

情感分析和 观点挖掘

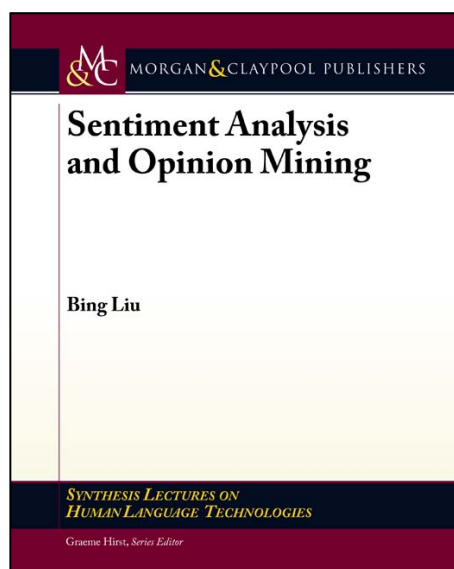
2012年4月22日

刘冰

liub@cs.uic.edu

草稿 : 由于复制编辑，出版版本略有不同

刘冰。情感分析与观点挖掘，*Morgan & Claypool Publishers*，2012年5月。



目录

前言.....	5
情感分析：一个迷人的问题	7
1.1 情感分析应用	8
1.2 情感分析研究	10
1.2.1 不同层次的分析	10
1.2.2 情感词典及其问题	12
1.2.3 自然语言处理问题.....	13
1.3 观点垃圾邮件检测	14
1.4 未来展望	14
情感分析的问题	16
2.1 问题定义	17
2.1.1 观点定义	17
2.1.2 情感分析任务	21
2.2 观点总结	24
2.3 不同类型的观点.....	25
2.3.1 正常和比较观点	25
2.3.2 明确和隐含观点	26
2.4 主观性和情感	27
2.5 作者和读者立场.....	29
2.6 总结	29
文档情感分类.....	30
3.1 使用监督学习的情感分类	31
3.2 使用无监督学习的情感分类.....	34
3.3 情感评级预测.....	36
3.4 跨领域情感分类	38
3.5 跨语言情感分类.....	41
3.6 总结	43
句子主观性和情感分类.....	44

情感分析与观点挖掘

4.1 主观性分类.....	45
4.2 句子情感分类	49
4.3 处理条件句	51
4.4 处理讽刺句.....	52
4.5 跨语言主观性和情感分类	53
4.6 使用话语信息进行情感分类	55
4.7 总结	56
基于方面的情感分析.....	58
5.1 方面情感分类.....	59
5.2 观点和组合语义的基本规则.....	62
5.3 方面提取	67
5.3.1 查找频繁名词和名词短语	68
5.3.2 使用观点和目标关系	71
5.3.3 使用监督学习	71
5.3.4 使用主题模型	73
5.3.5 映射隐含方面	77
5.4 识别资源使用方面	78
5.5 同时扩展意见词典和方面 提取	79
5.6 将方面分组为类别.....	81
5.7 实体、意见持有者和时间提取	84
5.8 共指消解和词义消歧	86
5.9 摘要	88
情感词典生成	90
6.1 基于词典的方法.....	91
6.2 基于语料库的方法	95
6.3 有益和无益的事实	99
6.4 摘要	100
观点总结	102
7.1 基于方面的观点总结.....	102
7.2 改进基于方面的观点总结	105
7.3 对比视角总结	107
7.4 传统总结.....	108
7.5 摘要	108

情感分析与观点挖掘

比较观点分析.....	110
8.1 问题定义	110
8.2 识别比较句子.....	113
8.3 识别首选实体	115
8.4 摘要	117
观点搜索和检索	118
9.1 网络搜索 vs. 观点搜索	118
9.2 现有的观点检索技术	119
9.3 总结	122
观点垃圾邮件检测.....	123
10.1 垃圾邮件和垃圾邮件类型.....	124
10.1.1 有害的虚假评论	125
10.1.2 个人和团体垃圾邮件.....	125
10.1.3 数据、特征和检测类型.....	126
10.2 监督式垃圾邮件检测.....	127
10.3 无监督式垃圾邮件检测	130
10.3.1 基于非典型行为的垃圾邮件检测.....	130
10.3.2 使用评论图进行垃圾邮件检测	133
10.4 团体垃圾邮件检测	134
10.5 总结	135
评论的质量	136
11.1 作为回归问题的质量	136
11.2 其他方法.....	138
11.3 总结	140
总结性评论.....	141
参考文献	143

前言

观点在几乎所有人类活动中都是核心，是我们行为的关键影响因素。我们的信念和对现实的感知，以及我们的选择，在很大程度上取决于他人如何看待和评价世界。因此，当我们需要做决策时，我们经常寻求他人的意见。这不仅适用于个人，也适用于组织。

情感分析和观点挖掘是研究观点及其相关概念（如情感、评价、态度和情绪）的主题。该领域的产生和快速发展与网络上的社交媒体（如评论、论坛讨论、博客、微博、Twitter和社交网络）的产生时间相吻合，因为人类历史上第一次有了大量以数字形式记录的带有观点的数据。自2000年初以来，情感分析已成为自然语言处理中最活跃的研究领域之一。它也被广泛研究于数据挖掘、网络挖掘和文本挖掘。事实上，由于对商业和社会的重要性，它已从计算机科学扩展到管理科学和社会科学。近年来，围绕情感分析的工业活动也蓬勃发展。许多初创公司涌现出来。许多大公司建立了自己的内部能力。情感分析系统已在几乎所有的商业和社会领域找到了应用。

本书的目标是深入介绍这个迷人的问题，并对领域中的所有重要研究主题和最新发展进行全面调查。作为证据，本书涵盖了来自所有主要会议和期刊的400多个参考文献。

尽管该领域处理的是常被认为是非结构化数据的自然语言文本，但本书采用了结构化的方法来介绍问题，旨在弥合非结构化和结构化世界，并促进观点的定性和定量分析。这对于实际应用至关重要。在本书中，我首先定义问题，以便为问题提供抽象或结构。

从抽象中，我们自然会看到其关键子问题。随后的章节讨论了解决这些子问题的现有技术。

本书适合对社交媒体分析有兴趣的学生、研究人员和从业人员。讲师可以在自然课程中轻松使用它。

情感分析与观点挖掘

语言处理，社交媒体分析，文本挖掘和数据挖掘。
讲义幻灯片也可以在线获取。

致谢

我要感谢我的前任和现任学生——Zhiyuan Chen, Xiaowen Ding, Geli Fei, Murthy Ganapathibhotla, Mingqing Hu, Nitin Jindal, Huayi Li, Arjun Mukherjee, Guang Qiu（来自浙江大学的访问学生），William Underwood, Andrea Vaccari, Zhongwu Zhai（来自清华大学的访问学生）和 Lei Zhang——多年来为本书做出了许多研究思路的贡献。与许多研究人员的讨论也有助于本书的形成：Malu G. Castellanos, Dennis Chong, Umesh Dayal, Eduard Dragut, Riddhiman Ghosh, Natalie Glance, Meichun Hsu, Jing Jiang, Birgit König, Xiaoli Li, Tieyun Qian, Gang Xu, Philip S. Yu, Clement Yu和ChengXiang Zhai。我也非常感谢两位匿名审稿人。尽管他们的日程很忙，但他们非常仔细地阅读了本书，并给了我许多很好的建议。我在改进本书时考虑了每一个建议。在出版方面，我要感谢编辑Graeme Hirst博士和Morgan & Claypool Publishers的总裁兼首席执行官Michael Morgan先生，他们设法按时完成了所有工作，并给了我许多宝贵的建议。最后，我最衷心地感谢我的家人：Yue, Shelley和Kate，他们在许多方面给予了帮助。

第1章

情感分析：一个迷人的问题

情感分析，也称为观点挖掘，是研究分析人们对产品、服务、组织、个人、问题、事件、主题及其属性等实体的意见、情感、评价、评估、态度和情绪的领域。它代表了一个广阔的问题空间。还有许多不同的名称和稍微不同的任务，例如情感分析、观点挖掘、观点提取、情感挖掘、主观性分析、情感分析、情感分析、评论挖掘等。然而，它们现在都属于情感分析或观点挖掘的范畴。在工业界，情感分析这个术语更常用，但在学术界，情感分析和观点挖掘都经常被使用。它们基本上代表了同一个研究领域。

情感分析这个术语可能首次出现在（Nasukawa和Yi，2003）中，而观点挖掘这个术语首次出现在（Dave，Lawrence和Pennock，2003）中。然而，关于情感和观点的研究出现得更早（Das和Chen，2001；Morinaga等，2002；Pang，Lee和Vaithyanathan，2002；Tong，2001；Turney，2002；Wiebe，2000）。在本书中，我们将情感分析和观点挖掘这两个术语互换使用。为了简化表述，在本书中，我们将使用观点一词来表示观点、情感、评价、评估、态度和情绪。然而，这些概念并不等同。我们将在需要时加以区分。观点本身的含义仍然非常广泛。情感分析和观点挖掘主要关注表达或暗示积极或消极情感的观点。

尽管语言学和自然语言处理（NLP）有着悠久的历史，在2000年之前几乎没有关于人们观点和情感的研究。从那时起，这个领域变得非常活跃的研究领域。这有几个原因。首先，它在几乎每个领域都有广泛的应用。情感分析周围的产业也因商业应用的普及而蓬勃发展。这为研究提供了强大的动力。其次，它提供了许多具有挑战性的研究问题，这些问题以前从未被研究过。本书将系统地定义和讨论这些问题，并描述解决这些问题的最新技术。第三，对于

人类历史上第一次，我们现在在社交媒体和网络上拥有大量的带有观点的数据。没有这些数据，很多研究将不可能进行。毫不奇怪，情感分析的开始和快速发展与社交媒体的发展相吻合。事实上，情感分析现在正处于社交媒体研究的中心位置。

因此，情感分析的研究不仅对自然语言处理有重要影响，还可能对管理科学、政治学、经济学和社会科学产生深远影响，因为它们都受到人们观点的影响。尽管情感分析研究主要始于2000年初，但在解释隐喻、情感形容词、主观性、观点和情感方面也有一些早期工作（Hatzivassiloglou和McKeown，1997年；Hearst，1992年；Wiebe，1990年；Wiebe，1994年；Wiebe，Bruce和O'Hara，1999年）。本书既是一本最新全面的入门教材，也是对该主题的调查。

1.1 情感分析应用

观点在几乎所有人类活动中都是核心，因为它们是我们行为的关键影响因素。每当我们需要做决策时，我们都想知道他人的观点。在现实世界中，企业和组织总是希望了解消费者或公众对其产品和服务的意见。个体消费者在购买产品之前也想了解现有用户的意见，在政治选举中做出投票决策之前也想知道他人对政治候选人的意见。过去，当个人需要意见时，他/她会向朋友和家人询问。当组织或企业需要公众或消费者的意见时，它们会进行调查、民意调查和焦点小组讨论。

获取公众和消费者的意见长期以来一直是市场营销、公共关系和政治竞选公司的重要业务。

随着社交媒体（例如评论、论坛讨论、博客、微博、Twitter、评论和社交网络网站上的帖子）在网络上的爆炸性增长，个人和组织越来越多地使用这些媒体中的内容进行决策。如今，如果一个人想购买消费品，不再局限于向朋友和家人询问意见，因为网络上有许多用户评论和公共论坛上的讨论。对于一个组织来说，不再需要进行调查、民意调查和焦点小组，以收集公众意见，因为这样的信息已经公开可得。然而，在网络上找到和监控观点网站，并提取其中的信息仍然是一个挑战。

情感分析与观点挖掘

由于各种各样的网站的泛滥，这是一项艰巨的任务。每个网站通常包含大量的意见文本，这些文本在长篇博客和论坛帖子中并不总是容易理解。普通人阅读者往往难以识别相关网站，并从中提取和总结意见。因此，需要自动化的情感分析系统。

近年来，我们目睹了社交媒体上的观点性帖子如何重塑业务，影响公众情绪，深刻影响我们的社会和政治体系。这些帖子还动员了大众参与政治变革，例如2011年发生在某些阿拉伯国家的变革。因此，收集和研究网络上的观点已经成为必要。当然，观点性文件不仅存在于网络上（称为外部数据），许多组织还拥有自己的内部数据，例如从电子邮件和呼叫中心收集的客户反馈或组织进行的调查结果。

由于这些应用，近年来工业活动蓬勃发展。情感分析应用已经扩展到几乎所有可能的领域，从消费产品、服务、医疗保健和金融服务到社会事件和政治选举。我自己实施了一个名为“观点解析器”的情感分析系统，并在初创公司的各个领域的项目上工作。仅在美国就至少有40-60家初创公司在这个领域。许多大公司也建立了自己的内部能力，例如微软、谷歌、惠普、SAP和SAS。这些实际应用和工业兴趣为情感分析的研究提供了强大的动力。

除了现实生活应用之外，还发表了许多应用导向的研究论文。例如，在（Liu等，2007年）中，提出了一个情感模型来预测销售业绩。在（McGlohon, Glance和Reiter，2010年）中，使用评论来对产品和商家进行排名。在（Hong和Skiena，2010年）中，研究了NFL博彩线和博客以及Twitter上的公众意见之间的关系。在（O'Connor等，2010年）中，将Twitter情感与公众舆论调查联系起来。在（Tumasjan等，2010年）中，也应用了Twitter情感来预测选举结果。在（Chen等，2010年）中，作者研究了政治立场。在（Yano和Smith，2010年）中，报道了一种预测政治博客评论数量的方法。在（Asur和Huberman，2010年；Joshi等，2010年；Sadikov, Parameswaran和Venetis，2009年）中，使用Twitter数据、电影评论和博客来预测电影票房收入。在（Miller等，2011年）中，研究了社交网络中的情感流动。在（Mohammad和Yang，2011年）中，使用邮件中的情感来发现不同性别在情感轴上的差异。在（Mohammad，2011年）中，追踪了小说和童话故事中的情绪。在（Bollen, Mao和Zeng，2011年）中，使用Twitter情绪来

预测股市。在（Bar-Haim等人，2011; Feldman等人，2011）中，通过微博识别了专业投资者，并对股票进行了情感分析。在（Zhang和Skiena，2010）中，使用博客和新闻的情感来研究交易策略。在（Sakunkoo和Sakunkoo，2009）中，研究了在线书评中的社交影响。在（Groh和Hauffa，2011）中，使用情感分析来表征社交关系。在（Castellanos等人，2011）中，还报道了一个综合的情感分析系统和一些案例研究。我们自己的团队在Twitter上追踪了关于电影的意见，并预测了非常准确的票房收入。我们只是使用我们的观点解析器系统来分析每部电影的正面和负面意见，没有使用额外的算法。

1.2 情感分析研究

如上所述，广泛的现实生活应用只是情感分析成为一个热门研究问题的部分原因。作为自然语言处理研究的一个高度具有挑战性的主题，情感分析涵盖了许多新颖的子问题，稍后我们将看到。此外，在2000年之前，无论是在自然语言处理还是在语言学方面，都几乎没有研究。部分原因是那之前，数字形式的意见文本很少。自2000年以来，该领域迅速发展成为自然语言处理中最活跃的研究领域之一。它也被广泛研究于数据挖掘、网络挖掘和信息检索领域。事实上，它已经从计算机科学扩展到管理科学（Archak, Ghose和Ipeirotis，2007; Chen和Xie，2008; Das和Chen，2007; Dellarocas, Zhang和Awad，2007; Ghose, Ipeirotis和Sundararajan，2007; Hu, Pavlou和Zhang，2006; Park, Lee和Han，2007）。

1.2.1 不同层次的分析

现在我将根据现有研究的粒度水平对主要研究问题进行简要介绍。总体而言，情感分析主要在三个层面上进行研究：

文档级别：在这个级别上的任务是对整个观点文档进行分类，判断其表达的情感是积极的还是消极的（Pang, Lee和Vaithyanathan，2002; Turney，2002）。例如，给定一个产品评论，系统确定该评论是否表达了对产品的整体积极或消极观点。这个任务通常是

情感分析与观点挖掘

被称为文档级情感分类。这个分析层次假设每个文档都表达了对一个实体（例如一个产品）的观点。因此，它不适用于评估或比较多个实体的文档。

句子级别：该任务在这个级别上针对句子，并确定每个句子是否表达了积极、消极或中立的观点。

中立通常意味着没有观点。这个分析级别与主观性分类（Wiebe, Bruce和O'Hara, 1999）密切相关，它区分了表达事实信息的句子（称为客观句子）和表达主观观点的句子（称为主观句子）。然而，我们应该注意到，主观性并不等同于情感，因为许多客观句子可能暗示了观点，例如，“我们上个月买了这辆车，雨刷掉了。”研究人员还分析了从句（Wilson, Wiebe和Hwa, 2004），但从句级别还不够，例如，“苹果在这个糟糕的经济环境中做得非常好。”

实体和方面级别：无论是文档级别还是句子级别分析都无法发现人们喜欢和不喜欢的具体内容。

方面级别执行更精细的分析。方面级别早先被称为特征级别（基于特征的观点挖掘和总结）（Hu和Liu，2004年）。与其关注语言结构

（文档、段落、句子、从句或短语），方面级别

直接关注观点本身。它基于一个观点的想法

包括情感（积极或消极）和目标（观点的对象）。

没有确定其目标的观点有限。认识到观点目标的重要性也有助于我们更好地理解情感分析问题。例如，尽管句子“尽管服务不是很好，我仍然喜欢这家餐厅”明显具有积极的语气，但我们不能说这个句子完全是积极的。事实上，这个句子对餐厅是积极的（强调），但对其服务是消极的（不强调）。在许多应用中，观点目标由实体和/或它们的不同方面描述。因此，这个级别的分析目标是发现实体和/或它们的方面上的情感。例如，句子“iPhone的通话质量很好，但电池寿命很短”评估了iPhone的两个方面，通话质量和电池寿命。对于iPhone的通话质量的情感是积极的，但对于其电池寿命的情感是消极的。iPhone的通话质量和电池寿命是观点目标。基于这个级别的分析，可以生成关于实体和它们的方面的观点的结构化摘要，将非结构化文本转化为结构化数据并可用于各种定性和定量分析。

文档级别和句子级别的分类已经完成

情感分析与观点挖掘

非常具有挑战性。方面级别更加困难。它包括几个子问题，在第2章和第5章中我们将讨论。

为了使事情更有趣和具有挑战性，存在两种类型的观点，即常规观点和比较观点(Jindal and Liu, 2006b)。常规观点仅对特定实体或实体的某个方面表达情感，例如“可口可乐味道非常好”，表达了对可口可乐味道方面的积极情感。比较观点基于共享方面比较多个实体，例如“可口可乐的味道比百事可乐好”，基于它们的味道（一个方面）比较可口可乐和百事可乐，并表达了对可口可乐的偏好(见第8章)。

1.2.2 情感词典及其问题

毫不奇怪，情感的最重要指标是情感词，也称为观点词。这些是常用的用于表达积极或消极情感的词语。例如，好、美妙和令人惊叹是积极情感词，而坏、差和可怕是消极情感词。除了单个词语，还有短语和习语，例如“花费某人一大笔钱”。情感词和短语对情感分析至关重要。这样的词语和短语列表被称为情感词典（或观点词典）。

多年来，研究人员设计了许多算法来编译这样的词典。我们将在第6章讨论这些算法。

尽管情感词和短语对于情感分析很重要，但仅仅使用它们是远远不够的。这个问题要复杂得多。

换句话说，我们可以说情感词典是情感分析所必需的，但并不足够。

下面，我们重点介绍几个问题：

1. 在不同的应用领域中，一个积极或消极的情感词可能具有相反的倾向。例如，“糟糕”通常表示负面情感，例如，“这个相机糟糕透了”，但它也可能表示积极情感，例如，“这个吸尘器真的很棒。”
2. 包含情感词的句子可能不表达任何情感。

这种现象在几种类型的句子中经常发生。

疑问句和条件句是两种重要的类型，例如，“你能告诉我哪款索尼相机好吗？”和“如果我能在店里找到一台好相机，我会买下它。”这两个句子都包含情感词“好”，但都不表达对任何特定相机的积极或消极意见。然而，并不是所有的条件句或疑问句都不表达情感，例如，“有人知道如何修理这台糟糕的打印机吗”和“如果你

情感分析与观点挖掘

如果你正在寻找一辆好车，选择丰田凯美瑞。”我们将在第四章讨论这样的句子。

3. 带有或不带有情感词的讽刺句子很难处理，例如，“多么好的车！两天就坏了。”讽刺在消费者对产品和服务的评论中并不常见，但在政治讨论中非常常见，这使得政治观点很难处理。我们将在第四章讨论这样的句子。

4. 许多没有情感词的句子也可以暗示观点。

这些句子中的许多实际上是用来表达一些事实信息的客观句子。

同样，这种句子有很多类型。这里我们只给出两个例子。句子“这台洗衣机用水很多”暗示了对洗衣机的负面情感，因为它使用了很多资源（水）。句子“睡了两天后，床垫中间凹陷了”表达了对床垫的负面意见。这个句子是客观的，因为它陈述了一个事实。所有这些句子都没有情感词。

这些问题都带来了重大挑战。事实上，这只是一些困难问题的一部分。更多内容将在第5章讨论。

1.2.3 自然语言处理问题

最后，我们不能忘记情感分析是自然语言处理的一个问题。它涉及到自然语言处理的各个方面，例如共指消解、否定处理和词义消歧，这增加了更多的困难，因为这些在自然语言处理中还没有解决的问题。然而，也有必要认识到情感分析是一个高度限制的自然语言处理问题，因为系统不需要完全理解每个句子或文档的语义，只需要理解其中的一些方面，即积极或消极的情感以及它们的目标实体或主题。从这个意义上说，情感分析为自然语言处理研究人员提供了一个很好的平台，可以在自然语言处理的各个方面取得实质性的进展，并有可能产生巨大的实际影响。在本书中，我将描述核心问题和当前最先进的算法。我希望利用本书吸引来自自然语言处理其他领域的研究人员加入力量，共同努力解决这个问题。

在这本书之前，有一本多作者的卷册“在文本中计算态度和情感”由Shanahan, Qu和Wiebe (2006) 编辑，还有一篇由Pang和Lee (2008) 撰写的调查文章/书籍。这两本书都有很好的内容。然而，它们在该领域的发展相对较早。自那时以来，由于过去5年中更加积极的研究，取得了显著的进展。

情感分析与观点挖掘

研究人员现在对问题的整个谱系、结构和核心问题有了更好的理解。已经提出了许多新的（正式的）模型和方法。研究不仅加深了，而且也显著扩大了。该领域的早期研究主要集中在对文档或句子中表达的情感或主观性进行分类，这对大多数实际应用来说是不够的。

实际应用通常需要更深入和细致的分析。

由于该领域的成熟度，这本书也以结构化的形式写成，从问题的定义角度来看，问题现在更加明确，不同的研究方向围绕着这个定义统一起来。

1.3 观点垃圾邮件检测

社交媒体的一个关键特点是它使得世界上任何地方的任何人都可以自由表达他/她的观点和意见，而不需要透露他/她的真实身份，也不用担心不良后果。因此，这些观点非常有价值。然而，这种匿名性也是有代价的。它使得那些隐藏动机或恶意意图的人可以轻易操纵系统，给人们留下他们是独立的公众成员并发布虚假观点来推广或诋毁目标产品、服务、组织或个人的印象，而不需要透露他们的真实意图，或者他们秘密为之工作的人或组织。这样的个体被称为观点垃圾邮件发送者，他们的活动被称为观点垃圾邮件发送（Jindal和Liu，2008；Jindal和Liu，2007）。

观点垃圾信息已成为一个重大问题。除了在评论和论坛讨论中提供虚假意见的个人外，还有一些商业公司专门为客户撰写虚假评论和虚假博客。新闻中报道了几起高调的虚假评论案例。检测此类垃圾信息活动非常重要，以确保网络上的观点是可信赖的有价值信息来源。与提取正面和负面观点不同，观点垃圾信息检测不仅仅是一个自然语言处理问题，因为它涉及对人们发布行为的分析。因此，它也是一个数据挖掘问题。第10章将讨论当前最先进的检测技术。

1.4 未来展望

在这本书中，我们探讨了迷人的主题。尽管这本书涉及

情感分析与观点挖掘

自然语言文本，通常被称为非结构化数据，我采用了一种结构化的方法来撰写这本书。下一章将正式地define这个问题，这让我们能够看到问题的结构。从这个定义中，我们将看到情感分析的关键任务。在随后的章节中，描述了执行这些任务的现有技术。由于我的研究、咨询和创业经验，这本书不仅讨论了关键的研究概念，还从应用的角度来看待技术，以帮助领域的从业者。然而，我必须道歉，当我谈论工业系统时，我不能透露公司的名称或其系统，部分原因是因为我的咨询/商业协议，部分原因是因为情感分析市场变化迅速，我所了解的公司在你阅读本书时可能已经改变或改进了他们的算法。我不想给他们和我自己制造麻烦。

尽管我试图涵盖本书中的所有主要思想和技术，但这已经成为一项不可能的任务。在过去的十年中，关于这个主题已经发表了大量的研究论文（可能超过1000篇）。

尽管大多数论文出现在自然语言处理的会议和期刊中，但也有很多论文发表在数据挖掘、网络挖掘、机器学习、信息检索、电子商务、管理科学和其他领域。因此，几乎不可能编写一本涵盖每篇发表论文中的思想的书籍。如果您的好想法或技术被忽视了，我很抱歉。然而，将这本书出版在Morgan & Claypool的综合讲座系列中的一个主要优势是，作者可以随时添加新的或更新的材料到这本书中，因为印刷是按需的。

因此，如果您发现一些重要的想法没有被讨论，请不要犹豫告诉我，我将非常乐意包含在内。

最后，以下领域的背景知识对于阅读本书将非常有帮助：自然语言处理（Indurkha和Damerau, 2010；Manning和Schutze, 1999），机器学习（Bishop, 2006；Mitchell, 1997），数据挖掘（Liu, 2006和2011），以及信息检索（Manning, Raghavan和Schutze, 2008）。

第2章

情感分析的问题

在本章中，我们定义了情感分析或观点挖掘问题的一个抽象。从研究的角度来看，这个抽象给我们提供了一个问题的陈述，并使我们能够看到构成情感分析问题的一组丰富的相互关联的子问题。常常有人说，如果我们不能对一个问题进行结构化，那么我们可能不理解这个问题。因此，定义的目标是从复杂和令人生畏的非结构化自然语言文本中抽象出一个结构。它们还作为一个共同的框架，统一各种现有的研究方向，并通过利用子问题之间的相互关系，使研究人员能够设计出更强大和准确的解决方案技术。从实际应用的角度来看，这些定义让从业者看到在一个实际系统中需要解决哪些子问题，它们之间的关系如何，以及应该产生什么输出。

与事实信息不同，观点和情感具有一个重要的特征，即它们是主观的。因此，重要的是要检查来自许多人的观点集合，而不仅仅是来自一个人的单一观点，因为这样的观点只代表了那个单个人的主观看法，通常不足以应用。由于网络上有大量的观点，需要某种形式的观点摘要（Hu和Liu，2004年）。问题定义说明了可能需要什么样的摘要。除了问题定义，本章还将讨论几个相关概念，如主观性和情感。

请注意，在本章以及整本书中，我主要使用评论和评论中的句子作为示例来介绍思想和定义关键概念，但这些思想和定义适用于所有形式的正式和非正式观点文本，如新闻文章、推文（Twitter帖子）、论坛讨论、博客和Facebook帖子。由于产品评论具有高度的聚焦性和丰富的观点，它们使我们能够更清楚地看到不同的问题，而不是从其他形式的观点文本中。从概念上讲，它们之间没有区别。区别主要是表面的，以及处理它们的难度程度。例如，Twitter帖子（推文）是短的（最多140个字符）和非正式的，并且使用许多互联网俚语和表情符号。实际上，由于长度限制，分析Twitter帖子更容易。

作者通常直截了当。因此，实现高情感分析准确性通常更容易。评论也更容易，因为它们专注度高，没有多余的信息。论坛讨论可能是最难处理的，因为用户可以讨论任何事情，并且彼此之间也可以互动。在难度上，还有不同应用领域的维度。对产品和服务的意见通常更容易分析。由于复杂的主题和情感表达，讨论社会和政治问题要困难得多，还有讽刺和反讽。

2.1 问题定义

正如在第1章开头提到的，情感分析主要研究表达或暗示积极或消极情感的观点。

因此，本节在这个背景下定义了问题。

2.1.1 观点定义

我们使用以下关于佳能相机的评论来介绍问题（每个句子都有一个关联的ID号以便参考）：

发布者：约翰·史密斯 日期：2011年9月10日“(1)我六个月前买了一台佳能G12相机。(2)我非常喜欢它。(3)画质很棒。(4)电池寿命也很长。(5)然而，我的妻子认为它对她说太重了。”

从这篇评论中，我们注意到几个重要的观点：

1. 这篇评论有很多关于佳能G12相机的观点，既有积极的，也有消极的。第二句表达了对佳能相机整体的积极观点。第三句表达了对其画质的积极观点。第四句表达了对其电池寿命的积极观点。第五句表达了对相机重量的消极观点。从这些观点中，我们可以得出以下重要观察：

观察：一个观点由两个关键组成部分构成：一个目标 g 和一个关于目标的情感 s ，即， (g, s) ，

其中 g 可以是任何实体或实体的方面，关于其已经表达了观点，而 s 是一个积极的、消极的或中立的情感，或者是表示强度/强度的数值评分。

情感分析与观点挖掘

情感的评价（例如，1到5星）。积极、消极和中性被称为情感（或观点）取向（或极性）。

例如，句子（2）中的观点目标是佳能 *G12*，句子（3）中的观点目标是佳能 *G12* 的图片质量。在文献中，目标也被称为话题。

2. 这篇评论有两个人的观点，被称为观点来源或观点持有者（Kim和Hovy, 2004; Wiebe, Wilson和Cardie, 2005）。句子（2）、（3）和（4）中的观点持有者是评论的作者（“John Smith”），但对于句子（5），观点持有者是作者的妻子。
3. 评论的日期是2011年9月10日。这个日期在实践中很重要，因为人们经常想知道观点如何随时间变化以及观点趋势。

我们现在可以定义观点为一个四元组。

定义(观点): 一个观点是一个四元组，

$$(g, s, h, t),$$

其中 g 是观点（或情感）的目标， s 是关于目标的情感， h 是观点持有者， t 是表达观点的时间。

尽管这个定义非常简洁，但在实践中可能不容易使用，特别是在产品、服务和品牌的在线评论领域，因为目标的完整描述可能很复杂，甚至可能不在同一个句子中出现。例如，在句子（3）中，观点的目标实际上是“佳能G12的画质”，但句子中只提到了“画质”。在这种情况下，观点的目标不仅仅是“画质”，因为如果不知道该句子是在评估佳能G12相机的画质，那么仅凭句子（3）中的观点是没有多大用处的。在实践中，目标通常可以被分解并以结构化的方式描述，具有多个层次，这极大地促进了观点的挖掘和后续使用挖掘结果。例如，“佳能G12的画质”可以被分解为一个实体和实体的属性，并表示为一对，

（佳能-G12，图片质量）

让我们使用术语实体来表示已经评估的目标对象。

实体可以定义如下（胡和刘，2004年；刘，2006年和2011年）。

定义（实体）：一个实体 e 是一个产品、服务、主题、问题、人物、组织或事件。它用一对 $e: (T, W)$ 来描述，其中 T 是一个层次结构的部分、子部分等等， W 是 e 的一组属性。

情感分析与观点挖掘

每个部分或子部分也有自己的一组属性。

示例1：相机的一个特定型号是一个实体，例如佳能G12。它有一组属性，例如图片质量、尺寸和重量，以及一组部件，例如镜头、取景器和电池。电池也有自己的一组属性，例如电池寿命和电池重量。一个主题也可以是一个实体，例如税收增加，它有“给穷人的税收增加”、“给中产阶级的税收增加”和“给富人的税收增加”等部分。这个定义基本上描述了基于部分关系的层次分解实体。根节点是实体的名称，例如上面评论中的佳能G12。所有其他节点都是部分和子部分等。可以对任何节点和节点的任何属性表达观点。

例子2：在我们上面的例子评论中，句子（2）整体上表达了对佳能G12相机的积极观点。句子（3）对相机的画质属性表达了积极观点。显然，人们也可以对相机的部件或组件表达观点。

这个实体作为一个层次结构需要一个嵌套关系来表示，这对于应用程序来说通常太复杂了。主要原因是因为自然语言处理是一项非常困难的任務，识别实体的不同层次上的部件和属性非常困难。大多数应用程序也不需要如此复杂的分析。因此，我们将层次结构简化为两个层次，并使用术语“方面”来表示部件和属性。在简化的树中，根节点仍然是实体本身，但第二层（也是叶子层）节点是实体的不同方面。这个简化的框架通常在实际情感分析系统中使用。

请注意，在研究文献中，实体也被称为对象，而方面也被称为特征（如产品特征）。然而，这里的特征可能会与机器学习中使用的特征混淆，其中一个特征表示数据属性。为了避免混淆，方面在近年来变得更加流行。请注意，一些研究人员还使用方面，属性和主题这些术语，在特定应用中，实体和方面也可能根据应用领域的惯例被称为其他名称。

在分解观点目标后，我们可以重新定义一个观点（Hu and Liu, 2004; Liu, 2010）。

定义（观点）：一个观点是一个五元组，

$(\text{实体 } i, \text{方面 } ij, \text{情感 } ijk, \text{持有者 } k, \text{时间 } l)$ ，

其中实体 i 是一个实体的名称，方面 ij 是实体 i 的一个方面，情感 ijk 是实体 i 的方面 j 上的情感，持有者 k 是观点的持有者，时间 l 是观点由持有者 k 表达的时间。情感 ijk 可以是积极的、消极的或

情感分析与观点挖掘

中性，或以不同的强度/程度表达，例如，1到5星，大多数网站上的评论都是这样。当观点是关于实体本身作为一个整体时，特殊的方面GENERAL用于表示它。在这里， e_i 和 a_{ij} 一起表示观点目标。

关于这个定义的一些重要说明如下：

1. 在这个定义中，我们特意使用下标来强调五个元组中的五个信息必须相互对应。
也就是说，观点 s_{ijkl} 必须由观点持有者 h_k 关于实体 e_i 的方面 a_{ij} 在时间 t_l 上给出。任何不匹配都是错误的。
2. 这五个组成部分是必不可少的。一般情况下，如果缺少其中任何一个，都会有问题。例如，如果我们没有时间组件，我们将无法根据时间分析实体上的观点，而这在实践中通常非常重要，因为两年前的观点和昨天的观点是不同的。没有观点持有者也是有问题的。例如，在句子“市长受到市民的喜爱，但遭到州政府的批评”中，两个观点持有者“市民”和“州政府”显然对应用程序非常重要。
3. 该定义涵盖了大多数但并非所有可能的观点语义方面，这可能是任意复杂的。例如，它不包括“取景器和镜头太近”的情况，这表达了对两个部分距离的观点。它也不包括观点的上下文，例如“这辆车对一个高个子来说太小了”，这并不意味着这辆车对每个人来说都太小。

这里的上下文是“高个子”。还要注意，在实体的原始定义中，它是一个部分、子部分等的层次结构。每个部分都可以有自己的属性集。由于简化，五元组表示可能会导致信息丢失。例如，“墨水”是打印机的一个部分/组件。在打印机评论中，有人写道“这台打印机的墨水很贵。”这并不意味着打印机很贵（这表示了方面价格）。如果一个人不关心墨水的任何属性，这句话只是对墨水的负面观点，这是打印机实体的一个方面。然而，如果一个人还想研究关于墨水不同方面的观点，例如价格和质量，墨水需要被视为一个单独的实体。然后，五元组表示仍然适用，但需要保存部分-整体关系。当然，从概念上讲，我们也可以使用嵌套关系扩展观点目标的表示。尽管存在一些限制，该定义涵盖了观点的基本信息，对于大多数应用来说已经足够。正如我们上面提到的，过于复杂的定义会使问题变得极其困难。

