了评价目标及对应的情感词，情感词修饰评价目标。因此，本章采用基于领域评论语料的关联度计算方法，计算评价目标和情感词在句子级别的共现关联度。以评论句子为单位，采用点互信息（PMI）公式计算评价目标 f_i 和情感词 o_j 之间的关联度值，如公式(5.1)所示：

$$\bar{E}_{ij} = PMI(f_i, o_j) = \log_2 \frac{p(f_i, o_j)}{p(f_i)p(o_j)} \quad (5.1)$$

其中元素 $\bar{E}_{ij} \in \bar{E}$ 表示 f_i 和 o_j 在句子级别的共现关联度； $p(f_i, o_j)$ 表示 f_i 和 o_j 在评论句子中的共现概率， $p(f_i)$ 和 $p(o_j)$ 分别表示 f_i 和 o_j 的独立出现概率。

评价目标与情感词在文档级别的共现。采用向量空间模型（VSM）表示评价目标集合 F 和情感词集合 O ，采用词频-逆文档频率（TF-IDF）公式分别计算评价目标和情感词在用户评论文本中的权重。定义 W_F 表示 F 在评论文本集合 T 上的权重矩阵，定义 W_O 表示 O 在 T 上的权重矩阵，则 W_F 和 W_O 分别表示评价目标之间和情感词之间在文档级别的共现关联。定义 F 和 O 在文档级别的共现关联矩阵如下^[151]：

$$\tilde{E} = W_F^T \cdot W_O \quad (5.2)$$

其中元素 $\tilde{E}_{ij} \in \tilde{E}$ 表示 f_i 和 o_j 在文档级别的共现关联度。

综合句子级别的共现关联矩阵 \bar{E} 和文档级别的共现关联矩阵 \tilde{E} ，定义 F 与 O 之间的关联矩阵 E 为：

$$E = \alpha \bar{E} + (1 - \alpha) \tilde{E} \quad (5.3)$$

其中，元素 $E_{ij} \in E$ 表示 f_i 和 o_j 之间的关联度，参数 $0 \leq \alpha \leq 1$ 用于平衡句子级别共现关联度和文档级别共现关联度的相对权重^[151]。

5.5 基于约束的联合聚类算法

5.5.1 基于约束的联合聚类建模

给定领域评论文本集合 T ，已知抽取的用户评价目标集合 $F = \{f_1, \dots, f_i, \dots, f_M\}$ ，情

感词集合 $O = \{o_1, \dots, o_j, \dots, o_N\}$ ，其中 M 表示评价目标的数量， N 表示情感词的数量。将评价目标集合 F 与情感词集合 O 之间的关联表示为图模型如下：定义二部图 $G = (F, O, E)$ ，其中 E 为评价目标与情感词之间的关联矩阵，如公式(5.4)所示，其中 E_{ij} 表示评价目标 f_i 和情感词 o_j 之间的关联度。则评价目标和情感词的联合聚类旨在基于关联矩阵 E ，对评价目标集合 F 按照主题或语义方面聚类，对情感词集合 O 按照与评价目标类的关联聚类。

$$E = [E_{ij}]_{M \times N} = \begin{bmatrix} E_{1,1} & \cdots & E_{1,N} \\ \vdots & \vdots & \vdots \\ E_{M,1} & \cdots & E_{M,N} \end{bmatrix} \quad (5.4)$$

基于约束的联合聚类假定结点之间存在着先验的约束关系。通过定义和抽取评价目标之间和情感词之间的正向和反向约束关系，能够为联合聚类过程提供先验性指导知识。评价目标之间的正向约束定义为两者属于相同评价目标类的先验关系，反向约束定义为两者属于不同评价目标类的先验关系。定义评价目标之间的正向约束关系矩阵 D_F 如下：

$$[D_F]_{ij} = \begin{cases} 1, & f_i \text{与} f_j \text{具有正向约束关系} \\ 0, & \text{未知情况} \end{cases} \quad (5.5)$$

其中，评价目标 f_i 和 $f_j \in F$ ， $[D_F]_{ij} = 1$ 表示 f_i 与 f_j 属于相同评价目标类的正向约束关系， $[D_F]_{ij} = 0$ 表示未知情况。定义评价目标之间的反向约束关系矩阵 R_F 如下：

$$[R_F]_{ij} = \begin{cases} 1, & f_i \text{与} f_j \text{具有反向约束关系} \\ 0, & \text{未知情况} \end{cases} \quad (5.6)$$

其中， $[R_F]_{ij} = 1$ 表示 f_i 与 f_j 属于不同评价目标类的反向约束关系。同理，情感词之间的正向约束定义为两者关联于相同评价目标类的先验关系，反向约束则定义为两者分别关联于不同评价目标类的先验关系。分别定义情感词之间的正向约束关系矩阵 D_O 和反向约束关系矩阵 R_O 如下：

$$[D_O]_{ij} = \begin{cases} 1, & o_i \text{与} o_j \text{具有正向约束关系} \\ 0, & \text{未知情况} \end{cases} \quad (5.7)$$

$$[R_O]_{ij} = \begin{cases} 1, & o_i \text{与} o_j \text{具有反向约束关系} \\ 0, & \text{未知情况} \end{cases} \quad (5.8)$$

其中， $[D_O]_{ij} = 1$ 表示 o_i 与 o_j 关联于相同评价目标类的正向约束关系， $[R_O]_{ij} = 1$ 表示 o_i 与 o_j 分别关联于不同的评价目标类的反向约束关系。

图 5.4 示例了评价目标与情感词之间的关联空间、评价目标之间的约束关系空间 C_F 以及情感词之间的约束关系空间 C_O 。其中, C_F 包含评价目标之间的正向约束关系矩阵 D_F 和反向约束关系矩阵 R_F , C_O 包含情感词之间的正向约束关系矩阵 D_O 和反向约束关系矩阵 R_O 。

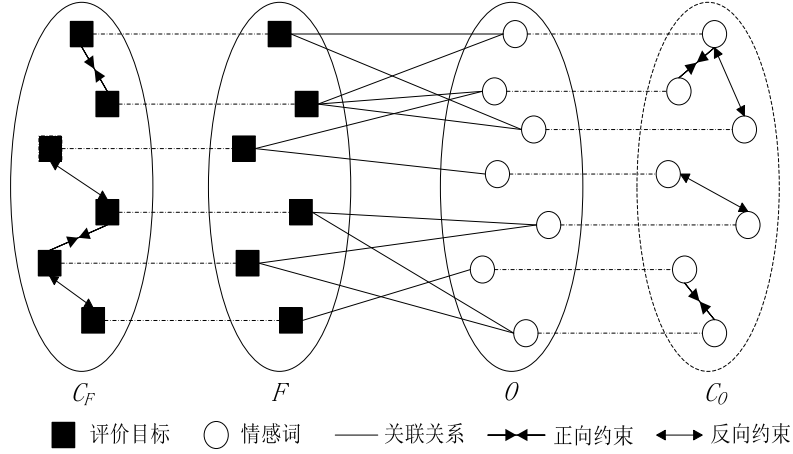


图 5.4 评价目标与情感词之间的关联空间及约束关系空间示例

假设评价目标集合 F 中包含 k 个评价目标类, 情感词集合 O 中包含 l 个关联于评价目标类的情感词类。基于约束的联合聚类旨在满足评价目标之间约束关系矩阵 D_F 和 R_F 以及情感词之间约束关系矩阵 D_O 和 R_O 的条件下, 将 F 划分为 k 个评价目标类, 将 O 划分为 l 个关联于评价目标类的情感词类。为了计算评价目标类与情感词类之间的匹配对应关系, 我们取 $l = k$ 。

5.5.2 约束关系的定义和抽取

5.5.2.1 评价目标约束关系的定义和抽取

以评价目标的词法结构和在评论文本中的上下文信息为基础, 定义了评价目标之间的正向约束和反向约束关系。我们将评价目标之间包含共同词的词法关系定义为正向约束关系, 正向约束关系从词法角度衡量了评价目标之间属于相同评价目标类的先验关系。通过对用户评论习惯和评论文本上下文结构的分析, 我们抽取评价目标之间的反向约束关系, 反向约束关系从上下文的角度衡量了评价目标之间属于不同评价目标类的先验关系。

本文在第 3 章 3.3.1 节中详细描述了评价目标之间正向和反向约束关系的定义和抽取规则, 在此不再赘述。

5.5.2.2 情感词约束关系的定义和抽取

针对一个评价目标，用户的情感倾向完全依赖于个人的主观判断。由于情感词仅仅关联于修饰的评价目标对象，直接构建情感词之间的约束关系比较困难。因此，由文献^[167, 168]中根据命名实体重合度抽取文档约束关系的启发，我们间接地利用情感词对评价目标的关联关系抽取情感词之间的约束关系。已知 FS_i 和 FS_j 分别表示情感词 o_i 和 o_j 关联的评价目标集合，定义 FS_i 与 FS_j 之间的重合度为：

$$Overlap(FS_i, FS_j) = \frac{|FS_i \cap FS_j|}{|FS_i| + |FS_j| - |FS_i \cap FS_j|} \quad (5.9)$$

定义参数 λ 为抽取情感词约束关系的阈值参数，如果 o_i 和 o_j 分别关联的评价目标集合的重合度 $Overlap(FS_i, FS_j) \geq \lambda$ 时，则定义 o_i 和 o_j 之间具有正向约束关系。在实验过程中， λ 的取值为 0.8。此外，由于已知的情感词与评价目标之间的直接关联关系并不完备，无法直接定义和抽取情感词之间的反向约束关系，抽取的准确率偏低。为了避免引入过多的误差和噪声，在本章中，我们仅引入情感词之间的正向约束关系。

5.5.3 约束联合聚类的求解算法

由于同时定义和抽取了评价目标约束关系和情感词约束关系，我们采取将基于约束的联合聚类求解转化为残值平方和(Sum of Squared Residues)最小化问题^[170]的求解策略。已知关联矩阵 E ，其中行维度表示评价目标，列维度表示情感词。定义集合 I 包含属于同一个评价目标类的行结点，集合 J 包含属于同一个情感词类的列结点，则 I 和 J 唯一确定子矩阵 $E_{IJ} \in E$ 称为联合类。已知元素 $E_{ij} \in E$ ，定义 E_{ij} 在联合类 E_{IJ} 上的残值(Residue) H_{ij} ^[166, 170]如下：

$$H_{ij} = E_{ij} - E_{Ij} - E_{iJ} + E_{IJ} \quad (5.10)$$

其中 $E_{IJ} = \frac{\sum_{i \in I, j \in J} E_{ij}}{|I| \cdot |J|}$ ， $E_{Ij} = \frac{\sum_{i \in I} E_{ij}}{|I|}$ ， $E_{iJ} = \frac{\sum_{j \in J} E_{ij}}{|J|}$ ， $|I|$ 为集合 I 的基数。定义 $H = [H_{ij}]_{M \times N}$ 为 E 对应的残值矩阵。则联合聚类的优化目标函数定义为：

$$\|H\|^2 = \sum_{I, J} \|H_{IJ}\|^2 = \sum_{I, J} \sum_{i \in I, j \in J} H_{ij}^2 \quad (5.11)$$

基于约束的联合聚类的求解目标是在满足先验约束条件的前提下，使得优化目标函数值最小化。我们采用文献^[166, 170]提出的通用约束联合聚类求解框架，在初始化行聚类划分矩阵和列聚类划分矩阵之后，迭代地优化行聚类和列聚类直至收敛。每次迭

代过程中，在满足先验约束条件的前提下，将行/列结点划分到距离最近的行/列类簇中，使得目标函数值逐步递减。

5.6 实验验证和结果讨论

5.6.1 数据准备和评价标准

实验部分采用 Amazon MP3⁸和 OpinRankDataset 数据集作为领域评论语料，对本章提出的基于约束的评价目标和情感词联合聚类效果进行验证。Amazon MP3 数据集包含从 Amazon 网站抓取的关于 MP3 产品的用户评论文本以及详细的评论元数据信息^[174]。OpinRankDataset 数据集包含从 Tripadvisor 和 Edmunds 网站上收集的关于 Cars 和 Hotels 领域的完整用户评论文本^[148]。这两个数据集的规模相对较大，从而能够尽量避免数据稀疏带来的影响。我们选择其中 MP3 和 Cars 领域的评论语料作为验证数据集，相关的统计信息如表 5.1 所示。

表 5.1 验证数据集统计

领域	评论文本数	评论句子数	平均句子数	平均词汇数
MP3	31,000	308,020	10	188
Cars	42,288	344,105	8	124

5.6.2 实验结果分析

5.6.2.1 候选评价目标和情感词抽取

首先从领域评论语料中抽取候选的用户评价目标和情感词。在相关文献^[109, 110, 175-177]中，评价目标抽取仍然是重要的研究问题。在对候选评价目标集合进行停用词和不规则表达方式处理之后，我们衡量基于频率阈值和基于命名实体识别的过滤策略分别对提高抽取准确率的效果。由于实验语料规模较大，抽取的候选评价目标数量较多，采用手工方式对每一个评论文本进行标注无法实现。因此，我们主要考虑抽取结果的准确率。同时，也为了尽可能地保证召回率，评价目标的频率阈值 T_F 设置为 20，情感词的频率阈值 T_O 设置为 10。我们对采用不同过滤策略后得到的候选评价目标集合进行统计，检查每一个候选项的正确性，并对最终得到的评价目标集合按照语义方面进行归类。表 5.2 给出了两种评论领域中候选评价目标和情感词抽取结果统计，表 5.3 给出了不同过滤策略对候选评价目标抽取准确率的影响。

⁸ <http://timan.cs.uiuc.edu/downloads.html>.

表 5.2 评价目标和情感词抽取结果统计

领域	MP3	Cars
评价目标数	1,538	1,503
情感词数	921	770
评价目标类别数	28	26

表 5.3 不同过滤策略的候选评价目标抽取准确率统计

领域	MP3	Cars
基于频率阈值	76.51%	81.09%
命名实体识别	42.67%	54.82%
频率+命名实体识别	89.78%	91.43%

如表 5.3 所示, 基于频率阈值的过滤策略能够去除大量低频的噪声名词, 从而有效地降低了评价目标空间的高维性和稀疏性。采用命名实体识别的过滤策略能够去除一些与评论对象相关的品牌、机构等名词。仅仅采用命名实体识别过滤策略时, 由于候选评价目标的基数较大, 导致准确率偏低。当综合利用这两种过滤策略时, 取得了较好的评价目标抽取准确率。

5.6.2.2 约束关系抽取结果分析

分别统计具有正向和反向约束关系的评价目标结点和情感词结点在评价目标集合 F 和情感词集合 O 中所占的比例。表 5.4 和 5.5 分别给出了在两种评论领域中评价目标约束关系和情感词约束关系的抽取结果统计。

表 5.4 评价目标约束关系的抽取结果统计

领域	MP3	Cars
正向约束关系比例	14.73%	13.29%
反向约束关系比例	16.35%	22.57%
两种约束关系比例之和	31.08%	35.86%

如表 5.4 所示, 在两种领域中具有正向约束关系的评价目标比例近似, 而由于 Cars 领域的评论文本规模相对较大, 抽取得到的反向约束关系比例也相对较高。由于评价目标的正向约束关系主要基于词法分析抽取, 具有正向约束关系的评价目标比例直接依赖于候选评价目标集合的规模和分布, 而并不直接受到领域评论文本规模的影响。

而反向约束关系基于评价目标在句子和评论文本单元中的分布关系抽取，具有反向约束关系的评价目标比例直接依赖于领域评论文本的规模。

表 5.5 情感词约束关系的抽取结果统计

领域	MP3	Cars
正向约束关系比例	20.78%	12.93%

由表 5.5 可知，在 Cars 领域中具有正向约束关系的情感词比例小于在 MP3 领域中具有正向约束关系的情感词比例。由于仅仅根据与评价目标的关联关系抽取情感词之间的先验约束关系，具有约束关系的情感词比例并不直接受到领域评论文本规模的影响。在实验过程中，我们仅利用情感词之间的正向约束关系。

5.6.2.3 评价目标聚类结果分析

实验部分采用 Rand Index^[88, 178]作为评价目标聚类效果的衡量标准。已知评价目标集合 F 和类别数 K ，人工标注的标准类划分为 $P = \{p_1, \dots, p_j, \dots, p_K\}$ ，基于约束的联合聚类算法将 F 划分为 K 个独立的结果类 $Q = \{q_1, \dots, q_i, \dots, q_K\}$ 。Rand 指数的准确率衡量了聚类结果 Q 和标准类划分 P 之间的一致性程度。Rand 指数计算公式如下：

$$Rand(P, Q) = \frac{a + b}{M(M-1)/2} \quad (5.12)$$

其中， M 为评价目标数量， $M(M-1)/2$ 表示 F 中评价目标之间的两两配对数目， a 表示配对的评价目标同时被 P 和 Q 划分到相同产品特征类的配对数， b 表示配对的评价目标同时被 P 和 Q 划分到不同产品特征类的配对数。