4. 这个定义提供了一个将非结构化文本转化为结构化数据的框架。上面的五元组基本上是一个数据库模式，基于这个模式，提取的观点可以放入数据库表中。
然后可以使用整套数据库管理系统（DBMS）和OLAP工具对观点进行丰富的定性、定量和趋势分析。
5. 这里定义的观点只是一种类型的观点，称为常规观点。另一种类型是比较观点（Jindal和Liu, 2006b; Liu, 2006和2011），它需要一个不同的定义。第2.3节将讨论不同类型的观点。第8章定义和分析了比较观点。在本节的其余部分，我们只关注常规观点。为了简单起见，我们只称它们为观点。

2.1.2 情感分析任务

有了这个定义，我们现在可以介绍情感分析的目标和关键任务（Liu, 2010; Liu, 2006和2011）。

情感分析的目标：给定一个意见文档 d ，发现其中的所有

意见五元组 $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ d 。

关键任务来源于五元组的五个组成部分。第一个组成部分是实体。也就是说，我们需要提取实体。这个任务类似于信息抽取中的命名实体识别（NER）（Hobbs and Riloff, 2010; Mooney and Bunescu, 2005; Sarawagi, 2008）。因此，提取本身就是一个问题。在提取之后，我们还需要对提取的实体进行分类。在自然语言文本中，人们经常以不同的方式写同一个实体。例如，Motorola可能被写作Mot、Moto和Motorola。我们需要识别它们都指代同一个实体。

定义（实体类别和实体表达式）：实体类别表示一个唯一的实体，而实体表达式是文本中表示实体类别的实际单词或短语。

在特定应用中，每个实体类别（或简称实体）应该有一个唯一的名称。将实体表达式分组为实体类别的过程称为实体分类。

现在我们来看实体的方面。问题基本上与实体相同。例如，相机的图片、图像和照片是同一个方面。因此，我们需要提取方面表达式并对其进行分类。

定义（方面类别和方面表达式）：实体的方面类别表示实体的一个独特方面，而方面

情感分析与观点挖掘

表达式是文本中表示方面类别的实际词语或短语。

在特定应用中，每个方面类别（或简称方面）也应该有一个唯一的名称。将方面表达式分组为方面类别（方面）的过程称为方面分类。

方面表达式通常是名词和名词短语，但也可以是动词、动词短语、形容词和副词。以下定义对我们有用（Hu和Liu，2004年）。

定义（明确的方面表达）：被称为名词和名词短语的方面表达被称为明确的方面表达。

例如，在“这台相机的画质很好”中，“画质”是一个明确的方面表达。

定义（隐含的方面表达）：不是名词或名词短语的方面表达被称为隐含的方面表达。

例如，在“这台相机很贵”中，“贵”是一个隐含的方面表达，它暗示了方面价格。许多隐含的方面表达是用来描述或限定某些特定方面的形容词和副词，例如昂贵（价格）和可靠地（可靠性）。它们也可以是动词和动词短语，例如，“我可以轻松安装软件。”中的“安装”表示方面安装。隐含的方面表达不仅仅是形容词、副词、动词和动词短语；它们也可以非常复杂，例如，“这台相机不容易放进口袋。”这里，“放进口袋”表示方面大小（和/或形状）。

观点定义中的第三个组成部分是情感。这个任务将情感分类为积极、消极或中性，同时第四和第五个组成部分是观点持有者和时间。它们还需要被提取和分类为实体和方面。请注意，观点持有者（Bethard等人，2004年；Choi等人，2005年；Kim和Hovy，2004年）（在（Wiebe，Wilson和Cardie，2005年）中也称为观点来源）可以是表达观点的个人或组织。对于产品评论和博客，观点持有者通常是帖子的作者。观点持有者对于新闻文章来说更加重要，因为它们经常明确说明持有观点的个人或组织。

然而，在某些情况下，识别观点持有者也可能在社交媒体中很重要，例如识别广告商或引用公司广告的人的观点。

基于以上讨论，我们可以定义一个实体模型和一个观点文档模型（Liu，2006年和2011年）。

情感分析与观点挖掘

实体模型: 一个实体 e_i 被自身和一个有限的一组方面 $A_i = \{a_{i1}, a_{i2}, \dots, a_{in}\}$ 表示。 e_i 可以用其有限集合中的任何一个实体表达 $\{ee_{i1}, ee_{i2}, \dots, ee_{is}\}$ 。实体 e_i 的每个方面 $a_{ij} \in A_i$ 可以用其有限集合中的任何一个方面表达 $\{ae_{ij1}, ae_{ij2}, \dots, ae_{ijm}\}$ 。

观点文档模型: 一个观点文档 d 包含对一组实体 $\{e_1, e_2, \dots, e_r\}$ 和它们的一部分方面的观点, 来自一组观点持有者 $\{h_1, h_2, \dots, h_p\}$ 在某个特定时间点。

最后, 总结一下, 给定一组观点文档 D , 情感分析包括以下6个主要任务。

任务1(实体提取和分类): 提取 D 中的所有实体表达式, 并将同义实体表达式归类或分组为实体簇 (或类别)。每个实体表达式簇表示一个唯一的实体 e_{io} 。

任务2(方面提取和分类): 提取实体的所有方面表达式, 并将这些方面表达式归类为簇。每个实体 e_i 的方面表达式簇表示一个唯一的方面 a_{ijo} 。

任务3(观点持有者提取和分类): 从文本或结构化数据中提取观点持有者, 并对它们进行分类。

这个任务类似于上面的两个任务。

任务4 (时间提取和标准化): 提取给出观点的时间并标准化不同的时间格式。这个任务也类似于上面的任务。

任务5 (方面情感分类): 确定对方面 a_{ij} 的观点是积极的、消极的还是中立的, 或者为方面分配一个数值情感评级。

任务6 (观点五元组生成): 根据上述任务的结果, 在文档 d 中生成所有表达的观点五元组 $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ 。这个任务看起来很简单, 但实际上在很多情况下非常困难, 正如下面的示例4所示。

基于这个框架的情感分析 (或观点挖掘) 通常被称为基于方面的情感分析 (或观点挖掘), 或者在 (Hu and Liu, 2004; Liu, Hu and Cheng, 2005) 中被称为基于特征的情感分析 (或观点挖掘)。

现在我们使用一个示例博客来说明任务 (每个句子都有一个句子ID) 和分析结果。

"示例4: "发布者: 大约翰 "日期: 2011年9月15日

"(1)昨天我买了一台三星相机, 我的朋友带了一台佳能相机。"(2)在过去的一周里, 我们都经常使用这些相机。"(3)我的三星相机拍出来的照片并不是很好, 而且电池寿命也很短。

情感分析与观点挖掘

"(4)我的朋友对他的相机非常满意，喜欢它的画质。"(5)我想要一台能拍出好照片的相机。"(6)我打算明天退掉它。

"任务1应该提取实体表达式“三星”，“三星”和“佳能”，并将“三星”和“三星”归为一组，因为它们代表同一实体。"任务2应该提取方面表达式“画质”，“照片”和“电池寿命”，并将“画质”和“照片”归为一组，因为它们在相机中是同义词。"任务3应该找到句子（3）中意见的持有者是大约翰（博客作者），句子（4）中意见的持有者是大约翰的朋友。"任务4还应该找到博客发布的时间为2011年9月15日。任务5应该发现句子（3）对三星相机的画质和电池寿命都持有负面意见。"句子（4）对佳能相机整体和画质都持有正面意见。"句子（5）表面上表达了正面意见，但实际上并不是。为了为句子（4）生成意见五元组，我们需要知道“他的相机”和“它”指的是什么。任务6最终应该生成以下四个意见五元组：（三星，画质，负面，大约翰，2011年9月15日）

（三星，电池寿命，负面，大约翰，2011年9月15日）

（佳能，一般，积极，大约翰的朋友，2011年9月15日）

（佳能，图片质量，积极，大约翰的朋友，2011年9月15日）

2.2 观点总结

与事实信息不同，观点本质上是主观的。一个观点来自单个观点持有者通常不足以采取行动。在大多数应用中，需要分析来自大量人的观点。这表明需要某种形式的观点总结。尽管观点总结可以采用多种形式，例如结构化总结（见下文）或简短文本总结，但总结的关键组成部分应包括对不同实体及其方面的观点，并且还应具有定量的角度。定量角度尤为重要，因为对一个产品持积极态度的人占20%与占80%是非常不同的。我们将在第7章进一步讨论这个问题。

上述定义的观点五元组实际上提供了一个很好的信息来源，也为生成定性和定量摘要提供了一个框架。一种常见的摘要形式是基于方面的观点摘要（或特征-based观点摘要）（Hu和Liu，2004；Liu，Hu和Cheng，2005）。在过去的几年中

数码相机 1:

方面：常规

积极： 105 <个别评论句子>
消极： 12 <个别评论句子>

方面：画质

积极： 95 <个别评论句子>
消极： 10 <个别评论句子>

方面：电池寿命

积极： 50 <个别评论句子>
消极： 9 <个别评论句子>

...

图2.1。基于方面的观点总结。

，关于观点摘要已经进行了大量的研究。
其中大部分与这个框架有关（参见第7章）。

让我们用一个例子来说明这种形式的摘要，这个例子是在（Hu和Liu，2004; Liu, Hu和Cheng，2005）中提出的。我们总结了一组关于一台数码相机的评论，称为数码相机1。摘要看起来像图2.1中的那样，它被称为结构化摘要，与从一个或多个长文档生成的短文档的传统文本摘要形成对比。在图中，GENERAL代表相机本身（实体）。105条评论表达了对相机的积极意见，而12条评论表达了消极意见。画质和电池寿命是两个相机方面。95条评论表达了对画质的积极意见，而10条评论表达了消极意见。<个别评论句子>是一个指向给出观点的句子和/或整个评论的链接。有了这样的摘要，人们可以很容易地了解现有客户对相机的感受。如果有人对特定方面和额外细节感兴趣，他/她可以通过跟随<个别评论句子>链接来查看实际的观点句子或评论。

2.3 不同类型的观点

到目前为止，我们讨论的观点类型被称为常规观点（Liu，2006年和2011年）。另一种类型被称为比较观点（Jindal和Liu，2006b）。实际上，我们还可以根据它们在文本中的表达方式对观点进行分类，明确观点和隐含（或暗示）观点。

2.3.1 正常和比较观点

常规观点：常规观点通常被简称为

情感分析与观点挖掘

这是文献中的一种观点，并且它有两个主要子类型（Liu，2006年和2011年）：直接观点：直接观点是指直接表达的观点对实体或实体方面的观点，例如，“画质很好。”间接观点：间接观点是指基于实体或实体方面对其他实体的影响间接表达的观点。这种子类型经常出现在医学领域中。

例如，“注射药物后，我的关节感觉更糟”描述了药物对“我的关节”的不良影响，间接给药物带来了负面观点或情感。在这种情况下，实体是药物，方面是对关节的影响。

当前研究主要关注直接观点。它们更容易处理。间接观点通常更难处理。例如，在药物领域，人们需要知道某种期望的和期望的状态是在使用药物之前还是之后。例如，“因为我的关节疼痛，医生给我开了这种药”这句话并没有对药物表达情感或观点，因为“疼痛的关节”（负面）发生在使用药物之前。

比较观点：比较观点表达了两个或多个实体之间的相似性或差异性关系，以及基于实体的某些共享方面的观点持有者的偏好（Jindal和Liu，2006a；Jindal和Liu，2006b）。例如，“可口可乐比百事可乐更好喝”和“可口可乐最好喝”表达了两个比较观点。比较观点通常使用形容词或副词的比较级或最高级形式来表达，尽管并不总是如此（例如，更喜欢）。比较观点也有许多类型。我们将在第8章中讨论并定义它们。

2.3.2 明确和隐含观点

明确的观点：明确的观点是一个主观陈述，给出了一个常规或比较的观点，例如，

“可口可乐很好喝，”和

“可口可乐比百事可乐更好喝。”

隐含（或暗示）的观点：隐含的观点是一个客观陈述

，暗示了一个常规或比较的观点。这样一个客观陈述通常表达了一个可取或不可取的事实，例如，

“我一周前买了这个床垫，现在已经凹陷了，”和

“诺基亚手机的电池寿命比三星手机长。”

明确的观点比隐含的观点更容易检测和分类。

目前的研究大部分集中在明确的观点上。相对而言，

对隐含观点的研究工作较少 (Zhang and Liu, 2011b)。在一个稍微不同的方向上, (Greene and Resnik, 2009) 研究了句法选择对隐含情感的影响。例如, 对于同一个故事, 不同的标题可以暗示不同的情感。

2.4 主观性和情感

有两个与情感和观点密切相关的重要概念, 即主观性和情感。

定义 (句子主观性): 客观句子提供关于世界的一些事实信息, 而主观句子表达个人的感受、观点或信仰。

一个例子是“iPhone是苹果产品”。一个主观句子的例子是“我喜欢iPhone”。主观表达有很多形式, 例如意见、指控、欲望、信仰、怀疑和猜测。有些研究人员混淆了主观性和持有观点。

通过持有观点, 我们指的是文档或句子表达或暗示了积极或消极的情感。这两个概念并不等同, 尽管它们有很大的交集。确定一个句子是主观还是客观的任务被称为主观性分类 (参见第4章)。在这里, 我们应该注意以下事项:

- 一个主观句子可能不表达任何情感。例如, “我认为他回家了”是一个主观句子, 但并不表达任何情感。例4中的句子 (5) 也是主观的, 但它对任何事情都没有给出积极或消极的情感。
- 客观句子可能由于可取和不可取的事实而暗示观点或情感 (Zhang and Liu, 2011b)。例如, 下面两个陈述一些事实的句子明确暗示了消极的情感 (这是隐含的观点) 关于它们各自的产品, 因为这些事实是不可取的:

“耳机两天就坏了。”

“我一个星期前买的床垫, 已经凹陷了”

除了明确表达观点的主观表达式外, 还研究了许多其他类型的主观性, 尽管没有那么广泛, 例如情感、判断、欣赏、推测、避险、观点、争论、一致和不一致、政治立场 (Alm, 2008; Ganter and Strube, 2009; Greene and Resnik, 2009; Hardisty, Boyd-Graber and Resnik, 2010; Lin et al., 2006; Medlock and Briscoe, 2007; Mukherjee and Liu,

情感分析与观点挖掘

2012; Murakami and Raymond, 2010; Neviarouskaya, Prendinger and Ishizuka, 2010; Somasundaran and Wiebe, 2009)。其中许多也可能暗示情感。

定义（情绪）：情绪是我们主观的感受和思想。

情绪已经在多个领域进行了研究，例如心理学、哲学和社会学。这些研究非常广泛，从生理反应的情感反应（例如心率变化、血压、出汗等），面部表情、手势和姿势，到个体心理状态的不同类型的主观体验。科学家已经将人们的情绪分为一些类别。然而，研究人员之间仍然没有一套共同认可的基本情绪。根据（Parrott, 2001），人们有六种基本情绪，即爱，喜悦，惊讶，愤怒，悲伤和恐惧，这些情绪可以进一步细分为许多次要和三级情绪。每种情绪也可以有不同的强度。

情绪与情感密切相关。情感或观点的强度通常与特定情绪的强烈程度相关，例如，喜悦和愤怒。在情感分析中，我们研究的观点大多是评价（尽管不总是如此）。根据消费者行为研究，评价可以广泛分为两类：理性评价和情感评价（Chaudhuri, 2006）。

理性评价：这些评价来自于理性推理、具体信念和功利主义态度。例如，以下句子表达了理性评价：“这部手机的声音很清晰”，“这辆车物有所值”和“我对这辆车很满意”。

情感评价：这些评价来自于对实体的非具体和情感反应，深入人们的内心。

例如，以下句子表达了情感评价：“我喜欢iPhone”，“我对他们的服务人员非常生气”和“这是有史以来最好的车。”

为了在实践中利用这两种评估类型，我们可以设计5个情感评级，即情感负面(-2)、理性负面(-1)、中立(0)、理性正面(+1)和情感正面(+2)。在实践中，中立通常意味着没有表达意见或情感。

最后，我们需要注意情感和观点的概念明显不相等。理性观点不表达情感，例如“这部手机的声音很清晰”，而许多情感句子不对任何事物表达观点/情感，例如“见到你在这里我感到非常惊讶”。

更重要的是，情感可能没有目标，只是人们内心的感受，例如“我今天很伤心”。

2.5

作者和读者的立场

我们可以从两个角度来看待观点，即表达观点的作者（观点持有者）和阅读观点的读者。例如，有人写道“房价下跌，对经济不利”，显然，这位作者谈论了房价下跌对经济的负面影响。然而，读者可以以两种方式理解这句话。对于卖家来说，这确实是负面的，但对于买家来说，这可能是一个好消息。另一个例子，有人写道“我很高兴谷歌的股价今天飙升”。如果读者昨天以亏损的价格卖出了谷歌股票，他不会很高兴，但如果读者昨天买了很多谷歌股票，他几乎肯定会和这句话的作者一样高兴。

我不知道有关这个问题的任何报道研究。在当前的研究或应用中，研究人员要么忽视这个问题，要么在他们的分析中假设一个立场。通常情况下，观点持有者被认为是消费者或普通公众，除非另有说明（例如，美国总统）。如果产品制造商或服务提供商的观点被明确标记，则被视为广告；如果没有明确标记（例如与消费者的观点混合），则被视为虚假观点。

2.6 摘要

本章定义了情感分析背景下的观点概念，情感分析的主要任务以及观点总结的框架。除此之外，还介绍了与观点高度相关但并不等同的主观性和情绪两个相关且重要的概念。现有的研究大多集中在它们与观点的交叉点上（尽管并不总是如此）。然而，我们应该意识到所有这些概念及其定义都相当模糊和主观。例如，至今仍没有一套所有研究人员都同意的情绪集合。观点本身也是一个广泛的概念。情感分析主要处理评价类型的观点或暗示积极或消极情感的观点。如果你对本章的所有内容都不完全同意，我不会感到惊讶。本章的目标是给出情感分析及其相关问题的一个相当精确的定义。我希望我在某种程度上取得了成功。

第三章

文档情感分类

从这一章开始，我们将讨论当前的主要研究方向或主题以及它们的核心技术。情感分类可能是最广泛研究的主题之一（也可以参考调查（Pang和Lee, 2008））。它的目标是将一个意见文档分类为表达积极或消极的意见或情感。这个任务通常也被称为文档级情感分类，因为它将整个文档视为基本信息单元。关于这个主题的大多数研究论文都对在线评论进行分类。因此，我们也将问题定义在评论的背景下，但这个定义也适用于其他类似的背景。

问题定义：给定一个评估实体的意见文档 d ，确定意见持有者关于实体的整体情感 s ，即确定在五元组中关于GENERAL方面的 s 表达

$$(_, GENERAL, s, _, _),$$

其中实体 e ，意见持有者 h 和意见的时间 t 被假设为已知或不相关（不重要）。

基于值类型的两种表达式。如果值类型是分类值，例如正面和负面，则是一个分类问题。如果值类型是数值或在给定范围内的有序分数，例如1到5，则问题变为回归问题。

为了确保任务在实践中有意义，现有研究做出以下隐含假设（Liu, 2010）：

假设：情感分类或回归假设意见文档 d （例如产品评论）对单个实体 e 表达意见，并包含来自单个意见持有者 h 的意见。

在实践中，如果一个意见文档评估多个实体，则对这些实体的情感可能不同。例如，意见持有者可能对某些实体持积极态度，对其他实体持消极态度。因此，在这种情况下，将一个情感倾向分配给整个文档是没有实际意义的。如果多个意见持有者在一个文档中表达意见，因为他们的意见可能不同，这也没有多大意义。

这个假设适用于产品和服务的评论，因为每个评论通常都集中在评估一个单一的产品或服务上，并且是由一个单一的评论者撰写的。

然而，这个假设可能不适用于论坛和博客文章，因为在这样的文章中，作者可能会对多个实体表达观点，并使用比较句进行比较。然而，这个假设可能不适用于论坛和博客文章，因为在这样的文章中，作者可能会对多个实体表达观点，并使用比较句进行比较。

首先，我们讨论分类问题，以预测分类类别标签，然后是回归问题，以预测评分。大多数现有的文档级分类技术使用监督学习，尽管也有无监督方法。情感回归主要使用监督学习进行。最近，还出现了几个扩展研究，尤其是跨领域情感分类（或领域适应）和跨语言情感分类，这也将详细讨论。

3.1 使用监督学习的情感分类

情感分类通常被定义为一个两类分类问题，即正面和负面。通常使用产品评论作为训练和测试数据。由于在线评论通常由评论者分配评分，例如1-5星级，所以可以通过评分来确定正面和负面类别。例如，4或5星级的评论被认为是正面评论，而1到2星级的评论被认为是负面评论。大多数研究论文不使用中性类别，这使得分类问题变得更加容易，但也可以使用中性类别，例如将所有3星级的评论分配给中性类别。

情感分类本质上是一个文本分类问题。

传统的文本分类主要是对不同主题的文档进行分类，例如政治、科学和体育。在这样的分类中，与主题相关的词语是关键特征。然而，在情感分类中，表示积极或消极观点的情感或意见词更加重要，例如很棒、优秀、令人惊讶、可怕、糟糕、最差等。由于这是一个文本分类问题，可以应用任何现有的监督学习方法，例如朴素贝叶斯分类和支持向量机（SVM）（Joachims, 1999; Shawe-Taylor和Cristianini, 2000）。

Pang、Lee和Vaithyanathan（2002）是第一篇采用这种方法将电影评论分类为积极和消极两类的论文。研究表明，在分类中使用单个词语（词袋）作为特征，无论是朴素贝叶斯还是SVM，都表现得相当好，尽管作者还尝试了其他一些特征选项。

情感分析与观点挖掘

在随后的研究中，许多研究人员尝试了更多的特征和学习算法。与其他监督式机器学习应用一样，情感分类的关键在于构建一组有效特征。一些示例特征包括：术语及其频率。这些特征是单个词（unigram）及其n-gram与相关频率计数。它们也是传统基于主题的文本分类中最常用的特征。

在某些情况下，还可以考虑词语位置。还可以应用信息检索中的TF-IDF加权方案。与传统文本分类一样，这些特征在情感分类中也被证明非常有效。

词性。每个词的词性也可能很重要。不同词性的词可能会被不同对待。例如，研究表明形容词是观点的重要指标。因此，一些研究人员将形容词视为特殊特征。

然而，人们也可以使用所有词性标签及其n-gram作为特征。请注意，在本书中，我们使用标准的*Penn Treebank*词性标签，如表3.1所示（Santorini, 1990）。*Penn Treebank*网站位于<http://www.cis.upenn.edu/~treebank/home.html>。

情感词和短语。情感词是一种语言中用来表达积极或消极情感的词语。例如，好、美妙和令人惊叹是积极情感词，而坏、差和可怕是消极情感词。大多数情感词是形容词和副词，但名词（例如，垃圾、废物和废话）和动词（例如，讨厌和喜爱）也可以用来表达情感。除了单个词语外，还有情感短语和习语，例如，花费某人一条手臂和一条腿。

意见规则。除了情感词和短语之外，还有很多其他表达或语言组合可以用来表达或暗示情感和观点。我们将在第5.2节列举和讨论一些这样的表达方式。

情感转变者。这些表达式用于改变情感倾向，例如从积极到消极或反之亦然。

否定词是情感转变者中最重要类别。例如，句子“我不喜欢这个相机”是消极的。还有其他几种类型的情感转变者。我们也将第5.2节中讨论它们。这些转变者也需要小心处理，因为并非所有出现的词都意味着情感的改变。例如，“不仅……而且”中的“不”不会改变情感倾向。

句法依赖。研究人员还尝试使用从解析或依赖树生成的基于词依赖的特征。

情感分析与观点挖掘

表3.1.宾州树库词性（POS）标签

标签	描述	标签	描述
CC	并列连词	PRP\$	所有格代词
CD	基数	RB	副词
DT	限定词	RBR	副词，比较级
EX	存在 <i>there</i>	RBS	副词，最高级
FW	外来词	RP	小品词
IN	介词 从属连词	SYM	符号
JJ	形容词	TO	到
JJR	形容词，比较级	UH	感叹词
JJS	形容词，最高级	VB	动词，基本形式
LS	列表项标记	VBD	动词，过去式
MD	情态动词	VBG	动词，动名词或现在分词
NN	名词，单数或不可数	VBN	动词，过去分词
NNS	名词，复数	VBP	动词，非第三人称单数 现在时
NNP	专有名词，单数	VBZ	动词，第三人称单数现在时
NNPS	专有名词，复数	WDT	Wh-限定词
PDT	前限定词	WP	Wh-代词
POS	所有格结束	WP\$	所有格疑问代词
PRP	人称代词	WRB	疑问副词

研究人员除了使用标准的机器学习方法外，还提出了几种专门用于情感分类的定制技术，例如基于积极和消极评论中的单词的评分函数(Dave, Lawrence和Pennock, 2003)，以及使用手动编译的领域特定单词和短语的聚合方法(Tong, 2001)。

大量的论文已经在文献中发表。在这里，我们简要介绍它们。在(Gamon, 2004)中，对客户反馈数据进行了分类，这些数据通常与评论相比较短且嘈杂。在(Pang和Lee, 2004)中，采用了在图上工作的最小割算法来帮助情感分类。在(Mullen和Collier, 2004; Xia和Zong, 2010)中，句法关系与传统特征一起使用。在(Kennedy和Inkpen, 2006; Li等, 2010)中，采用了上下文情感和情感转移词进行分类。

在(Cui, Mittal和Datar, 2006)中，报告了一项评估，当时有几种可用的情感分类算法。在(Ng, Dasgupta和Arifin, 2006)中，使用了一些语言知识源进行分类。在(Abbasi, Chen和Salem, 2008)中，提出了一种基于遗传算法的特征选择方法，用于不同语言的情感分类。在(Li, Zhang和Sindhwani, 2009)中，提出了一种非负矩阵分解方法。在(Dasgupta和Ng, 2009; Li等, 2011; Zhou, Chen和Wang, 2010)中，使用了半监督学习和/或主动学习。

进行了学习。在(Kim, Li和Lee, 2009)和(Paltoglou和Thelwall, 2010)中, 研究了不同的IR术语加权方案, 并进行了情感分类比较。在(Martineau和Finin, 2009)中, 提出了一种新的术语加权方案Delta TFIDF。在(Qiu等人, 2009)中, 采用了基于词典的自我监督方法。在(He, 2010)中, 利用了标记特征(而不是标记文档)进行分类。在(Mejova和Srinivasan, 2011)中, 作者探索了各种特征定义和选择策略。在(Nakagawa, Inui和Kurohashi, 2010)中, 提出了一种基于依赖树的分类方法, 该方法使用了带有隐藏变量的条件随机场(CRF)(Lafferty, McCallum和Pereira, 2001)。在(Bickerstaffe和Zukerman, 2010)中, 报告了一种考虑类间相似性的分层多分类器。在(Li等人, 2010)中, 利用了个人(我, 我们)和非人称(他们, 它, 这个产品)句子来帮助分类。在(Yessenalina, Choi和Cardie, 2010)中, 使用了自动生成的注释者理由来帮助分类。在(Yessenalina, Yue和Cardie, 2010)中, 提出了多级结构模型。在(Wang等人, 2011)中, 作者提出了一种基于图的标签方法来分类Twitter帖子情感, 并且在(Kouloumpis, Wilson和Moore, 2011)中, 还利用了微博中使用的非正式和创造性语言的语言特征和特征。在(Maas等人, 2011)中, 作者使用了可以捕捉单词的一些潜在方面的词向量来帮助分类。在(Bespalov等人, 2011)中, 基于监督的潜在n-gram分析进行了情感分类。在(Burfoot, Bird和Baldwin, 2011)中, 对国会辩论进行了分类。在(Becker和Aronson, 2010)中, 作者通过他们的心理语言学和心理物理学实验表明, 情感分类应该关注文本的最后部分。在(Liu等人, 2010)中, 比较了博客和评论情感分类的不同语言特征。在(Tokuhisa, Inui和Matsumoto, 2008)中, 研究了对话话语的情感分类。首先对三个类别(积极、消极和中性)进行了情感分类, 然后将积极和消极话语分类为10个情感类别。

3.2 使用无监督学习的情感分类

由于情感词通常是情感分类的主要因素, 可以想象情感词和短语可以以无监督的方式用于情感分类。该方法

情感分析与观点挖掘

在 (Turney, 2002) 中是一种技术。它基于一些固定的句法模式进行分类, 这些模式可能用于表达观点。

句法模式是基于词性标记 (POS) 组成的。

(Turney, 2002) 中给出的算法包括三个步骤: 步骤1: 如果两个

连续的词的词性标记符合表3.2中的任何模式, 则提取这两个词。例如, 模式2表示如果第一个词是副词, 第二个词是形容词, 并且第三个词 (不提取) 不是名词, 则提取两个连续的词。例如, 在句子“这个钢琴发出美妙的声音”中,

“美丽的声音”被提取出来, 因为它满足第一个模式。使用这些模式的原因是因为JJ、RB、RBR和RBS这些词经常表达观点。名词或动词作为上下文, 因为在不同的上下文中, JJ、RB、RBR和RBS这些词可能表达不同的情感。

例如, 形容词 (JJ) “不可预测”在汽车评论中可能有负面情感, 如“不可预测的转向”, 但在电影评论中可能有积极情感, 如“不可预测的剧情”。第二步: 使用点互信息 (PMI) 测量来估计提取短语

的情感倾向 (SO) :

$$PMI(\text{术语}_1, \text{术语}_2) = \log_2 \left(\frac{\Pr(\text{术语}_1 \text{ 术语}_2)}{\Pr(\text{术语}_1) \Pr(\text{术语}_2)} \right) \quad (1)$$

PMI衡量了两个术语之间的统计依赖程度。

这里, $\Pr(\text{术语}_1 \wedge \text{术语}_2)$ 是术语₁和术语₂的实际共现概率, 而 $\Pr(\text{术语}_1) \Pr(\text{术语}_2)$ 是两个术语在它们在统计上独立的情况下的共现概率。短语的情感倾向 (SO) 是基于它与积极参考词“优秀”和消极参考词“糟糕”的关联计算的:

$$SO(\text{短语}) = PMI(\text{短语}, \text{“优秀”}) - PMI(\text{短语}, \text{“差劲”}) \quad (2)$$

概率是通过向搜索引擎发出查询并收集命中数来计算的。对于每个搜索查询, 搜索引擎通常会给出与查询相关的文档数量, 即命中数。因此, 通过同时搜索这两个术语以及分别搜索它们,

表3.2.提取两个词组的POS标签模式

	第一个词	第二个词	第三个词 (未提取)
1	JJ	NN或NNS	任何内容
2	RB, RBR或RBS	JJ	既不是NN也不是NNS
3	JJ	JJ	既不是NN也不是NNS
4	NN或NNS	JJ	既不是NN也不是NNS
5	RB, RBR或RBS	VB, VBD, VBN或VBG	任何内容

情感分析与观点挖掘

可以估计方程 (1) 中的概率。在 (Turney, 2002) 中, 使用了 Alta Vista 搜索引擎, 因为它具有 NEAR 运算符, 可以将搜索限制在包含这些词的文档中, 这些词之间的距离不超过十个词。让命中数 (查询) 表示返回的命中数。方程 (2) 可以重写为:

$$SO(\text{短语}) = \log_2 \left(\frac{\text{命中}(\text{短语} \text{ 附近 "优秀"})}{\text{命中}(\text{短语} \text{ 附近 "差劲"})} \right) \quad (3)$$

步骤3: 给定一篇评论, 算法计算评论中所有短语的平均 SO , 并根据平均 SO 是否为正向来分类评论, 否则分类为负向。

来自不同领域的评论的最终分类准确率从汽车评论的84%到电影评论的66%不等。

另一种无监督方法是基于词典的方法, 它使用带有相关方向和强度的情感词和短语的词典, 并结合强化和否定来计算每个文档的情感得分 (Taboada等, 2011年)。这种方法最初用于句子和方面级情感分类 (Ding, Liu和Yu, 2008年; Hu和Liu, 2004年; Kim和Hovy, 2004年)。

3.3 情感评级预测

除了对正面和负面情感进行分类外, 研究人员还研究了预测评论评分 (例如, 1-5 星级) 的问题 (Pang和Lee, 2005年)。在这种情况下, 问题可以被定义为回归问题, 因为评分是有序的, 尽管并非所有研究人员都使用回归技术解决这个问题。Pang和Lee (2005年) 尝试了 SVM 回归、SVM 多类别分类使用一对多 (OVA) 策略, 以及一种称为度量标签的元学习方法。结果显示, 基于 OVA 的分类明显比其他两种方法差。这是可以理解的, 因为数值评分不是分类值。Goldberg和Zhu (2006年) 通过将评分预测建模为基于图的半监督学习问题来改进这种方法, 该方法使用了带有评分的标记 (有评分) 和未标记的 (没有评分) 评论。未标记的评论也是需要预测评分的测试评论。在图中, 每个节点是一个文档 (评论), 两个节点之间的链接是两个文档之间的相似度值。较大的相似度权重意味着两个文档倾向于具有相同的情感评分。

情感分析与观点挖掘

本文尝试了几种不同的相似度方案。该算法还假设最初已经有一个单独的学习器预测了未标记文档的数值评分。基于图的方法只通过解决一个优化问题来修正评分，以使得整个图在评分和链接权重方面都更加平滑。

Qu、Ifrim和Weikum（2010）引入了一种基于观点的文档表示方法，以捕捉带有观点的n-gram的强度，这与传统的基于词袋的表示方法不同。每个观点都是一个三元组，包括情感词、修饰词和否定词。例如，在“不太好”的例子中，“好”是情感词，“很”是修饰词，“不”是否定词。对于两类情感分类（积极和消极），观点修饰词并不重要，但对于评分预测来说，它非常重要，否定词的影响也是如此。开发了一种约束岭回归方法，从领域无关的评级评论语料库中学习每个观点的情感得分或强度。学习的关键思想是利用可用的观点词典和评论评级。为了将回归模型转移到新给定的领域相关应用中，该算法推导出一组关于观点得分的统计数据，然后将它们作为额外特征与标准单词一起用于评分预测。在这项工作之前，（Liu和Seneff，2009）提出了一种基于句子解析成分层表示的方法来提取副词-形容词-名词短语（例如，“非常好的车”）。他们根据启发式方法为副词、状语和否定词对情感程度进行评分，该方法基于这些词在评论评级中的贡献。与上述工作不同，这项工作中没有涉及学习过程。

斯奈德和巴齐莱（2007年）并不是预测每个评论的评分，而是研究了预测每个方面的评分的问题。对于这个任务，一个简单的方法是使用标准的回归或分类技术。然而，这种方法没有充分利用用户在不同方面上的判断之间的依赖关系。了解这些依赖关系对于准确的预测是有用的。因此，本文提出了两个模型，即方面模型（用于处理各个方面）和一致性模型（用于模拟各个方面之间的评分一致性）。这两个模型在学习过程中进行了组合。用于训练的特征包括每个评论的词汇特征，如一元组和二元组。

Long, Zhang和Zhu（2010）采用了与（Pang和Lee，2005）类似的方法，但使用了贝叶斯网络分类器来预测评分。

评论中的每个方面。为了获得更准确的结果，他们不是对每个评论进行预测，而是专注于仅预测一部分选择的评论中的方面评分，这些评论全面评估了各个方面。显然，这些评论的估计结果应该比其他评论的估计结果更准确，因为其他评论没有足够的信息。评论选择方法使用了基于科尔莫哥洛夫复杂性的信息度量。选定评论的方面评分预测使用了机器学习。训练的特征仅来自与方面相关的句子。方面提取的方法与（Hu和Liu，2004）类似。

3.4 跨领域情感分类

已经证明情感分类对从训练数据中提取的领域非常敏感。使用来自一个领域的观点文档训练的分类器在另一个领域的测试数据上通常表现不佳。原因是在不同的领域中用于表达观点的词甚至语言结构可能非常不同。更糟糕的是，一个领域中的同一个词可能表示积极，但在另一个领域中可能表示消极。因此，需要领域适应或迁移学习。现有的研究主要基于两种设置。