为了评估基于约束的联合聚类算法对评价目标聚类的准确率，引入两种对比的基线方法：K-均值(K-means)是传统聚类算法的典型代表，为了验证传统聚类算法对评价目标聚类的有效性，引入 K-均值作为一种基线方法；为了验证我们引入的约束关系对提高评价目标聚类准确率的贡献，采用不带任何先验约束条件的基于信息论的联合聚类算法^[164](Co-clustering)作为另一种基线方法。本章提出的基于约束的联合聚类方法表示为 Cons+Co-clustering。表 5.6 给出了不同聚类算法在两种评论领域上对评价目标聚类的准确率比较。

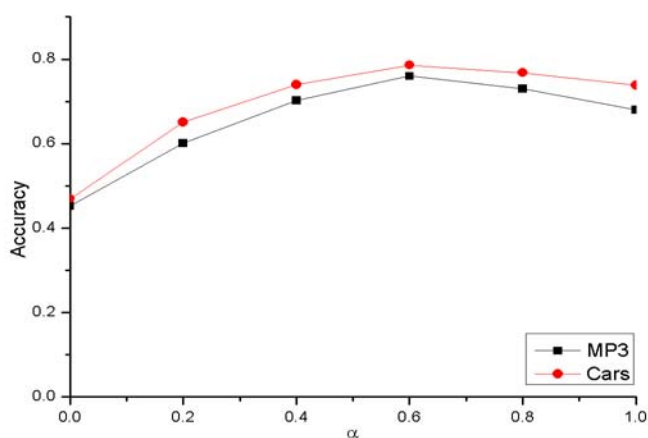
表 5.6 不同聚类算法的准确率比较

领域	MP3	Cars
K-means	0.694	0.712
Co-clustering	0.728	0.759
Cons+Co-clustering	0.761	0.786

如表 5.6 所示, 与 K-means 和 Co-clustering 算法相比, Cons+Co-clustering 算法在两种评论领域上都取得了最高的准确率值, 评价目标聚类结果与人工标注标准的一致性达到最高。其中, K-means 算法对评价目标的聚类效果最差, 在两种评论领域上都取得了最低的准确率值。Co-clustering 算法对评价目标的聚类准确率高于 K-means 算法。与 Co-clustering 算法相比, Cons+Co-clustering 算法的聚类准确率取得了明显的提高, 在 MP3 领域上的准确率提高了 3.3%, 在 Cars 领域上提高了 2.7%。

5.6.2.4 参数影响分析

本节分析参数 α 的取值对评价目标聚类准确率的影响。 α 的取值范围从 0 到 1, 以 0.2 为步长逐步增加。针对不同的 α 取值, 基于约束的联合聚类算法在两种评论领域上对评价目标聚类的准确率变化如图 5.5 所示。

图 5.5 α 值对评价目标聚类准确率的影响

如图 5.5 所示, 当 $\alpha = 0.6$ 时基于约束的联合聚类算法在两种评论领域上几乎都取得了最高的准确率值。当 $\alpha = 1$ 时, 联合聚类仅仅利用了评价目标与情感词在句子级别的共现关联度, 在两种评论领域上的聚类准确率均高于仅利用在文档级别的共现关联度时 ($\alpha = 0$) 的聚类结果。因此, 句子级别的共现关联度能够直接计算情感词与评价目标的相互关联关系, 对于聚类准确率的贡献大于文档级别的共现关联度。在实验过

程中，取参数 α 的经验值为 0.6。

5.6.2.5 匹配修饰关系识别分析

在对评价目标和情感词联合聚类以后，我们进一步识别评价目标类与情感词类之间的匹配修饰关系。我们统计评论语料中评价目标与情感词之间具有直接依存关系的出现频率，如果某一评价目标类与情感词类中词项之间直接依存的频率最高，则判定该评价目标类与情感词类构成匹配修饰关系，情感词类中包含的所有情感词都能够匹配修饰对应的评价目标类。如果情感词与评价目标之间不存在局部的直接修饰关系，则识别为全局领域的隐含修饰关系。为了衡量约束联合聚类算法识别评价目标与情感词之间匹配修饰关系的效果，引入评价目标与情感词之间的直接共现关系作为基线方法。采用准确率（Precision）作为评价指标，抽取评价目标与情感词之间存在的直接依存关系（包括“amod”、“nsubj”等类型）构建评估集。对于情感词与评价目标之间的隐含修饰关系，采用人工方式判定其正确性。同时，对于“good”、“terrible”等通用情感词，无论搭配任何评价目标类别都判定为正确的匹配关系。表 5.7 给出了在两种评论领域中评价目标与情感词之间匹配修饰关系的抽取结果和准确率统计。

表 5.7 评价目标与情感词匹配修饰关系抽取结果和准确率统计

领域	匹配修饰关系数量		准确率（Precision）	
	MP3	Cars	MP3	Cars
直接共现方法	61,593	58,382	69.12%	72.53%
约束联合聚类方法	80,756	55,740	77.68%	80.26%

如表 5.7 所示，直接共现方法识别评价目标与情感词的近邻关系，具有较大的偶然性，导致准确率偏低。基于约束的联合聚类方法能够识别评价目标与情感词之间的隐含修饰关系，与直接共现方法相比，在 MP3 和 Cars 领域上匹配修饰关系识别的准确率分别提高了 8.56% 和 7.73%。

5.7 本章小结

本章引入了评价目标与情感词之间的关联关系，对评价目标和情感词进行联合聚类建模，将评价目标和情感词联合抽取以及匹配修饰关系计算问题转化为基于约束的半监督学习过程。通过定义和抽取评价目标之间以及情感词之间的约束关系，为联合聚类过程提供先验性的指导知识。在满足先验约束条件下，同步迭代地对评价目标和

情感词进行联合聚类，并利用评论语料中评价目标与情感词之间先验的直接依存关系计算评价目标类与情感词类之间的匹配修饰关系。

实验结果表明，基于约束的联合聚类方法有效地解决了评价目标和情感词抽取以及它们之间的匹配修饰关系识别问题，取得了比基线方法更高的准确率。本章提出的方法本质上属于无监督学习策略，避免了有监督学习通常面临的标注训练语料缺乏和无法解决跨领域问题；同时通过定义和抽取先验的约束知识，增强算法的领域性，解决有效关联信息不足问题。

第6章 总结和展望

随着 Web 评论文本的爆炸式增长, 文本意见挖掘已经成为智能信息处理、数据挖掘、计算语言学等领域的热点研究课题。与传统的信息抽取、文本分类和自动摘要等技术相比, 文本意见挖掘更加接近人工智能目标, 并可以与上述技术相结合, 广泛地使用在互联网的各种智能应用中。通过挖掘评论文本中的主观性信息, 快速了解人们对客观事物和主题的态度和观点, 为我们提供了自动分析和理解互联网群体智能的途径。评论文本意见挖掘技术具有巨大的社会和经济价值, 已经广泛地应用于电子商务、商业智能、舆情分析和企业管理等领域。

本文针对 Web 评论文本的细粒度意见挖掘面临的特征空间较大、数据稀疏、有效特征不足、自动化程度不高和领域依赖性问题, 从细粒度意见元素抽取、自适应评价目标聚类、领域情感词典自动构建、评价目标和情感词联合聚类四个方面开展研究工作。具体来说, 本文的贡献主要表现在以下四个方面:

(1) 针对细粒度意见挖掘面临的特征缺乏以及多类型特征的有效融合问题, 从序列标注学习的角度出发, 提出了融合句法语义结构特征的意见元素抽取方法。在词汇级别和上下文级别标注特征的基础上引入句法语义结构特征, 可有效地利用长距离依存句法结构关系, 解决了有效标注特征的稀疏问题。将评价目标和情感词抽取任务转化为序列标注学习过程, 采用条件随机场 CRFs 模型构建统一的抽取框架, 解决了多种标注特征的有效融合问题。利用语义词汇库和分布上下文计算评价目标之间的领域相似度, 然后采用图剪枝算法进行评价目标分类。实验结果表明本文提出的意见元素抽取方法有效地提高了评价目标的抽取和分类效果。

(2) 针对计算评价目标之间关联度的领域依赖性和有效关联信息不足问题, 提出了一种基于约束的谱聚类方法, 将评价目标聚类任务转化为基于约束的半监督学习过程。通过对评价目标自身词法结构、用户评论习惯以及评论文本局部上下文进行分析, 定义和抽取评价目标之间的先验约束关系, 并采用约束传播将局部约束关系传播到全局评价目标空间, 从而增强了评价目标之间的领域关联性, 解决了评价目标之间相似度的领域依赖性问题。采用约束谱聚类算法, 在聚类操作之前将评价目标映射到低维特征向量空间, 从而降低了聚类特征空间的高维性和稀疏性。实验结果表明, 本文提出的基于约束的谱聚类方法有效解决了评价目标聚类问题, 取得了比基线方法更