第一种设置需要新领域的少量标记训练数据（Aue和Gamon，2005）。第二种设置不需要新领域的标记数据（Blitzer，Dredze和Pereira，2007；Tan等，2007）。原始领域具有标记的训练数据通常被称为源领域，用于测试的新领域被称为目标领域。

在（Aue和Gamon，2005）中，作者提出了在没有大量标记数据的情况下将情感分类器转移到新领域的方法。他们尝试了四种策略：（1）在其他领域的标记评论的混合数据上进行训练，并在目标领域进行测试；（2）在上述训练分类器的基础上，限制特征集仅包括目标领域中观察到的特征；（3）使用具有可用标记数据的领域的分类器集合，并在目标领域进行测试；（4）将少量标记数据与目标领域中的大量未标记数据相结合（这是传统的半监督学习设置）。前三种策略使用了SVM，而半监督学习中使用了EM（Nigam等，2000）。他们的实验表明，策略（4）表现最佳，因为它能够利用目标领域中的标记和未标记数据。

在 (Yang, Si和Callan, 2006) 中, 提出了一种基于特征选择的简单策略, 用于句子级分类的迁移学习。他们的方法首先使用来自两个领域的两个完全标记的训练集来选择在两个领域中排名靠前的特征。这些选定的特征被认为是领域无关的特征。然后, 使用这些特征构建的分类器被应用于任何目标/测试领域。在 (Tan等, 2007) 中提出了另一种简单的策略, 该策略首先使用来自源领域的标记数据训练基本分类器, 然后使用分类器标记目标领域中的一些信息性示例。基于目标领域选择的示例, 学习一个新的分类器, 最后将其应用于目标领域中的测试用例。

在 (Blitzer, Dredze和Pereira, 2007) 中, 作者使用了一种称为结构对应学习 (SCL) 的方法进行领域适应, 该方法早在 (Blitzer, McDonald和Pereira, 2006) 中提出。给定来自源领域的标记评论以及来自源领域和目标领域的未标记评论, SCL首先选择一组在两个领域中频繁出现且对源标签也是良好预测器的特征 (该论文选择了与源标签的互信息最高的特征)。这些特征被称为枢轴特征, 它们代表了两个领域的共享特征空间。然后, 计算每个枢轴特征与两个领域中其他非枢轴特征的相关性。这产生了一个相关矩阵 W , 其中第 i 行是与第 i 个枢轴特征的非枢轴特征的相关值的向量。直观地, 正值表示这些非枢轴特征与源领域或新领域中的第 i 个枢轴特征呈正相关。这建立了两个领域之间的特征对应关系。然后, 使用奇异值分解 (SVD) 来计算低维线性逼近 θ (W 的前 k 个左奇异向量的转置)。用于训练和测试的最终特征集是原始特征集 x 与 θx 的组合, 产生 k 个实值特征。使用组合特征和源领域中的标记数据构建的分类器应该在源领域和目标领域都能工作。

潘等人 (潘等人, 2010年) 在高层提出了一种类似于 SCL 的方法。该算法适用于只有标记示例在源域中和未标记示例在目标域中的情况。它通过使用谱特征对齐

(SFA) 算法将来自不同领域的领域特定词汇对齐到统一的聚类中, 借助于领域无关词汇作为桥梁来弥合领域之间的差距。领域无关词汇类似于 (Blitzer, Dredze和Pereira, 2007年) 中的枢纽词, 并且可以类似地选择。SFA 首先构建一个双分图, 其中领域无关词汇作为一组节点, 领域特定词汇作为另一组节点。如果领域特定词汇共现, 则将其与领域无关词汇链接。共现可以定义为在同一文档中或在窗口内共同出现。

情感分析与观点挖掘

链接权重是它们共同出现的频率。然后，在二分图上应用谱聚类算法，将领域特定和领域无关的词语对齐到一组特征聚类中。这个想法是，如果两个领域特定的词语在图中与更常见的领域无关的词语有连接，它们往往以更高的概率被对齐或聚类在一起。同样地，如果两个领域无关的词语在图中与更常见的领域特定的词语有连接，它们往往以更高的概率被对齐在一起。对于最终的跨领域训练和测试，所有的数据示例都用这些聚类和原始特征的组合表示。

在同样的思路下，He、Lin和Alani（2011）使用联合主题建模来识别意见主题（类似于上述工作中的聚类）从两个领域中桥接它们。覆盖两个领域的结果主题被用作增加原始特征集的附加特征进行分类。在（Gao和Li，2011）中，主题建模也被用来基于领域术语对应和术语共现两个领域中找到一个共同的语义空间。然后，这个共同的语义空间被用来学习一个分类器，该分类器被应用于目标领域。Bollegala、Weir和Carroll（2011）提出了一种方法，利用来自多个源领域的标记和未标记数据自动创建一个情感敏感的同义词词典，以找到在不同领域中表达相似情感的词语之间的关联。然后，使用创建的同义词词典扩展原始特征向量来训练一个二元情感分类器。在（Yoshida等，2011）中，作者提出了一种从多个源领域到多个目标领域的转移方法，通过识别领域相关和领域无关的词语情感。在（Andreevskaia和Bergler，2008）中，提出了一种使用两个分类器的集成方法。第一个分类器是使用字典构建的，第二个分类器是使用少量领域内训练数据构建的。

在（Wu，Tan和Cheng，2009）中，提出了一种基于图的方法，该方法使用了在相似性图上的标签传播思想（Zhu和Ghahramani，2002）来进行转移。在图中，每个文档都是一个节点，两个节点之间的每个链接都是使用两个文档的余弦相似度计算的权重。最初，旧领域中的每个文档都有一个标签得分+1（正面）或-1（负面），并且根据正常情感分类器为新领域中的每个文档分配一个标签得分，该分类器可以从旧领域中学习。然后，算法通过在旧领域中找到 k 个最近邻居和在新领域中找到 k 个最近邻居来迭代更新每个新领域文档的标签得分 i 。使用邻居标签得分和链接权重的线性组合来为节点 i 分配一个新的得分。当标签得分收敛时，迭代过程停止。新领域文档的情感倾向由它们的标签得分确定。

Xia和Zong (2011) 发现, 在不同领域中, 某些词性 (POS) 标签的特征通常是与领域相关的, 而其他一些则是与领域无关的。基于这一观察, 他们提出了一种基于POS标签的集成模型, 以整合具有不同类型POS标签的特征, 以提高分类性能。

3.5 跨语言情感分类

跨语言情感分类意味着在多种语言中对观点文档进行情感分类。跨语言分类有两个主要动机。首先, 来自不同国家的研究人员希望在自己的语言中构建情感分析系统。然而, 大部分研究都是用英语进行的。在其他语言中, 很少有可以用来快速构建良好情感分类器的资源或工具。自然而然的问题是, 是否可以利用自动化机器翻译能力和现有的英语情感分析资源和工具来帮助在其他语言中构建情感分析系统。第二个动机是在许多应用中, 公司希望了解并比较不同国家关于其产品和服务的消费者意见。如果他们在英语中拥有情感分析系统, 他们希望通过翻译快速构建其他语言的情感分析系统。

一些研究人员已经研究了这个问题。目前的工作主要集中在文档级别的情感分类, 以及句子级别的主观性和情感分类。除了 (Guo et al., 2010) 中的研究外, 方面级别的工作还很有限。在本节中, 我们将重点关注跨语言文档级别的情感分类。下一章的第4.5节将重点关注句子级别。

在 (Wan, 2008) 中, 作者利用英文情感资源对中文评论进行分类。算法的第一步是使用多个翻译器将每个中文评论翻译成英文, 产生不同的英文版本。然后, 它使用基于词典的方法对每个翻译的英文版本进行分类。该词典包括一组积极词汇、一组消极词汇、一组否定词汇和一组增强词汇。然后, 算法根据否定词汇和增强词汇考虑评论中的词汇情感得分。如果最终得分小于0, 则评论为负面, 否则为正面。对于每个评论的最终分类, 它使用各种集成方法 (例如平均值、最大值、加权平均值、投票) 结合不同翻译版本的得分。

情感分析与观点挖掘

等等。如果还有中文词典，可以将相同的技术应用于中文版本。其结果也可以与那些英文翻译的结果相结合。结果表明集成技术是有效的。Brooke、Tofiloski和Taboada（2009）还尝试了从源语言（英语）翻译到目标语言（西班牙语），然后使用基于词典的方法或机器学习进行目标语言文档情感分类。

在（Wan，2009）中，提出了一种利用标注的英文语料库以监督方式对中文评论进行分类的协同训练方法。没有使用中文资源。在训练中，输入包括一组带标签的英文评论和一组无标签的中文评论。标记的英文评论被翻译成带标签的中文评论，无标签的中文评论被翻译成无标签的英文评论。因此，每个评论都与英文版本和中文版本相关联。每个评论的英文特征和中文特征被视为评论的两个独立且冗余的视图。然后，应用支持向量机的协同训练算法来学习两个分类器。

最后，这两个分类器被合并成一个单一的分类器。在分类阶段，每个未标记的中文评论首先被翻译成英文评论，然后应用学习的分类器将评论分类为积极或消极。

Wei和Pal（2010）提出了一种用于跨语言情感分类的迁移学习方法。由于机器翻译仍然远离完美，为了最小化翻译中引入的噪音，他们提出使用结构对应学习（SCL）方法（Blitzer，Dredze和Pereira，2007）在前一节中讨论的英语和中文两种语言中找到一小部分共享的核心特征。为了缓解数据和特征稀疏性问题，他们向搜索引擎发出查询，以找到与核心特征集中高度相关的其他特征，然后使用新发现的特征为训练创建额外的伪示例。

Boyd-Graber和Resnik（2010）将主题建模方法扩展为多语言评论的监督潜在狄利克雷分配（SLDA）（Blei和McAuliffe，2007），用于评论评级预测。SLDA能够在主题建模中考虑每个评论的用户评级。扩展模型MLSLDA同时使用来自多种语言的文档创建主题。由此产生的多语言主题在各种语言中是全局一致的。为了在主题建模中连接不同语言的主题词，该模型使用了不同语言的对齐WordNet或词典。

在 (Guo等, 2010) 中, 提出了一种基于主题模型的方法, 将不同语言中的一组给定方面表达式分组为方面簇 (类别), 用于对来自不同国家的观点进行基于方面的情感比较 (也参见第5.3.4节)。

在 (Duh, Fujino和Nagata, 2011) 中, 作者对跨语言情感分类的研究发表了自己的观点。根据他们的分析, 他们声称领域不匹配不是由机器翻译 (MT) 错误引起的, 即使是完美的MT也会导致准确性下降。它还认为, 跨语言适应问题在NLP中与其他 (单语) 适应问题有质的不同, 因此应考虑新的适应算法。

3.6 摘要

文档级别的情感分类提供了对实体、主题或事件的整体意见。它已经被许多研究者研究过。然而, 这种分类水平在应用中存在一些缺点:

- 在许多应用中, 用户需要了解更多细节, 例如消费者喜欢和不喜欢实体的哪些方面。在典型的观点文档中, 提供了这些细节, 但文档情感分类不会为用户提取它们。
- 文档情感分类不容易适用于非评论类的内容, 例如论坛讨论、博客和新闻文章, 因为许多此类帖子可以评估多个实体并进行比较。在许多情况下, 很难确定一个帖子是否实际评估了用户感兴趣的实体, 以及该帖子是否表达了任何观点, 更不用说确定对它们的情感了。

文档级别的情感分类不能执行这样的细粒度任务, 这些任务需要深入的自然语言处理。事实上, 在线评论不需要情感分类, 因为几乎所有评论都已经有用户分配的星级评分。实际上, 论坛讨论和博客需要情感分类来确定人们对不同实体 (例如产品和服务) 和主题的观点。

第四章

句子主观性和情感分类

正如前一章所讨论的，对于大多数应用来说，文档级情感分类可能过于粗糙。现在我们转向句子级别，即对每个句子中表达的情感进行分类。

然而，文档级别和句子级别的分类之间没有根本区别，因为句子只是短文档。

研究人员通常对句子级别分析做出的一个假设是，一个句子通常包含一个观点（尽管在许多情况下并不正确）。一个文档通常包含多个观点。让我们从一个例子评论开始：

“我两周前买了一部摩托罗拉手机。一开始一切都很好。声音很清晰，电池寿命很长，尽管有点笨重。然后，它昨天突然停止工作了。”

第一句话只是陈述事实，没有表达观点。其他所有句子都表达了明确或隐含的情感。请注意，没有观点通常被视为中立。

问题定义：给定一个句子 x ，确定 x 是否表达了积极、消极或中立（或无）观点。

五元组 (e, a, s, h, t) 的定义在这里不适用，因为句子级别的分类是一个中间步骤。在大多数应用中，人们需要知道观点的目标。仅仅知道一个句子表达了积极或消极的观点，而不知道观点所涉及的实体/方面，是有限的。然而，句子级别的分类仍然有用，因为在许多情况下，如果我们知道一个句子中谈论的实体和实体方面，这一步骤可以帮助确定对实体及其方面的观点是积极的还是消极的。

句子情感分类可以作为一个三类分类问题或两个独立分类问题来解决。在后一种情况下，第一个问题（也称为第一步）是分类一个句子是否表达了观点。第二个问题（也称为第二步）然后将那些观点句子分类为积极和消极类别。第一个问题通常被称为主观性分类，它确定一个句子是否表达了主观信息或事实（客观）信息（Hatzivassiloglou

和Wiebe, 2000; Riloff, Patwardhan和Wiebe, 2006; Riloff和Wiebe, 2003; Wiebe等, 2004; Wilson, Wiebe和Hwa, 2004; Wilson, Wiebe和Hwa, 2006; Yu和Hatzivassiloglou, 2003)。客观句被认为不表达情感或观点。正如我们之前讨论的那样, 这可能会引起问题, 因为客观句也可能暗示观点。

例如, “然后, 昨天它停止工作了”在上面的评论中是一个客观句, 但由于不良事实, 它暗示了对手机的负面情感。因此, 更合适的第一步是将每个句子分类为有观点或无观点, 而不管它是主观还是客观。然而, 由于常见的做法, 我们在本章中仍然使用术语主观性分类。在下面, 我们首先讨论句子级主观性分类的现有工作, 然后是情感分类。

4.1 主观性分类

主观性分类将句子分为两类, 主观和客观 (Wiebe, Bruce和O'Hara, 1999)。客观句子表达一些事实信息, 而主观句子通常表达个人观点和意见。事实上, 主观句子可以表达许多类型的信息, 例如意见、评价、情绪、信仰、推测、判断、指控、立场等 (Quirk等, 1985; Wiebe, Bruce和O'Hara, 1999)。其中一些表示积极或消极的情感, 而另一些则不表示。早期的研究将主观性分类作为一个独立的问题解决, 即不用于情感分类的目的。在最近的研究中, 一些研究人员将其视为情感分类的第一步, 通过使用它来删除被认为不表达或暗示任何观点的客观句子。

大多数现有的主观性分类方法都基于监督学习。例如, (Wiebe, Bruce和O'Hara, 1999)中报道的早期工作使用朴素贝叶斯分类器和一组二进制特征进行主观性分类, 例如句子中是否存在代词、形容词、基数词、除了will之外的情态动词和除了not之外的副词。随后的研究还使用了其他学习算法和更复杂的特征。

在 (Wiebe, 2000) 中, Wiebe提出了一种无监督的主观性分类方法, 该方法仅使用句子中的主观表达来确定句子的主观性。由于没有完整的这类表达式集合, 它提供了一些种子, 然后使用分布相似性 (Lin, 1998) 来找到相似的单词, 这些单词也是

很可能是主观指标。然而，通过这种方式找到的词具有较低的精确度和较高的召回率。然后，使用(Hatzivassiloglou and McKeown, 1997)中的方法和(Hatzivassiloglou and Wiebe, 2000)中的可分级性来过滤错误的主观表达。我们将在第6.2节中讨论(Hatzivassiloglou and McKeown, 1997)中的方法。可分级性是一种语义属性，使单词可以出现在比较结构中，并接受作为强调词或减弱词的修饰表达式。可分级形容词以不同的程度表达属性，相对于明确提及的规范或由修改的名词隐含提供的规范（例如，一个小行星通常比一个大房子大得多）。使用手动编制的副词和名词短语的种子列表（例如一点点，非常，有点和非常）找到了可分级形容词，这些副词和名词短语经常用作评分修饰语。这样的可分级形容词是主观性的良好指标。

在（Yu和Hatzivassiloglou，2003年）中，Yu和Hatzivassiloglou使用句子相似性和朴素贝叶斯分类器进行了主观性分类。句子相似性方法基于这样的假设，即主观或观点句子与其他观点句子相比更类似于事实句子。他们在（Hatzivassiloglou等人，2001年）中使用SIMFINDER系统来测量基于共享词、短语和WordNet同义词集的句子相似性。对于朴素贝叶斯分类，他们使用了诸如词（unigram）、二元组、三元组、词性、情感词的存在、情感词序列的极性计数（例如，“++”表示两个连续的正向词）、以及词性与情感信息的计数（例如，“JJ+”表示正向形容词）等特征，还使用了编码头动词、主语以及它们的直接修饰语的情感（如果有的话）的特征。这项工作还进行情感分类，以确定主观句子是积极的还是消极的，我们将在下一节讨论。

在应用监督学习中的一个瓶颈是手动标注大量训练样本所需的人工努力。为了节省手动标注的工作量，在(Riloff and Wiebe, 2003)中提出了一种自动标注训练数据的自举方法。该算法首先使用两个高精度分类器(HP-Subj和HP-Obj)自动识别一些主观和客观句子。高精度分类器使用词汇项列表(单词或 n -gram)，这些列表是很好的主观性线索。如果一个句子包含两个或更多强烈的主观线索，HP-Subj将把它分类为主观句子。如果没有强烈的主观线索，HP-Obj将把一个句子分类为客观句子。这些分类器将具有很高的精确度，但召回率较低。然后将提取的句子添加到

情感分析与观点挖掘

训练数据用于学习模式。这些模式（构成下一次迭代中的主观性分类器）然后用于自动识别更多的主观和客观句子，这些句子随后被添加到训练集中，算法的下一次迭代开始。

对于模式学习，提供了一组句法模板来限制要学习的模式类型。下面是一些示例句法模板和示例模式。

句法模板	示例模式
<主语> 被动动词	<主语> 被满意
<主语> 主动动词	<主语> 抱怨
主动动词 <宾语>	认可 <宾语>
名词 助动词 <宾语>	事实是 <宾语>
被动动词 介词 <名词短语>	对 <名词短语> 感到担心

Wiebe和Riloff（2005）使用这些发现的模式生成了一种基于规则的方法，用于生成主观性分类的训练数据。基于规则的主观分类器将句子分类为主观性，如果它包含两个或更多强烈的主观线索（否则，它不会标记该句子）。相反，基于规则的客观分类器寻找线索的缺失：如果句子中没有强烈的主观线索，并且满足其他几个条件，它将句子分类为客观性。该系统还使用信息提取系统AutoSlog-TS（Riloff，1996）学习关于客观句子的新模式，该系统基于一些固定的句法模板找到模式。基于规则的分类器生成的数据被用来训练朴素贝叶斯分类器。在（Wiebe等，2004）中还报道了一项相关研究，该研究对主观性分类使用了更全面的特征或主观线索集。

Riloff, Patwardhan和Wiebe（2006）研究了不同特征之间的关系。他们定义了单词、n-gram和词法-句法模式之间的包含关系。如果一个特征被另一个特征包含，那么被包含的特征就不再需要。这可以去除许多冗余的特征。

在（Pang和Lee，2004）中，提出了一种基于最小割的算法来将每个句子分类为主观或客观。该算法基于观点文档（例如评论）的句子图进行操作。首先，根据局部标记一致性（这会产生两个句子之间的关联分数）和传统分类方法产生的每个句子的主观性得分（这会产生每个句子的得分），构建图。局部标记一致性意味着彼此相邻的句子更有可能具有相同的类别标签（主观或客观）。最小割方法能够改善基于单个句子的主观性分类，因为它考虑了局部的情况。

情感分析与观点挖掘

标签一致性。这项工作的目的实际上是从评论中删除客观句子，以改善文档级情感分类。

Barbosa和Feng（2010）根据传统特征对推文（Twitter上的帖子）的主观性进行分类，并包括一些Twitter特定线索，如转发、标签、链接、大写字母、表情符号以及感叹号和问号。对于主观推文的情感分类，也使用了相同的特征集。

有趣的是，在（Raaijmakers和Kraaij, 2008）中发现，子词的字符n-gram而不是词n-gram也可以很好地进行情感和主观性分类。例如，对于句子“这辆车很棒”，子词字符二元组是th, hi, is, ca, ar, ro, oc, ck, ks。在（Raaijmakers, Truong和Wilson, 2008）和（Wilson和Raaijmakers, 2008）中，进行了词n-gram、字符n-gram和音素n-gram的实验和比较，用于主观性分类。学习算法使用了Booster（Schapire和Singer, 2000）。

令人惊讶的是，他们的实验表明字符n-gram表现最好，音素n-gram与词n-gram表现相似。

Wilson, Wiebe和Hwa（2004）指出一句话中可能包含主观和客观从句。准确定位这样的从句很有用。确定主观性的强度也很有用。他们提出了一项自动主观性分类的研究，通过个别从句中表达的主观性强度对句子的从句进行分类，最多可达四个层次（中性、低、中、高）。中性表示没有主观性。因此，强度分类包含了将句子分类为主观或客观的任务。作者采用了监督学习。他们的特征包括表示主观性的词语和短语，以及从依赖解析树生成的句法线索。

Benamara等人（2011）进行了四类主观性分类，其中S表示主观和评价性（情感可以是积极或消极），OO表示在客观句子或句子片段中暗示的积极或消极观点，O表示没有观点的客观句子，SN表示主观但非评价性（没有积极或消极情感）。这种分类更加完整，符合我们之前的讨论，也符合第2.4节中所示的情况，即主观句子可能没有评价性（积极或消极情感），而客观句子也可能暗示情感。

关于句子的主观性分类的附加工作也已经在阿拉伯语（Abdul-Mageed, Diab和Korayem, 2011）和乌尔都语（Mukund和Srihari, 2010）上进行了不同的机器学习基础上完成

算法使用通用和语言特定的特征。

4.2 句子情感分类

如果一个句子被分类为主观的，我们确定它是否表达了积极或消极的观点。监督学习可以像文档级情感分类一样应用，词典方法也可以。在讨论现有算法之前（一些算法不使用主观性分类步骤），让我们指出在该领域的大部分研究中存在的一种隐含假设。

句子级情感分类的假设：一个句子表达了一个观点持有者的单一情感。

这个假设适用于只有一个情感的简单句子，例如“这台相机的画质很棒。”然而，对于复合和复杂的句子，一个句子可能表达多个情感。例如，句子“这台相机的画质很棒，电池寿命也很好，但取景器对于这样一台好相机来说太小了，”同时表达了积极和消极的情感（或者说有混合情感）。对于“画质”和“电池寿命”，句子是积极的，但对于“取景器”来说是消极的。对于整个相机来说，它也是积极的（这是第2.1节中的总体方面）。

对于主观句子的情感分类，Yu和Hatzivassiloglou（2003）使用了与（Turney, 2002）中类似的方法，该方法已在第3.2节中讨论过。与（Turney, 2002）中使用一个正面种子词和一个负面种子词不同，这项工作使用了一大组种子形容词。此外，该工作不使用PMI，而是使用修改后的对数似然比来确定每个形容词、副词、名词和动词的正面或负面倾向。为了为每个句子分配一个倾向，它使用了其单词的平均对数似然得分。使用训练数据选择了两个阈值，并应用于确定句子的正面、负面或中性倾向。（Hatzivassiloglou和Wiebe, 2000）也在研究了同样的问题，考虑了可分级形容词。

在（Hu和Liu, 2004）中，Hu和Liu提出了一种基于词典的算法，用于方面级情感分类，但该方法也可以确定句子的情感倾向。它基于使用引导策略生成的情感词典，其中包括一些给定的正面和负面情感词种子以及WordNet中的同义词和反义词关系。我们将讨论生成情感词典的各种方法。

情感分析与观点挖掘

第6章中的情感词典。通过对句子中所有情感词的情感分数求和来确定句子的情感倾向。积极的词语被赋予+1的情感分数，消极的词语被赋予-1的情感分数。否定词和相反词（例如，但是、和、然而）也被考虑在内。在（Kim and Hovy, 2004）中，也使用了类似的方法。他们编制情感词典的方法也类似。然而，他们通过将句子中情感词的分数相乘来确定句子的情感倾向。同样，积极的词语被赋予+1的情感分数，消极的词语被赋予-1的情感分数。作者还尝试了另外两种聚合情感分数的方法，但它们效果较差。在（Kim and Hovy, 2007; Kim and Hovy, 2004; Kim et al., 2006）中，使用了监督学习来识别几种特定类型的观点。在（Nigam and Hurst, 2004）中，Nigam和Hurst应用了领域特定的词典和浅层自然语言处理方法来评估句子的情感倾向。

在（Gamon等人，2005年）中，使用了一种半监督学习算法来从一小部分标记句子和一大部分未标记句子中学习。该学习算法基于期望最大化（EM）算法，使用朴素贝叶斯作为基本分类器（Nigam等人，2000年）。这项工作进行了三类分类，即正面、负面和“其他”（没有观点或混合观点）。

在（McDonald等人，2007年）中，作者提出了一个类似于条件随机场（CRF）的分层序列学习模型，用于同时学习和推断句子级和文档级的情感。在训练数据中，每个句子都带有情感标签，每个完整的评论也带有情感标签。他们表明，同时学习两个级别可以提高两个级别的分类准确性。在（Täckström和McDonald，2011年）中，报道了一种仅从文档级别标注学习但可以进行句子和文档级情感分类的方法。因此，该方法是部分监督的。在（Täckström和McDonald，2011年）中，将完全监督模型和部分监督模型集成起来进行多级情感分类。

在（Hassan, Qazvinian和Radev，2010）中，提出了一种方法来识别在线讨论中参与者的态度。由于该论文只对讨论接收者感兴趣，算法只使用带有第二人称代词的句子段落。它的第一步是使用监督学习找到带有态度的句子。特征是使用马尔可夫模型生成的。它的第二步是确定态度的倾向（积极或消极），它使用了类似于（Ding, Liu和Yu，2008）的基于词典的方法。

不同的是，当句子中存在冲突的情感词时，它利用依赖树中的最短路径来确定倾向，而（Ding, Liu和Yu, 2008）使用词语距离（见第5.1节）。

在（Davidov, Tsur和Rappoport, 2010）中，研究了Twitter帖子（或推文）的情感分类。每个推文基本上都是一个句子。

作者采用了一种监督学习的方法。除了传统的特征之外，该方法还使用了标签、表情符号、标点符号和它们的频繁模式。这些特征被证明非常有效。

4.3 处理条件句

现有的句子级主观性分类或情感分类的研究大多集中在解决一般问题，而没有考虑到不同类型的句子可能需要非常不同的处理方法。Narayanan、Liu和Choudhary（2009）认为，很难找到一种适用于所有情况的解决方案，因为不同类型的句子以非常不同的方式表达情感。可能需要一种分而治之的方法，即对不同类型的句子进行重点研究。他们的论文重点研究了条件句，这些句子具有一些独特的特点，使得系统很难确定它们的情感倾向。

条件句是描述假设情况及其后果的句子。这样的句子通常包含两个从句：条件从句和结果从句，它们彼此依赖。它们之间的关系对于句子是否表达积极或消极情感有重要影响。一个简单的观察是，情感词（例如好、美丽、糟糕）本身无法区分观点句和非观点句，例如“如果有人制造出可靠的汽车，我会买它”和“如果你的诺基亚手机不好，买这款三星手机”。第一句并没有对任何特定汽车表达情感，尽管“可靠”是一个积极的情感词，但第二句对三星手机持积极态度，并没有对诺基亚手机表达观点（尽管诺基亚手机的所有者可能对它持消极态度）。因此，用于确定非条件句中的情感的方法对条件句不起作用。提出了一种监督学习方法来处理这个问题，使用了一组语言特征，例如情感词/短语及其位置、情感词的词性标记、时态模式、条件连接词等。

另一种难以处理的句子是疑问句。例如，

“有人能告诉我在哪里可以找到一部好的诺基亚手机吗？”明显没有对任何特定手机表达意见。然而，“有人能告诉我如何修理这部糟糕的诺基亚手机吗？”对诺基亚手机持有负面意见。

据我所知，目前还没有关于这个问题的研究。我认为，为了更准确的情感分析，我们需要针对不同类型的句子采取不同的处理方法。在这个方向上还需要进行更深入的研究。

4.4 处理讽刺句子

讽刺是一种复杂的言语行为形式，说话者或写作者说或写的是与他们实际意思相反的内容。讽刺已经在语言学、心理学和认知科学中进行了研究（Gibbs和Colston，2007；Gibbs，1986；Kreuz和Caucci，2007；Kreuz和Glucksberg，1989；Utsumi，2000）。在情感分析的背景下，这意味着当一个人说一些积极的话时，实际上是指负面的，反之亦然。

讽刺是非常难处理的。一些初步的工作已经在(González-Ibáñez, Muresan and Wacholder, 2011; Tsur, Davidov and Rappoport, 2010)中完成。根据我自己的经验，讽刺的句子在产品和服务的评论中并不常见，但在关于政治的在线讨论和评论中非常频繁。

(Tsur, Davidov and Rappoport, 2010)提出了一种半监督学习方法来识别讽刺。它使用了一小组标记的句子（种子），但没有使用未标记的示例。相反，它通过网络搜索自动扩展了种子集。作者认为，讽刺的句子经常与其他讽刺的句子一起出现在文本中。对于每个种子训练集中的句子，执行自动化的网络搜索作为查询。然后，系统为每个种子示例收集了最多50个搜索引擎片段，并将收集到的句子添加到训练集中。这个丰富的训练集然后用于学习和分类。对于学习，它使用了两种类型的特征，基于模式的特征和基于标点符号的特征。模式简单地是一系列高频词的有序序列。还设计了两个标准来删除过于一般化和过于具体化的模式。这些模式类似于数据挖掘中的序列模式（Liu, 2006和2011）。基于标点符号的特征包括句子中的“！”、“？”和引号的数量，以及大写/全部大写单词的数量。对于分类，采用了基于 k 近邻的方法。然而，这项工作没有进行情感分类。它只是将讽刺和非讽刺的句子分开。

González-Ibáñez、Muresan和Wacholder（2011）的研究在情感分析使用Twitter数据的背景下研究了这个问题，即

区分讽刺的推文和直接传达积极或消极观点的非讽刺推文（中性话语不考虑）。

同样，采用了监督学习方法，使用了SVM和逻辑回归。作为特征，他们使用了一元词和一些基于词典的信息。基于词典的特征包括（i）词类（Pennebaker等，2007）；（ii）WordNet情感（WNA）（Strapparava和Valitutti，2004）；以及（iii）一系列插入词（例如，啊，哦，是的）和标点符号（例如，！，？）。还使用了表情符号和ToUser（标记推文是否是回复另一个推文，由<@user>表示）等特征。针对三分类（讽刺、积极和消极）的实验结果表明，这个问题非常具有挑战性。最好的准确率只有57%。再次强调，这项工作没有将讽刺句子分类为积极和消极类别。

4.5 跨语言主观性和情感分类

就像在文档级别的跨语言情感分类中一样，研究人员还研究了句子级别的跨语言主观性分类和情感分类。同样，研究侧重于利用英语中的大量资源和工具以及自动翻译来帮助构建其他语言的情感分析系统，这些语言资源和工具较少。当前研究提出了三种主要策略：

- (1) 将目标语言的测试句子翻译成源语言，并使用源语言分类器对其进行分类。
- (2) 将源语言训练语料库翻译成目标语言，并在目标语言中构建基于语料库的分类器。
- (3) 将源语言的情感或主观性词典翻译成目标语言，并在目标语言中构建基于词典的分类器。

Kim和Hovy（2006）尝试了两种实验：（1）将德语电子邮件翻译成英语，并应用英语情感词来确定情感倾向，（2）将英语情感词翻译成德语，并使用德语情感词分析德语电子邮件。Mihalcea、Banea和Wiebe（2007）也尝试将英语主观性词和短语翻译成目标语言。事实上，他们实际上尝试了两种跨语言主观性分类的翻译策略。首先，他们通过翻译使用英语主观性词典来为新语言（在他们的案例中是罗马尼亚语）建立主观性词典。类似于（Riloff和Wiebe，2003）中的基于规则的主观性分类器

情感分析与观点挖掘

然后应用于将罗马尼亚句子分类为主观和客观类别。精确度还不错，但召回率很差。其次，他们使用手动翻译的平行语料库，生成了一种新语言的主观性标注语料库。他们首先使用一些现有工具自动将语料库中的英语句子分类为主观和客观类别，然后利用平行语料库中的句子级对齐将主观性类别标签投影到罗马尼亚句子上。然后，在罗马尼亚语中构建了一个基于监督学习的主观性分类器，用于分类罗马尼亚句子。在这种情况下，结果比第一种方法要好。

然而，需要注意的是，平行语料库的翻译是手动完成的。

在（Banea等人，2008）中，报告了三组实验。首先，将源语言（英语）的标记语料库自动翻译成目标语言（罗马尼亚语）。然后，将源语言中的主观性标签映射到目标语言的翻译版本中。

其次，源语言文本被自动标记为主观性然后翻译成目标语言。在这两种情况下，使用带有目标语言主观性标签的翻译版本来训练目标语言的主观性分类器。第三，目标语言被翻译成源语言，然后使用主观性分类工具对自动翻译的源语言文本进行分类。分类后，标签被映射回目标语言。然后使用生成的带标签语料库来训练目标语言的主观性分类器。最终的分类结果对于这三种策略来说非常相似。

在（Banea, Mihalcea和Wiebe, 2010）中，通过将标记的英语语料库翻译成其他5种语言，进行了广泛的跨语言句子级主观性分类实验。首先，实验表明使用翻译后的语料库进行训练在所有5种语言中都能取得相当好的效果。将不同语言的翻译版本与原始的英语版本结合起来形成一个训练语料库，也可以改善原始的英语主观性分类本身。其次，该论文证明通过将单语分类器的预测结果进行多数投票的方式，可以生成一个高精度的句子级主观性分类器。

（Bautin, Vijayarenu和Skiena, 2008）中的技术还将目标语言的文档翻译成英语，并使用基于英语词典的方法来确定包含实体的每个句子的情感倾向。这篇论文实际上是在方面级别上工作的。情感分类方法与（Hu和Liu, 2004）中的方法类似。

在（Kim, Li和Lee, 2010）中，引入了一个称为多语言可比性的概念

用于评估多语言主观性分析系统。通过多语言可比性，他们指的是具有相同主观含义的一对多语言文本的分类结果的一致性水平。他们使用平行语料库研究了源语言和目标语言的分类结果之间的一致性，使用了Cohen的Kappa系数。对于目标语言分类，他们尝试了几种现有的基于翻译的跨语言主观性分类方法。他们的结果表明，使用从英语翻译到目标语言的语料库训练的分类器在主观性分类和多语言可比性方面表现良好。

在（Lu等人，2011）中，尝试了一个稍微不同的问题。该论文假设源语言和目标语言都有一定数量的情感标记数据可用，并且还有一个无标记的平行语料库。他们的方法可以同时改进两种语言的情感分类。该方法是基于最大熵的EM算法，通过将无标记平行文本中的情感标签视为未观察到的潜在变量，并最大化语言特定标记数据的正则化联合似然以及平行文本的推断情感标签来联合学习两个单语情感分类器。在学习过程中，它利用了两个句子或文档之间平行（即彼此的翻译）的直觉应该表现出相同的情感。

4.6 利用话语信息进行情感分类

大多数现有的文档级和句子级情感分类工作都不使用句子之间或同一句子中的话语信息。在（Asher, Benamara和Mathieu, 2008; Somasundaran, Ruppenhofer和Wiebe, 2008）中研究了话语级别的情感注释。Asher, Benamara和Mathieu（2008）使用了五种修辞关系：对比，更正，支持，结果和延续，并附有情感信息进行注释。Somasundaran, Ruppenhofer和Wiebe（2008）提出了一个称为“观点框架”的概念。观点框架的组成部分是观点及其目标之间的关系。

在（Somasundaran et al., 2009）中，Somasundaran等人基于观点框架注释进行情感分类。所使用的分类算法是集体分类（Bilgic, Namata和Getoor,

2007年），它在图上进行分类。节点是需要分类的句子（或其他表达），链接是关系。在话语背景中，它们是与情感相关的话语关系。这些关系可以用来生成一组用于学习的关系特征。

每个节点本身也生成一组本地特征。关系特征允许一个节点的分类影响其他节点的分类，在集体分类方案中。在（Zhou等，2011年）中，单个复合句中的话语信息被用来对句子进行情感分类。例如，这个句子

“尽管福吉莫里受到国际社会的批评，但他受到国内人民的喜爱，因为人们讨厌腐败的统治阶级”，尽管这句话中有更多的负面观点词（参见第4.7节），但它是一个积极的句子。本文使用模式挖掘来寻找分类的话语模式。

在（Zirn等人，2011年）中，作者提出了一种分类话语段的方法。每个段落都表达了一个单一的（积极或消极）观点。

马尔可夫逻辑网络被用于分类，它不仅可以利用情感词典，还可以利用局部/邻近的话语背景。

4.7 摘要

句子级主观性分类和情感分类比文档级情感分类更进一步，因为它更接近观点目标和目标上的情感。它可以被视为整体情感分析任务中的一个中间步骤。然而，对于许多实际应用，它仍然存在一些缺点：

- 在大多数应用中，用户需要知道额外的细节，即喜欢和不喜欢的实体或实体的方面。就像文档级别一样，句子级别的分析仍然没有做到这一点。
- 虽然有人可能会说，如果我们知道观点目标（例如实体和方面，或主题），我们可以将句子的情感倾向分配给句子中的目标。然而，这是不够的：（1）许多复杂的句子对不同的目标有不同的情感，例如“尝试Chrome，因为Firefox经常崩溃”。

“在这个糟糕的经济环境中，苹果表现得非常好。”在最后一句中，即使在从句级别的分类也是不够的。我们需要去观点目标或者方面级别。

- （2）尽管一个句子可能有整体上的积极或消极的语气，但它的一些组成部分可能表达相反的观点。例如，一些研究人员认为下面的句子是积极的

情感分析与观点挖掘

(Neviarouskaya, Prendinger和Ishizuka, 2010; Zhou等, 2011):

“尽管失业率高，经济状况良好。”这个句子的整体语气确实是积极的，或者作者试图强调积极的一面，但它确实包含了对失业率的消极情绪，我们不能忽视。如果我们进行方面级别的情感分析，问题就解决了。也就是说，这个句子对整体经济是积极的，但对失业率是消极的。

- (3) 句子级情感分类无法处理比较句中的观点，例如，“可口可乐比百事可乐更好喝。”在这种情况下，我们需要不同的方法来提取和分析比较观点，因为它们与常规观点有着非常不同的含义。

尽管这个句子明确表达了一个观点，但我们不能简单地将其分类为积极、消极或中性。我们将在下一章讨论方面级情感分

析，并在第8章讨论比较观点分析。

第5章

基于方面的情感分析

按照章节的自然进展，本章应该侧重于短语和词级情感分类，因为前两章是关于文档和句子级分类的。然而，我们将把这个话题留到下一章。在本章中，我们将重点关注基于方面的情感分析，因为现在是处理第2章中定义的完整问题的时候了，许多短语和词的情感取决于方面的上下文。

正如我们在前两章中讨论的那样，仅仅在文档级别或句子级别对观点文本进行分类通常是不够的，因为它们无法识别观点目标或为这些目标分配情感。即使我们假设每个文档评估一个实体，一个关于该实体的积极观点文档并不意味着作者对该实体的所有方面都持有积极观点。同样，一个消极观点文档并不意味着作者对所有事物都持有消极观点。为了进行更完整的分析，我们需要发现各个方面并确定每个方面的情感是积极还是消极。

为了提取这些细节，我们需要进入方面级别，这意味着我们需要第2章的完整模型，即基于方面的情感分析（或观点挖掘），在（Hu和Liu，2004）中也被称为基于特征的观点挖掘。请注意，如第2章所讨论的，观点目标被分解为实体及其方面。在结果中，方面GENERAL用于表示实体本身。因此，基于方面的情感分析涵盖了实体和方面。它还引入了一系列需要更深入的自然语言处理能力并产生更丰富结果的问题。

回想一下，在方面级别上，目标是在给定的文档 d 中发现每个五元组 $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ 。为了实现这个目标，需要执行六个任务。本章主要关注以下两个核心任务。这些任务已经得到研究人员的广泛研究。其他任务也会被涉及，但相对较简略。

1. 方面提取：该任务提取已经评估过的方面。

例如，在句子“这部手机的语音质量很棒”中，方面是“语音质量”，表示为“这部手机”的实体。请注意，“这部手机”在这里并不表示一般的方面，因为评估的只是它的语音质量。然而，句子“我喜欢这部手机”

情感分析与观点挖掘

评估的是整个手机，即表示为“这部手机”的实体的一般方面。请记住，每当我们谈论一个方面时，我们必须知道它属于哪个实体。在下面的讨论中，为了简单起见，我们经常省略实体。

2. 方面情感分类：该任务确定不同方面的观点是积极的、消极的还是中立的。在上面的第一个例子中，“语音质量”方面的观点是积极的。在第二个例子中，对一般方面的观点也是积极的。

请注意，在某些应用中，可能会给出意见目标，因为用户只对这些特定目标感兴趣（例如，宝马和福特品牌）。在这种情况下，我们不需要执行实体或方面提取，只需要确定目标上的情感。

5.1 方面情感分类

我们首先研究第二个任务，即确定句子中每个方面表达的情感倾向。有两种主要方法，即监督学习方法和基于词典的方法。

对于监督学习方法，第4章讨论的用于句子级和从句级情感分类的基于学习的方法是适用的。在（Wei和Gulla，2010）中，还提出了一种分层分类模型。然而，关键问题是如何确定每个情感表达的范围，即它是否涵盖了句子中感兴趣的方面。目前的主要方法是使用解析来确定依赖关系和其他相关信息。例如，在（Jiang等，2011）中，使用依赖解析器生成一组与分类相关的方面依赖特征。在（Boiy和Moens，2009）中也使用了类似的方法，根据特征在解析树中相对于目标方面的位置进行加权。对于比较句子，可以使用“than”或其他相关词来分割句子（Ding，Liu和Zhang，2009；Ganapathibhotla和Liu，2008）。

监督学习依赖于训练数据。正如我们在第3.4节中讨论的那样，从一个领域的标记数据中训练的模型或分类器在另一个领域中通常表现不佳。尽管领域自适应（或迁移学习）已经被研究人员研究过（第3.4节），但该技术仍然远未成熟，目前的方法主要用于文档级情感分类，因为文档较长且包含更多用于分类的特征，而不是单个句子或从句。

因此，监督学习在大量应用领域上存在困难。

情感分析与观点挖掘

基于词典的方法可以避免一些问题 (Ding, Liu and Yu, 2008; Hu and Liu, 2004), 并且在许多领域中表现出色。这些方法通常是无监督的。它们使用情感词典 (其中包含一系列情感词、短语和习语)、复合表达式、观点规则 (第5.2节) 和 (可能的话) 句子解析树来确定句子中每个方面的情感倾向。它们还考虑情感转移、但从句 (见下文) 和其他可能影响情感的结构。当然, 基于词典的方法也有其缺点, 我们将在后面讨论。将这种方法扩展到处理比较句将在第8.2节中讨论。下面, 我们介绍一种简单的基于词典的方法, 以了解这种方法的特点。该方法来自于 (Ding, Liu and Yu, 2008), 它有四个步骤。在这里, 我们假设实体和方面已知。它们的提取将在第5.3节中讨论。

1. 标记情感词和短语: 对于包含一个或多个方面的每个句子, 此步骤标记句子中的所有情感词和短语。每个积极词都被赋予+1的情感分数, 每个消极词都被赋予 -1的情感分数。例如, 我们有句子“这部手机的声音质量不好, 但电池寿命很长。”在此步骤之后, 句子变为“这部手机的声音质量不好[+1], 但电池寿命很长”, 因为“好”是一个积极的情感词 (句子中的方面已用斜体表示)。请注意, “长”在这里不是一个情感词, 因为它本身通常不表示积极或消极的情感, 但我们可以在此上下文中推断出它的情感。事实上, “长”可以被视为一个依赖上下文的情感词, 我们将在第6章中讨论。在下一节中, 我们将看到一些其他可以给出或暗示积极或消极情感的表达方式。

2. 应用情感转换词: 情感转换词 (也称为情感转移词) 是可以改变情感倾向的词语和短语。有几种类型的这种转换词。

否定词如不、从不、没有、没有人、没有地方、既不和不能是最常见的类型。这一步将我们的句子转换为“这部手机的声音质量不好[-1], 但电池寿命很长”, 因为有否定词“不”。我们将在下一节讨论几种其他类型的情感转换词。请注意, 并非每次出现情感转换词都会改变情感倾向, 例如“不仅……而且”。这些情况需要小心处理。也就是说, 这些特殊用法和模式需要事先识别出来。

3. 处理but从句: 表示相反意义的词语或短语需要特殊处理, 因为它们经常改变情感倾向。

情感分析与观点挖掘

英语中最常用的相反词是“but”。包含相反词或短语的句子通过以下规则处理：如果无法确定一侧的观点，则相反词（例如but）之前和之后的情感倾向是相反的。规则中的if条件用于判断相反词和短语是否总是表示观点变化，例如“Car-x很棒，但是Car-y更好”。经过这一步骤，上述句子变为“这部手机的语音质量不好[-1]，但是电池寿命很长[+1]”（在but从句末尾添加[+1]）。请注意，我们可以推断出“长”对于“电池寿命”是积极的。除了but之外，诸如“除了”，“除了”，“除了”等短语也具有相反的意义，并以相同的方式处理。与否定一样，并非每个but都表示相反，例如“不仅……而且”。这种包含“but”的非but短语也需要事先识别。

4. 聚合意见：此步骤将一个意见聚合函数应用于得到的情感分数，以确定句子中每个方面的最终倾向。假设句子为 s ，其中包含一组方面 $\{a_1, \dots, a_m\}$ 和一组情感词或短语 $\{sw_1, \dots, sw_n\}$ ，它们的情感分数是从步骤1-3得到的。每个方面 a_i 在 s 中的情感倾向由以下聚合函数确定：

$$score(a_i, s) = \sum_{sw_j \in s} \frac{sw_j.so}{dist(sw_j, a_i)}, \quad (5)$$

其中 sw_j 是 s 中的情感词/短语， $dist(sw_j, a_i)$ 是方面 a_i 和情感词 sw_j 在 s 中的距离。 $sw_j.so$ 是 sw_j 的情感分数。使用乘法逆元来给与离方面 a_i 较远的情感词较低的权重。如果最终分数为正，则对 s 中的方面 a_i 持有正面意见。如果最终分数为负，则对该方面的情感持有负面意见。否则，情感是中性的。

这个简单的算法在许多情况下表现得非常好。它能够处理“苹果在这个糟糕的经济环境中表现得非常好”的句子，没有任何问题。请注意，还有许多其他的观点聚合方法。例如，(Hu and Liu, 2004) 简单地将句子或句子片段中所有情感词的情感得分相加。Kim, and Hovy (2004) 使用了词语情感得分的乘法。其他研究人员也采用了类似的方法 (Wan, 2008; Zhu et al., 2009)。

为了使这种方法更加有效，我们可以确定每个情感词的范围，而不是使用上述的词语距离。在

这种情况下，需要进行解析以找到依赖关系，就像上面讨论的有监督方法一样。我们还可以自动发现上下文相关词语的情感倾向，比如上面的“长”。更多细节将在第6章中给出。实际上，上述简单方法可以在许多方向上得到改进。例如，Blair-Goldensohn等人(2008)将基于词典的方法与有监督学习相结合。Kessler和Nicolov(2009)尝试了四种不同的策略来确定每个方面/目标的情感（包括一种排名方法）。他们还根据大量手动注释的数据展示了一些有趣的统计数据，说明为什么很难将情感词与它们的目标联系起来。

除了方面情感分类研究之外，研究人员还研究了方面情感评级预测问题，这主要是在主题建模的背景下与方面提取一起完成的，我们在第5.3.4节中讨论。

如上所述，除了情感词和短语之外，还有许多其他类型的表达方式可以传达或暗示情感。它们中的大多数也更难处理。下面，我们列举了一些被称为观点的基本规则(Liu, 2010)。

5.2 观点的基本规则和组合语义

观点规则表达了一个蕴含着积极或消极情感的概念。它可以是简单的带有暗示情感的个别情感词，也可以是需要常识或领域知识来确定其方向的复合表达式。本节描述了其中一些规则。表示这些规则的一种方式是使用组合语义的思想(Dowty, Wall and Peters, 1981; Montague, 1974)，它指出复合表达式的含义是其构成成分的含义和它们组合的句法规则的函数。下面，我们首先在概念层面上描述规则，而不考虑它们在实际句子中的表达方式，因为其中许多规则可以用多种方式表达，也可能与领域和上下文有关。之后，我们转向表达层面，讨论了在情感分析的背景下关于组合语义的当前研究，该研究旨在将多个输入成分表达式结合起来，为复合表达式推导出整体情感倾向。

规则使用类似BNF形式的形式主义来呈现。规则来自（Liu, 2010）。

情感分析与观点挖掘

- | | |
|-------|----------|
| 1. 积极 | ::= P |
| 2. | PO |
| 3. | 情感转移词 N |
| 4 | 情感转移词 NE |
| 5. 消极 | ::= N |
| 6. | NE |
| 7. | 情感转移词 P |
| 8. | 情感转移词 PO |

非终结符P和PO表示两种类型的积极情感表达。P表示一个原子积极表达，一个词或一个短语，而PO表示由多个表达式组成的积极表达。

同样，非终结符N和NE也表示两种类型的消极情感表达。“情感转移词 N”和“情感转移词 NE”分别表示N和NE的否定，而“情感转移词 P”和“情感转移词 PO”表示P和PO的否定。

我们需要注意，这些并不是以实际的BNF形式表达的，而是以一种伪语言表达一些抽象概念。很难准确地指定它们，因为在实际的句子中，情感转移器可能以许多不同的形式出现在N、NE、P或PO之前或之后，并且在情感转移器和积极（或消极）情感表达之间可能有词语。POSITIVE和NEGATIVE是用来确定句子中目标/方面的观点的最终情感。

情感转移器（或情感转移器（Polanyi和Zaenen, 2004））：否定词如not、never、none、nobody、nowhere、neither和cannot是最常见的情感转移器类型。情态助动词（例如would、should、could、might、must和ought）是另一种类型，例如“刹车可以改进”，它可能改变情感的方向，但并不总是如此。一些预设项目是另一种类型。这种情况对于像barely和hardly这样的副词是典型的，比较“它工作”和“它几乎不工作”。“工作”表示积极，但“几乎不工作”则不是：它预设了更好的预期。

像失败、忽略、忽视这样的词表现得相似，例如，“这个相机让我失望。”此外，讽刺常常也改变方向，例如，“多么好的车，它在第一天就无法启动。”虽然手动识别这种转变可能不难，但是通过自动化系统在实际句子中正确地识别和处理它们是具有挑战性的（参见第4.4节）。此外，规则11-14也可以看作是情感转变器。我们将它们单独呈现，因为它们还涵盖了比较观点。请注意，一些研究人员还研究了否定的应用范围（Ikeda等，2008; Jia, Yu和Meng, 2009; Li等，2010; Morante, Schrauwen和Daelemans, 2011）。我们

情感分析与观点挖掘

在讨论情感组合时，将更多地讨论情感转变器。

我们现在定义N，NE，P和PO，它们不包含情感转变词。我们根据它们的特定特征将这些表达式分为六个概念类别。

1. 情感词或短语：这是最简单也是最常用的类别，其中仅凭情感词或短语就可以暗示对方面的积极或消极意见，例如，“好”在“声音质量很好”中。这些词或短语被简化为P和N。9. P

::= 一个积极的情感词或短语

10. N ::= 一个消极的情感词或短语

同样，右侧的细节没有指定（这也适用于所有后续规则）。目前的大部分研究只使用这个类别中的词和短语。

2. 减少和增加观点项的数量（N和P）：这组规则与否定（或情感转变）规则3、4、7和8类似。它们表达了减少或增加与观点项（通常是名词和名词短语）相关联的数量可以改变情感的方向。例如，在句子中

“这种药显著减轻了我的疼痛，”“疼痛”是一个负面情绪词，而“疼痛”的减轻表示该药物具有良好的效果。

因此，减轻疼痛意味着对该药物持有积极的观点。减少的概念也适用于去除和消失，例如，“我服用药物后我的疼痛消失了。”然后我们有以下规则：

11. PO ::= 减少或降低 N

12. | 增加或提高 P

13. NE ::= 减少或降低 P

14. | 增加或提高 N

请注意，规则12和14不改变情感的方向，但它们可以改变观点的强度。在句子中，表示减少或增加的实际词语或短语可能出现在N或P之前或之后，例如，“我服用药物后我的疼痛减轻了，”和“这种药物减轻了我的疼痛。”3. 积极或消极潜在项目的高、低、增加和减少数量：对于某些项目，它们的小值/数量是负面的，而它们的大值/数量是积极的，例如，“电池寿命很短”和“电池寿命很长。”我们称这样的项目为积极潜在项目(PPI)。这里的“电池寿命”是一个积极的潜在项目。对于其他一些方面，它们的小值/数量是积极的，而它们的大值/数量是负面的，例如，“这部手机很贵”和

情感分析与观点挖掘

“索尼降低了相机的价格。”这些项目被称为负面潜在项目（NPI）。 “成本”和“价格”是负面潜在项目。

正面和负面潜在项目本身并不意味着观点，即“电池寿命”和“成本”，但当它们被数量形容词或数量变化词或短语修饰时，可能会暗示积极或消极的情感。以下规则涵盖了这些情况：15. PO ::= 没有低于或减少数量的NPI

- 16. | 大量或增加数量的PPI
- 17. NE ::= 没有低于或减少数量的PPI
- 18. | 大量或增加数量的NPI
- 19. NPI ::= 一个负面潜在项目
- 20. PPI ::= 一个正面潜在项目

在（Wen和Wu，2011）中，提出了一种引导和分类方法来发现中文中的PPI和NPI。

4. 可取或不可取的事实：以上规则都包含了一些主观表达。但客观表达也可以暗示积极或消极的情感，因为它们可以描述可取和不可取的事实。这样的句子通常不使用任何情感词。例如，“我和妻子在床垫上睡了两个星期后，我看到中间有座山”这句话明显暗示了对床垫的负面意见。然而，“山”这个词本身并不带有任何观点。因此，我们有以下两个规则：

- 21. P ::= 可取事实
- 22. N ::= 不可取事实

5. 偏离规范或期望值范围：在某些应用领域中，物品的值有一个期望范围或规范。如果值偏离了正常范围，就是负面的，例如，“我服用药物后，我的血压升到了410。”这样的句子通常也是客观句子。因此，我们有以下规则：

- 23. P ::= 在所需值范围内
- 24. N ::= 偏离所需值范围

6. 生产和消耗资源和废物：如果一个实体生产大量的资源，那是可取的（或积极的）。如果它消耗大量的资源，那是不可取的（或消极的）。例如，电力是一种资源。句子“这台电脑消耗了很多电力”对电脑的功耗给出了负面意见。同样，如果一个实体产生大量的废物，那是消极的。如果它消耗大量的废物，那是积极的。这些给出了以下规则：

情感分析与观点挖掘

- 25. P ::= 生产大量或更多的资源
- 26. | 不产生、少量或更少的废物
- 27. | 不消耗、少量或更少的资源
- 28. | 消耗大量或更多的废物
- 29. N ::= 不产生、少量或更少的资源
- 30. | 产生一些或更多的废物
- 31. | 消耗大量或更多的资源
- 32. | 不消耗、少量或更少的废物

这些概念规则可以以许多不同的形式出现在实际句子中，使用不同的词语和短语，在不同的领域中它们也可能以不同的方式表现。因此，它们很难被识别。如果不能识别它们，规则就无法应用。

这组概念规则绝不是统治意见或情感的完整集合。事实上，还有其他规则，并且通过进一步的研究，可能会发现更多规则。还需要注意的是，像个别情感词一样，句子中出现任何规则并不总是意味着意见。例如，“我想要一辆可靠性高的汽车”并不表达对任何特定汽车的积极或消极意见，尽管“高可靠性”满足规则16。处理这类句子可能需要更复杂的规则或话语层面分析。

我们现在讨论应用组合性原理来表达上述规则的现有工作。最常研究的组合规则是与情感反转相关的规则，这些规则是情感转移词和积极或消极情感词的组合，例如，“不”和POS(“好”) => NEG(“不好”)。我们在上面详细讨论了它们。另一种主要类型由规则11到14表示，例如，“减少”和NEG(“痛苦”) => POS(“减少痛苦”)。

这些组合规则可以表达一些观点规则，也可以表达某些其他表达层面的情感组合。除了上述两种组合类型，Moilanen和Pulman (2007) 还引入了情感冲突，当多个情感词同时出现时使用，例如，“非常好”。通过对它们分配的相对权重进行排名，可以实现冲突解决，这些权重决定了哪个成分在情感方面更重要。

在(Neviarouskaya, Prendinger和Ishizuka, 2010)中，引入了六种组合规则，即情感反转、聚合、传播、支配、中和和强化。情感反转与我们上面讨论的相同。聚合类似于上面的情感冲突，但定义不同。如果形容词-名词、名词-名词、副词-形容词、副词-动词短语的情感方向相反，则将具有主导极性的短语分配给混合极性，例如POS(‘美丽’)和NEG(‘战斗’) => POSneg(‘美丽的战斗’)。

当短语/从句中使用了“传播”或“转移”类型的动词，并且需要确定具有先前中性极性的参数的情感时，应用传播规则，例如PROP-POS（“钦佩”）和“这种行为”=> POS（“这种行为”）；“先生。”支配规则适用于具有支配性极性的短语/从句，例如POS（“喜欢”）和“这个电影”=> POS（“这个电影”）；“他们的表演。”

X”和TRANS(“支持”)和NEG(“犯罪业务”) => NEG(‘X先生’). 统治的规则是：（1）如果一个动词和一个从句中的宾语的极性方向相反，则动词的极性占优势（例如，NEG(“欺骗”) & POS(“希望”) => NEG(“欺骗希望”)); （2）如果一个复合句使用并列连词“但是”连接从句，则连接词后面的从句的态度特征占主导地位（例如，‘NEG(“整夜爬山很辛苦”), 但是 POS(“早上旅行者得到了壮丽的景色”).’ => 整个句子为POS）。中性化规则适用于介词修饰语或条件运算符与情感陈述相关的情况，例如，“尽管”和NEG(“担忧”) => NEUT(“尽管担忧”). 强化或削弱情感分数（强度）的规则，例如，Pos_score(“高兴”) < Pos_score(“非常高兴”). 可以在（Choi和Cardie, 2008; Ganapathibhotla和Liu, 2008; Min和Park, 2011; Nakagawa, Inui和Kurohashi, 2010; Nasukawa和Yi, 2003; Neviarouskaya, Prendinger和Ishizuka, 2009; Polanyi和Zaenen, 2004; Socher等, 2011; Yessenalina和Cardie, 2011）中找到其他相关作品。

正如我们所看到的，一些观点规则没有通过组合来表达，例如，那些涉及资源使用的规则（规则25-32）。

然而，可以通过三元组在一定程度上表达它们（Zhang和Liu, 2011a）。期望和不希望的事实或值范围也没有被包括在内（规则21-24）。实际上，它们与组合并没有直接关联，因为它们本质上是上下文或领域隐含的情感术语，需要在领域语料库中发现（Zhang和Liu, 2011b）。

5.3 方面提取

现在我们转向方面提取，它也可以被视为一项信息提取任务。然而，在情感分析的背景下，问题的一些特定特征可以促进提取。关键特征是观点总是有一个目标。目标通常是从句子中提取的方面或主题。因此，识别每个观点表达及其目标从一个句子中是重要的。然而，

情感分析与观点挖掘

我们还应该注意到，一些观点表达可以扮演两个角色，即表示积极或消极情感并暗示一个（隐含的）方面（目标）。例如，在“这辆车很贵”中，“贵”是一个情感词，同时也表示了方面价格。我们将在第5.3.5节讨论隐含的方面。在这里，我们将重点讨论显式方面提取。有四种主要方法：

1. 基于频繁名词和名词短语的提取
2. 利用观点和目标关系的提取
3. 使用监督学习的提取
4. 使用主题建模的提取

由于现有的关于方面提取（更准确地说，方面表达提取）的研究主要是在在线评论中进行的，我们也使用评论上下文来描述这些技术，但并没有阻止它们被用于其他形式的社交媒体文本。

在网络上有两种常见的评论格式。

格式1 – 优点、缺点和详细评论：评论者首先分别描述一些简明的优点和缺点，然后写一篇详细/完整的评论。这种评论的示例见图5.1。

格式2 – 自由格式：评论者自由书写，即没有简明的优点和缺点。

从格式1的评论中提取优点和缺点（不包括详细评论，与格式2相同）

是从完整评论中提取方面的特殊情况，也相对容易。在（Liu,

Hu和Cheng, 2005）中，提出了一种基于顺序学习方法的特定方法从优点和缺点中提取方面，该方法还利用了优点和缺点的一个关键特征，即它们通常非常简短，由短语或句子片段组成。每个片段通常只包含一个方面。句子片段可以用逗号、句号、分号、连字符、&、和等分隔。这个观察有助于提取算法的更准确执行。

由于相同的基本技术可以应用于正面和负面评论以及完整的文本，从现在开始我们将不再区分它们，而是专注于不同的方法。

5.3.1 查找频繁名词和名词短语

该方法从给定领域的大量评论中找到明确的方面表达，这些表达是名词和名词短语。胡和刘（2004）使用了一种数据挖掘算法。名词和名词短语（或组）

情感分析与观点挖掘

我的单反相机放在架子上

由camerafun4拍摄。 Aug 09 '04

优点：拍照效果好，易于使用，非常小巧

缺点：电池使用时间短；附带的内存有限。

在购买这台佳能A70之前，我从未使用过数码相机。

我一直使用单反相机...[阅读完整评论](#)

图5.1.一个格式1评论的例子。

很棒的相机。 ， 2004年6月3日

评论者：jprice174来自亚特兰大，佐治亚州。

去年在购买这台相机之前，我做了很多研究... 离开我心爱的尼康35mm单反相机有点痛苦，但我要去意大利，需要一台更小、更轻便的数码相机。

这台相机拍出来的照片太棒了。“自动”功能大部分时间都能拍出很棒的照片。而且使用数码相机，如果照片拍得不好也不会浪费胶卷...图5.2.一个格式2评论的例子。

通过词性标注器进行了识别。它们的出现频率被计数，并且只保留频繁出现的。可以通过实验来决定一个频率阈值。这种方法有效的原因是当人们评论实体的不同方面时，他们使用的词汇通常会收敛。因此，那些经常被谈论的名词通常是真实而重要的方面。评论中的无关内容通常是多样的，即在不同的评论中它们是非常不同的。因此，那些不常见的名词很可能是非方面或不太重要的方面。尽管这种方法非常简单，但实际上非常有效。一些商业公司正在使用这种方法并进行了一些改进。

该算法的精度在（Popescu和Etzioni，2005年）中得到了改进。他们的算法试图去除那些可能不是实体方面的名词短语。它通过计算短语与某些与实体类别相关的部分关系判别器之间的点互信息（PMI）得分来评估每个发现的名词短语，例如相机类。相机类的部分关系判别器包括“相机的”，“相机具有”，“相机配有”等，这些用于通过在Web上搜索来查找相机的组件或部件。PMI度量是3.2节中的简化版本：

$$PMI(a, d) = \frac{\text{击中}(a \wedge d)}{\text{击中}(a) \text{击中}(d)}, \quad (4)$$

其中 a 是使用频率方法识别出的候选方面， d 是一个判别器。使用Web搜索找到了击中次数

情感分析与观点挖掘

个别术语以及它们的共现。这种方法的思路很清晰。如果候选观点的PMI值太低，它可能不是产品的组成部分，因为 a 和 d 不经常共现。该算法还使用WordNet的is-a层次结构（列举了不同种类的属性）和形态学线索（例如，“-iness”，“-ity”后缀）来区分组件/部件和属性。

Blair-Goldensohn等人（2008年）通过考虑主要出现在带有情感的句子或一些表示情感的句法模式中的名词短语，对频繁名词和名词短语的方法进行了改进。

应用了几个过滤器来删除不太可能的方面，例如删除没有足够提及的方面以及已知情感词。他们还将方面折叠到词干级别，并通过手动调整的加权和对它们在带有情感的句子中的频率以及情感短语/模式类型的考虑，对发现的方面进行了排序，其中出现在短语中的权重更大。使用情感句子与第5.3.2节中的方法相关。

（Ku, Liang和Chen, 2006）也采用了基于频率的方法。

作者称这些发现的术语为主要主题。他们的方法还利用了文档级别和段落级别考虑术语的TF-IDF方案。Moghaddam和Ester（2010年）通过额外的基于模式的过滤器来增强基于频率的方法，以删除一些非方面术语。他们的工作还预测了方面的评级。

Scaffidi等人（2007年）将评论语料库中提取的频繁名词和名词短语的频率与通用英语语料库中的出现率进行比较，以识别真实的方面。

Zhu等人（2009年）提出了一种基于Cvalue测量方法（Frantzi, Ananiadou和Mima, 2000年）的多词方面提取方法。Cvalue方法也基于频率，但它考虑了多词术语的频率、 t 的长度以及包含 t 的其他术语。然而，Cvalue只能帮助找到一组候选项，然后使用一组给定的种子方面的自举技术对其进行改进。改进的思想是基于每个候选项与种子的共现。

Long, Zhang和Zhu（2010年）基于频率和信息距离提取方面（名词）。他们的方法首先使用基于频率的方法找到核心方面词。然后，它使用（Cilibrasi和Vitanyi, 2007年）中的信息距离找到与方面相关的其他词，例如，对于方面价格，它可能找到“\$”和“美元”。然后，所有这些词都用于选择最多讨论特定方面的评论。

5.3.2 使用观点和目标关系

由于观点有目标，它们显然是相关的。它们的关系可以被利用来提取观点目标，因为情感词通常是已知的。这种方法在（Hu和Liu，2004年）中用于提取不常见的方面。思路如下：相同的情感词可以用来描述或修改不同的方面。如果一个句子没有常见的方面，但有一些情感词，那么将提取每个情感词最近的名词或名词短语。由于在（Hu和Liu，2004年）中没有使用解析器，因此“最近”函数近似表示情感词和名词或名词短语之间的依赖关系，这通常效果很好。例如，在以下句子中，“这个软件太棒了。”

如果我们知道“棒”是一个情感词，那么“软件”将被提取为一个方面。即使单独应用，这个想法在实践中也非常有用。（Blair-Goldensohn等人，2008年）中的情感模式方法使用了类似的思路。此外，这种基于关系的方法也是发现观点文档中重要或关键方面（或主题）的有用方法，因为如果没有人对某个方面或主题表达任何观点或情感，那么它很可能不重要。

在（Zhuang, Jing和Zhu，2006）中，使用了依赖解析器来识别此类依赖关系以进行方面提取。Somasundaran和Wiebe（2009）采用了类似的方法，Kobayashi等人也是如此（Kobayashi等人，2006）。这个依赖思想进一步推广为双向传播方法，用于同时提取情感词和方面（Qiu等人，2011）（将在第5.5节讨论）。在（Wu等人，2009）中，使用了短语依赖解析器而不是普通的依赖解析器来提取名词短语和动词短语，这些短语形成候选方面。然后，系统采用语言模型来过滤掉那些不太可能的方面。请注意，普通的依赖解析器仅识别单个词的依赖关系，而短语依赖解析器识别短语的依赖关系，这对于方面提取可能更合适。使用依赖关系的想法已经被许多研究人员用于不同的目的（Kessler和Nicolov，2009）。

5.3.3 使用监督学习

方面提取可以看作是一般信息提取问题的特例

提取问题。过去已经提出了许多基于监督学习的算法用于信息提取（Hobbs和Riloff, 2010; Mooney和Bunescu, 2005; Sarawagi, 2008）。最主要的方法是基于顺序学习（或顺序标记）。由于这些是监督技术，它们需要手动标记的训练数据。也就是说，需要在语料库中手动注释方面和非方面。目前最先进的顺序学习方法是隐马尔可夫模型（HMM）（Rabiner, 1989）和条件随机场（CRF）（Lafferty, McCallum和Pereira, 2001）。Jin和Ho（2009）应用了一个词汇化的HMM模型来学习提取方面和观点表达式的模式。Jakob和Gurevych（Jakob和Gurevych, 2010）使用了CRF。他们在来自不同领域的评论句子上训练了CRF，以实现更独立于领域的提取。还使用了一组独立于领域的特征，例如标记、词性标记、句法依赖、词距和观点句子。Li等人（2010）集成了两种CRF变体，即Skip-CRF和Tree-CRF，用于提取方面和观点。与原始的CRF不同，Skip-CRF和Tree-CRF使CRF能够利用结构特征。CRF也在（Choi and Cardie, 2010）中使用。Liu, Hu和Cheng（2005）和Jindal和Liu（2006b）使用了顺序模式规则。这些规则是基于顺序模式挖掘考虑标签（或类别）的。

人们还可以使用其他监督方法。例如，(Kobayashi, Inui和Matsumoto, 2007)中的方法首先使用依赖树找到候选的方面和观点词对，然后采用树结构分类方法来学习和分类候选对是否是方面和评价关系。方面从得分最高的对中提取出来。学习中使用的特征包括上下文线索、统计共现线索等。Yu等人(2011)使用了一种部分监督学习方法，称为单类支持向量机(Manevitz和Yousef, 2002)来提取方面。使用单类支持向量机，只需要标记一些正例，即方面，而不是非方面。在他们的案例中，他们只从评论格式2的优点和缺点中提取方面，如(Liu, Hu和Cheng, 2005)中所述。他们还对这些同义方面进行了聚类，并根据它们在整体评论评分中的频率和贡献对方面进行了排序。Ghani等人(2006)在方面提取中使用了传统的监督学习和半监督学习。Kovelamudi等人(2011)使用了一种监督方法，但也利用了维基百科中的一些相关信息。

5.3.4 使用主题模型

近年来，统计主题模型已经成为一种从大量文本文档中发现主题的原则性方法。主题建模是一种无监督学习方法，它假设每个文档由多个主题的混合组成，每个主题是单词的概率分布。主题模型基本上是一个文档生成模型，它指定了一种通过概率过程生成文档的方法。

主题建模的输出是一组单词聚类。每个聚类形成一个主题，并且是文档集合中单词的概率分布。

有两个主要的基本模型，pLSA（概率潜在语义分析）（Hofmann, 1999）和LDA（潜在狄利克雷分配）（Blei, Ng和Jordan, 2003；Griffiths和Steyvers, 2003；Steyvers和Griffiths, 2007）。从技术上讲，主题模型是基于贝叶斯网络的一种图模型类型。尽管它们主要用于对文本集合建模和提取主题，但它们可以扩展到同时对许多其他类型的信息进行建模。例如，在情感分析的背景下，可以设计一个联合模型同时对情感词和主题进行建模，因为观点的每个方面都有一个目标。对于不熟悉主题模型、图模型或贝叶斯网络的读者，除了阅读主题建模文献外，Christopher M. Bishop的《模式识别与机器学习》（Bishop, 2006）是一个很好的背景知识来源。