好的实验效果。

(3) 针对情感词典构建面临的算法领域适应性差、自动构建程度不高和准确率低下问题, 提出了基于约束标签传播的领域情感词典自动构建方法。针对情感词的领域性, 利用组块依存知识和先验通用情感词典抽取领域候选情感词和短语。通过分析领域评论文本中局部上下文情感的一致性和转折性, 以及情感词的词法关系, 定义和抽取情感词之间的上下文和词法约束关系, 额外地获取在特定领域中情感词倾向之间的依赖关系; 采用约束传播算法能够有效地将局部约束关系传播到全局情感词空间, 从而解决了局部约束关系的稀疏性问题; 最后使用半监督的标签传播算法, 在融合先验约束知识的同时, 计算候选词的情感倾向性, 构建领域情感词典。实验结果表明, 约束标签传播方法有效提高了领域情感词典自动构建的准确率, 受种子词覆盖的影响较小, 并具有良好的计算效率和收敛性能。

(4) 针对细粒度意见挖掘中评价目标和情感词抽取以及匹配修饰关系计算问题, 提出了评价目标和情感词联合聚类方法。在该任务中, 传统有监督方法通常面临着需要标注大量领域训练语料的问题, 而无监督方法又存在着数据稀疏和领域依赖引起的准确率低下问题。因此, 本文引入评价目标与情感词的关联关系, 对评价目标和情感词进行约束联合聚类建模, 将评价目标和情感词联合抽取以及匹配修饰关系计算问题转化为基于约束的半监督学习过程。通过定义和抽取评价目标之间以及情感词之间的约束关系, 为联合聚类过程提供先验性的指导知识。在满足先验约束的条件下, 同步迭代地对评价目标和情感词进行联合聚类, 并计算评价目标类与情感词类之间的匹配修饰关系。实验结果表明, 基于约束的联合聚类方法有效解决了评价目标和情感词抽取以及它们之间的匹配修饰关系判定问题, 取得了比基线方法更高的聚类准确率。

本文针对目前 Web 评论文本细粒度意见挖掘存在的问题, 在以上四个方面提高了原有方法的性能, 取得了一些积极有效的成果。然而随着研究的深入, 仍然存在一些问题需要解决, 未来的研究工作将主要围绕以下几个方向展开:

(1) 隐式评价目标和情感识别。由于自然语言表达的多样性, 有时作者的主观评论意见并没有在评论中明确提出, 而是隐式地蕴涵在上下文中。进一步挖掘隐式评价目标和情感有助于更加全面地了解作者对商品或主题的主观评价。然而隐式评价目标和情感的识别是非常困难的, 这涉及到深层次的自然语言理解和语义推断。因此研究隐式评价目标和情感识别问题是下一步的研究重点。

(2) 细粒度的词语情感倾向性分析。在词语和短语级别的意见挖掘中, 一些词

汇和短语的情感倾向完全依赖于匹配的评价目标，修饰的评价目标不同，表达的情感倾向性也不同，甚至针对不同的词义也具有不同的情感倾向性。例如在智能手机领域，形容词“大”在用来修饰“屏幕”特征时，表达了用户的褒义评价，而在用来修饰“电池”特征时，则表达了用户的贬义评价。因此，如何区分基于特征和词义的词语情感倾向性是未来工作的一个主要方向。

（3）面向社交媒体的意见挖掘。随着近几年来社交媒体的兴起，面向微博的意见挖掘已经引起了学术界和工业界的广泛关注，特别是针对微博中出现的热点话题讨论，如何有效地进行话题的自动识别和倾向性分析，以及研究话题传播过程中网民情感倾向的变化过程等都是重要的研究课题。此外，微博数据具有文本内容短小、时效性强以及更加口语化的表达的特点，这对目前常用的意见挖掘方法提出了挑战。因此需要我们针对新的社交媒体环境研究有效的意见挖掘模型和方法。

参考文献

- [1]Liu B. Sentiment analysis and subjectivity[J]. Indurkha N, and Damerau FJ. Handbook of Natural Language Processing, Second Edition. CRC Press, Taylor & Francis Group, 2010: 627-666.
- [2]江鹏. 文本意见挖掘关键技术研究[D]. 北京: 北京理工大学, 2011.
- [3]Moghaddam S, Ester M. Opinion Mining in Online Reviews: Recent Trends. Simon Fraser University Tutorial at WWW2013, 2013.
- [4]Hatzivassiloglou V, McKeown KR. Predicting the semantic orientation of adjectives[C]. Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Madrid, Spain. Association for Computational Linguistics, 1997: 174-181.
- [5]Wiebe JM, Bruce RF, O'Hara TP. Development and use of a gold-standard data set for subjectivity classifications[C]. Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Association for Computational Linguistics, 1999: 246-253.
- [6]Hatzivassiloglou V, Wiebe JM. Effects of adjective orientation and gradability on sentence subjectivity[C]. Proceedings of the 18th conference on Computational linguistics - Volume 1, Saarbrücken, Germany. Association for Computational Linguistics, 2000: 299-305.
- [7]Wiebe J. Instructions for annotating opinions in newspaper articles. Department of Computer Science Technical Report TR-02-101, University of Pittsburgh, 2002.
- [8]Turney PD. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews[C]. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, Pennsylvania. Association for Computational Linguistics, 2002: 417-424.
- [9]Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques[C]. Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002: 79-86.
- [10]Ghose A, Ipeirotis PG, Sundararajan A. Opinion mining using econometrics: A case study on reputation systems[C]. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic. Association for Computational Linguistics, 2007: 416-423.
- [11]Chen H, Zimbra D. AI and opinion mining[J]. IEEE Intelligent Systems, 2010, 25(3): 74-80.
- [12]O'Connor B, Balasubramanyan R, Routledge BR, et al. From tweets to polls: Linking text sentiment

to public opinion time series[C]. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC. 2010, 11: 122-129.

[13]Joshi M, Penstein-Rosé C. Generalizing dependency features for opinion mining[C]. Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Suntec, Singapore. Association for Computational Linguistics, 2009: 313-316.

[14]Lu Y, Zhai C. Opinion integration through semi-supervised topic modeling[C]. Proceedings of the 17th international conference on World Wide Web, Beijing, China. ACM, 2008: 121-130.

[15]Hu M, Liu B. Mining opinion features in customer reviews[C]. Proceedings of the 19th national conference on Artificial intelligence, San Jose, California. AAAI Press, 2004: 755-760.

[16]Ye Q, Shi W, Li Y. Sentiment classification for movie reviews in Chinese by improved semantic oriented approach[C]. Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06). IEEE Computer Society, 2006.

[17]Boiy E, Hens P, Deschacht K, et al. Automatic Sentiment Analysis in On-line Text[C]. Proceedings ELPUB2007 Conference on Electronic Publishing, Vienna, Austria. 2007: 349-360.

[18]Durant KT, Smith MD. Mining sentiment classification from political web logs[C]. Proceedings of Workshop on Web Mining and Web Usage Analysis of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (WebKDD-2006), Philadelphia, PA, 2006.

[19]Pang B, Lee L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts[C]. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Barcelona, Spain. Association for Computational Linguistics, 2004: 271.

[20]徐军, 丁宇新, 王晓龙. 使用机器学习方法进行新闻的情感自动分类[J]. 中文信息学报, 2007, 21(6): 95-100.

[21]Tan S, Cheng X, Wang Y, et al. Adapting naive Bayes to domain adaptation for sentiment analysis[C]. Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval, Toulouse, France. Springer-Verlag, 2009: 337-349.

[22]Zagibalov T, Carroll J. Automatic seed word selection for unsupervised sentiment classification of Chinese text[C]. Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1, Manchester, United Kingdom. Association for Computational Linguistics, 2008: 1073-1080.

[23]Taboada M, Brooke J, Tofiloski M, et al. Lexicon-Based Methods for Sentiment Analysis[J]. Computational Linguistics, 2011, 37(2): 267-307.