直观地说，主题模型中的主题是情感分析上下文中的方面。因此，可以应用主题建模来提取方面。然而，这里也存在一些差异。也就是说，主题可以涵盖方面词和情感词。对于情感分析来说，它们需要被分开。可以通过扩展基本模型（如LDA）来同时建模方面和情感来实现这种分离。下面，我们将概述当前在情感分析中使用主题模型提取方面和执行其他任务的研究。需要注意的是，主题模型不仅可以发现方面，还可以对同义方面进行分组。

Mei等人（Mei et al., 2007）提出了一种用于情感分析的联合模型。具体而言，他们构建了一个基于方面（主题）模型、正面情感模型和负面情感模型的方面-情感混合模型，并借助一些外部训练数据进行学习。他们的模型基于pLSA。研究人员提出的大多数其他模型都基于LDA。

在（Titov和McDonald, 2008）中，作者表明全局主题模型如LDA（Blei, Ng和Jordan, 2003）可能不适用于检测

方面。原因是LDA依赖于主题分布差异和文档中词的共现来识别主题和词在每个主题中的概率分布。然而，针对特定类型产品的评论等观点文档非常同质化，意味着每个文档都谈论相同的方面，这使得全局主题模型无效，只能用于发现实体（例如，不同品牌或产品名称）。然后，作者提出了多粒度主题模型。全局模型发现实体，而本地模型使用几个句子（或滑动文本窗口）来发现方面作为一个文档。在这里，每个发现的方面都是一个一元语言模型，即，一个词的多项分布。表达相同或相关方面的不同词会自动分组到同一个方面下。

然而，这种技术不能将方面和情感词分开。

Branavan等人（2008年）提出了一种方法，该方法利用了Pros and Cons评论格式1中的方面描述作为关键词来帮助在详细评论文本中找到方面。他们的模型由两部分组成。

第一部分根据分布相似性将Pros and Cons中的关键词聚类到一些方面类别中。第二部分构建了一个主题模型，对评论文本中的主题或方面进行建模。他们的最终图形模型同时对这两部分进行建模。这两部分是基于这样一个思想进行整合的：模型使得隐藏主题在评论文本中的分配与评论中的关键词所代表的主题相似，但也允许文档中的一些词来自于关键词所代表的其他主题。这种耦合的灵活性使得模型能够在关键词不完整的情况下有效学习，同时仍然鼓励关键词聚类与评论文本支持的主题相一致。然而，这种方法仍然不能将方面和情感分开。

Lin和He（2009）通过扩展LDA提出了一个联合主题情感模型，其中方面词和情感词仍然没有明确分离。Brody和Elhadad（2010）建议首先使用主题模型识别方面，然后仅考虑形容词来识别方面特定的情感词。Li, Huang和Zhu（2010）提出了两个联合模型，Sentiment-LDA和Dependency-sentiment-LDA，以找到具有正面和负面情感的方面。它不独立地找到方面，也不分离方面词和情感词。赵等人（赵等人，2010）提出了MaxEnt-LDA（最大熵和LDA组合）混合模型，以共同发现方面词和方面特定的意见词，可以利用句法特征来帮助分离方面和情感词。通过一个指示变量（也称为开关变量）来实现联合建模，该变量被绘制

从一组参数控制的多项式分布中抽取 指示变量确定句子中的单词是方面词、观点词还是背景词 最大熵被用来学习使用标记训练数据的变量的参数

在 (Sauper, Haghighi和Barzilay, 2011) 中还提出了一种联合模型, 该模型仅适用于从评论中提取的短片段, 例如
“电池寿命是我找到的最好的。”它将主题建模与隐马尔可夫模型 (HMM) 相结合, 其中HMM对具有类型 (方面词、情感词或背景词) 的单词序列建模 他们的模型与 (Griffiths等人, 2005) 中提出的HMM-LDA相关, 该模型也对单词序列进行建模 在 (Liu等人, 2007) , (Lu和Zhai, 2008) 和 (Jo和Oh, 2011) 中还采用了联合主题建模方法的变体

在 (Mukherjee和Liu, 2012年) 中, 提出了一个半监督联合模型, 允许用户为某些主题/方面提供一些种子方面术语, 以引导推理以产生符合用户需求的方面分布。

另一种使用主题建模的工作线路旨在将方面与观点/情感评分关联起来, 即基于方面和评分的联合建模来预测方面评分。Titov和McDonald (2008年) 提出了一个从评论中发现方面并从评论中提取支持每个方面评分的文本证据的模型。Lu, Zhai和Sundaresan (2009年) 定义了从eBay.com的短评论中的评级方面摘要的问题。他们的方面提取基于一个名为结构化pLSA的主题模型。该模型可以对短评论中短语的依赖结构进行建模。为了预测评论中每个方面的评分, 它结合了评论的整体评分和基于所有评论的学习分类器对方面的分类结果。Wang等人 (2010年) 提出了一个概率评分回归模型来为方面分配评分。他们的方法首先使用一些给定的种子方面, 使用启发式自举方法找到更多的方面词。然后, 它使用提出的概率评分回归模型来预测方面评分, 该模型也是一个图形模型。该模型利用评论评分, 并假设评论的整体评分是其方面评分的线性组合。

模型参数使用最大似然 (ML) 估计器和EM风格算法进行估计。

基于HMM-LDA的复合主题模型 (Griffiths et al., 2005) 提出了一系列联合模型, 这些模型还考虑了词序列和词袋。因此, 这些模型可以捕捉到类似于句法结构和语义依赖的特征

在 (Sauper, Haghighi和Barzilay, 2011) 中。它们能够发现潜在的方面及其相应的情感评分。Moghaddam和Ester (2011) 还提出了一个联合主题模型, 用于查找和分组方面, 并推导出它们的评分。

尽管主题建模是一种基于概率推理的原则性方法, 并且可以扩展到模拟许多类型的信息, 但它在实际情感分析应用中存在一些限制。一个主要问题是它需要大量的数据和大量的调整才能达到合理的结果。更糟糕的是, 大多数主题建模方法使用吉布斯采样, 在不同运行中产生稍微不同的结果, 这使得参数调整时间很长。虽然对于主题建模来说, 从大型文档集合中找到那些非常普遍的主题或方面并不困难, 但要找到那些在局部频繁但在全局不那么频繁的方面并不容易。这些局部频繁的方面通常对应用程序最有用, 因为它们很可能与用户感兴趣的特定实体最相关。这些非常普遍和频繁的方面也可以很容易地通过前面讨论的方法找到。这些方法可以找到较少频繁的方面, 而无需大量的数据。简而言之, 当前主题建模方法的结果通常对许多实际情感分析应用来说不够细致或具体。对于用户来说, 了解文档集合的一些高层次想法更有用。

话虽如此, 主题建模是一个强大而灵活的建模工具。从概念上和数学上来说, 它也非常好。我期待继续的研究将使其更加实用。一个有前途的研究方向是在模型中融入更多现有的自然语言和领域知识。在这个方向上已经有一些初步的工作(Andrzejewski和Zhu, 2009; Andrzejewski, Zhu和Craven, 2009; Mukherjee和Liu, 2012; Zhai等, 2011)。我们将在第5.6节中讨论它们。然而, 我认为它们仍然过于统计中心化, 并且有其自身的局限性。如果我们能更多地向自然语言和知识中心化转变, 可能会更有成果, 以实现更加平衡的方法。

另一个方向是将主题建模与其他技术相结合, 以克服其缺点。

除了上面和前三节中讨论的主要方法之外, 还有其他关于方面提取的工作。例如, Yi等人(2003)使用混合语言模型和似然比来提取产品方面。Ma和Wan(2010)使用中心理论和监督学习。Meng和Wang(2009)从产品中提取方面。

规格，这些是结构化数据。Kim和Hovy（2006）使用了语义角色标注。Stoyanov和Cardie（2008）利用了共指消解。Toprak, Jakob和Gurevych（2010）设计了一个全面的基于方面的观点注释方案。早期的注释只是部分的，主要是为了个别论文的特殊需求。Carvalho等人（2011）对一系列政治辩论进行了方面和其他信息的注释。

5.3.5 隐含方面的映射

在（Hu和Liu，2004）中，识别出了两种类型的方面，即显式方面和隐含方面。然而，它只处理了显式方面。回顾第2.1节，我们将以名词和名词短语形式表达的方面称为显式方面，例如，“这台相机的画质很好”。所有其他表示方面的表达式都称为隐含方面。隐含方面表达式有很多种类型。形容词和副词可能是最常见的类型，因为大多数形容词描述实体的某些特定属性或特性，例如，“昂贵”描述了“价格”，“漂亮”描述了“外观”。隐含方面也可以是动词。总的来说，隐含方面表达式可能非常复杂，例如，“这台相机不容易放进口袋里”，“放进口袋里”表示了“大小”这个方面。

尽管明确的方面提取已经得到了广泛研究，但在将隐含的方面映射到它们的明确方面方面上，研究还很有限。

在（Su等人，2008年）中，提出了一种聚类方法，用于将被假定为情感词的隐含方面表达式映射到它们对应的明确方面。该方法利用了句子中形成共现对的明确方面和情感词之间的相互增强关系。这样的一对可能表明情感词描述了该方面，或者该方面与情感词相关联。该算法通过迭代地分别对明确方面集合和情感词集合进行聚类来找到映射。在每次迭代中，在对一个集合进行聚类之前，使用另一个集合的聚类结果来更新集合的成对相似度。集合中的成对相似度由集合内相似度和集合间相似度的线性组合确定。两个项的集合内相似度是传统相似度。

两个项目的互集相似度是基于方面和情感词之间的关联程度计算的。关联（或互相强化关系）使用二分图进行建模。如果一个方面和一个观点词在同一上下文中出现，则它们会被连接起来。

句子。链接也根据共现频率进行加权。
在迭代聚类之后，最强的 n 链接形成了方面和情感词组之间的映射。

在 (Hai, Chang和Kim, 2011) 中，提出了一种两阶段共现关联规则挖掘方法，用于将隐含的方面（也被假设为情感词）与明确的方面匹配。在第一阶段，该方法生成涉及每个情感词作为条件和明确方面作为结果的关联规则，在语料库的句子中频繁共现。在第二阶段，它对规则结果（明确方面）进行聚类，为上述每个情感词生成更强大的规则。对于应用或测试，给定一个没有明确方面的情感词，它找到最佳的规则簇，然后将簇的代表词指定为最终确定的方面。

5.4

识别资源使用方面

如第4.3节所讨论的，研究人员通常试图以一种普遍的方式解决问题，并且在许多情况下基于一种简化的观点。在方面提取和方面情感分类的背景下，并不总是情感词和方面词对是重要的。如第5.2节所示，现实世界比那复杂多样。

在这里，我们以资源使用为例，展示了在基于方面的情感分析中可能需要采用分而治之的方法。

在许多应用中，资源使用是一个重要的方面，例如，“这个洗衣机用水很多。”这里水的使用是洗衣机的一个方面，这句话表达了消耗太多资源是不可取的负面意见。这个句子中没有观点词。

因此，发现被称为资源术语的资源词和短语对于情感分析是重要的。在第5.2节中，我们介绍了涉及资源的一些观点规则。我们在下面重现了其中两个：

1. P ::= 不消耗、少量或更少的资源
2. N ::= 消耗大量或更多的资源

在 (Zhang和Liu, 2011a) 中，提出了一种提取资源术语的方法。例如，在上面的例子中，“水”应该被提取为一个资源术语。该论文基于一个二分图来制定问题，并提出了一个迭代算法来解决问题。该算法基于以下观察结果：

观察：在句子中表达的关于
资源使用的情感或观点通常由以下三元组决定，

(动词, 量词, 名词术语) ,

其中 名词术语 是一个名词或名词短语

例如, 在“这台洗衣机使用了很多水”中, “使用”是主要动词, “很多”是一个量词短语, “水”是表示一个资源的名词。该方法使用这样的三元组来帮助识别领域语料库中的资源。该模型使用循环定义来反映资源使用动词(例如, 消耗)和资源术语(例如, 水)之间的特殊强化关系, 基于二分图。量词在计算中没有被使用, 但被用来识别候选动词和资源术语。该算法假设给定了一个量词列表, 这个列表不多, 可以手动编制。基于循环定义, 使用类似于(Kleinberg, 1999)中的HITS算法的迭代算法来解决问题。为了开始迭代计算, 使用一些全局种子资源来找到并评分一些强资源使用动词。然后, 这些分数被应用为任何应用领域的迭代计算的初始化。当算法收敛时, 将识别出一个候选资源术语的排序列表。

5.5 同时进行观点词典扩展和方面提取

正如第2章中提到的, 观点总是有一个目标。这个属性已经被几位研究人员在方面提取中利用(参见第5.3.2节)。在(Qiu等, 2009年; Qiu等, 2011年)中, 它被用来同时提取情感词和方面, 通过利用情感和目标之间的某些句法关系, 以及一小组种子情感词(不需要种子方面)进行提取。该方法基于自举。请注意, 情感词生成本身就是一项重要任务(参见第6章)。

由于情感/观点与它们的目标(或方面)之间的关系, 情感词可以通过已识别的方面来识别, 而方面可以通过已知的情感词来识别。提取的情感词和方面被用于识别新的情感词和新的方面, 这些新的情感词和方面再次用于提取更多的情感词和方面。当找不到更多的情感词或方面时, 这个传播过程就结束了。由于这个过程涉及情感词和方面的传播, 该方法被称为双重传播。

提取规则基于情感词和方面之间的特定依赖关系。依赖语法(Tesniere, 1959年)

情感分析与观点挖掘

被采用来描述这些关系。使用的依赖解析器是minipar (Lin, 2007年)。

还施加了一些约束。情感词被认为是形容词，而方面是名词或名词短语。情感词和方面之间的依赖关系包括mod, pnm, subj, s, obj, obj2和desc, 而情感词和方面本身的关系只包含连词关系conj。OA-Rel表示情感词和方面之间的关系, OO-Rel表示情感词之间的关系, AA-Rel表示方面之间的关系。OA-Rel, OO-Rel或AA-Rel中的每个关系都是一个三元组 $\langle \text{POS}(w_i), R, \text{POS}(w_j) \rangle$, 其中 $\text{POS}(w_i)$ 是单词 w_i 的词性标记, R 是上述依赖关系之一。

提取过程使用基于规则的方法。例如, 在“佳能G3拍出了很棒的照片”中, 形容词“很棒”通过依赖于名词“照片”进行解析, 通过 mod进行公式化, 表示为OA-Rel $\langle JJ, \text{mod}, NNS \rangle$ 。如果我们知道“很棒”是一个情感词, 并且给定规则“情感词直接通过 mod依赖的名词被视为一个方面”, 我们可以提取“照片”作为一个方面。同样, 如果我们知道“照片”是一个方面, 我们可以使用类似的规则提取“很棒”作为观点词。传播执行四个子任务:

1. 使用情感词提取方面
2. 使用提取的方面提取方面
3. 使用提取的方面提取情感词
4. 使用给定的和提取的观点词提取情感词

OA-Rels用于任务(1)和(3), AA-Rels用于任务(2), OO-Rels用于任务(4)。分别为这四个子任务定义了四种规则(如表5.1所示)。在表中, o (或 a) 表示输出(或提取的)情感词(或方面)。 $\{O\}$ (或 $\{A\}$) 是已知情感词(或方面)的集合, 可以是给定的或提取的。 H 表示任意单词。 $\text{POS}(O \text{ (或 } A))$ 和 $O \text{ (或 } A) - \text{Dep}$ 分别表示单词 O (或 A) 的POS标签和依赖关系。 $\{JJ\}$ 和 $\{NN\}$ 分别是潜在情感词和方面的POS标签集合。 $\{JJ\}$ 包含 JJ , JJR 和 JJS ; $\{NN\}$ 包含 NN 和 NNS 。 $\{MR\}$ 由依赖关系组成, 即集合 $\{\text{mod}, \text{pnm}, \text{subj}, s, \text{obj}, \text{obj2}$ 和 $\text{desc}\}$ 。 $\{CONJ\}$ 仅包含 conj 。箭头表示依赖关系。例如, $O \rightarrow O - \text{Dep} \rightarrow A$ 表示 O 通过关系 $O - \text{Dep}$ 依赖于 A 。具体来说, 使用 R_1 提取方面(a)使用情感词(O), 使用 R_2 提取观点词(o)使用方面(A), 使用 R_3 提取方面(a)使用提取的方面(A_i), 使用 R_4 提取情感词(o)使用已知情感词(O_i)。

情感分析与观点挖掘

	观察	输出	示例
R1 ₁ (OA-Rel)	$O \rightarrow O-Dep \rightarrow A$ s.t. $O \in \{O\}$, $O-Dep \in \{MR\}$, $POS(A) \in \{NN\}$	$a = A$	这部手机有一个 <u>好的</u> “屏幕”。 好 $\rightarrow mod \rightarrow$ 屏幕
R1 ₂ (OA-Rel)	$O \rightarrow O-Dep \rightarrow H \leftarrow A-Dep \leftarrow A$ s.t. $O \in \{O\}$, $O/A-Dep \in \{MR\}$, $POS(A) \in \{NN\}$	$a = A$	“iPod”是 <u>最好的</u> mp3播放器。 最好的 $\rightarrow mod \rightarrow$ 播放器 $\leftarrow subj \leftarrow$ iPod
R2 ₁ (OA-Rel)	$O \rightarrow O-Dep \rightarrow A$ s.t. $A \in \{A\}$, $O-Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	与屏幕作为已知词，好作为 提取词的R1 ₁
R2 ₂ (OA-Rel)	$O \rightarrow O-Dep \rightarrow H \leftarrow A-Dep \leftarrow A$ s.t. $A \in \{A\}$, $O/A-Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	与R1 ₂ 相同与iPod是已知的 词和最好的作为提取词。
R3 ₁ (AA-Rel)	$A_{i(j)} \rightarrow A_{i(j)}-Dep \rightarrow A_{j(i)}$ 满足 $A_j \cap \emptyset \in \{A\}$, $A_i \cap \emptyset-Dep \in \{CONJ\}$, $POS(A_{i(j)}) \in \{NN\}$	$a = A_{i(j)}$	播放器是否能够播放带有 音频和“视频”?? 视频 \rightarrow 连词 \rightarrow 音频
R3 ₂ (AA-Rel)	$A_i \rightarrow A_i-Dep \rightarrow H \leftarrow A_j-Dep \leftarrow A_j$ 满足 $A_i \in \{A\}$, $A_i-Dep = A_j-Dep$ 或 ($A_i-Dep =$ 主语且 $A_j-Dep =$ 宾语), $POS(A_j) \in \{NN\}$	$a = A_j$	佳能“G3”拥有一支很棒的镜头。 镜头 \rightarrow 物体 \rightarrow 拥有 \leftarrow 主语 \leftarrow G3
R4 ₁ (OO-关系)	$O_{i(j)} \rightarrow O_{i(j)}-依存 \rightarrow O_{j(i)}$ 满足 $O_{j(i)} \in \{O\}$, $O_{i(j)}-依存 \in \{CONJ\}$, $POS(O_{i(j)}) \in \{JJ\}$	$o = O_{i(j)}$	这台相机很棒， <u>而且</u> “易用”。 易用 \rightarrow 并列 \rightarrow 很棒
R4 ₂ (OO-关系)	$O_i \rightarrow O_i-依存 \rightarrow H \leftarrow O_j-依存 \leftarrow O_j$ 满足 $O_i \in \{O\}$, $O_i-依存 = O_j-依存$ 或者 ($O_i / O_j-依存 \in \{pnmod, mod\}$), $POS(O_j) \in \{JJ\}$	$o = O_j$	如果你想买一个性感，“酷”的 ，可用的mp3播放器，你可以 选择iPod。 性感 $\rightarrow mod \rightarrow$ 播放器 $\leftarrow mod \leftarrow$ 酷

表5.1。方面和观点词提取的规则。第1列是规则ID，第2列是观察到的关系（第1行）和必须满足的约束条件（第2-4行），第3列是输出，第4列是一个示例。在每个示例中，下划线的单词是已知的单词，双引号中的单词是提取出来的单词。相应的实例化关系给出在示例的正下方。

这种方法最初是为英语设计的，但也已经被用于中文在线讨论（Zhai等，2011年）。这种方法也可以通过使用大型情感词典来仅找到方面进行简化。为了实际应用，关系集合可以显著扩展。此外，与其使用基于单词的依存句法分析，使用短语级别的依存句法分析可能更好，因为许多方面是短语（Wu等，2009年）。Zhang等（2010年）通过添加更多关系并使用图方法对提取的方面进行排序，改进了这种方法。

5.6 将方面分组为类别

在提取方面之后，需要将方面表达式（表示方面的实际词语和短语）分组为同义方面类别。每个类别代表一个独特的方面。像任何写作一样，人们经常

使用不同的词和短语来描述相同的方面。例如，对于手机，“通话质量”和“语音质量”指的是同一个方面。将来自同一方面的这种方面表达式分组对于观点分析至关重要。尽管WorldNet和其他同义词词典在一定程度上有所帮助，但它们远远不够，因为许多同义词是依赖于领域的（Liu, Hu和Cheng, 2005）。例如，在电影评论中，“电影”和“图片”是同义词，但在相机评论中它们不是同义词，因为“图片”更有可能与“照片”同义，而“电影”更有可能与“视频”同义。许多方面表达式是由多个词组成的短语，这些短语不能很容易地通过词典处理。此外，还需要注意的是，描述相同方面的许多方面表达式不是通用的或领域特定的同义词。例如，“昂贵”和“便宜”都可以表示方面价格但它们不是彼此的同义词（而是反义词）或价格的同义词。

Carenini, Ng和Zwart（2005）提出了第一种处理这个问题的方法。他们的方法基于使用字符串相似性，同义词和使用WordNet测量的词汇距离定义的几个相似性度量。该方法需要为特定领域给出一个方面的分类法。它根据相似性将每个发现的方面表达式合并到分类法中的一个方面节点中。基于数码相机和DVD评论的实验显示了有希望的结果。在（Yu等，2011）中，提出了一种更复杂的方法，该方法还使用公开可用的产品方面层次结构/分类法和实际产品评论来生成最终的方面层次结构。还使用了一组距离度量，但与优化策略相结合。

在（Zhai等，2010）中，提出了一种半监督学习方法，将方面表达式分组到一些用户指定的方面类别中。为了反映用户需求，他/她首先为每个类别标记了少量种子。然后，系统使用半监督学习方法将其余的方面表达式分配到合适的类别中，该方法使用带标签和未标签的示例。该方法使用了（Nigam等，2000）中的期望最大化（EM）算法。该方法还使用了两个先验知识来为EM提供更好的初始化：（1）共享一些公共词汇的方面表达式很可能属于同一组，例如“电池寿命”和“电池功率”，以及（2）在字典中是同义词的方面表达式很可能属于同一组，例如“电影”和“图片”。这两个知识帮助EM产生更好的分类结果。在（Zhai等，2011）中，使用软约束来帮助标记一些示例，即共享词汇和词汇相似性（Jiang和Conrath，1997）。学习方法还使用了

EM，但它消除了要求用户提供种子的需要。请注意，一般的概念相似性和同义词发现的自然语言处理研究在这里也是相关的（Mohammad和Hirst, 2006；Wang和Hirst, 2011）。

在（Guo等, 2009）中，提出了一种称为多层次潜在语义关联的方法。在第一层，将方面表达式中的所有单词（每个方面表达式可以有多个单词）使用LDA分组成一组概念/主题。结果用于构建方面表达式的潜在主题结构。例如，我们有四个方面表达式“day photos”，“day photo”，“daytime photos”和“daytime photo”。如果LDA将单词“day”和“daytime”分组到主题10中，将“photo”和“photos”分组到主题12中，系统将把所有四个方面表达式分组到一个组中，称为“topic10-topic12”，这被称为潜在主题结构。在第二层，根据第一层产生的潜在主题结构和评论中的上下文片段，再次使用LDA对方面表达式进行分组。继续以上例子，在“topic10-topic12”中的“day photos”，“day photo”，“daytime photos”和“daytime photo”与它们周围的单词组成一个文档。LDA在这样的文档上运行以产生最终结果。在（Guo等, 2010）中，还使用了类似的思想将来自不同语言（或国家）的方面分组为方面类别，这可以用于比较不同语言（或国家）的不同方面上的观点。

在第5.3.4节中讨论的主题建模方法实际上同时以无监督的方式执行方面表达发现和分类，因为主题建模基本上是将文档集中的术语进行聚类。最近，一些算法还提出使用领域知识或约束来指导主题建模，以产生更好的主题聚类（Andrzejewski, Zhu和Craven, 2009）。约束的形式是必须链接和不能链接。在聚类中，必须链接约束指定两个数据实例必须在同一个聚类中。不能链接约束指定两个数据实例不能在同一个聚类中。然而，该方法可能导致不能链接约束的编码呈指数增长，因此在处理大量约束时可能会遇到困难。

Zhai等人（2011）的约束-LDA采用了不同但启发式的方法。约束-LDA不将约束视为先验，而是在Gibbs采样中使用约束来偏置单词的主题分配的条件概率。该方法可以处理大量的必须链接和不能链接约束。约束也可以放宽，即将其视为软约束（而不是硬约束），可能无法满足。对于方面分类，约束-LDA使用以下约束：

Must-link: 如果两个方面表达式 a_i 和 a_j 共享一个或多个单词, 则它们形成一个必连, 即它们很可能属于同一个主题或类别, 例如, “电池功率”和“电池寿命”。

Cannot-link: 如果同一个句子中的两个方面表达式 a_i 和 a_j , 则它们形成一个不能连。这个约束的原因是人们通常不会在同一个句子中重复相同的方面, 例如, “我喜欢这个相机的画质、电池寿命和变焦。”

在(Mukherjee and Liu, 2012)中, 领域知识以某些用户提供的种子方面词的形式出现在某些主题(或方面)上。因此, 得到的模型是半监督的。该模型还将方面词和情感词分开。(Andrzejewski, Zhu and Craven, 2009)中的模型或约束-LDA方法并不这样做。

5.7 实体、观点持有者和时间提取

实体、观点持有者和时间提取是命名实体识别(NER)的经典问题。NER在多个领域中得到了广泛研究, 例如信息检索、文本挖掘、数据挖掘、机器学习和自然语言处理, 以信息提取的名义进行研究(Hobbs和Riloff, 2010; Mooney和Bunescu, 2005; Sarawagi, 2008)。

信息提取有两种主要方法: 基于规则和统计。早期的提取系统主要基于规则(例如, (Riloff, 1993))。统计方法通常基于隐马尔可夫模型(HMM)(Rabiner, 1989)(Jin和Ho, 2009)和条件随机场。

(CRF)(Lafferty, McCallum和Pereira, 2001)。HMM和CRF都是有监督的方法。由于该领域的先前工作, 情感分析和观点挖掘的具体工作并不广泛。因此, 我们不会进一步讨论它。在(Sarawagi, 2008)中, 可以看到关于信息提取任务和算法的综合调查。在这里, 我们只讨论情感分析应用中的一些具体问题。

在大多数使用社交媒体的应用程序中, 我们不需要从文本中提取意见持有者和发布时间, 因为意见持有者通常是评论、博客或讨论帖子的作者, 他们的登录ID是已知的, 尽管他们在现实世界中的真实身份是未知的。提交帖子的日期和时间也是已知的, 并显示在网页上。它们可以使用结构化数据提取技术(Liu, 2006和2011)从页面中抓取。在某些情况下, 意见持有者可能在实际文本中, 并且需要被提取出来。我们在下面讨论这个问题。

在这里，我们首先讨论情感分析背景下的具体命名实体提取问题。在典型的情感分析应用程序中，用户通常希望找到关于一些竞争实体（例如竞争产品或品牌）的意见。然而，他/她通常只能提供一些名称，因为有很多不同的品牌和型号。即使对于同一个实体，网络用户可能以许多不同的方式来写实体。例如，“Motorola”可能被写成“Moto”或“Mot”。因此，对于一个系统来自动从语料库（例如评论、博客和论坛讨论）中发现它们是很重要的。这种提取的主要要求是提取的实体必须与用户提供的实体类型相同（例如手机品牌和型号）。

在（Li等人，2010年）中，Li等人将问题形式化为一个集合扩展问题（Ghahramani和Heller，2006年；Pantel等人，2009年）。问题的陈述如下：给定一个特定类别 C 的种子实体集合 Q 和一个候选实体集合 D ，我们希望确定 D 中的哪些实体属于 C 。也就是说，我们根据种子示例集合 Q 来“扩展”类别 C 。尽管这是一个分类问题，但在实践中，通常将其作为一个排名问题来解决，即根据实体属于 C 的可能性对 D 中的实体进行排名。

解决这个问题的经典方法是基于分布相似性（Lee，1999年；Pantel等人，2009年）。该方法通过比较每个候选实体的周围词与种子实体的相似性，然后根据相似性值对候选实体进行排名。在（Li等人，2010年）中，证明了这种方法的不准确性。使用S-EM算法（Liu等人，2002年）从正例和无标签示例（PU学习）中学习明显更好。为了应用PU学习，使用给定的种子自动提取包含一个或多个种子的句子。这些句子中种子的周围词作为种子的上下文。其余的句子被视为无标签示例。实验结果表明，S-EM优于机器学习技术贝叶斯集合（Ghahramani和Heller，2006年），后者也明显优于分布相似性度量。

关于情感分析背景下的观点持有者提取，已有几位研究人员进行了调查。提取主要是在新闻文章中进行的。Kim和Hovy（2004）将个人和组织视为唯一可能的观点持有者，并使用命名实体标记器进行识别。Choi，Breck和Cardie（2006）使用条件随机场（CRF）进行提取。为了训练CRF，他们使用了诸如周围单词、周围单词的词性、语法角色、情感词等特征。在（Kim和Hovy，2006）中，该方法首先在句子中生成所有可能的持有者候选，即所有名词短语，包括普通名词短语、命名实体和代词。

然后，它解析句子并从解析树中提取一组特征。然后，一个学习的最大熵（ME）模型根据ME模型得到的分数对所有持有者候选进行排序。系统选择得分最高的候选者作为句子中的观点持有者。Johansson和Moschitti（2010）使用了一组特征的SVM。Wiegand和Klakow（2010）使用了卷积核，Lu（2010）应用了依赖解析器。

在（Ruppenhofer, Somasundaran和Wiebe, 2008）中，作者讨论了使用自动语义角色标注（ASRL）来识别观点持有者的问题。他们认为ASRL是不够的，还需要考虑其他语言现象，如话语结构。Kim和Hovy（2006）早先也使用了语义角色标注来达到这个目的。

5.8 共指消解和词义消歧

尽管我们在这一部分只讨论了共指消解和词义消歧，但我们真正想强调的是情感分析背景下的自然语言处理问题和难题。大多数这类问题在情感分析中尚未得到研究。

共指消解在自然语言处理领域得到了广泛研究。它指的是确定句子或文档中指向同一事物的多个表达式的问题，即它们具有相同的“指称”。例如，在“我两天前买了一部iPhone。它看起来很漂亮。”中，第二个句子中的“它”指的是iPhone，它是一个实体，第四个句子中的“它们”指的是“通话”，它是一个方面。过去两天我打了很多电话。它们很棒。”第二个句子中的“它”指的是iPhone，它是一个实体，第四个句子中的“它们”指的是“通话”，它是一个方面。识别这些共指关系对于基于方面的情感分析显然非常重要。如果我们不解决这些关系，而只是独立考虑每个句子中的观点，我们就会失去回忆。也就是说，虽然我们知道第二个和第四个句子表达了观点，但我们不知道观点是关于什么的。因此，从这段文本中我们将得不到有用的观点，但实际上，它对iPhone本身有一个积极的观点，也对通话质量有一个积极的观点。

Ding和Liu（2010）提出了实体和方面共指解决问题。该任务旨在确定代词所指的实体和/或方面。该论文采用了监督学习方法。有趣的关键点是设计和测试了两个与观点相关的特征，这表明情感分析被用于

情感分析与观点挖掘

共指解决的目的。第一个特征基于常规句子和比较句子的情感分析，以及情感一致性的思想。考虑以下句子，“诺基亚手机比这个摩托罗拉手机好。它也很便宜。”我们的常识告诉我们，“它”指的是“诺基亚手机”，因为在第一句中，“诺基亚手机”的情感是积极的（比较积极），但对于“摩托罗拉手机”是消极的（比较消极），而第二句是积极的。因此，我们得出结论，“它”指的是“诺基亚手机”，因为人们通常以一致的方式表达情感。不太可能“它”指的是“摩托罗拉手机”。然而，如果我们将“它也很贵”改为“它也很贵”，那么“它”现在应该指的是“摩托罗拉手机”。为了获得这个特征，系统需要具备确定常规句子和比较句子中表达的积极和消极观点的能力。

第二个特征考虑了哪些实体和方面被哪些观点词修改。考虑以下句子，“昨天我买了一部诺基亚手机。声音质量很好。它也很便宜。”问题是“它”指的是什么，“声音质量”还是“诺基亚手机”。显然，我们知道“它”指的是“诺基亚手机”，因为“声音质量”不能便宜。为了获得这个特征，系统需要确定哪些情感词通常与哪些实体或方面相关联。这些关系必须从语料库中挖掘出来。这两个特征是当前通用共指消解方法不考虑的语义特征。这两个特征可以帮助提高共指消解的准确性。

在（Stoyanov和Cardie，2006）中，Stoyanov和Cardie提出了源共指消解问题，即确定观点持有者（源）的哪些提及指的是同一实体。作者在（Ng和Cardie，2002）中使用了现有的共指消解特征。

然而，他们并不仅仅使用监督学习，而是使用了部分监督聚类。

Akkaya，Wiebe和Mihalcea（2009）研究了主观词义消歧（SWSD）。任务是自动确定语料库中哪些词实例是带有主观意义的，哪些是带有客观意义的。目前，大多数主观性或情感词典是以单词列表的形式编制的，而不是单词含义（意义）。

然而，许多单词既有主观意义又有客观意义。错误命中-用于客观意义的主观性线索-是主观性和情感分析中的一个重要错误来源。作者们在语料库上建立了一个监督SWSD模型，将主观词典的成员消歧为具有主观意义或客观意义。该算法依赖于用于词义消歧（WSD）的常见机器学习特征。然而，性能要比

完整的WSD性能在相同数据上表现出色，这表明SWSD任务是可行的，并且主观性提供了一种自然的粗粒度感觉分组。他们还表明，SWS D随后可以帮助主观性和情感分析。

5.9 摘要

方面级情感分析通常是实际应用所需的详细级别。大多数工业系统都是这样构建的。尽管研究界已经做了大量的工作，并且也建立了许多系统，但问题仍然远未解决。

每个子问题仍然非常具有挑战性。正如一位首席执行官所说，“我们的情感分析和其他人一样糟糕”，这是对当前情况和问题的困难的一个很好的描述。

最突出的两个问题是方面提取和方面情感分类。这两个问题的准确率都不高，因为现有算法仍然无法处理复杂的句子，需要超过情感词和简单解析，或者处理暗示观点的事实句子。我们在第5.2节的基本观点规则中讨论了其中一些问题。

总体而言，我们似乎遇到了一个长尾问题。虽然情感词可以处理大约60%的情况（在某些领域更多，在其他领域更少），但其余的情况千差万别、数量众多且不常见，这使得统计学习算法很难学习到模式，因为它们没有足够的训练数据。事实上，人们表达积极或消极观点的方式似乎有无限多种。每个领域似乎都有一些特殊之处。在（Wu等人，2011年）中，提出了一种更复杂的基于图的观点表示方法，这需要更复杂的解决方案。

到目前为止，研究界主要关注电子产品、酒店和餐厅的观点。这些领域相对较容易（虽然并不容易），如果能够专注于每个领域并处理其特殊情况，就可以达到相当高的准确性。当涉及其他领域时，例如床垫和油漆，情况变得更加困难，因为在这些领域中，许多事实陈述都暗示了观点。政治是另一个棘手的问题。在这里，当前的方面提取算法只取得了有限的成功，因为很少有政治问题（方面）可以用一两个词来描述。由于事实报道和主观观点的复杂混合以及大量使用讽刺句子，政治情感的确定也更加困难。

情感分析与观点挖掘

就社交媒体类型而言，从事基于方面的情感分析的研究人员主要关注产品/服务评论和来自Twitter的推文。这些形式的数据也更容易处理（虽然不容易），因为评论富含观点且没有多余的信息，而推文非常简短且通常直截了当。

然而，其他形式的观点文本，如论坛讨论和评论，要处理起来要困难得多，因为它们与各种非观点内容混合在一起，通常涉及多个实体和用户互动。这引出了另一个我们迄今尚未讨论的重要问题。这就是数据噪声的问题，对此的研究有限。

几乎所有形式的社交媒体都非常嘈杂（评论除外），充斥着各种拼写、语法和标点错误。大多数自然语言处理工具，如词性标注器和解析器，需要干净的数据才能准确执行。因此，在进行任何分析之前，需要进行大量的预处理。有关一些预处理任务和方法，请参阅（Dey和Haque, 2008）的文献。

为了取得重大进展，我们仍然需要新颖的想法和研究广泛的领域。成功的算法很可能是机器学习、领域知识和自然语言知识的良好融合。

第6章

情感词典生成

现在应该很清楚，传达积极或消极情感的词语和短语对于情感分析至关重要。本章讨论如何编制这样的词语列表。在研究文献中，情感词也被称为观点词、极性词或带有观点的词语。积极情感词用于表达一些期望的状态或品质，而消极情感词用于表达一些不希望的状态或品质。积极情感词的例子有美丽的、精彩的和令人惊叹的。消极情感词的例子有糟糕的、可怕的和差劲的。除了单个词语，还有情感短语和习语，例如“花费某人一条手臂和一条腿”。总体上，它们被称为情感词典（或观点词典）。为了方便表述，从现在开始，当我们说情感词时，我们指的是单个词语和短语。

情感词可以分为两种类型，基本类型和比较类型。上面的所有例子词都属于基本类型。比较类型的情感词（包括最高级类型）用于表达比较和最高级观点。这类词的例子有更好、更差、最好、最差等，它们是其基本形容词或副词的比较和最高级形式，例如好和坏。与基本类型的情感词不同，比较类型的情感词不表达对一个实体的常规观点，而是对多个实体的比较观点，例如“百事可乐比可口可乐更好喝”。这句话并不表达任何一种饮料好或者不好的观点，只是说相比于可口可乐，百事可乐更好喝。我们将在第8章进一步讨论比较和最高级情感词。本章仅关注基本类型的情感词。

研究人员提出了许多方法来编译情感词汇。

三种主要方法是：手动方法，基于词典的方法，以及基于语料库的方法。手动方法工作量大且耗时，因此通常不单独使用，而是与自动化方法结合作为最终检查，因为自动化方法会出错。下面，我们讨论两种自动化方法。除此之外，我们还将讨论涉及观点的事实陈述问题，这在研究界很大程度上被忽视了。

6.1 基于词典的方法

使用词典编译情感词汇是一种明显的方法，因为大多数词典（例如WordNet（Miller et al., 1990））列出了每个词的同义词和反义词。因此，在这种方法中，一个简单的技术是根据词典的同义词和反义词结构，使用一些种子情感词进行引导。具体而言，该方法的工作方式如下：首先手动收集一小组具有已知正面或负面倾向的情感词（种子），这非常容易。然后，算法通过在WordNet或其他在线词典中搜索它们的同义词和反义词来扩展这个集合。新发现的词被添加到种子列表中。下一次迭代开始。当找不到更多新词时，迭代过程结束。这种方法在（Hu and Liu, 2004）中被使用。完成该过程后，使用手动检查步骤来清理列表。Valitutti, Strapparava和Stock（2004）也使用了类似的方法。Kim和Hovy（2004）试图清理结果词汇（以消除错误）并使用概率方法为每个词赋予情感强度。Mohammad, Dunne和Dorr（2009）还利用了许多生成反义词的词缀模式，如 X 和 $\text{dis}X$ （例如，honest-dishonest），以增加覆盖范围。

在(Kamps等人, 2004)提出了一种更复杂的方法，该方法使用基于WordNet距离的方法来确定给定形容词的情感倾向。术语 t_1 和 t_2 之间的距离 $d(t_1, t_2)$ 是连接 t_1 和 t_2 的最短路径的长度。形容词术语 t 的方向由其与两个参考（或种子）术语good和bad的相对距离确定，即 $SO(t) = (d(t, \text{bad}) - d(t, \text{good})) / d(\text{good}, \text{bad})$ 。如果 $SO(t) > 0$ ，则 t 为正向，否则为负向。 $SO(t)$ 的绝对值表示情感的强度。沿着类似的思路，Williams和Anand（2009）研究了为每个词分配情感强度的问题。

在（Blair-Goldensohn等人, 2008年）中，提出了一种不同的引导方法，该方法使用了一个正向种子集、一个负向种子集和一个中性种子集。该方法基于一个有向加权语义图，其中相邻节点是WordNet中单词的同义词或反义词，并且不属于种子中性集。中性集用于阻止情感通过中性词的传播。边的权重是根据不同类型的边（即同义词或反义词边）的缩放参数预先分配的。然后，使用（Zhu和Ghahramani, 2002年）中的修改版本的标签传播算法对每个单词进行评分（给出情感值）。在开始时，每个正向种子词的得分为+1，每个负向种子词的得分为-1，所有

其他词语得分为0。在传播过程中,分数会进行修订。当传播在若干次迭代后停止时,经过对数缩放的最终分数被分配给词语作为它们积极或消极程度的度量。

在(Rao和Ravichandran, 2009)中,尝试了三种基于图的半监督学习方法,以分离给定正向种子集、负向种子集和从WordNet中提取的同义词图中的正向和负向词语。这三种算法是Mincut (Blum和Chawla, 2001), Randomized Mincut (Blum等, 2004)和标签传播 (Zhu和Ghahramani, 2002)。结果显示, Mincut和Randomized Mincut的F得分更好,但标签传播的精确度明显更高,召回率较低。

Hassan和Radev (2010)提出了一个基于马尔可夫随机游走模型的词相关性图,用于为给定的词语产生情感估计。它首先使用WordNet的同义词和上位词构建词相关性图。然后定义了一种称为平均命中时间 $h(i|S)$ 的度量,并用于衡量从节点 i 到一组节点(词语) S 的距离,即一个随机行走者从状态 $i \notin S$ 开始,进入状态 $k \in S$ 的平均步数。给定一组正向种子词 S^+ 和一组负向种子词 S^- ,为了估计给定词语 w 的情感倾向,它计算命中时间 $h(w|S^+)$ 和 $h(w|S^-)$ 。如果 $h(w|S^+)$ 大于 $h(w|S^-)$,则将该词语分类为负向,否则为正向。在(Hassan等, 2011)中,该方法被应用于找到外语词语的情感倾向。为此,创建了一个包含英语词语和外语词语的多语言词图。不同语言的词语根据它们在字典中的含义相连。

基于图的其他方法包括(Takamura, Inui and Okumura, 2005)和(Takamura, Inui and Okumura, 2007; Takamura, Inui and Okumura, 2006)。

在(Turney and Littman, 2003)中,使用与(Turney, 2002)相同的基于PMI的方法来计算给定单词的情感倾向。具体来说,它通过计算单词与一组积极词汇(好,不错,优秀,积极,幸运,正确和优越)的关联强度减去与一组消极词汇(糟糕,讨厌,贫穷,消极,不幸,错误和劣质)的关联强度来计算单词的情感倾向。关联强度使用PMI来衡量。

Esuli和Sebastiani (2005)使用有监督学习将单词分类为积极和消极类别。给定一组 P 正面种子词和一组 N 负面种子词,首先使用在线词典(例如WordNet)中的同义词和反义词关系扩展这两个种子集合,以

生成扩展集合 P' 和 N' ，形成训练集。然后，算法使用 $P' \cup N'$ 中每个术语的词典中的所有解释生成特征向量。然后使用不同的学习算法构建二元分类器。该过程也可以迭代运行。也就是说，新识别出的积极和消极术语及其同义词和反义词被添加到训练集中，可以构建更新的分类器等等。在 (Esuli和Sebastiani, 2006) 中，作者还包括了类别客观。为了扩展客观种子集合，除了使用同义词和反义词外，还使用了上位词。然后尝试了不同的策略来进行三类分类。在 (Esuli和Sebastiani, 2006) 中，基于上述方法的分类器委员会被用于构建SentiWordNet，这是一个词汇资源，其中WordNet的每个同义词集与三个数值分数 $Obj(s)$ ， $Pos(s)$ 和 $Neg(s)$ 相关联，描述了同义词集中包含的术语的客观性、积极性和消极性。Kim和Hovy (2006) 的方法也从三个种子集合开始，分别是积极、消极和中性词。然后在WordNet中找到它们的同义词。然而，扩展集合中存在许多错误。然后，该方法使用贝叶斯公式计算每个词与每个类别（积极、消极和中性）的接近程度，以确定该词的最可能的类别。

Andreevskaya和Bergler (2006) 提出了一种更复杂的自举方法，使用几种技术扩展初始的正面和负面种子集，并清理扩展集（删除非形容词和同时出现在正面和负面集中的词）。此外，他们的算法还使用非重叠的种子子集执行多次自举过程。每次运行通常会找到略有不同的情感词集。然后，根据单词在运行中作为正面词和负面词被发现的次数，计算每个词的重叠得分。然后根据模糊集合理论将得分归一化为[0, 1]。

在 (Kaji和Kitsuregawa, 2006; Kaji和Kitsuregawa, 2007) 中，使用了许多启发式方法从HTML文档中构建情感词典，这些方法基于Web页面布局结构。例如，Web页面中的表格可能有一列明确表示周围文本的正面或负面倾向（例如，优点和缺点）。这些线索可以用来从大量的Web页面中提取大量候选的正面和负面观点句子。然后从这些句子中提取形容词短语，并根据它们在正面和负面句子集中的出现统计信息分配情感倾向。

Velikovich等人 (2010年) 还提出了一种构建情感的方法

情感分析与观点挖掘

词典使用网页。它基于短语相似性图上的图传播算法。它再次假设输入是一组正面种子短语和一组负面种子短语。短语图中的节点是从40亿个网页中提取的所有n-gram的候选短语。只选择了2000万个候选短语，使用了几个启发式方法，例如词边界的频率和互信息。然后为每个候选短语构建一个上下文向量，该向量是在40亿个文档中短语的所有提及上聚合的大小为6的词窗口。通过计算候选短语的上下文向量的余弦相似度来构建边集。如果边 (v_i, v_j) 不是与节点 v_i 或 v_j 相邻的25条最高权重边之一，则将其丢弃。边的权重设置为相应的余弦相似度值。使用图传播方法计算每个短语的情感，作为到种子词的所有最佳路径的聚合。

在(Dragut等人, 2010)中, 提出了另一种非常不同的基于WordNet的引导方法。给定一组种子词, 作者们提出了一组复杂的推理规则, 通过演绎过程确定其他词的情感倾向。也就是说, 该算法以已知情感倾向的词(种子)作为输入, 并生成具有情感倾向的同义词集。具有推断出的情感倾向的同义词集可以进一步用于推断其他词的极性。

彭和朴(2011)提出了一种使用约束对称非负矩阵分解(CSNMF)的情感词典生成方法。

该方法首先使用引导方法在字典中找到一组候选情感词, 然后使用大型语料库为每个词分配极性分数。因此, 该方法同时使用了字典和语料库。徐、孟和王(2010)还提出了几种使用字典和语料库寻找情感词的综合方法。他们的方法基于相似性图中的标签传播(Zhu和Ghahramani, 2002)。

总之, 我们注意到使用基于词典的方法的优点是可以轻松快速找到大量带有情感倾向的词语。尽管结果列表可能有很多错误, 但可以进行手动检查来清理它, 这需要时间(对于母语为中文的人来说并不糟糕, 只需要几天时间), 但这只是一次性的工作。主要的缺点是通过这种方式收集到的词语的情感倾向是普遍的或领域和上下文无关的。换句话说, 很难使用基于词典的方法找到情感词语的领域或上下文相关的倾向。正如之前讨论的, 许多情感词语具有上下文相关的特点。

方向。例如，对于一个扬声器电话，如果很安静，通常是负面的。然而，对于一辆汽车来说，如果很安静，是积极的。安静的情感方向是与领域或上下文相关的。下面的基于语料库的方法可以帮助解决这个问题。

6.2 基于语料库的方法

基于语料库的方法已经应用于两个主要场景：（1）从领域语料库中给定一个已知（通常是通用的）情感词汇种子列表，发现其他情感词汇及其情感倾向；（2）使用领域语料库将通用情感词汇适应到新的情感词汇，用于领域中的情感分析应用。然而，问题比仅仅构建一个特定领域的情感词汇更加复杂，因为在同一个领域中，同一个词在一个上下文中可能是积极的，但在另一个上下文中可能是消极的。下面，我们讨论一些已有的研究作品，试图解决这些问题。需要注意的是，虽然基于语料库的方法也可以用于构建通用情感词汇，如果有一个非常大且非常多样的语料库可用，但基于词典的方法通常更加有效，因为词典包含了所有的词汇。

Hazivassiloglou和McKeown（1997）提出了一个关键且早期的想法。作者使用了一个语料库和一些种子形容词情感词汇，在语料库中找到了额外的情感形容词。他们的技术利用了一组关于连接词的语言规则或约定，以识别语料库中更多的形容词情感词汇及其情感倾向。其中一条规则是关于连词“和”的，它指出连接的形容词通常具有相同的情感倾向。例如，在句子“这辆车漂亮且宽敞”中，如果“漂亮”被认为是积极的，可以推断出“宽敞”也是积极的。这是因为人们通常在连词两侧表达相同的情感。

以下句子不太可能，“这辆车很漂亮但很难开。””如果改为“这辆车很漂亮但是开起来很难。”则更可接受。”还与其他连接词设计了规则，例如OR、BUT、EITHER-OR和NEITHER-NOR。这个想法被称为情感一致性。实际上，它并不总是一致的。因此，还应用了学习步骤来确定两个并列形容词是否具有相同或不同的方向。首先，形成了一个图，其中包含形容词之间的相同方向和不同方向的链接。然后，在图上进行聚类，产生两组词：积极和消极。

Kanayama和Nasukawa（2006）通过引入

情感分析与观点挖掘

句内（在一个句子内部）和句间（在相邻句子之间）情感一致性的概念，他们称之为连贯性。句内一致性与上述思想类似。句间一致性只是将这个思想应用于相邻的句子。也就是说，通常连续的句子中表达了相同的情感倾向。

情感变化通过转折表达词如但是和然而来表示。还提出了一些标准来确定是否将一个词加入到积极或消极词汇中。这项研究基于日本的文本，用于寻找领域相关的情感词和它们的倾向性。其他相关工作包括(Kaji and Kitsuregawa, 2006; Kaji and Kitsuregawa, 2007)。

虽然找到领域特定的情感词和它们的倾向性是有用的，但在实践中还不够。Ding, Liu和Yu (2008)表明，在不同的语境中，同一领域的许多词可能具有不同的倾向性。事实上，这种现象已经在第5.2节的基本规则中描述了意见。例如，在相机领域中，词“长”在以下两个句子中明显表达了相反的意见：“电池寿命很长”（积极）和“对焦需要很长时间”（消极）。

这种情况经常发生在量词中，例如，长,短,大,小等等。然而，并不总是如此。例如，在一篇汽车评论中，句子“这辆车非常安静”是积极的，但是句子“车里的音响系统非常安静”是消极的。因此，仅仅找到领域相关的情感词和它们的倾向是不够的。作者发现，方面和表达情感的词语都很重要。然后，他们提出使用(方面,情感词)这对作为观点上下文，例如，(\“电池寿命\”, \“长\”)。他们的方法确定情感词和它们的倾向以及它们修改的方面。在确定一对是积极还是消极时，仍然应用上述句内和句间情感一致性规则。

在(Ganapathibhotla and Liu, 2008)的工作中，采用了相同的上下文定义，但用于分析比较句。Wu and Wen (2010)解决了中文中类似的问题。然而，他们只关注形容词是量词的情况，例如大,小,低和高。他们的方法基于句法模式，如(Turney, 2002)，并且使用网络搜索命中次数来解决问题。Lu et al. (2011)也使用了相同的上下文定义。与(Ding, Liu and Yu, 2008)中的情况类似，他们假设方面集合已知。他们将每对情感词分配为积极或消极的问题，形式化为一个带有多个约束条件的优化问题。目标函数和约束条件基于一般用途的情感词典、每篇评论的整体情感评分、同义词等线索设计。

情感分析与观点挖掘

以及反义词，以及连词“和”规则，“但是”规则和“否定”规则。在某种程度上，(Takamura, Inui and Okumura, 2007; Turney, 2002)中的方法也可以被视为一种隐式方法用于寻找特定上下文观点，但他们没有使用情感一致性的概念。相反，他们使用网络来找到他们的方向。

然而，我们应该注意到，所有这些上下文定义仍然不足以涵盖所有情况，正如第5.2节中讨论的观点的基本规则所示，即，许多上下文可能更加复杂，例如，消耗大量资源。

沿着类似的思路，Wilson, Wiebe, and Hoffmann (2005)研究了短语或表达的上下文主观性和情感。

上下文情感意味着尽管词汇表中的一个词或短语被标记为积极或消极，但在句子表达的上下文中，它可能没有情感或具有相反的情感。在这项工作中，首先在语料库中标记了主观表达，即包含给定主观性词汇的表达。请注意，主观性词汇与情感词汇略有不同，因为主观性词汇可能只包含指示主观性而不包含情感的词，例如感觉和思考。该工作的目标是对包含主观性线索实例的给定表达进行上下文情感分类。该论文采用了有监督学习方法，分为两个步骤。在第一步中，确定表达是否主观或客观。在第二步中，确定主观表达是积极的、消极的、两者都是还是中立的。两者都表示既有积极情感又有消极情感。中立仍然包括在内，因为第一步可能会犯错误，导致一些中立表达未被识别。对于主观性分类，使用了大量丰富的特征，包括词特征、修饰特征(依赖特征)、结构特征(基于依赖树的模式)、句子特征和文档特征。对于情感分类的第二步，使用了诸如词标记、词先前情感、否定、由极性修饰、连词极性等特征。对于这两个步骤，使用了机器学习算法Boosting AdaBoost.HM (Schapire and Singer, 2000)来构建分类器。

在(Choi和Cardie, 2008)中也进行了与表达级情感分类相关的工作，其中作者对在多角度问答(MPQA)语料库中进行了标注的表达式进行了分类(Wiebe, Wilson和Cardie, 2005)。实验了基于词典的分类和监督学习。在(Breck, Choi和Cardie, 2007)中，作者使用条件随机场(CRF) (Lafferty,

情感分析与观点挖掘

McCallum和Pereira, 2001年)。

将通用词典适应到领域特定的新词典中进行情感分类的问题在 (Choi和Cardie, 2009年) 中进行了研究。他们的技术通过利用领域中的表达级别极性来适应通用情感词典的词级极性, 反过来, 适应后的词级极性被用来改进表达级别极性。词级和表达级别的极性关系被建模为一组约束条件, 并使用整数线性规划来解决问题。这项工作假设存在一个给定的通用极性词典 L , 以及一个可以根据 L 中的词来确定观点表达 el 的极性的极性分类算法 $f(el, L)$ 。Jijkoun, Rijke和Weerkamp (2010年) 也提出了一种将通用情感词典适应到特定主题的方法。

Du等人 (2010年) 研究了将情感词典从一个领域 (不是通用的词典) 适应到另一个领域的问题。作为输入, 该算法假设有一组领域内带有情感标签的文档, 一组来自这些领域内文档的情感词汇, 以及一组领域外文档。任务是使领域内的情感词典适应领域外的文档。该研究使用了两个思想。首先, 如果一个文档包含许多正面 (或负面) 词汇, 则该文档应该是积极的 (或消极的), 如果一个词汇出现在许多正面 (或负面) 文档中, 则该词汇应该是积极的 (或消极的)。这些是相互强化的关系。其次, 即使两个领域可能服从不同的分布, 仍然有可能识别出它们之间的共同部分 (例如, 相同的词汇具有相同的倾向)。

情感词典的适应问题是通过信息瓶颈框架来解决的。同样的问题也在 (Du和Tan, 2009年) 中得到了解决。

在一个稍微不同的话题上, Wiebe和Mihalcea (2006) 调查了基于语料库为词义分配主观性标签的可能性。进行了两项研究。第一项研究调查了手动为WordNet词义分配标签 (主观、客观或两者都有) 的标注者之间的一致性。第二项研究评估了一种自动为词义分配主观性标签/分数的方法。该方法基于分布相似性 (Lin, 1998)。他们的工作表明, 主观性是可以与词义相关联的属性, 并且词义消歧可以直接受益于主观性注释。在 (Akkaya, Wiebe和Mihalcea, 2009) 中报告了随后的工作。Su和Markert (2008) 还研究了这个问题, 并进行了主观性识别的案例研究。在 (Su和Markert, 2010) 中, 他们进一步研究了这个问题, 并将其应用于跨语言环境。

Brody和Diakopoulos (2011) 研究了微博中单词的延长（例如，*sloooooo w*）。他们表明，延长与主观性和情感强烈相关，并提出了一种自动利用这种关联性来检测领域情感和情绪词的方法。

最后，Feng、Bose和Choi (2011) 研究了产生内涵词典的问题。内涵词典与情感词典不同，后者涉及明确或隐含地表达情感的词语，而前者涉及与特定情感极性经常相关的词语，例如，*award*和 *promotion*具有积极内涵，*cancer*和 *war*具有消极内涵。提出了一种基于相互强化的基于图的方法来解决这个问题。

6.3 可取和不可取的事实

到目前为止，我们讨论的情感词和表达主要是指表示积极或消极观点的主观词和表达。

然而，正如前面提到的，在某些领域或上下文中，许多客观词和表达也可以暗示观点，因为它们可以代表这些领域或上下文中的可取或不可取的事实。

在 (Zhang和Liu, 2011b) 中，提出了一种方法来识别在特定领域中既是方面又暗示情感的名词和名词短语。这些名词和名词短语本身并不表示情感，但在领域上下文中，它们可能代表可取或不可取的事实。例如，“山谷”和“山”在一般情况下没有任何情感内涵，即它们是客观的。然而，在床垫评论领域中，它们经常暗示负面观点，如“一个月内，在床垫中间形成了一个山谷。”在这里，“山谷”暗示了对床垫质量方面的负面情感。识别这些方面的情感倾向非常具有挑战性，但对于在这些领域中进行有效的情感分析至关重要。

算法(Zhang and Liu, 2011b)基于以下思想:虽然这些句子通常是客观的，没有明确的情感，但在某些情况下，作者/评论者也可能给出明确的情感，例如，

“一个月内，在床垫中间形成了一个凹陷，这太糟糕了。”这句话的上下文表明“凹陷”可能不是令人满意的。请注意，这项工作假设给定了一组被视为名词和名词短语的方面。然而，这种方法的问题是，那些没有暗示情感的方面(名词和名词短语)也可能出现在一些积极或消极的情感背景中，例如，“语音质量”在“语音质量很差”中。为了区分这两种情况，

情感分析与观点挖掘

使用了以下观察结果。

观察: 对于本身没有积极或消极含义的正常方面, 人们可以表达不同的意见, 即既可以是积极的也可以是消极的。例如, 对于“语音质量”这个方面, 人们可以说“声音质量好”和“声音质量差”。然而, 对于代表可取或不可取事实的方面, 它们通常只有一个情感, 要么是积极的, 要么是消极的, 而不是两者兼有。例如, 以下两个句子同时出现的可能性很小: “形成了一个糟糕的凹陷”和“形成了一个好的凹陷”。

考虑到这一观察结果, 该方法包括两个步骤: 1. 候选识别: 该步骤确定每个名词方面的周围情感上下文。如果一个方面在负面 (或正面) 情感上下文中出现的频率明显高于在正面 (或负面) 情感上下文中出现的频率, 则可以推断其极性为负面

(或正面)。因此, 该步骤生成了一个具有正面观点的候选方面列表和一个具有负面观点的候选方面列表。

2. 修剪: 该步骤根据上述观察结果修剪了这两个列表。

这个想法是, 当一个名词方面被正面和负面情感词直接修饰时, 它不太可能是一个有观点的方面。

使用了两种类型的直接依赖关系。

类型1: $O \rightarrow O-Dep \rightarrow F$

这意味着 O 通过关系 $O-Dep$ 依赖于 F , 例如, “这台电视有很好的画质。”类型2:

$O \rightarrow O-Dep \rightarrow H \leftarrow F-Dep \leftarrow F$

这意味着 O 和 F 都依赖于 H 通过关系 $O-Dep$ 和 $F-Dep$ 分别, 例如, “床垫的弹簧不好。”

其中 O 是情感词, $O-Dep / F-Dep$ 是依赖关系。 F

是名词方面。 H 表示任何词。 对于第一个例子, 给定方面

“画质”, 我们可以确定其修饰情感词

“好。”对于第二个例子, 给定方面“弹簧”, 我们可以得到其

修饰情感词“坏。”这里 H 是词“are”。

这项工作只是解决问题的第一次尝试。 它的准确性仍然不高。 还需要进行更深入的研究。

6.4 摘要

由于许多研究人员的贡献, 已经构建了几个通用的主观性、情感和情绪词典, 其中一些也是公开可用的, 例如,

情感分析与观点挖掘

- General Inquirer词典 (Stone, 1968) :
(http://www.wjh.harvard.edu/~inquirer/spreadsheet_guide.htm)
- 情感词典 (Hu和Liu, 2004) :
(<http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>)
- MPQA主观性词典 (Wilson, Wiebe和Hoffmann, 2005) :
(http://www.cs.pitt.edu/mpqa/subj_lexicon.html)
- SentiWordNet (Esuli和Sebastiani, 2006) :
(<http://sentiwordnet.isti.cnr.it/>)
- 情绪词典 (Mohammad和Turney, 2010) :
(<http://www.purl.org/net/emolex>)

然而，即使有这么多的研究，领域和上下文相关的情感仍然非常具有挑战性。最近的研究还使用了词向量和矩阵来捕捉情感词的上下文信息 (Maaset al., 2011; Yessenalina和Cardie, 2011)。事实性词语和表达暗示观点的词语几乎没有被研究过 (参见第6.3节)，但它们对许多领域非常重要。

最后，我们注意到拥有情感词典 (即使具有特定领域的倾向性) 并不意味着词典中的词在特定句子中总是表达观点/情感。例如，在“我正在寻找一辆好车买”中，“好”并不表达对任何特定车辆的积极或消极观点。

第7章

观点总结

如第2章所讨论的，在大多数情感分析应用中，由于观点的主观性，需要研究来自许多人的观点，因为仅仅看一个人的观点通常是不够的。需要某种形式的总结。第2章指出，观点五元组为观点总结提供了基本信息。这样的总结被称为基于方面的总结（或特征-基于总结），并在（Hu and Liu, 2004; Liu, Hu and Cheng, 2005）中提出。大部分观点总结研究使用相关的思想。这个框架在工业界也被广泛应用。例如，微软必应和谷歌产品搜索的情感分析系统使用这种形式的总结。输出的总结可以是结构化形式（参见第7.1节）或作为短文本文档的非结构化形式。

一般来说，观点总结可以被看作是一种多文档文本总结的形式。文本总结在自然语言处理领域得到了广泛研究（Das, 2007）。然而，观点总结与传统的单一文档或多文档总结（关于事实信息）有很大的不同，因为观点总结通常集中在实体、方面和对它们的情感，并且还具有定量的一面，这是基于方面的观点总结的本质。传统的单一文档总结通过提取一些“重要”的句子，从长文本中生成一个简短的文本。传统的多文档总结找出文档之间的差异并丢弃重复的信息。它们都没有明确地捕捉到文档中讨论的不同主题/实体及其方面，也没有定量的一面。在传统文本总结中，句子的“重要性”通常是每个系统中使用的总结算法和度量来进行操作定义的。

另一方面，观点总结可以在概念上进行定义。因此，摘要是有结构的。即使对于短文本文档的输出摘要，它们仍然具有一些明确的结构。

7.1 基于方面的观点摘要

情感分析与观点挖掘

基于方面的观点摘要具有两个主要特点。首先，它捕捉了观点的本质：观点目标（实体及其方面）和对它们的情感。其次，它是定量的，这意味着它给出了持有正面或负面观点的人数或百分比。定量方面对于观点的主观性质至关重要。由第2.1节中的观点五元组生成的观点摘要是一种结构化摘要形式。我们在第2.2节中描述了该摘要。为了完整起见，这里再次呈现。图7.1显示了关于数码相机的基于方面的摘要（Hu和Liu，2004）。方面GENERAL代表了关于相机整体（即实体）的观点。对于每个方面（例如，画质），它显示了持有正面和负面观点的人数。〈个别评论句子〉链接到实际句子（或完整评论或博客）。这种结构化摘要也可以进行可视化（Liu，Hu和Cheng，2005）。图7.2（A）使用条形图来可视化图7.1中的摘要。在图中，X轴上方的每个条形图显示了顶部给出的方面的正面观点数量。X轴下方对应的条形图显示了相同方面的负面观点数量。点击每个条形图，我们可以看到个别句子和完整评论。显然，还有其他可视化方式。

例如，Microsoft Bing搜索和Google产品搜索的条形图使用每个方面上积极意见的百分比。

比较几个实体的观点总结更加有趣（Liu，Hu和Cheng，2005）。图7.2（B）显示了两个相机的视觉观点比较。我们可以看到消费者如何在不同的方面维度上看待它们。

事实上，观点五元组可以提供更多形式的结构化摘要。例如，如果提取时间，可以显示不同方面观点的趋势。即使不使用情感，也可以看到每个方面提及的热度（频率），这给用户提供了人们最关注的方面的想法。事实上，使用五元组，可以使用全面的数据库和OLAP工具对数据进行各种定性和定量分析。例如，在汽车领域的一个实际情感分析应用中，首先挖掘了个别汽车的观点五元组。然后，用户比较了关于小型汽车、中型汽车、德国汽车和日本汽车等的情感。此外，情感分析结果还被用作数据挖掘的原始数据。用户运行了一个聚类算法，并找到了一些有趣的市场细分。例如，发现有一部分顾客总是谈论汽车的外观漂亮和流畅驾驶等，而另一部分顾客则更多地谈论后座和行李箱空间等。显然，

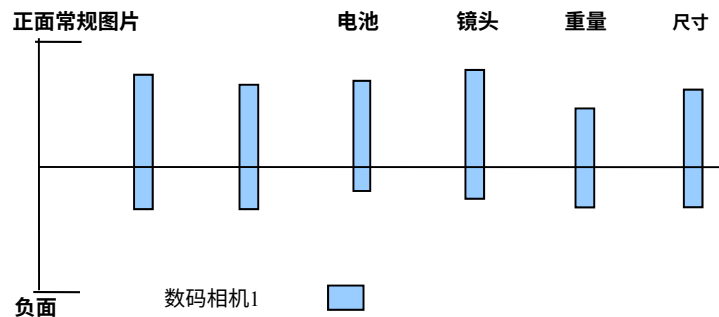
数码相机 1:

方面: 常规		
积极:	105	<个别评论句子>
负面:	12	<个别评论句子>
方面: 画质		
积极:	95	<个别评论句子>
负面:	10	<个别评论句子>
方面: 电池寿命		
积极:	50	<个别评论句子>
负面:	9	<个别评论句子>

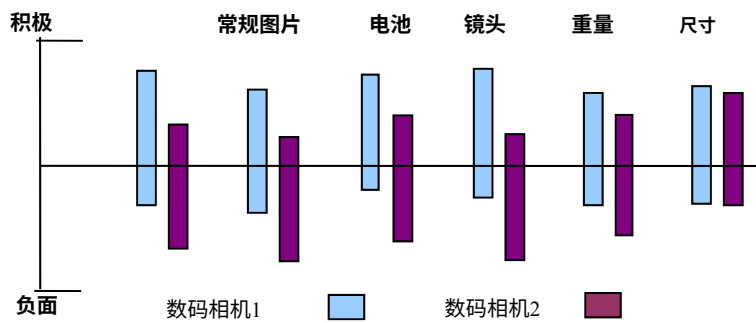
...