- [24]王根, 赵军. 基于多重冗余标记 CRFs 的句子情感分析研究[J]. 中文信息学报, 2007, 21(5).
- [25]Fu G, Wang X. Chinese sentence-level sentiment classification based on fuzzy sets[C]. Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Beijing, China. Association for Computational Linguistics, 2010: 312-319.
- [26]Li L, Yao T. Kernel-based Sentiment Classification for Chinese Sentence[C]. Proceedings of the Sixth International Conference on Advanced Language Processing and Web Information Technology (ALPIT 2007), IEEE Computer Society, 2007: 27-32.
- [27]Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis[J]. Comput. Linguist., 2009, 35(3): 399-433.
- [28]Esuli A, Sebastiani F. Sentiwordnet: A publicly available lexical resource for opinion mining[C]. Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06), 2006: 417-422.
- [29]Baccianella S, Esuli A, Sebastiani F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining[C]. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA), May 2010.
- [30]朱嫣岚, 闵锦, 周雅倩等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报, 2006, 20(1).
- [31]王素格, 李德玉, 魏英杰等. 基于同义词的词汇情感倾向判别方法[J]. 中文信息学报, 2009, 23(5): 68-74.
- [32]宋晓雷, 王素格, 李红霞等. 基于概率潜在语义分析的词汇情感倾向判别[J]. 中文信息学报, 2011, 25(2).
- [33]郭冲, 王振宇. 面向细粒度意见挖掘的情感本体树及自动构建[J]. 中文信息学报, 2013, 27(5).
- [34]Wei W, Gulla JA. Sentiment Learning on Product Reviews via Sentiment Ontology Tree[C]. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden. Association for Computational Linguistics, 2010: 404-413.
- [35]Lau RYK, Lai CCL, Ma J, et al. Automatic Domain Ontology Extraction for Context-Sensitive Opinion Mining[C]. Proceedings of the Thirtieth International Conference on Information Systems(ICIS), 2009: 35-53.
- [36]Cheng X. Automatic Topic Term Detection and Sentiment Classification for Opinion Mining[D]. Master Thesis. Saarbrücken, Germany: The University of Saarland, 2007.
- [37]Lin C, He Y, Everson R, et al. Weakly-supervised Jointly Sentiment-Topic Detection from Text[J].

IEEE Transactions on Knowledge and Data Engineering, 2012, 24(6): 1134-1145.

[38]Moghaddam S, Ester M. Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews[C]. Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada. ACM, 2010: 1825-1828.

[39]施寒潇. 细粒度情感分析研究[D]. 苏州: 苏州大学, 2013.

[40]Daumé III H, Marcu D. Domain Adaptation for Statistical Classifiers[J]. Journal of Artificial Intelligence Research (JAIR), 2006, 26: 101-126.

[41]Jiang J. A literature survey on domain adaptation of statistical classifiers[J]. URL: <http://sifaka.cs.uiuc.edu/jiang4/domainadaptation/survey>, 2008.

[42]Pan SJ, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[43]Blitzer J, Dredze M, Pereira F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification[C]. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic. Association for Computational Linguistics, 2007: 440-447.

[44]Tan S, Cheng X. Improving SCL model for sentiment-transfer learning[C]. Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers, Boulder, Colorado. Association for Computational Linguistics, 2009: 181-184.

[45]Wu Q, Tan S, Cheng X. Graph ranking for sentiment transfer[C]. Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Suntec, Singapore. Association for Computational Linguistics, 2009: 317-320.

[46]朱俭. 中文文本情感分析关键技术研究[D]. 北京: 北京理工大学, 2010.

[47]Li T, Sindhwani V, Ding C, et al. Knowledge transformation for cross-domain sentiment classification[C]. Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, Boston, MA, USA. ACM, 2009: 716-717.

[48]Liu K, Zhao J. Cross-domain sentiment classification using a two-stage method[C]. Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009: 1717-1720.

[49]Jakob N, Gurevych I. Extracting opinion targets in a single- and cross-domain setting with conditional random fields[C]. Proceedings of the 2010 Conference on Empirical Methods in Natural

Language Processing, Cambridge, Massachusetts. Association for Computational Linguistics, 2010: 1035-1045.

[50]Glorot X, Bordes A, Bengio Y. Domain adaptation for large-scale sentiment classification: A deep learning approach[C]. Proceedings of the 28th International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 2011.

[51]Bollegala D, Weir D, Carroll J. Cross-Domain Sentiment Classification using a Sentiment Sensitive Thesaurus[J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(8).

[52]Li S, Xue Y, Wang Z, et al. Active learning for cross-domain sentiment classification[C]. Proceedings of the Twenty-Third international joint conference on Artificial Intelligence. AAAI Press, 2013: 2127-2133.

[53]Morinaga S, Yamanishi K, Tateishi K, et al. Mining product reputations on the Web[C]. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, Edmonton, Alberta, Canada. ACM, 2002: 341-349.

[54]Hu M, Liu B. Mining and summarizing customer reviews[C]. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA. ACM, 2004: 168-177.

[55]Liu B, Hu M, Cheng J. Opinion observer: analyzing and comparing opinions on the Web[C]. Proceedings of the 14th international conference on World Wide Web, Chiba, Japan. ACM, 2005: 342-351.

[56]Jin W, Ho HH, Srihari RK. OpinionMiner: a novel machine learning system for web opinion mining and extraction[C]. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France. ACM, 2009: 1195-1204.

[57]Su Q, Xu X, Guo H, et al. Hidden sentiment association in chinese web opinion mining[C]. Proceedings of the 17th international conference on World Wide Web, Beijing, China. ACM, 2008: 959-968.

[58]Jiang P, Zhang C, Fu H, et al. An approach based on tree kernels for opinion mining of online product reviews[C]. 2010 IEEE 10th International Conference on Data Mining (ICDM). IEEE Computer Society, 2010: 256-265.

[59]Qi L, Chen L. A linear-chain CRF-based learning approach for web opinion mining[C]. Proceedings of the 11th international conference on Web information systems engineering, Hong Kong, China.

Springer-Verlag, 2010: 128-141.

[60]Brody S, Elhadad N. An unsupervised aspect-sentiment model for online reviews[C]. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, California. Association for Computational Linguistics, 2010: 804-812.

[61]Yang B, Cardie C. Joint inference for fine-grained opinion extraction[C]. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria. Association for Computational Linguistics, 2013: 1640-1649.

[62]姚天昉, 娄德成. 汉语语句主题语义倾向分析方法的研究[J]. 中文信息学报, 2007, 21(5): 73-79.

[63]章剑锋, 张奇, 吴立德等. 中文观点挖掘中的主观性关系抽取[J]. 中文信息学报, 2008, 22(2): 55-59.

[64]郝博一, 夏云庆, 邬晓钧等. 基于泛化和繁殖的自举式意见目标抽取方法[J]. 清华大学学报(自然科学版), 2009, 49(S1): 1333-1338.

[65]徐冰, 赵铁军, 王山雨等. 基于浅层句法特征的评价对象抽取研究[J]. 自动化学报, 2011, 37(10): 1241-1247.

[66]张莉, 钱玲飞, 许鑫. 基于核心句及句法关系的评价对象抽取[J]. 中文信息学报, 2011, 25(3): 23-29.

[67]Dave K, Lawrence S, Pennock DM. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews[C]. Proceedings of the 12th international conference on World Wide Web, Budapest, Hungary. ACM, 2003: 519-528.

[68]Gamon M, Aue A, Corston-Oliver S, et al. Pulse: Mining customer opinions from free text[C]. Proceedings of the 6th international conference on Advances in Intelligent Data Analysis, Madrid, Spain. Springer-Verlag, 2005: 121-132.

[69]Yi J, Niblack W. Sentiment Mining in WebFountain[C]. Proceedings of the 21st International Conference on Data Engineering. IEEE Computer Society, 2005: 1073-1083.

[70]Popescu AM, Nguyen B, Etzioni O. OPINE: Extracting product features and opinions from reviews[C]. Proceedings of HLT/EMNLP on interactive demonstrations. Association for Computational Linguistics, 2005: 32-33.

[71]Wilson T, Hoffmann P, Somasundaran S, et al. OpinionFinder: A system for subjectivity analysis[C].

Proceedings of HLT/EMNLP on Interactive Demonstrations, Vancouver, British Columbia, Canada. Association for Computational Linguistics, 2005: 34-35.

[72]Scaffidi C, Bierhoff K, Chang E, et al. Red Opal: product-feature scoring from reviews[C]. Proceedings of the 8th ACM conference on Electronic commerce, San Diego, California, USA. ACM, 2007: 182-191.

[73]Guo H, Zhu H, Guo Z, et al. OpinionIt: a text mining system for cross-lingual opinion analysis[C]. Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada. ACM, 2010: 1199-1208.

[74]Zhang W, Xu H, Wan W. Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis[J]. Expert Systems with Applications, 2012, 39(11): 10283-10291.