图7.1.基于方面的观点摘要。

第一部分主要由年轻人组成，而第二部分主要由有家庭和孩子的人组成。这些见解非常重要。它们使用户能够看到不同客户群体的观点。



(A) 数码相机的基于方面的观点摘要的可视化



(B) 两台数码相机的观点比较的可视化

图7.2.基于方面的观点摘要的可视化

其他研究人员也采用了这种结构化摘要的形式，用于总结电影评论（Zhuang, Jing和Zhu, 2006），总结中文观点文本（Ku, Liang和Chen, 2006），以及总结服务

评论 (Blair-Goldensohn等, 2008年)。然而, 我们应该注意到基于方面的摘要不一定要以这种结构化形式呈现。它也可以以基于相同思想的文本文档的形式呈现。在接下来的部分, 我们讨论其他相关研究。

7.2 基于方面的 观点摘要的改进

研究人员已经提出了几种改进和完善基本的基于方面的摘要。Carenini, Ng和Pauls (2006年) 提出将基于方面的摘要与两种传统的文本摘要方法相结合, 即句子选择 (或提取) 和句子生成。我们首先讨论与句子选择方法的集成。他们的系统首先使用 (Hu和Liu, 2004年) 中的方法从特定实体 (例如产品) 的评论中识别出方面表达式。然后, 它将方面表达式映射到作为实体的本体树组织的一些给定方面类别。然后, 根据它们的情感强度对树中的这些方面进行评分。还提取包含方面表达式的句子。然后, 根据句子中方面的得分对每个这样的句子进行评级。如果多个句子具有相同的句子评级, 则使用传统的基于质心的句子选择方法来打破平局 (Radev等, 2003年)。所有相关句子都附加到本体中对应的方面上。然后根据句子得分和方面在本体树中的位置选择每个方面的句子作为最终摘要。与句子生成方法的集成类似。首先, 根据本体中的出现频率、情感强度和位置, 使用一种度量方法对本体中的方面进行评分。还应用了一种算法来选择本体树中的方面。

然后计算出方面的正面和负面情感。基于选择的方面及其情感, 语言生成器生成可以定性和定量的摘要句子。进行了用户评估以评估两种集成方法的有效性。结果显示它们的表现相当, 但原因不同。句子选择方法提供了更多样化的语言和更多细节, 而句子生成方法则提供了对评论的更好情感概述。

在(Tata and Di Eugenio, 2010)中, Tata和Eugenio类似于(Hu and Liu, 2004), 为每个方面和每个情感 (正面或负面) 首先选择了一个观点摘要。

情感分析与观点挖掘

代表性句子的群组。该句子应该提及最少的方面（因此代表性句子是专注的）。然后，他们使用给定的领域本体映射句子到本体节点来对句子进行排序。本体基本上编码了关键领域概念及其关系。句子按照树的顺序进行排序和组织，以便以概念上连贯的方式出现在段落中。

Lu等人（2010）还使用在线实体和方面本体来组织和总结意见。他们的方法与上述两种方法相关，但也有所不同。他们的系统首先选择捕捉主要观点的方面。选择是通过频率、观点覆盖（无冗余）或条件熵来完成的。然后，根据一种连贯性度量对方面及其对应的句子进行排序，该度量试图优化排序，使其最好地遵循原始帖子中方面出现的顺序。

Ku、Liang和Chen（2006）进行了博客观点摘要，并生成了两种类型的摘要：简要摘要和详细摘要，基于提取的主题（方面）和对主题的情感。对于简要摘要，他们的方法选择具有最多正面或负面句子的文档/文章，并使用其标题来代表正面主题或负面主题句子的整体摘要。对于详细摘要，它列出了情感度较高的正面主题和负面主题句子。

Lerman, Blair-Goldensohn和McDonald（2009）以稍微不同的方式定义了观点总结。给定一组包含关于某个实体的观点的文档 D （例如，评论），观点总结系统的目标是生成一个代表平均观点并涉及其重要方面的摘要 S 。

本文提出了三种不同的模型来对产品的评论进行总结。所有这些模型都从评论中选择一些句子。第一个模型称为情感匹配（SM），它提取句子，使得摘要的平均情感尽可能接近实体评论的平均情感评分。第二个模型称为情感匹配+方面覆盖（SMAC），它构建一个在最大程度上涵盖重要方面并与实体的整体情感匹配的摘要。第三个模型称为情感-方面匹配（SAM），它不仅试图涵盖重要方面，还试图用适当的情感来涵盖它们。进行了对三种类型摘要的人类用户的全面评估。发现尽管SAM模型是最好的，但与其他模型相比并没有显著优势。

在 (Nishikawa等人, 2010b) 中, 提出了一种更复杂的总结技术, 通过选择和排序来自多个评论的句子, 考虑最终总结的信息量和可读性, 生成传统的文本总结。信息量被定义为每个方面-情感对的频率之和。

可读性被定义为句子的自然顺序, 通过测量序列中所有相邻句子的连通性之和来衡量。然后通过优化来解决这个问题。在 (Nishikawa等人, 2010a) 中, 作者进一步使用整数线性规划的形式研究了这个问题。在 (Ganesan, Zhai和Han, 2010) 中, 使用基于图模型的方法生成了一个抽象的意见摘要。在 (Yatani等人, 2011) 中, 形容词-名词对被提取为一个摘要。

7.3 对比视角总结

一些研究人员还通过找到对比观点来研究总结意见的问题。例如, 一个评论者可能会对iPhone的语音质量给出积极的意见, 说“iPhone的语音质量真的很好”, 但另一个评论者可能会说相反的话, “我的iPhone的语音质量很差。”这样的对比可以让读者直接比较不同的意见。

Kim和Zhai (2009) 提出并研究了这个问题。给定一个积极的句子集和一个消极的句子集, 这项工作通过从这些集合中提取一组 k 对比句子对来进行对比观点总结。如果句子 x 和句子 y 关于同一主题方面, 但情感取向相反, 那么一对有观点的句子 (x, y) 被称为对比句子对。所选择的 k 句子对还必须很好地代表积极和消极的句子集。作者将总结问题形式化为一个优化问题, 并基于几个相似性函数来解决它。

Paul、Zhai和Girju (2010) 也在研究这个问题。他们的算法生成了一个宏观多视角摘要和一个微观多视角摘要。宏观多视角摘要包含多个句子集, 每个集合代表一个不同的观点。微观多视角摘要包含一组对比句子对 (每对由两个代表不同观点的句子组成)。该算法分为两个步骤。在第一步中, 它使用主题建模方法对主题 (方面) 和情感进行建模和挖掘。在第二步中, 提出了一种随机游走公式 (类似于PageRank (Page et al., 1999)) 来评分

基于它们的代表性和相互对比性，从相反的观点中提取句子和句子对。沿着类似的思路，Park、Lee和Song（2011）报道了另一种在新闻文章中生成对立观点的方法。

在（Lerman和McDonald，2009）中，Lerman和McDonald提出了一个不同的对比总结问题。他们希望产生关于两个不同产品的观点的对比摘要，以突出它们之间的观点差异。他们的方法是联合建模这两个总结任务，并在优化中明确考虑到它希望这两个摘要形成对比的事实。

7.4 传统总结

一些研究人员还以传统方式研究了观点总结，例如，生成一个简短的文本摘要，对其进行有限或不考虑方面（或主题）和情感。在（Beineke等，2003）中提出了一种监督学习方法来选择评论中的重要句子。在（Seki等，2006）中提出了一种段落聚类算法，也选择了一组重要句子。

在（Wang和Liu，2011）中，作者研究了对话中观点的抽取式总结（选择重要句子）。他们尝试了传统的句子排序和基于图的方法，还考虑了主题相关性、情感和对话结构等其他特征。

这种传统摘要的一个弱点是它们只有有限或没有考虑目标实体和方面以及对它们的情感。

因此，它们可能选择与情感或任何方面无关的句子。另一个问题是缺乏定量的观点，这在实践中通常很重要，因为十个人中有一个人讨厌某件事与十个人中有五个人讨厌某件事是非常不同的。

7.5 摘要

观点总结仍然是一个活跃的研究领域。大多数观点总结方法产生的简短文本摘要并没有关注积极和消极

观点的比例。未来的研究可以解决这个问题，同时也能产生可读性强的文本。我们应该注意到观点总结

情感分析与观点挖掘

研究不能独立进行，因为它在很大程度上依赖于情感分析、方面或主题提取和情感分类等其他研究领域的结果和技术。所有这些研究方向都需要相互配合。最后，我们还应该注意，根据第7.1节的结构化摘要，可以使用一些预定义的句子模板基于柱状图中显示的内容生成自然语言句子。例如，图7.2(B)中的第一根柱状图可以总结为“70%的人对数字相机1总体上持积极态度。”然而，这可能不是最适合人们阅读愉悦的句子。

第8章

比较观点分析

除了直接表达对实体及其方面的积极或消极意见外，人们还可以通过比较类似的实体来表达观点。

这种观点被称为比较观点（Jindal和Liu，2006a；Jindal和Liu，2006b）。比较观点与常规观点相关，但也有所不同。它们不仅具有不同的语义含义，还具有不同的句法形式。例如，一个典型的常规观点句子是“这部手机的音质很棒”，而一个典型的比较观点句子是“诺基亚手机的音质比苹果手机的音质更好”。这个比较句子并没有说任何手机的音质是好还是坏，只是简单地进行了比较。由于这种差异，比较观点需要不同的分析技术。

与常规句子一样，比较句子可以有观点或无观点。上面的比较句子是有观点的，因为它明确地表达了作者的比较情感，而句子

“iPhone比普通诺基亚手机宽1英寸”没有表达情感。

在本章中，我们首先定义问题，然后介绍一些现有的解决方法。我们还应该注意到，实际上有两种基于比较的主要观点类型：比较观点和最高级观点。在英语中，它们通常使用形容词或副词的比较或最高级形式来表达，但并不总是如此。

然而，在本章中，我们将它们一起研究，并且只是通称为一般的比较观点，因为它们的语义意义和处理方法是相似的。

8.1 问题定义

比较句表达了基于多个实体的相似性或差异性的关系。有几种类型的比较。

它们可以分为两个主要类别：可分级比较和不可分级比较（Jindal和Liu，2006a；Kennedy，2005）。

可分级比较：这种比较表达了被比较实体之间的排序关系。它有三个子类型：非相等可分级比较：它表达了一种类型的关系

情感分析与观点挖掘

大于或小于根据一些共同的方面对一组实体进行排名，而不是另一组实体例如，“可口可乐的味道比百事可乐好。”这种类型还包括偏好，例如，“我更喜欢可口可乐而不是百事可乐。”

2. 等同比较：它表示一种关系，即基于一些共同的方面，两个或多个实体是相等的例如，“可口可乐和百事可乐的味道相同。”
3. 最高级比较：它表示一种关系，即一个实体在所有其他实体中排名更高或更低例如，“可口可乐在所有软饮料中味道最好。”

不可分级比较：这种比较表示两个或多个实体之间的关系，但不对其进行分级。有三个主要的子类型：₁ 实体 A 与实体 B 相似或不同，基于一些共同的方面例如，“可口可乐的味道与百事可乐不同。”₂ 实体 A 具有方面 a_1 ，而实体 B 具有方面 a_2 （通常可以互换），例如，“台式电脑使用外部扬声器，而笔记本电脑使用内部扬声器。”

3. 实体 A 具有方面 a ，但实体 B 没有，例如，“诺基亚手机配有耳机，而 iPhone 则没有。”

在本章中，我们只关注可分级比较。非分级比较也可以表达观点，但它们通常更加微妙和难以识别。

在英语中，比较通常使用比较级词（也称为比较级）和最高级词（也称为最高级）来表达。

比较级通过在其基本形容词和副词后面添加后缀 *-er* 形成，而最高级通过在其基本形容词和副词后面添加后缀 *-est* 形成。例如，在“诺基亚手机的电池寿命比摩托罗拉手机长”中，“longer”是形容词“long”的比较级形式。“longer”（和“than”）在这里也表示这是一个比较句。在“诺基亚手机的电池寿命是最长的”中，“longest”是形容词“long”的最高级形式，并且它表示这是一个最高级句子。

我们称这种比较级和最高级为类型 1 的比较级和最高级。请注意，为了简单起见，我们经常使用 比较级 来表示比较级和最高级，如果没有明确指定最高级。

然而，两个或更多音节且不以 *y* 结尾的形容词和副词不通过添加 *-er* 或 *-est* 来形成比较级或最高级。相反地，更、最、更少和最少在这类词之前使用，例如，更美丽。

我们称这种比较级和最高级为类型 2 的比较级和最高级。类型 1 和类型 2 都被称为常规比较级和最高级。

情感分析与观点挖掘

英语还有不规则比较级和最高级，即更、最、更少、最少、更好、最好、更坏、最坏、更远和最远，它们不遵循上述规则。然而，它们的行为与类型1的比较级类似，因此被归为类型1。

这些标准的比较级和最高级只是一些表明比较的词语，实际上，还有许多其他词语和短语可以用来表达比较，例如，更喜欢和更优越。例如，句子“iPhone的语音质量优于Blackberry”表示iPhone的语音质量更好，更受欢迎。在（Jindal和Liu，2006a）中，编制了这样的词语和短语列表

（这绝不是完整的）。由于这些词语和短语通常与类型1的比较级类似，它们也被归为类型1。所有这些词语和短语以及上述标准的比较级和最高级被统称为比较关键词。

在非等级可比较的比较中使用的比较关键词可以根据它们是否表示增加或减少的数量进一步分为两类，这对情感分析很有用。

- 增加比较：这种比较表示数量增加，例如，更多和更长。
- 减少比较：这种比较表示数量减少，例如，更少和更少。

挖掘比较观点的目标（Jindal和Liu，2006b；Liu，2010）：给定一个观点文档 d ，在 d 中发现所有形式为比较观点六元组的内容：

$$(E_1, E_2, A, PE, h, t),$$

其中 E_1 和 E_2 是基于它们共享的方面 A 进行比较的实体集合（在句子中， E_1 中的实体出现在 E_2 中的实体之前）， $PE (\in \{E_1, E_2\})$ 是观点持有者的首选实体集合 h ， t 是表达比较观点的时间。对于最高级比较，如果一个实体集合是隐含的（在文本中没有给出），我们可以使用一个特殊集合 U 来表示它。对于等式比较，我们可以使用特殊符号 EQUAL 作为 PE 的值。

例如，考虑由Jim于2011年9月25日写的比较句“佳能的画质比LG和索尼的好。”提取的比较观点是：

$$(\{佳能\}, \{LG, 索尼\}, \{画质\}, \{佳能\}, Jim, 2011年9月25日)$$

实体集合 E_1 是 $\{佳能\}$ ，实体集合 E_2 是 $\{LG, 索尼\}$ ，它们共同比较的方面集合 A 是 $\{画质\}$ ，首选实体集合是

{佳能}, 观点持有者 h 是 Jim, 这个比较观点的写作时间 t 是 2011 年 9 月 25 日。

请注意, 由于使用了集合, 上述表示可能不容易放入数据库中, 但可以轻松转换为没有集合的多个元组, 例如, 上述基于集合的六元组可以扩展为两个元组: (佳能, LG, 画质, 佳能, Jim, 2010 年 12 月 25 日)

(佳能, 索尼, 图片质量, 佳能, 吉姆, 2010 年 12 月 25 日) 与挖掘常规

观点类似, 挖掘比较观点需要提取实体、方面、观点持有者和时间。使用的技术也是类似的。实际上, 这些任务通常对于比较句子来说更容易, 因为实体通常位于比较关键词的两侧, 而方面也靠近。然而, 为了识别首选实体集的情感分析, 需要使用不同的方法, 我们将在第 8.3 节中讨论。我们还需要识别比较句子本身, 因为并不是所有包含比较关键词的句子都表达了比较, 而且许多比较关键词和短语很难识别 (Jindal 和 Liu, 2006b)。在下面, 我们只关注研究两个比较观点情感分析的特定问题, 即识别比较句子和确定首选实体集。

8.2 识别比较句子

尽管大多数比较句子包含比较和最高级关键词, 例如更好、更优和最好, 但许多包含这些词的句子并不是比较句子, 例如“我再同意你更多也不行。”在 (Jindal 和 Liu, 2006a) 中, 他们证明几乎每个比较句子都有一个表示比较的关键词 (一个词或短语)。使用一组关键词, 根据他们的数据集, 可以识别出 98% 的比较句子 (召回率=98%), 精确度为 32%。关键词包括: 1. 比较形容词 (JJR) 和比较副词 (RBR), 例如更、更少、更好和以 *-er* 结尾的词。这些只被计算为两个关键词。

2. 最高级形容词 (JJS) 和最高级副词 (RBS), 例如, 最, 最少, 最好, 以及以 *-est* 结尾的词。这些也被计算为只有两个关键词。
3. 其他非标准的指示性词和短语, 例如偏爱, 击败, 胜过, 超过, 表现优异, 更喜欢, 领先, 比, 更优, 更劣, 第一名, 面对等。这些词在关键词的数量中被单独计算。

由于仅凭关键词就能实现较高的召回率，它们可以用来过滤掉那些不太可能是比较句的句子。我们只需要提高剩下句子的准确性。

在（Jindal和Liu，2006a）中还观察到比较句具有涉及比较关键词的强烈模式，这并不令人意外。这些模式可以用作学习中的特征。为了发现这些模式，在（Jindal和Liu，2006a）中采用了类顺序规则（CSR）挖掘。类顺序规则挖掘是一种特殊的顺序模式挖掘（Liu，2006和2011）。每个训练示例是一对 (s_i, y_i) ，其中 s_i 是一个序列， y_i 是一个类标签，即 $y_i \in \{\text{比较}, \text{非-比较}\}$ 。该序列是从一个句子生成的。使用训练数据，可以生成CSR_s。

对于分类模型构建，使用具有高条件概率的CSR规则的左侧序列模式作为特征。使用朴素贝叶斯进行模型构建。在（Yang和Ko，2011）中，研究了相同的问题，但是在韩语环境中。使用的学习算法是基于转换的学习，它生成规则。

将比较句分类为四种类型：在识别出比较句之后，算法还将其分类为四种类型，即非等级可分，等价，最高级和不可分。对于这个任务，（Jindal和Liu，2006a）表明关键词和关键短语作为特征已经足够。支持向量机给出了最好的结果。

Li等人（2010）研究了识别比较问题和被比较的实体（他们称之为比较器）的问题。与上述作品不同，本文没有确定比较的类型。对于比较句的识别，他们还使用了顺序模式/规则。然而，他们的模式是不同的。他们同时确定一个问题是否是比较问题以及被比较的实体。例如，问题句“哪个城市更好，纽约还是芝加哥？”满足顺序模式 $\langle \text{which } NV \text{ is better, } \$C \text{ or } \$C ? \rangle$ ，其中 $\$C$ 是一个实体。使用基于（Ravichandran和Hovy，2002）中思想的弱监督学习方法来学习这些模式。该算法基于引导，从用户给定的模式开始。从这个模式中，算法提取一组初始种子实体（比较器）对。对于每个实体对，从问题集合中检索包含该对的所有问题，并将其视为比较问题。从比较问题和实体对中，学习和评估所有可能的顺序模式。学习过程是传统的

泛化和特化过程。句子中与\$C匹配的任何单词或短语都是实体。
(Jindal和Liu, 2006b) 和 (Yang和Ko, 2011) 也提取了比较的实体。我们将在第8.4节中讨论它们。其他信息提取算法也适用于这里。

8.3 识别首选实体

与常规意见不同，对于比较意见句子来说，将整个句子进行情感分类并没有太多意义，因为这样的句子并不表达直接的积极或消极意见。相反，它通过基于共享方面对实体进行排名来比较多个实体，从而给出了比较意见。也就是说，它使用比较来表达实体的优先顺序。由于大多数比较句子比较两组实体，对于一个带有观点的比较句子的分析意味着识别首选的实体集。然而，为了应用目的，可以将正面意见分配给首选集中实体的方面，并将负面意见分配给非首选集中实体的方面。请注意，与常规句子一样，对于一个比较句子是否表达观点仍然有意义，但是对于这种分类的研究很少。下面我们只描述一种识别首选实体集的方法。

该方法在 (Ding, Liu和Zhang, 2009) 和 (Ganapathibhotla和Liu, 2008) 中提出，基本上将基于词典的方法扩展到常规观点的基于方面的情感分类到比较观点。因此，它需要一个用于比较观点的情感词典。与基本类型的观点词类似，我们可以将比较观点词分为两类：

1. 通用比较情感词：对于类型1的比较，该类别包括诸如更好，更差等词语，这些词通常具有与领域无关的积极或消极情感。在涉及此类词语的句子中，通常很容易确定首选的实体集。对于由形容词/副词之前添加更多，更少，最或最少形成的类型2的比较，首选的实体集由这两个词确定。应用以下规则：

比较负面 ::= 增加比较 N
| 减少比较 P
比较积极 ::= 增加比较 P
| 减少比较 N

情感分析与观点挖掘

这里，P（分别为N）表示基本类型的积极（消极）情感词或短语。上述第一条规则表示增加比较（例如，更多）和消极情感词（例如，糟糕）的组合意味着消极的比较观点（在左边）。其他规则具有类似的含义。请注意，上述四条规则已在第5.2节中作为观点的基本规则进行了讨论。2. 上下文相关的比较情感词：在类型1比较中，这些词包括更高、更低等。例如，

“诺基亚手机的电池寿命比摩托罗拉手机长”对“诺基亚手机”有一个比较积极的情感，对“摩托罗拉手机”有一个比较消极的情感，即“诺基亚手机”在电池寿命方面更受青睐。然而，如果没有领域知识，很难知道“更长”对于电池寿命是积极的还是消极的。这个问题与常规观点相同，也已包含在第5.2节的基本规则中。在这里，“电池寿命”是一个积极的潜在项目（PPI）。

对于第二类比较，情况类似。然而，在这种情况下，比较词（更，最，较少或至少），形容词/副词和语境都对确定偏好很重要。如果我们知道比较词是增加还是减少的比较（这很容易，因为只有四个），则可以通过应用（1）中的四个规则来确定观点。

如第6.2节所讨论的，（语境，情感词）对形成观点语境很重要。为了确定一对是积极还是消极的，（Ganapathibhotla和Liu，2008）的算法使用了大量的外部数据。它使用了大量的产品评论中的正面和负面观点。这个想法是确定语境和情感词在正面评论中更相关还是在负面评论中更相关。如果它们在正面评论中更相关，则情感词很可能是积极的。否则，它很可能是消极的。然而，由于正面和负面评论很少使用比较观点，所以比较句子中的语境观点词必须转换为其基本形式，可以使用WordNet和英语比较形成规则来完成这一转换。这种转换很有用，因为有以下观察结果。

观察：如果一个形容词或副词的基本形式是积极的（或消极的），那么它的比较级或最高级形式也是积极的（或消极的），例如，好，更好和最好。

在转换后，这些词被手动分类为增加和减少的比较级。对于上下文相关的观点

词，比较级词也可以转换为它们的基本形式。

在确定情感词和它们的倾向性之后，确定哪个实体集更受青睐是相当简单的。没有否定词，如果比较级是积极的（或消极的），那么比 *than* 之前（或之后）的实体更受青睐。否则，比 *than* 之后（或之前）的实体更受青睐。更多详细信息可以在（Ding, Liu和Zhang, 2009; Ganapathibhotla和Liu, 2008）中找到。

8.4 摘要

尽管已经有一些现有的工作，但比较句子的研究还没有像情感分析的许多其他主题那样广泛进行。

仍然需要进一步的研究。其中一个困难的问题是如何识别许多类型的非标准或隐含的比较句子，例如，

“我非常高兴我的*iPhone*和我旧的丑陋的*Droid*完全不同。”

如果不识别它们，进一步的情感分析就很难进行。

除了识别比较句子及其类型外，一些研究人员还研究了比较实体、比较方面和比较词的提取。Jindal和Liu（2006b）使用了标签顺序规则挖掘，这是一种基于顺序模式的监督学习方法。Yang和Ko（2011）应用了最大熵和SVM学习算法来提取比较实体和比较谓词，即被比较的方面。如第8.2节所述，（Li等人，2010）中的顺序模式已经可以用于识别比较问题中的比较实体。然而，他们的工作在某种程度上受限于仅适用于简单的比较问题。在（Fiszman等人，2007）中，作者研究了在生物医学文本中识别哪个实体在比较句子中具有更多某些方面的问题，但他们没有分析比较中的观点。

第9章

观点搜索和检索

由于网络搜索已被证明是一个有价值的网络服务，因此很容易想象观点搜索也将非常有用。两种典型的观点搜索查询包括：

1. 查找关于特定实体或实体的某个方面的公众观点，例如查找关于数码相机或相机画质的客户观点，以及查找关于政治问题或候选人的公众观点。
2. 查找某个人或组织（即观点持有者）对特定实体或实体（或主题）的观点，例如查找巴拉克·奥巴马对堕胎的观点。这种类型的搜索特别适用于新闻文章，其中明确说明了表达观点的个人或组织。

对于第一种类型的查询，用户可以简单地提供实体的名称或方面的名称以及实体的名称。对于第二种类型的查询，用户可以提供观点持有者的名称和实体或主题的名称。

9.1 网络搜索与观点搜索

与传统的网络搜索类似，观点搜索也有两个主要任务：1) 检索与用户查询相关的文档/句子；2) 对检索到的文档或句子进行排序。然而，也存在一些主要的区别。

在检索方面，观点搜索需要执行两个子任务：

1. 找到与查询相关的文档或句子。这是传统网络搜索或检索中唯一执行的任务。
2. 确定文档或句子是否对查询主题（实体和/或方面）表达了观点，以及观点是积极的还是消极的。这是情感分析的任务。传统搜索不执行这个子任务。

至于排序，传统的网络搜索引擎根据权威性和相关性得分对网页进行排序（Liu, 2006和2011）。基本前提是排名靠前的页面（理想情况下是第一页）包含足够的信息来满足用户的信息需求。这种范式是足够的

对于事实信息搜索，一个事实等于任意数量的相同事实。也就是说，如果第一页包含所需信息，则无需查看其他相关页面。对于观点搜索，这种范式只适用于第二类查询，因为观点持有者通常只对特定实体或主题有一个观点，并且该观点包含在单个文档或页面中。然而，对于第一类观点查询，这种范式需要进行修改，因为观点搜索的排名有两个目标。首先，它需要将那些具有高效用或信息内容的观点文档或句子排在前面（参见第11章）。其次，它需要反映积极和消极观点的自然分布。这第二个目标很重要，因为在大多数应用中，积极和消极观点的实际比例是关键信息。仅仅阅读排名靠前的结果，就像传统搜索一样，是有问题的，因为排名靠前的结果只代表一个观点持有者的观点。因此，在观点搜索中进行排名需要捕捉整个人群的积极和消极情感的自然分布。解决这个问题的一个简单方法是产生两个排名，一个用于积极观点，一个用于消极观点，并显示积极和消极观点的数量。

为每个观点搜索提供基于方面的摘要将会更好。然而，这是一个非常具有挑战性的问题，因为方面提取、方面分类和将实体与其方面关联起来都是非常具有挑战性的问题。如果没有对它们的有效解决方案，这样的摘要将不可能实现。

9.2 现有的观点检索技术

目前在观点检索方面的研究通常将任务视为一个两阶段的过程。在第一阶段，文档仅按主题相关性进行排名。

在第二阶段，候选相关文档根据其观点得分进行重新排名。观点得分可以通过机器学习的情感分类器（如SVM）或基于词典的情感分类器（使用情感词典和查询词-情感词接近度得分的组合）来获得。

更先进的研究同时考虑主题相关性和观点，并根据它们的综合得分进行排名。

为了给出观点搜索的味道，我们介绍一个示例系统（Zhang和Yu，2007），该系统是2007年TREC博客赛道的获胜者。

情感分析与观点挖掘

评估 (<http://trec.nist.gov/>)。任务正是观点搜索（或检索）。该系统有两个组成部分。第一个组件用于为每个查询检索相关文档。第二个组件用于将检索到的文档分类为有观点或无观点。有观点的文档进一步分为积极、消极或混合（包含积极和消极观点）。

检索组件：该组件执行传统的信息检索（IR）任务。它考虑关键词和概念。概念是命名实体（例如人名或组织名）或来自字典和其他来源的各种类型的短语（例如维基百科条目）。处理用户查询的策略如下（Zhang等人，2008年；Zhang和Yu，2007年）：首先识别和消除用户查询中的概念。然后使用其同义词扩展搜索查询。之后，它识别检索到的文档中的概念，并执行伪-反馈以自动从排名靠前的文档中提取相关词汇来扩展查询。最后，它使用概念和关键词计算每个文档与扩展查询的相似度（或相关度得分）。

观点分类组件：该组件执行两个任务：（1）将每个文档分类为两个类别之一，即有观点和无观点；（2）将每个有观点的文档分类为表达积极观点、消极观点或混合观点。对于这两个任务，系统使用有监督学习。对于第一个任务，它从评价网站（如`rateitall.com`和`epinions.com`）获取大量有观点（主观）的训练数据。这些数据还来自涉及消费品和服务以及政府政策和政治观点的不同领域。无观点的训练数据来自提供客观信息的网站，如维基百科。通过这些训练数据，构建了一个SVM分类器。

然后将该分类器应用于每个检索到的文档，具体步骤如下。首先将文档分割成句子。然后，SVM分类器将每个句子分类为有观点或无观点。如果一个句子被分类为有观点，它的强度也会被SVM确定并记录。如果至少有一个句子被分类为有观点，则认为该文档是有观点的。为了确保句子的观点是针对查询主题的，系统要求在其附近找到足够的查询概念/词语。文档中所有有观点的句子及其强度以及文档与查询的相似度被用来对文档进行排名。

情感分析与观点挖掘

为了确定一个有观点的文档表达了积极的、消极的还是混合的观点，构建了第二个分类器。训练数据是来自评论网站的评论评分（例如，rateitall.com）。低评分表示消极观点，而高评分表示积极观点。使用积极和消极的评论作为训练数据，构建了一个情感分类器，将每个文档分类为表达积极、消极或混合的观点。

在TREC评估中，还有其它方法来进行观点检索。

鼓励读者阅读TREC网站上的论文（<http://trec.nist.gov/>）。有关TREC评估的摘要，请参阅2006年TREC博客跟踪的概述论文（Ounis等，2006），2007年TREC博客跟踪的概述论文（Macdonald, Ounis和Soboroff，2007），以及2008年TREC博客跟踪的概述论文（Ounis, Macdonald和Soboroff，2008）。下面，我们讨论在其他论坛上发表的研究。

在（Eguchi和Lavrenko，2006）中，Eguchi和Lavrenko提出了一种基于生成语言建模的情感检索技术。在他们的方法中，用户需要提供一组表示特定主题兴趣的查询词，以及情感极性（方向）兴趣，可以表示为一组种子情感词或特定的情感方向（积极或消极）。他们的工作的一个主要进展是将情感相关模型和主题相关模型与从训练数据中估计的模型参数相结合，考虑到情感的主题依赖性。他们表明，明确建模主题和情感之间的依赖性比独立处理它们产生更好的检索结果。黄和克罗夫特（2009）也提出了类似的方法，他们使用主题依赖模型和观点相关模型对文档的相关性进行评分。

这两项工作都对主题相关性和情感相关性进行了线性组合，用于最终排名。在（Zhang和Ye，2008）中，作者使用了两个相关性分数的乘积。相关性公式也基于语言建模。

在（Na等人，2009年）中，提出了一种基于词典的方法来进行意见检索。他们还试图解决与领域相关的词典构建问题。提出了一种相关反馈式学习方法，用于生成特定查询的情感词典，该方法利用了对查询的一组排名靠前的文档。

Liu、Li和Liu（2009年）探索了各种词汇和情感特征以及不同的学习算法，用于识别带有观点的博客。他们还提出了结合观点分析和检索组件来检索相关和带有观点的博客的策略的结果。

Li等人（2010年）采用了一种不同的方法。他们的算法首先从文档的每个句子中找到主题和情感词对，然后建立一个二分图，将这些词对与包含这些词对的文档连接起来。

应用了基于图的排名算法HITS（Kleinberg，1999年）对文档进行排名，其中文档被视为权威，词对被视为中心。连接词对和文档的每个链接根据词对对文档的贡献进行加权。

在（Pang和Lee，2008）中，提出了一种用于评论搜索的简单方法。它只重新对基于主题的搜索结果进行排序，通过使用在初始搜索结果中出现的术语的稀有性定义的一种特殊度量。该度量的理论基础在论文中有解释。假设搜索引擎已经找到了好的结果，只需要重新排序以将评论放在前面。该方法是无监督的，不使用任何预先存在的词典。

9.3 摘要

如果像Google或Microsoft Bing这样的网络搜索引擎能够提供一种普遍的观点搜索服务，那将非常有用。尽管Google和Microsoft Bing已经为一些产品的评论提供了观点总结服务，但它们的覆盖范围仍然非常有限。对于那些没有被覆盖的实体和主题，要找到关于它们的观点并不容易，因为它们的观点散布在整个互联网上。还有一些大型知名的评论托管网站，如Amazon.com和Yelp.com。然而，它们也不能涵盖所有的实体和主题。对于那些没有被覆盖的实体或主题，要找到关于它们的观点仍然是一项艰巨的任务，因为各种网站的泛滥和识别相关观点的困难。在取得突破之前，仍然需要大量的研究工作。

"第10章

观点垃圾邮件检测

"社交媒体上的观点越来越被个人和组织用于购买决策、选举选择以及市场营销和产品设计。"积极的观点通常意味着对企业和个人的利润和声誉，不幸的是，这给了人们以强烈的动机来通过发布虚假观点或评论来推广或诋毁某些目标产品、服务、组织、个人甚至思想，而不揭示他们的真实意图，或者他们秘密工作的人或组织。这样的个人被称为观点垃圾邮件发送者，他们的活动被称为观点垃圾邮件发送（Jindal和Liu，2008；Jindal和Liu，2007）。

关于社会和政治问题的观点垃圾邮件发送甚至可能令人恐惧，因为它们可以扭曲观点并动员大众采取与法律或道德准则相悖的立场。可以说，随着社交媒体上的观点越来越多地被实践使用，观点垃圾邮件发送将变得越来越猖獗和复杂，这对于它们的检测提出了重大挑战。然而，为了确保社交媒体继续成为公众观点的可信来源，而不是充满虚假观点、谎言和欺骗，必须检测出这些观点垃圾邮件发送。

垃圾邮件检测在许多领域中都得到了研究。网络垃圾邮件和电子邮件垃圾邮件是最广泛研究的两种垃圾邮件类型。然而，观点垃圾邮件则非常不同。网络垃圾邮件有两种主要类型，即链接垃圾邮件和内容垃圾邮件（Castillo和Davison，2010；Liu，2006和2011）。

链接垃圾邮件是指在超链接上的垃圾邮件，在评论中几乎不存在。虽然在其他形式的社交媒体中广告链接很常见，但它们相对容易检测出来。内容垃圾邮件会在目标网页中添加流行（但无关）的词语，以欺骗搜索引擎使其与许多搜索查询相关，但这在观点发布中几乎不会发生。电子邮件垃圾邮件指的是未经请求的广告，在在线观点中也很少见。

挑战：观点垃圾邮件检测的关键挑战是与其他形式的垃圾邮件不同，很难，甚至不可能通过手动阅读来识别虚假观点，这使得难以找到观点垃圾邮件数据来帮助设计和评估检测算法。对于其他形式的垃圾邮件，人们可以相对容易地识别出来。

实际上，在极端情况下，仅仅通过阅读是逻辑上不可能识别垃圾邮件的。例如，一个人可以为一个好的产品写一个真实的评论。

餐厅并将其发布为一家糟糕餐厅的虚假评论，以此来宣传它。除非考虑评论文本本身之外的信息，否则无法检测到这种虚假评论，因为同一条评论不能同时是真实和虚假的。

本章以消费者评论为例研究这个问题。
在其他形式的社交媒体背景下，很少有研究进行。

10.1 垃圾邮件和垃圾邮件类型

在（Jindal和Liu，2008）中确定了三种垃圾评论：

类型1（假评论）：这些评论不是基于评论者真实使用产品或服务的经验，而是出于隐藏动机而写的。它们经常包含对某些目标实体（产品或服务）的不应得的积极意见，以促进这些实体和/或对某些其他实体的不公正或虚假的负面意见，以损害它们的声誉。

类型2（仅关于品牌的评论）：这些评论不针对具体的产品或服务进行评论，而只评论产品的品牌或制造商。尽管它们可能是真实的，但由于它们不针对特定的产品并且常常带有偏见，所以被视为垃圾邮件。例如，对于特定的惠普打印机的评论说：“我讨厌惠普。我从不购买他们的任何产品”。

类型3（非评论）：这些不是评论。主要有两个子类型：（1）广告和（2）其他不相关的文本，不包含观点（例如问题、答案和随机文本）。严格来说，它们不是观点垃圾邮件，因为它们不提供用户观点。

（Jindal和Liu，2008）已经证明，类型2和类型3的垃圾评论很少，并且相对容易使用监督学习进行检测。即使它们没有被检测出来，这也不是一个主要问题，因为人类读者在阅读过程中很容易发现它们。因此，本章重点讨论类型1的假评论。

虚假评论可以被视为一种特殊形式的欺骗（Hancock等人，2007年；Mihalcea和Strapparava，2009年；Newman等人，2003年；Pennebaker等人，2007年；Vrij，2008年；Zhou，Shi和Zhang，2008年）。然而，传统的欺骗通常指的是关于某些事实或一个人真实感受的谎言。

研究人员已经在文本中识别出许多欺骗信号。例如，研究表明，当人们撒谎时，他们倾向于与自己分离，并喜欢使用诸如你，她，他，他们等词，而不是我，我自己，我的等词。骗子还更频繁地使用与确定性相关的词语来隐藏“假”或强调“真实”。虚假评论在许多方面与谎言不同。

情感分析与观点挖掘

表10.1.虚假评论与产品质量

	积极的虚假评论	消极的虚假评论
优质产品	1	2
一般质量的产品	3	4
质量差的产品	5	6

首先，虚假评论者实际上喜欢使用我，我自己，我的等词，以给读者留下他们评论表达了他们真实的经历的印象。其次，虚假评论不一定是传统的谎言。例如，有人写了一本书，假装是读者，并写了一篇评论来推广这本书。

评论可能是作者的真实感受。此外，许多虚假的评论者可能从未使用过被评论的产品/服务，而只是试图对他们不了解的事物给出积极或消极的意见。他们没有说谎关于他们所知道的任何事实或他们的真实感受。

10.1.1 有害的虚假评论

并非所有虚假评论都同样有害。表10.1给出了不同类型虚假评论的概念视图。在这里，我们假设我们知道产品的真实质量。区域1、3和5中虚假评论的目标是推广产品。尽管区域1中的意见可能是真实的，但评论者没有透露他们的利益冲突或隐藏动机。

区域2、4和6中虚假评论的目标是损害产品的声誉。尽管区域6的评论中的意见可能是真实的，但评论者有恶意的意图。显然，区域1和6中的虚假评论并不是很有害，但区域2、3、4和5中的虚假评论非常有害。因此，虚假评论检测算法应该专注于识别这些区域中的评论。一些现有的检测算法已经使用这个想法，通过采用不同类型的评分偏差特征。请注意，好、坏和平均质量可以根据给予产品的评论的平均评分来定义。然而，如果存在许多垃圾邮件发送者或评论太少，这可能是无效的。

10.1.2 个人和团体垃圾邮件

虚假评论可能由许多类型的人撰写，例如朋友和家人、公司员工、竞争对手、提供虚假评论撰写服务的企业，甚至是真实的顾客（一些企业给予部分顾客折扣甚至全额退款的条件是顾客为他们撰写积极的评论）。在其他形式的社交媒体、公共或私人机构和政治组织中，可能雇佣人员发布信息，以秘密影响社交媒体对话并传播谎言和虚假信息。