[75]姚天昉, 聂青阳, 李建超等. 一个用于汉语汽车评论的意见挖掘系统[C]. 中文信息处理前沿进展-中国中文信息学会二十五周年学术会议论文集. 北京: 清华大学出版社, 2006.

[76]郝博一, 夏云庆, 郑方. OPINAX: 一个有效的产品属性挖掘系统[C]. 第四届全国信息检索与内容安全学术会议论文集 (上卷), 2008.

[77]Melville P, Gryc W, Lawrence RD. Sentiment analysis of blogs by combining lexical knowledge with text classification[C]. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France. ACM, 2009: 1275-1284.

[78]Saravabhotla K, Pingali P, Varma V. Sentiment classification: a lexical similarity based approach for extracting subjectivity in documents[J]. Information retrieval, 2011, 14(3): 337-353.

[79]Ding X, Liu B, Yu PS. A holistic lexicon-based approach to opinion mining[C]. Proceedings of the international conference on Web search and web data mining, Palo Alto, California, USA. ACM, 2008: 231-240.

[80]Popescu AM, Etzioni O. Extracting product features and opinions from reviews[C]. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada. Association for Computational Linguistics, 2005: 339-346.

[81]Zhao WX, Jiang J, Yan H, et al. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid[C]. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, Massachusetts. Association for Computational Linguistics, 2010: 56-65.

[82]Bagheri A, Saraee M, de Jong F. An Unsupervised Aspect Detection Model for Sentiment Analysis

of Reviews[M]. Natural Language Processing and Information Systems. Springer Berlin Heidelberg, 2013: 140-151.

[83]Yu J, Zha ZJ, Wang M, et al. Aspect ranking: identifying important product aspects from online consumer reviews[C]. Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, Portland, Oregon. Association for Computational Linguistics, 2011: 1496-1505.

[84]Wu Y, Zhang Q, Huang X, et al. Phrase dependency parsing for opinion mining[C]. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing - Volume 3, Singapore. Association for Computational Linguistics, 2009: 1533-1541.

[85]Qiu G, Liu B, Bu J, et al. Opinion word expansion and target extraction through double propagation[J]. Comput. Linguist., 2011, 37(1): 9-27.

[86]Hu M, Liu B. Opinion extraction and summarization on the web[C]. Proceedings of the 21st national conference on Artificial intelligence - Volume 2, Boston, Massachusetts. AAAI Press, 2006: 1621-1624.

[87]Zhai Z, Liu B, Xu H, et al. Constrained LDA for grouping product features in opinion mining[C]. Proceedings of the 15th Pacific-Asia conference on Advances in knowledge discovery and data mining - Volume Part I, Shenzhen, China. Springer-Verlag, 2011: 448-459.

[88]Du W, Tan S. An iterative reinforcement approach for fine-grained opinion mining[C]. Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Boulder, Colorado. Association for Computational Linguistics, 2009: 486-493.

[89]Moghaddam S, Ester M. The FLDA model for aspect-based opinion mining: addressing the cold start problem[C]. Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro, Brazil. International World Wide Web Conferences Steering Committee, 2013: 909-918.

[90]Kim SM, Hovy E. Determining the sentiment of opinions[C]. Proceedings of the 20th international conference on Computational Linguistics, Geneva, Switzerland. Association for Computational Linguistics, 2004: 1367.

[91]姚天昉, 程希文, 徐飞玉等. 文本意见挖掘综述[J]. 中文信息学报, 2008, 22(3): 71-80.

[92]徐琳宏, 林鸿飞, 赵晶. 情感语料库的构建和分析[J]. 中文信息学报, 2008, 22(1): 116-122.

[93]赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.

[94]Zhang Y, Zhu W. Extracting implicit features in online customer reviews for opinion mining[C].

- Proceedings of the 22nd international conference on World Wide Web companion, Rio de Janeiro, Brazil. International World Wide Web Conferences Steering Committee, 2013: 103-104.
- [95]Xu L, Liu K, Lai S, et al. Mining Opinion Words and Opinion Targets in a Two-Stage Framework[C]. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria. Association for Computational Linguistics, 2013: 1764-1773.
- [96]Ma T, Wan X. Opinion target extraction in Chinese news comments[C]. Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Beijing, China. Association for Computational Linguistics, 2010: 782-790.
- [97]Guo H, Zhu H, Guo Z, et al. Product feature categorization with multilevel latent semantic association[C]. Proceeding of the 18th ACM conference on Information and knowledge management, Hong Kong, China. ACM, 2009: 1087-1096.
- [98]Li T, Zhang Y, Sindhwani V. A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge[C]. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP - Volume 1, Suntec, Singapore. Association for Computational Linguistics, 2009: 244-252.
- [99]Li T, Sindhwani V, Ding C, et al. Bridging domains with words: Opinion analysis with matrix tri-factorizations[C]. Proceedings of the 2010 SIAM International Conference on Data Mining, 2010: 293-302.
- [100]Cambria E, Hussain A, Durrani T, et al. Towards a chinese common and common sense knowledge base for sentiment analysis[M]. Advanced Research in Applied Artificial Intelligence, Springer Berlin Heidelberg, 2012: 437-446.
- [101]Lafferty JD, McCallum A, Pereira FCN. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[C]. Proceedings of the Eighteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., 2001: 282-289.
- [102]McCallum A. Efficiently inducing features of conditional random fields[C]. Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., 2002: 403-410.
- [103]刘新兰. 基于 CRFs 的产品评论意见挖掘[D]. 北京: 北京理工大学, 2012.
- [104]Feldman R, Sanger J. The text mining handbook: advanced approaches in analyzing unstructured data[M]. Cambridge University Press, 2007.

- [105]陈天缘. 线性链条件随机场训练算法优化的研究[D]. 上海: 复旦大学, 2010.
- [106]Darroch JN, Ratcliff D. Generalized iterative scaling for log-linear models[J]. The annals of mathematical statistics, 1972, 43(5): 1470-1480.
- [107]Della Pietra S, Della Pietra V, Lafferty J. Inducing features of random fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(4): 380-393.
- [108]Sha F, Pereira F. Shallow parsing with conditional random fields[C]. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, Edmonton, Canada. Association for Computational Linguistics, 2003: 134-141.
- [109]Liu K, Xu L, Liu Y, et al. Opinion Target Extraction Using Partially-Supervised Word Alignment Model[C]. Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, AAAI Press, 2013: 2134-2140.
- [110]Liu K, Xu L, Zhao J. Opinion target extraction using word-based translation model[C]. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea. Association for Computational Linguistics, 2012: 1346-1356.
- [111]Toutanova K, Klein D, Manning CD, et al. Feature-rich part-of-speech tagging with a cyclic dependency network[C]. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, Edmonton, Canada. Association for Computational Linguistics, 2003: 173-180.
- [112]Huang S, Liu X, Peng X, et al. Fine-grained Product Features Extraction and Categorization in Reviews Opinion Mining[C]. 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW), IEEE Computer Society, 2012: 680-686.
- [113]颜伟, 荀恩东. 基于 WordNet 的英语词语相似度计算[C]. 第二届全国学生计算语言学研讨会论文集, 2004.
- [114]Miller GA. WordNet: a lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
- [115]Jiang JJ, Conrath DW. Semantic similarity based on corpus statistics and lexical taxonomy[J]. arXiv preprint cmp-lg/9709008, 1997.
- [116]Agirre E, Alfonseca E, Hall K, et al. A study on similarity and relatedness using distributional and

WordNet-based approaches[C]. Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2009: 19-27.

[117]Pedersen T, Patwardhan S, Michelizzi J. WordNet::Similarity: measuring the relatedness of concepts[C]. Demonstration Papers at HLT-NAACL 2004. Association for Computational Linguistics, 2004: 38-41.

[118]Zhai Z, Liu B, Xu H, et al. Clustering product features for opinion mining[C]. Proceedings of the fourth ACM international conference on Web search and data mining, Hong Kong, China. ACM, 2011: 347-354.

[119]Zhang Q, Wu Y, Li T, et al. Mining product reviews based on shallow dependency parsing[C]. Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, Boston, MA, USA. ACM, 2009: 726-727.

[120]Liu B. Web data mining: exploring hyperlinks, contents, and usage data[M]. Springer Verlag, 2007.

[121]赵妍妍, 秦兵, 车万翔等. 基于句法路径的情感评价单元识别[J]. 软件学报, 2011, 22(5).

[122]Carenini G, Ng RT, Zwart E. Extracting knowledge from evaluative text[C]. Proceedings of the 3rd international conference on Knowledge capture, Banff, Alberta, Canada. ACM, 2005: 11-18.

[123]Lin D, Wu X. Phrase clustering for discriminative learning[C]. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP - Volume 2, Suntec, Singapore. Association for Computational Linguistics, 2009: 1030-1038.

[124]Titov I, McDonald R. Modeling online reviews with multi-grain topic models[C]. Proceedings of the 17th international conference on World Wide Web, Beijing, China. ACM, 2008: 111-120.

[125]Raju S, Shishtla P, Varma V. A Graph Clustering Approach to Product Attribute Extraction[C]. Indian International Conference on Artificial Intelligence, 2009: 1438-1447.

[126]Mukherjee A, Liu B. Aspect extraction through Semi-Supervised modeling[C]. Proceedings of the 50th annual meeting of the association for computational Linguistics: Long Papers - Volume 1, Jeju Island, Korea. Association for Computational Linguistics, 2012: 339-348.

[127]Zhai Z, Liu B, Xu H, et al. Grouping product features using semi-supervised learning with soft-constraints[C]. Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China. Association for Computational Linguistics, 2010: 1272-1280.

- [128]Lu Z, Ip H. Constrained spectral clustering via exhaustive and efficient constraint propagation[M]. Computer Vision – ECCV 2010. Springer Berlin Heidelberg, 2010: 1-14.
- [129]Fu X, Liu G, Guo Y, et al. Multi-aspect Blog sentiment analysis based on LDA topic model and hownet lexicon[C]. Proceedings of the 2011 international conference on Web information systems and mining - Volume Part II, Taiyuan, China. Springer-Verlag, 2011: 131-138.
- [130]Fu X, Liu G, Guo Y, et al. Multi-aspect Sentiment Analysis for Chinese Online Social Reviews Based on Topic Modeling and HowNet Lexicon[J]. Knowledge-Based Systems, 2013, 37: 186-195.
- [131]Moreo A, Romero M, Castro JL, et al. Lexicon-based Comments-oriented News Sentiment Analyzer system[J]. Expert Systems with Applications, 2012, 39(10): 9166-9180.
- [132]Shi B, Chang K. Mining Chinese reviews[C]. Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), Hong Kong, 2006: 585-589.
- [133]Esuli A, Sebastiani F. Pageranking wordnet synsets: An application to opinion mining[C]. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic. Association for Computational Linguistics, 2007: 424-431.
- [134]Cruz FL, Troyano JA, Ortega FJ, et al. Automatic expansion of feature-level opinion lexicons[C]. Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Portland, Oregon. Association for Computational Linguistics, 2011: 125-131.
- [135]Kanayama H, Nasukawa T. Fully automatic lexicon expansion for domain-oriented sentiment analysis[C]. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia. Association for Computational Linguistics, 2006: 355-363.
- [136]Tan S, Wu Q. A random walk algorithm for automatic construction of domain-oriented sentiment lexicon[J]. Expert Systems with Applications, Pergamon Press, Inc., 2011, 38(10): 12094-12100.
- [137]Kamps J, Marx M, Mokken RJ, et al. Using wordnet to measure semantic orientations of adjectives[C]. Proceedings of LREC 2004, 2004: 1115-1118.
- [138]Esuli A, Sebastiani F. Random-walk models of term semantics: An application to opinion-related properties[C]. Proceedings of LTC 2007, 2007.
- [139]Turney PD. Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL[C]. Proceedings of the 12th European Conference on Machine Learning, Springer-Verlag, 2001: 491-502.
- [140]Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis[C]. Proceedings of the conference on Human Language Technology and Empirical Methods in

Natural Language Processing, Vancouver, British Columbia, Canada. Association for Computational Linguistics, 2005: 347-354.

[141]Qiu G, Liu B, Bu J, et al. Expanding domain sentiment lexicon through double propagation[C]. Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09), Pasadena, California, USA. Morgan Kaufmann Publishers Inc., 2009: 1199-1204.

[142]Rao D, Ravichandran D. Semi-supervised polarity lexicon induction[C]. Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Athens, Greece. Association for Computational Linguistics, 2009: 675-682.

[143]Weichselbraun A, Gindl S, Scharl A. Using games with a purpose and bootstrapping to create domain-specific sentiment lexicons[C]. Proceedings of the 20th ACM international conference on Information and knowledge management, Glasgow, Scotland, UK. ACM, 2011: 1053-1060.

[144]Du W, Tan S, Cheng X, et al. Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon[C]. Proceedings of the third ACM international conference on Web search and data mining, New York, New York, USA. ACM, 2010: 111-120.

[145]Bollegala D, Weir D, Carroll J. Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification[C]. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, Portland, Oregon. Association for Computational Linguistics, 2011: 132-141.

[146]Zhou D, Bousquet O, Lal TN, et al. Learning with local and global consistency[C]. Advances in neural information processing systems. MIT Press, 2003, 16: 321-328.

[147]Wang F, Zhang C. Label propagation through linear neighborhoods[C]. Proceedings of the 23rd international conference on Machine learning, Pittsburgh, Pennsylvania. ACM, 2006: 985-992.

[148]Ganesan K, Zhai CX. Opinion-based entity ranking[J]. Information retrieval, 2012, 15(2): 116-150.

[149]Kaji N, Kitsuregawa M. Building lexicon for sentiment analysis from massive collection of HTML documents[C]. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 2007: 1075-1083.

[150]Huang S, Niu Z, Shi Y. Product Features Categorization Using Constrained Spectral Clustering[C]. Proceedings of the 18th International Conference on Applications of Natural Language to Information Systems (NLDB 2013), Salford, UK. Springer-Verlag Berlin Heidelberg, 2013: 285-290.

[151]Fu X, Guo Y, Guo W, et al. Aspect and sentiment extraction based on information-theoretic

co-clustering[C]. Proceedings of the 9th International Symposium on Neural Networks - Volume Part II, Shenyang, China. Springer-Verlag Berlin Heidelberg, 2012: 326-335.

[152]Zhuang L, Jing F, Zhu XY. Movie review mining and summarization[C]. Proceedings of the 15th ACM international conference on Information and knowledge management, Arlington, Virginia, USA. ACM, 2006: 43-50.

[153]Cheng Y, Church GM. Bicustering of Expression Data[C]. Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology, AAAI Press, 2000: 93-103.

[154]Dhillon IS. Co-clustering documents and words using bipartite spectral graph partitioning[C]. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, California. ACM, 2001: 269-274.

[155]王跃. 联合聚类算法研究及应用[D]. 杭州: 浙江大学, 2012.

[156]吴湖, 王永吉, 王哲等. 两阶段联合聚类协同过滤算法[J]. 软件学报, 2010, 21(5): 1042-1054.

[157]Rege M, Dong M, Fotouhi F. Co-clustering documents and words using bipartite isoperimetric graph partitioning[C]. Proceedings of the Sixth International Conference on Data Mining (ICDM'06), Los Alamitos. IEEE Computer Society, 2006: 532-541.

[158]Shan H, Banerjee A. Bayesian co-clustering[C]. 2008 Eighth IEEE International Conference on Data Mining (ICDM'08), IEEE Computer Society, 2008.

[159]Greene D, Cunningham P. Spectral co-clustering for dynamic bipartite graphs[C]. Proceedings of the 1st Workshop on Dynamic Networks and Knowledge Discovery Barcelona, September 2010.

[160]Bichot CE. Co-clustering documents and words by minimizing the normalized cut objective function[J]. Journal of Mathematical Modelling and Algorithms, 2010, 9(2): 131-147.

[161]Xie S, Lu H, He Y. Multi-task co-clustering via nonnegative matrix factorization[C]. 2012 21st International Conference on Pattern Recognition (ICPR), 2012.

[162]Yan Y, Chen L, Tjhi WC. Fuzzy semi-supervised co-clustering for text documents[J]. Fuzzy Sets and Systems, 2013, 215: 74-89.

[163]Banerjee A, Dhillon I, Ghosh J, et al. A generalized maximum entropy approach to bregman co-clustering and matrix approximation[C]. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004: 509-514.

[164]Dhillon IS, Mallela S, Modha DS. Information-theoretic co-clustering[C]. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003: 89-98.

- [165]Pensa RG, Robardet C, Boulicaut JF. Constraint-driven co-clustering of 0/1 data[J]. *Constrained Clustering: Advances in Algorithms, Theory and Applications*, 2008: 145-170.
- [166]Pensa RG, Boulicaut JF. Constrained co-clustering of gene expression data[C]. *Proceedings of the 2008 SIAM International Conference on Data Mining (SDM'08)*, Atlanta, USA. 2008: 25-36.
- [167]Song Y, Pan S, Liu S, et al. Constrained co-clustering for textual documents[C]. *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI*, 2010.
- [168]Song Y, Pan S, Liu S, et al. Constrained Text Co-Clustering with Supervised and Unsupervised Constraints[J]. *IEEE Transactions on Knowledge and Data Engineering*. IEEE Computer Society, 2013, 25(6): 1227-1239.
- [169]Salunke A, Liu X, Rege M. Constrained co-clustering with non-negative matrix factorisation[J]. *International Journal of Business Intelligence and Data Mining*, 2012, 7(1): 60-79.
- [170]Pensa RG, Boulicaut JF, Cordero F, et al. Co-clustering numerical data under user-defined constraints[J]. *Statistical Analysis and Data Mining*, John Wiley & Sons, Inc., 2010, 3(1): 38-55.
- [171]Shi X, Fan W, Yu PS. Efficient semi-supervised spectral co-clustering with constraints[C]. *2010 IEEE 10th International Conference on Data Mining (ICDM'10)*, IEEE Computer Society, 2010: 1043-1048.
- [172]Hai Z, Chang K, Cong G. One seed to find them all: mining opinion features via association[C]. *Proceedings of the 21st ACM international conference on Information and knowledge management, Maui, Hawaii, USA. ACM*, 2012: 255-264.
- [173]Zhang L, Liu B, Lim SH, et al. Extracting and ranking product features in opinion documents[C]. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Beijing, China. Association for Computational Linguistics*, 2010: 1462-1470.
- [174]Wang H, Lu Y, Zhai C. Latent aspect rating analysis without aspect keyword supervision[C]. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, California, USA. ACM*, 2011: 618-626.
- [175]Li S, Wang R, Zhou G. Opinion Target Extraction Using a Shallow Semantic Parsing Framework[C]. *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- [176]Xu L, Liu K, Lai S, et al. Walk and learn: a two-stage approach for opinion words and opinion targets co-extraction[C]. *Proceedings of the 22nd international conference on World Wide Web companion, Rio de Janeiro, Brazil. International World Wide Web Conferences Steering Committee*,

2013: 95-96.

[177]Gindl S, Weichselbraun A, Scharl A. Rule-based opinion target and aspect extraction to acquire affective knowledge[C]. Proceedings of the 22nd international conference on World Wide Web companion, Rio de Janeiro, Brazil. International World Wide Web Conferences Steering Committee, 2013: 557-564.

[178]Rand WM. Objective criteria for the evaluation of clustering methods[J]. Journal of the American Statistical association, 1971, 66(336): 846-850.

[179]Lau RYK, Lai CL, Bruza PD, et al. Leveraging Web 2.0 data for scalable semi-supervised learning of domain-specific sentiment lexicons[C]. Proceedings of the 20th ACM international conference on Information and knowledge management, Glasgow, Scotland, UK. ACM, 2011: 2457-2460.

攻读学位期间发表论文与研究成果清单

- [1] 第一作者. Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. Knowledge-Based Systems, Jan 2014, Volume 56, pp 191-200. (SCI Indexing, 索引号: 000331160200016, Impact Factor: 4.104)
- [2] 第一作者. Product features categorization using constrained spectral clustering. 18th International Conference on Applications of Natural Language to Information Systems (NLDB 2013), Salford, UK, Jun 2013, Lecture Notes in Computer Science, Volume 7934, pp 285-290. (EI 索引号: 20134116830110)
- [3] 第一作者. Fine-grained Product Features Extraction and Categorization in Reviews Opinion Mining. 2012 IEEE 12th International Conference on Data Mining Workshops (ICDM 2012 Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction), Brussels Belgium, Dec 2012, pp 680-686. (EI 索引号: 20130615994844)
- [4] 第一作者. News Topic Detection based on Hierarchical Clustering and Named Entity. 2011 7th International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE 2011), Tokushima Japan, Nov 2011, pp 280-284. (EI 索引号: 20120914806023)
- [5] 第四作者. Building Enhanced Link Context by Logical Sitemap. 6th International Conference on Knowledge Science, Engineering and Management (KSEM 2013), Dalian China, Aug 2013, pp 36-47. (EI 索引号: 20133116565257)
- [6] 第三作者. Personalized Web Search Using Clickthrough Data and Web Page Rating [J]. Journal of computers, Oct 2012, Volume 7(10), pp 2578-2584. (EI 索引号: 20124415634491)
- [7] 第二作者. An study on personalized recommendation model based on search behaviors and resource properties. 2nd International Conference on Information Engineering and Computer Science (ICIECS 2010), Wuhan China, Dec 2010, pp 1-4. (EI 索引号: 20110813673840)
- [8] 第三作者. Design of a Web Public Opinion Analyzing System. 8th Japan-China International Workshop on Internet Technology and Control Applications(ITCA 2011), Tokyo Japan, Dec 2011.

发明专利和软件著作权

- [1] 一种基于查询语义和点击流数据的查询建议方法.中国专利. 公开号: CN102253982A, 2011-11-23. 已授权. 第三申请人
- [2] 一种基于约束关系的意见目标和情感词联合聚类方法.中国专利. 申请号: 201310701519.8, 2013-12-19. 已申请. 第二申请人(导师为第一申请人)
- [3] 国防科技工业网络信息舆情分析平台系统 (软件著作权)
- [4] 互联网舆情分析系统 (软件著作权)
- [5] 国防科技工业智能信息检索系统 (软件著作权)

攻读博士学位期间参加的科研项目

- [1] Web 文本意见挖掘关键技术研究 (国家自然科学基金项目, No: 61250010)
- [2] 面向 Web 主观性文本意见挖掘研究 (国家自然科学基金项目, No: 61370137)
- [3] 面向 Web 文本的属性和属性值知识获取方法研究 (国家自然科学基金项目, No: 61272361)
- [4] XXXXX 网络舆情监控分析系统

致谢

时光飞逝，六年的硕士和博士生涯即将结束。在这六年的时间里，快乐、兴奋、迷茫和痛苦交织着伴随我成长，这段经历将是我人生中最宝贵的财富。在博士论文完成之际，向所有关心和帮助过我的老师、同学、朋友和亲人表示最衷心的感谢。

首先要向我的导师牛振东教授表示衷心的感谢。从我进入实验室起，牛老师一直以自身的行动言传身教地影响着我，多年来的关心和指导使我受益良多。攻读博士学位期间牛老师不仅教给了我如何进行科学研究的理论和方法，更教会了处理问题 and 对待生活兢兢业业的态度。博士期间所取得的进步和成绩都和牛老师的教导密不可分。本论文从选题、具体的科学研究方法、论文撰写到最终定稿各个阶段都是在牛老师精心指导和严格要求下完成的。在此，谨向牛老师致以最真切的敬意。

另外，还要感谢实验室的樊孝忠、张春霞、刘辉、袁武和施重阳老师，他们在生活和学业上都给了我很多的关心和帮助。樊孝忠教授虽然年事已高，却坚持每天到实验室，亲身为我们学生做出了表率。特别感谢张春霞老师，在我请教问题时，张老师总是不惜指导，积极地帮助和鼓励我。同时也要感谢郭贵锁老师，郭老师对待教学和学生认真负责的精神和平易近人的作风，使我深受鼓舞和感染。

在本人攻读博士学位期间，得到了实验室很多同学和师弟师妹的大力帮助和支持。在我进入实验室的早期，已毕业的彭学平和束博师兄在学术研究和项目开发上给予了我很多的指导，带领我迈进了科研的门槛。陈杰博士和我同处一个项目组，很多工作我们能够相互沟通，相互协作。王坤山同学对我博士期间参与的重大工程项目付出甚多。刘新兰、林爽和石玉龙同学参与了我的早期部分研究工作。刘沙、王帅和王诗航同学承担了后期的项目开发和研究工作，付红萍、李维银和江小天同学则负责了最后博士毕业论文的一些翻译、修改和勘误工作。在此，对他们一并表示感谢。

在我两次脚踝扭伤期间，林爽、刘沙和李维银等同学在生活和学习上都为我提供了很大的帮助，不辞辛劳地陪同我奔波在学校和医院之间。还有已毕业的杨青师兄和张雅芬同学，曾经经常一起吃饭、打球和聊天，使我能够排解苦闷的情绪和心理压力。非常感谢你们。

感谢我的父母、哥哥和嫂子，你们给予了我最大的支持和鼓励。是你们无微不至的关怀给了我前进的动力，是你们默默的付出使我解除了后顾之忧，得以专心投入学业。每当我遭遇困难和挫折的时候，你们是最坚实的后盾，让我重拾信心与力量，使我能够一次又一次地克服困难，继续前进。

最后，再次感谢所有关心过我、帮助过我和教导过我的老师、同学、朋友和亲人。谢谢你们！