情感分析与观点挖掘

媒体、公共或私人机构和政治组织可能雇佣人员发布信息，以秘密影响社交媒体对话并传播谎言和虚假信息。

一般来说，垃圾邮件发送者可以单独工作，也可以有意或无意地作为一个团队的成员工作（这些活动通常非常秘密）。

个体垃圾邮件发送者：在这种情况下，垃圾邮件发送者不与任何人合作。他/她只是使用一个用户ID（例如一本书的作者）自己撰写虚假评论。

垃圾邮件组：有两个主要子情况（Mukherjee, Liu和Glance, 2012; Mukherjee等, 2011）。

- 一群垃圾邮件发送者（人）合谋推广一个目标实体和/或损害另一个实体的声誉。群组中的个别垃圾邮件发送者可能互相认识也可能不认识。
- 一个人注册多个用户ID并使用这些用户ID发送垃圾邮件。这些多个用户ID的行为就像一个合谋的群组。这种情况通常被称为假人操作。

群组垃圾邮件发送非常有害，因为由于群组中的成员数量庞大，它可以完全控制产品的情感，并且完全误导潜在客户，特别是在产品发布初期。尽管群组垃圾邮件发送者也可以被视为许多个体垃圾邮件发送者，但群组垃圾邮件发送具有一些特殊特征，我们将在第10.4节中看到。

我们还应该注意，垃圾邮件发送者有时会单独工作，有时会作为团队的一员。垃圾邮件发送者有时也可能是真实的评论者，因为他/她也作为消费者购买产品，并根据自己的真实经历撰写评论。所有这些复杂的情况使得观点垃圾邮件非常难以检测。

10.1.3 数据、特征和检测类型

用于评论垃圾邮件检测的主要数据类型有三种：

评论内容：每个评论的实际文本内容。从内容中，我们可以提取诸如词语和词性标记 n -gram以及其他句法和语义线索等语言特征，用于欺骗和谎言的检测。然而，仅仅依靠语言特征可能不足够，因为一个人可以相当容易地伪造一篇与真实评论非常相似的假评论。例如，一个人可以根据在一家好餐厅的真实经历为一家差餐厅写一篇假的正面评论。

关于评论的元数据：包括每个评论给出的星级评分等数据。

情感分析与观点挖掘

评论，评论者的用户ID，评论发布时间，撰写评论所花费的时间，评论者计算机的主机IP地址和MAC地址，评论者的地理位置以及在评论网站上的点击顺序。通过这些数据，我们可以挖掘出评论者和他们的评论的许多异常行为模式。

例如，通过评论评级，我们可以发现一个评论者只为一个品牌写正面评论，而为竞争品牌写负面评论。类似地，如果来自同一台计算机的多个用户ID发布了大量关于某个产品的正面评论，这些评论就是可疑的。此外，如果一个酒店的正面评论都来自酒店附近的地区，那么这些评论显然是不可信的。

产品信息：关于被评论实体的信息，例如产品描述和销量/排名。例如，一个产品销售不好但有很多正面评论，这是难以相信的。

这些类型的数据已被用于生成许多垃圾邮件特征。人们还可以将数据分类为公共数据和站点私有数据。通过公共数据，我们指的是显示在托管网站的评论页面上的数据，例如评论内容、评论者的用户ID以及评论发布的时间。通过私有数据，我们指的是网站收集但不在其评论页面上公开展示的数据，例如评论者计算机的IP地址和MAC地址，以及Cookie信息。

观点垃圾邮件检测：在评论环境中，观点垃圾邮件检测的最终目标是识别每个假评论、假评论者和假评论者群体。这三个概念明显相关，因为假评论是由假评论者编写的，而假评论者可以组成假评论者群体。一种类型的检测可以帮助检测其他类型。

然而，每种类型也都有其自身的特点，可以用于检测。

在接下来的两个部分中，我们将重点关注检测个别虚假评论和评论者，并在第10.4节中讨论垃圾评论者群体的检测。

10.2 监督式垃圾邮件检测

一般来说，观点垃圾检测可以被定义为一个具有两个类别的分类问题，即虚假和非虚假。有监督学习是自然适用的。然而，正如我们上面所描述的，一个关键的困难是很难，如果不是不可能的，通过手动阅读来可靠地识别虚假评论，因为垃圾评论者可以精心制作一个看起来和任何无辜评论一样的虚假评论（Jindal和Liu，2008年）。由于这个困难，目前没有可靠的虚假评论和非虚假评论数据可用

情感分析与观点挖掘

来训练一个机器学习算法来识别虚假评论。尽管存在这些困难，已经提出并评估了几种检测算法的不同方法。本节讨论了三种有监督学习方法。下一节描述了一些无监督方法。

由于没有用于学习的标记训练数据，Jindal和Liu（2008）利用了重复的评论。在他们对来自amazon.com的580万条评论和214万名评论者的研究中，发现了大量重复和近似重复的评论，这表明评论垃圾广泛存在。由于编写新评论可能很费力，许多垃圾评论者在不同产品上使用相同的评论或稍作修改的评论。

这些重复和近似重复可以分为四类：

1. 来自同一用户ID在同一产品上的重复评论
2. 来自不同用户ID在同一产品上的重复评论
3. 来自同一用户ID在不同产品上的重复评论
4. 来自不同用户ID在不同产品上的重复评论

第一类重复可能是评论者误点击多次提交按钮的结果（可以根据提交日期轻松检查）。然而，最后三类重复很可能是假的。因此，最后三类重复被用作假评论，其余评论被用作非假评论进行机器学习的训练数据。使用了三组特征：

以评论为中心的特征：这些特征与每个评论有关。示例特征包括评论的实际单词和 n -gram，品牌名称被提及的次数，观点词的百分比，评论长度以及有用反馈的数量。在许多评论网站（例如amazon.com），读者可以通过回答“您是否认为此评论有帮助？”这样的问题来提供反馈。评论者为中心的特征：这些特征与每个评论者有关。示例特征包括评论者给出的平均评分，评分的平均值和标准差，该评论者撰写的作为产品首次评论的评论数量与他/她撰写的总评论数量的比例，以及他/她是唯一评论者的情况的比例。

产品为中心的特征：这些特征与每个产品有关。示例特征包括产品的价格，产品的销售排名（amazon.com根据销售量为每个产品分配销售排名），以及产品评论评分的平均值和标准差。

逻辑回归被用于模型构建。实验结果显示了一些初步但有趣的结果。

情感分析与观点挖掘

- 负面异常评论（与产品平均评分有显著负偏差的评分）往往受到严重的垃圾邮件攻击。正面异常评论则没有受到严重的垃圾邮件攻击。
- 对于某些产品而言，唯一的评论很可能是假的。这可以通过卖家推销不受欢迎的产品而写假评论的倾向来解释。
- 排名靠前的评论者更有可能是假评论者。亚马逊根据其专有方法为每个评论者给出一个排名。分析显示，排名靠前的评论者通常写了大量的评论。撰写大量评论的人是自然的嫌疑人。一些顶级评论者撰写了数千甚至数万条评论，这对于普通消费者来说是不太可能的。
- 假评论可以获得良好的反馈，真实评论可以获得不好的反馈。这表明，如果根据有用反馈来定义评论的质量，人们可能会被假评论欺骗，因为垃圾邮件发送者可以轻松地编写出一个能够获得许多正面反馈的复杂评论。
- 销售排名较低的产品更容易被垃圾邮件攻击。这表明垃圾邮件活动似乎仅限于销售较低的产品，这是合理的，因为很难损害热门产品的声誉，而不受欢迎的产品需要一些推广。

需要再次强调，这些结果是初步的，因为（1）尚未确认这三种重复类型是否绝对是虚假评论，以及（2）许多虚假评论并非重复评论，在模型构建中被视为非虚假评论（Jindal和Liu，2008年）。

在（Li等，2011年）中，另一种监督学习方法被尝试用于识别虚假评论。在他们的案例中，从Epinions评论中构建了一个手动标记的虚假评论语料库。在Epinions中，用户可以通过给评论评分来评估评论的有用性。他们还可以对评论进行写作。作者通过阅读评论和评论来手动标记一组虚假或非虚假评论。为了学习，提出了几种类型的特征，这些特征与（Jindal和Liu，2008年）中的特征类似，但也有一些添加，例如主观和客观特征，积极和消极特征，评论者的个人资料，使用PageRank（Page等，1999年）计算的权威分数等。他们使用朴素贝叶斯分类进行学习，取得了有希望的结果。作者还尝试了一种半监督学习方法，利用了一个想法，即垃圾邮件发送者倾向于编写许多虚假评论。

在（Ott等人，2011年）中，也采用了监督学习。在这种情况下，作者使用了亚马逊机械土耳其人来众包虚假的酒店评论。

20家酒店。为了确保虚假评论的质量，采取了几项措施。例如，他们只允许每个土耳其人提交一次，土耳其人必须在美国，等等。土耳其人还被告知他们在酒店工作，他们的老板要求他们写虚假评论来推广酒店。真实的评论是从TripAdvisor网站获得的。作者尝试了几种分类方法，这些方法在相关任务中已经被使用，例如流派识别、心理语言学欺骗检测和文本分类。所有这些任务都有一些研究人员提出的现有特征。他们的实验表明，仅基于50/50虚假和非虚假类别分布的一元和二元模型的文本分类效果最好。传统的欺骗特征（Hancock等人，2007年；Mihalcea和Strapparava，2009年；Newman等人，2003年；Pennebaker等人，2007年；Vrij，2008年；Zhou，Shi和Zhang，2008年）效果不佳。然而，与先前的研究不同，这里使用的评估数据也不完美。

来自亚马逊机械土耳其的虚假评论可能不是真正的“假评论”，因为土耳其人不知道被评论的酒店，尽管他们被要求假装自己是酒店的员工。此外，使用50/50的虚假和非虚假数据进行测试可能无法反映真实情况的真实分布。类别分布对检测到的虚假评论的准确性有重要影响。

10.3 无监督垃圾邮件检测

由于手动标记训练数据的困难，仅使用有监督学习进行虚假评论检测是困难的。在本节中，我们讨论了两种无监督方法。类似于这些技术已经在许多评论托管网站中使用。

10.3.1 基于非典型行为的垃圾邮件检测

本小节描述了一些试图发现评论者异常行为以进行垃圾邮件检测的技术。例如，如果一个评论者对某个品牌写了所有的负面评论，而其他评论者对该品牌都持积极态度，并且对竞争品牌写了所有的正面评论，那么这个评论者自然是可疑的。

第一种技术来自（Lim等人，2010年），它基于不同的评论模式识别出几种异常的评论者行为模型，这些模型表明存在垃圾评论。每个模型都为评论者分配一个数值化的垃圾行为分数

情感分析与观点挖掘

通过测量评论者实施垃圾行为的程度来评估。然后将所有分数合并以产生最终的垃圾分数。因此，该方法侧重于寻找垃圾评论者或虚假评论，而不是虚假评论本身。垃圾行为模型包括：（a）针对产品：为了操纵评论系统，假设垃圾评论者会将大部分精力放在推广或针对

少数目标产品上。他们预计会密切监视这些产品，并在合适的时机通过撰写虚假评论来干扰评分。

（b）针对群体：这种垃圾行为模型定义了一组共享某些属性的产品在短时间内受到垃圾评论者操纵评分的模式。例如，垃圾评论者可能会在几个小时内针对某个品牌的多个产品进行操作。这种评分模式可以节省垃圾评论者的时间，因为他们不需要多次登录评论系统。为了达到最大的影响，针对这些目标产品群体的评分要么非常高，要么非常低。（c）一般评分偏差：真实的评论者应该给出与其他评分者相似的评分。由于垃圾评论者试图推广或贬低某些产品，他们的评分通常与其他评论者的评分相差很大。

（d）早期评级偏差：早期偏差捕捉到了一个在产品发布后不久就贡献了一个虚假评论的垃圾邮件发送者的行为。这样的评论很可能会引起其他评论者的注意，从而影响后续评论者的观点。

第二种技术还专注于寻找虚假评论者或垃圾邮件发送者(Jindal, Liu and Lim, 2010)。在这里，问题被定义为一个数据挖掘任务，用于发现意外的类关联规则。与传统的垃圾邮件检测方法（如上述的有监督和无监督方法）不同，这些方法首先手动识别一些启发式垃圾邮件特征，然后将它们用于垃圾邮件检测。这种技术是通用的，可以应用于解决一类问题，因为它与领域无关。

类关联规则是一种特殊类型的关联规则（Liu, Hsu and Ma, 1998），具有固定的类属性。用于挖掘类关联规则（CARs）的数据由一组数据记录组成，这些记录由一组正常属性 $A = \{A_1, \dots, A_n\}$ 和一个类属性 $C = \{c_1, \dots, c_m\}$ 描述，其中 m 是离散值，称为类标签。CAR规则的形式为： $X \rightarrow c_i$ ，其中 X 是来自属性 A 的条件集合， c_i 是 C 中的一个类标签。这样的规则计算条件概率 $\Pr(c_i | X)$ （称为置信度）和联合概率 $\Pr(X, c_i)$ （称为支持度）。

对于垃圾邮件检测应用程序，CAR挖掘的数据如下所示：每个评论形成一个带有一组属性的数据记录，例如，

情感分析与观点挖掘

评论者ID, 品牌ID, 产品ID, 和一个类别。该类别表示评论者对产品的情感, 积极, 消极, 或中立基于评论评级。在大多数评论网站 (例如 *amazon.com*) , 每个评论都有一个由评论者分配的1 (最低) 到5 (最高) 的评级。评级为4或5被分配为积极, 3为中立, 1或2为消极。一个发现的CAR规则可能是评论者对某个品牌的产品给出所有积极评级。(Jindal, Liu和Lim, 2010) 中的方法基于四个意外性定义找到四种类型的意外规则。这些意外规则代表评论者的非典型行为。

下面, 给出了每种意外性定义的示例行为。意外性定义非常复杂, 可以在 (Jindal, Liu和Lim, 2010) 中找到。

- 置信度意外性: 使用这个度量, 可以找到给某个品牌的产品都给出高分的评论者, 但大多数其他评论者对该品牌普遍持负面态度。
- 支持度意外性: 使用这个度量, 可以找到为单个产品撰写多篇评论的评论者, 而其他评论者只撰写一篇评论。
- 属性分布意外性: 使用这个度量, 可以发现大多数对某个品牌的产品给出正面评价的评论是由一个评论者撰写的, 尽管有很多评论者对该品牌的产品进行了评论。
- 属性意外性: 使用这个度量, 可以找到只对一个品牌撰写正面评价并只对另一个品牌撰写负面评价的评论者。

这种方法的优点是所有的意外度量都是在CARs规则上定义的, 因此与领域无关。这种技术可以在其他领域中使用, 以发现意外的模式。缺点是一些非典型行为无法被检测到, 例如与时间相关的行为, 因为类关联规则不考虑时间。

重要的是要注意, 发表论文中研究的行为都是基于其各自评论托管网站上的公开数据。正如前面提到的, 评论托管网站还收集了关于每个评论者及其在网站上活动的许多其他数据。

这些数据对一般公众不可见, 但对于垃圾邮件检测可能非常有用, 甚至可能比公开数据更有用。例如, 如果同一IP地址的多个用户ID发布了大量关于某个产品的正面评价, 则这些用户ID是可疑的。如果一个酒店的正面评价都来自该酒店附近的地区, 那么它们也是可疑的。一些评论托管网站已经在使用这些和其他技术。

他们利用内部数据的片段来检测虚假评论者和评论。

最后，吴等人（2010年）还提出了一种基于失真准则（而不是评论者行为）的无监督方法来检测虚假评论。这个想法是，虚假评论会扭曲一组实体的整体受欢迎程度排名。也就是说，随机删除一组评论不应过分扰乱实体的排名列表，而删除虚假评论应该显著改变或扭曲实体的排名，以揭示“真实”的排名。可以通过使用排名相关性比较删除前后的受欢迎程度排名来衡量这种失真。

10.3.2 使用评论图进行垃圾邮件检测

在（Wang等人，2011年）中，提出了一种基于图的方法来检测商店或商家评论中的垃圾邮件。这些评论描述了购买体验和对商店的评价。这项研究基于2010年10月6日从resellerratings.com爬取的所有评论的快照。

在删除没有评论的商店后，有343603个评论者写了408470条关于14561个商店的评论。

虽然可以从产品评论垃圾邮件检测中借鉴一些思路，但它们在商店评论的背景下不足够。例如，一个人对同一产品发表多篇评论是可疑的，但一个人对同一家商店发表多篇评论是正常的，因为可能有多次购买经历。此外，同一评论者对多个商店的近似重复评论是正常的，因为不同于不同的产品，不同的商店基本上提供相同类型的服务。因此，现有方法中提出的用于检测虚假产品评论和评论者的特征或线索并不都适用于检测商店评论的垃圾邮件发送者。因此，需要寻找一种更复杂和互补的框架。

本文使用了一个异构的评论图，包括评论者、评论和商店三种类型的节点，来捕捉它们之间的关系并建模垃圾邮件的线索。一个评论者节点与他/她撰写的每篇评论都有链接。如果评论是关于某个商店的，那么一个评论节点就与一个商店节点相连。一个商店通过该评论者对该商店的评论与一个评论者相连。每个节点还附带一组特征。例如，一个商店节点具有关于其平均评分、评论数量等特征。基于评论图，定义和计算了三个概念，即评论者的可信度、评论的诚实度和商店的可靠性。

如果评论者撰写了更多真实的评论，那么他/她更值得信赖；

如果一个商店有更多来自可信评论者的积极评论，那么它更可靠；如果一篇评论得到了许多其他真实评论的支持，那么它更诚实。此外，如果一篇评论的诚实度下降，它会影响评论者的可信度，进而影响他/她所评论的商店。

这些交织在一起的关系在评论图中得以揭示，并通过数学方式进行定义。提出了一种迭代计算方法来计算这三个值，然后用于对评论者、商店和评论进行排名。

那些排名靠前的评论者、商店和评论很可能参与评论垃圾信息的发布。通过与美国著名公司*Better Business Bureaus (BBB)* 收集的商业可靠性报告和警示公众商业或消费欺诈的分数进行人工评判来进行评估。

10.4 群组垃圾信息检测

在(Mukherjee等人, 2011)中提出了一个初始的垃圾邮件检测算法，在(Mukherjee, Liu和Glance, 2012)中进行了改进。该算法发现了一些可能合谋推广或贬低某些目标实体的垃圾邮件组。它分为两个步骤：

1. 频繁模式挖掘：首先，它对评论数据进行预处理，生成一组交易。每个交易代表一个唯一的产品，并包含所有评论该产品的评论者（他们的ID）。使用所有的交易，它进行频繁模式挖掘，找到一组频繁模式。每个模式基本上是一组同时评论一组产品的评论者。这样的一组被视为候选垃圾邮件组。使用频繁模式挖掘的原因如下：如果一组评论者只是合作一次来推广或贬低一个单一产品，基于他们的集体行为可能很难检测出来。然而，这些虚假评论者（尤其是那些被付费写作的人）不可能只为一个单一产品写一篇评论，因为这样他们赚不到足够的钱。相反，他们会在许多产品上工作，即写许多关于许多产品的评论，这也会暴露他们。频繁模式挖掘可以发现他们在多个产品上合作。
2. 基于一组群体垃圾指标对群体进行排名：在第一步中发现的群体可能并非全部是真正的垃圾邮件发送者群体。许多评论者在模式挖掘中仅仅是由于偶然性而被分组在一起。然后，这一步首先使用一组指标来捕捉不同类型的异常群体和个体成员行为。这些指标包括在短时间窗口内一起撰写评论，撰写评论

在产品发布后立即进行群组评论内容相似性，群组评级偏差等（Mukherjee, Liu和Glance, 2012年）。然后提出了一种关系模型，称为GSRank（群组垃圾邮件排名），以利用群组、个体群组成员和产品之间的关系来根据它们成为垃圾邮件群组的可能性对候选群组进行排名。然后使用迭代算法来解决这个问题。还手动标记了一组垃圾邮件群组，并用于评估所提出的模型，结果显示有希望。这种方法的一个弱点是由于模式挖掘中使用的频率阈值，如果一个群组没有多次合作（三次或更多次），则无法通过此方法检测到它。

这种方法是无监督的，因为它不使用任何手动标记的数据进行训练。显然，使用标记数据可以应用监督学习。事实上，(Mukherjee, Liu和Glance, 2012)描述了对几种最先进的监督分类、回归和学习排序算法的实验，但它们被证明效果较差。

10.5 总结

随着社交媒体在组织和个人的重要决策中的越来越广泛使用，观点垃圾信息也越来越普遍。对于许多企业来说，发布虚假观点或雇佣他人代替他们进行是一种廉价的营销和品牌推广方式。

尽管目前关于观点垃圾信息检测的研究仍处于早期阶段，但已经提出并在实践中使用了几种有效的算法。然而，垃圾信息发布者也变得越来越精明和小心地编写和发布虚假观点以避免被检测。事实上，我们已经看到了检测算法和垃圾信息发布者之间的一场军备竞赛。然而，我对设计更复杂的检测算法使垃圾信息发布者很难发布虚假观点持乐观态度。这样的算法很可能是综合性的方法，在检测过程中整合所有可能的特征或线索。

最后，我们应该注意到观点垃圾邮件不仅出现在评论中，还出现在其他形式的社交媒体中，如博客、论坛讨论、评论和Twitter帖子中。然而，到目前为止，在这些情境下进行的研究很少。

第11章

评论的质量

在本章中，我们讨论评论的质量。这个主题与观点垃圾邮件检测有关，但也不同，因为低质量的评论可能不是垃圾或虚假评论，而虚假评论可能不会被读者视为低质量的评论，因为正如我们在上一章中讨论的那样，通过阅读评论很难发现虚假评论。因此，如果冒名顶替者早期撰写并精心编写他们的评论，虚假评论也可能被视为有帮助或高质量的评论。

这个任务的目标是确定每个评论的质量、有用性或实用性（Ghose和Ipeirotis, 2007；Kim等, 2006；Liu等, 2007；Zhang和Varadarajan, 2006）。这是一个有意义的任务，因为在向用户展示评论时，希望根据质量或有用性对评论进行排序，将最有帮助的评论放在前面。事实上，许多评论聚合或托管网站多年来一直在实践这一点。他们通过要求读者对每个评论提供有用性反馈来获得每个评论的有用性或质量分数。例如，在amazon.com上，读者可以通过回答问题“这篇评论对您有帮助吗？”来表示他/她是否发现评论有帮助。”就在每个评论下面。所有回答者的反馈结果然后被汇总并显示在每个评论之前，例如，“16个人中有15个人发现以下评论有帮助。”尽管大多数评论托管网站已经提供了这项服务，但自动确定每个评论的质量仍然很有用，因为许多用户反馈可能需要很长时间才能积累。这就是为什么许多评论只有很少或没有反馈的原因。这对于新评论尤其如此。

11.1 质量作为回归问题

确定评论的质量通常被制定为回归问题。学习模型为每个评论分配一个质量分数，可以用于评论排名或评论推荐。在这个研究领域中，用于训练和测试的基准数据通常是用户对每个评论提供的有用反馈，正如我们上面讨论的那样，这些反馈在许多评论托管网站上都有提供。因此，与假评论检测不同，这里的训练和测试数据不是一个问题。

情感分析与观点挖掘

研究人员使用了许多类型的特征来构建模型。

在(Kim et al., 2006)中, 使用SVM回归来解决这个问题。特征集包括:

结构特征: 评论长度、句子数量、问句和感叹句的百分比, 以及HTML粗体标签和换行符
的数量。

词汇特征: 带有tf-idf权重的一元组和二元组。

句法特征: 解析的令牌中开放类别(即名词、动词、形容词和副词)的百分比, 名词的令牌百分比, 动词的令牌百分比, 第一人称动词的令牌百分比, 以及形容词或副词的令牌百分比。

语义特征: 产品方面和情感词汇。

元数据特征: 评论评级(星级数量)。

在(Zhang和Varadarajan, 2006)中, 作者还将问题视为回归问题。他们使用了类似的特征, 例如评论长度, 评论评级, 一些特定POS标签的计数, 情感词汇, tf-idf加权分数, wh-词, 产品方面提及, 与产品规格的比较, 与编辑评论的比较等。

与上述方法不同, (Liu等, 2008)考虑了三个主要因素, 即评论者的专业知识, 评论的及时性和基于POS标签的评论风格。提出了一个非线性回归模型来整合这些因素。这项工作侧重于电影评论。

在(Ghose和Ipeirotis, 2007; Ghose和Ipeirotis, 2010)中, 使用了三个额外的功能集, 即来自评论网站的评论者个人资料功能, 捕捉过去评论的有用性的评论者历史功能, 以及一组可读性功能, 即拼写错误和可读性研究中的可读性指数。为了学习, 作者尝试了回归和二元分类两种方法。

Lu等人(2010)从另一个角度研究了这个问题。他们调查了评论者的社交背景如何帮助提高基于文本的评论质量预测器的准确性。他们认为社交背景可以揭示关于评论者质量的大量信息, 进而影响他们的评论质量。具体而言, 他们的方法基于以下假设:

作者一致性假设: 同一作者的评论具有相似的质量。

信任一致性假设: 一个评论者 r_1 到另一个评论者 r_2 的链接是一种明确或隐含的信任陈述。评论者 r_1 只信任评论者 r_2

如果评论者 r_2 的质量至少与评论者 r_1 的质量相同。

共引用一致性假设：人们在信任他人方面是一致的。因此，如果两个评论者 r_1 和 r_2 都受到同一个第三个评论者 r_3 的信任，则他们的质量应该相似。

链接一致性假设：如果两个人在社交网络中相连（ r_1 信任 r_2 ，或 r_2 信任 r_1 ，或两者都有），则他们的评论质量应该相似。

这些假设被作为正则化约束添加到基于文本的线性回归模型中，以解决评论质量预测问题。在实验中，作者使用了来自Ciao (www.ciao.co.uk) 的数据，这是一个社区评论网站。在Ciao中，人们不仅为产品和服务撰写评论，还对其他人撰写的评论进行评级。此外，如果人们发现这些成员的评论一贯有趣和有帮助，他们可以将其添加到他们的信任成员网络或“信任圈子”中。显然，这种技术不适用于没有建立信任社交网络的网站。

11.2 其他方法

在 (O'Mahony和Smyth, 2009) 中，提出了一种分类方法来对有帮助和无帮助的评论进行分类。使用了许多特征：

声誉特征：评论者撰写的所有评论的有用性平均值 ($R1$) 和标准差 ($R2$)，评论者撰写的至少收到 T 个反馈的评论所占的百分比 ($R3$)，等等。

内容特征：评论长度 ($C1$)，评论文本中大写字母与小写字母的比例 ($C3$)，等等。

社交特征：评论者撰写的评论数量 ($SL1$)，所有评论者撰写的评论数量的平均值 ($SL2$) 和标准差 ($SL3$)，等等。

情感特征：评论的评分 ($ST1$)，评论者对其撰写的所有评论的评分的平均值 ($ST5$) 和标准差 ($ST6$)，等等。

在 (Liu等人, 2007年) 中，该问题也被定义为一个两类分类问题。然而，他们认为使用帮助性投票作为真实情况可能不合适，因为存在三种偏见：(1) 投票不平衡（很大比例的投票是有帮助的）；(2) 早期偏见（早期评论往往获得更多的投票）；(3) 内部圈子偏见（当某些评论获得很多投票时，它们在评论网站上排名较高，从而帮助它们获得更多的投票）。那些排名较低的评论获得很少的投票，但它们可能并不是质量低下的评论。然后，作者将评论分为4个类别：“最佳评论”、“好评”、“公平评论”和“差评”，根据评论是否讨论了产品的许多方面并提供了令人信服的观点进行手动标记，以生成黄金标准的训练和测试数据。在分类中，他们使用支持向量机 (SVM) 进行二元分类。只有“差评”类别被视为低质量类别，其他三个类别都被视为高质量类别。学习的特征包括信息性、主观性和可读性。每个特征都包含一个个体特征

这些特征帮助他们获得更多的投票)。那些排名较低的评论获得很少的投票,但它们可能并不是质量低下的评论。然后,作者将评论分为4个类别:“最佳评论”、“好评”、“公平评论”和“差评”,根据评论是否讨论了产品的许多方面并提供了令人信服的观点进行手动标记,以生成黄金标准的训练和测试数据。在分类中,他们使用支持向量机(SVM)进行二元分类。只有“差评”类别被视为低质量类别,其他三个类别都被视为高质量类别。学习的特征包括信息性、主观性和可读性。每个特征都包含一组个体特征。

Tsur和Rappoport (2009)使用了一种无监督的方法研究了书评的有用性,这与上述的有监督方法非常不同。该方法分为三个步骤。给定一组评论,首先识别评论中的一组重要术语。这些术语一起形成了表示虚拟最佳或核心评论的向量。然后,根据评论中发现的重要术语的出现情况,将每个实际评论映射或转换为该向量表示。之后,根据评论与虚拟评论之间的距离(两者都表示为向量),为每个评论分配一个排名分数。

在(Moghaddam, Jamali和Ester, 2012)中,提出了一个新的问题,即个性化评论质量预测以推荐有用的评论。所有上述方法都假设评论对所有用户/读者的有用性相同,而作者认为这是不正确的。为了解决这个新问题,他们提出了几个因子分解模型。这些模型基于观察到的评论评级取决于评论、评论者、评分者/用户和产品的一些潜在特征的假设。

本质上,该论文将问题视为个性化推荐问题。解决这个问题的提出的技术相当复杂。

关于这种推荐形式的一些背景知识可以在该书的第12章中找到(Liu, 2006和2011)。

所有上述方法都基于计算出的有用性或质量分数对评论进行排名。然而,Tsaparas, Ntoulas和Terzi (2011)认为这些方法没有考虑一个重要事实,即前几个高质量评论可能高度冗余,并重复相同的信息。在他们的工作中,他们提出了选择一组全面且数量较少的高质量评论,涵盖被评论实体的许多不同方面以及评论的不同观点的问题。他们将问题形式化为最大覆盖问题,并提出了解决该问题的算法。在(Lappas和

Gunopulos, 2010) 也研究了找到一小组评论覆盖所有产品方面的问题。

11.3 总结

总之，确定评论的有用性是一个重要的研究课题。

它对于拥有大量评论的产品和服务尤其有用。为了帮助读者快速获取高质量的意见，评论网站应该提供良好的评论排名。然而，我还想补充一些警告。首先，正如我们在关于意见搜索和检索的章节中讨论的那样，我们认为评论排名必须反映积极和消极意见的自然分布。仅仅因为某些评论具有高质量分数，就将所有积极（或消极）评论排在前面并不是一个好主意。（Tsaparas、Ntoulas和Terzi，2011）提出的冗余问题也是一个合理的关注点。在我看来，质量和（积极和消极观点的）分布都很重要。

其次，读者倾向于根据评论是否对产品的许多方面表达意见并且看起来真实来确定评论是否有帮助。一个垃圾评论者可以通过精心制作一篇与正常有帮助的评论非常相似的评论来满足这一要求。因此，仅仅使用有用反馈的数量来定义评论质量或作为唯一的真实性标准可能存在问题。此外，用户反馈也可能受到垃圾信息的影响。反馈垃圾信息是搜索广告中点击欺诈的一个子问题，其中一个人或机器人点击一些在线广告以给出真实客户点击的印象。在这种情况下，机器人或人类垃圾评论者可以点击有用反馈按钮来增加评论的有用性。

第12章

总结性评论

本书介绍了情感分析和观点挖掘领域，并对当前的最新技术进行了调查。由于存在许多具有挑战性的研究问题和各种实际应用，该领域的研究近年来非常活跃。它已经从计算机科学扩展到管理科学（Archak, Ghose和Ipeirotis, 2007; Chen和Xie, 2008; Das和Chen, 2007; Dellarcas, Zhang和Awad, 2007; Ghose, Ipeirotis和Sundararajan, 2007; Hu, Pavlou和Zhang, 2006; Park, Lee和Han, 2007），因为对产品的意见与利润密切相关。

本书首先定义了情感分析问题，为该领域中不同的研究方向提供了一个共同的框架。然后讨论了广泛研究的文档级情感分类主题，旨在确定一个意见文档（例如评论）是否表达了积极或消极的情感。接下来是句子级主观性和情感分类，它确定一个句子是否带有观点，并且如果是的话，是积极的还是消极的观点。然后描述了基于方面的情感分析，探索了问题定义的全部能力，并展示了情感分析是一个具有许多具有挑战性的子问题的多面问题。讨论了处理这些问题的现有技术。

之后，该书讨论了情感词典生成的问题。

涵盖了两种主要方法。接着是关于观点总结的章节，这是一种特殊形式的多文档摘要。然而，它与传统的多文档摘要也有很大的不同，因为观点总结可以以结构化的方式进行，这有助于定性和定量分析以及观点的可视化。第8章讨论了分析比较和最高级句子的问题。这些句子代表了一种不同类型的评价，与当前研究的重点——常规观点不同。第9章介绍了观点搜索或检索的主题。最后，我们在第10章讨论了观点垃圾邮件检测，并在第11章评估了评论的质量。

通过编写虚假评论和发布虚假评论，观点垃圾邮件成为越来越重要的问题，因为越来越多的人依赖于网络上的观点进行决策。为了确保这些观点的可信度，打击观点垃圾邮件是一项紧迫的任务。

情感分析与观点挖掘

和关键任务。

通过阅读这本书到目前为止，很容易看出情感分析在技术上非常具有挑战性。尽管研究界已经从许多不同的角度尝试了许多子问题，并且已经发表了大量的研究论文，但是没有一个子问题得到了令人满意的解决。我们对整个问题及其解决方案的理解和知识仍然非常有限。主要原因是这是一个自然语言处理任务，而自然语言处理没有简单的问题。另一个原因可能是我们进行研究的流行方式。我们可能过于依赖机器学习。

一些最有效的机器学习算法，例如支持向量机、朴素贝叶斯和条件随机场，产生的结果对人类来说是不可理解的，尽管它们可以帮助我们提高准确性，但我们对其如何以及为什么知之甚少，除了在手动特征工程过程中获得的一些肤浅知识。

话虽如此，过去十年我们确实取得了重大进展。这一点可以从提供情感分析服务的初创公司和已经建立起来的公司的数量之多看出来。行业对此类服务有着真实而巨大的需求，因为每个企业都想知道消费者如何看待他们自己和竞争对手的产品和服务。对于消费者来说也是如此，因为每当有人想购买某样东西时，都想知道现有用户的意见。这些实际需求和技术挑战将使这个领域在未来几年保持活力和生机。

在已经完成的工作基础上，我相信我们只需要进行更深入的研究，并构建综合系统来同时处理所有子问题，因为它们的相互作用可以帮助解决每个单独的子问题。我对整个问题在不久的将来能够得到令人满意的解决持乐观态度，以便广泛应用。

对于应用程序来说，完全自动化和准确的解决方案还没有出现。然而，可以设计出有效的半自动化解决方案。关键是要充分理解所有问题和陷阱，巧妙地管理它们，并确定哪些部分可以自动完成，哪些部分需要人类的帮助。在完全手动解决方案和完全自动化解决方案之间的连续过程中，随着时间的推移，我们可以越来越多地向自动化方向发展。我不认为很快会有一个万能解决方案。一个好的策略是在大量不同的应用领域上努力工作，了解每个领域，并逐渐设计出一个通用的解决方案。

参考文献

1. Abbasi, Ahmed, Hsinchun Chen, and Arab Salem. 在多种语言中的情感分析：网络论坛观点分类的特征选择。 *ACM Transactions on Information Systems (TOIS)*, 2008. **26**(3).2. Abdul-Mageed, Muhammad, Mona T. Diab, and Mohammed Korayem. 现代标准阿拉伯语的主观性和情感分析。在计算语言学协会第49届年会短论文集. 2011.
3. Akkaya, Cem, Janyce Wiebe, and Rada Mihalcea. 主观词义消歧。在2009年经验方法自然语言处理会议论文集 (*EMNLP-2009*) . 2009.
4. Alm, Ebba Cecilia Ovesdotter. 文本和语音中的情感, 2008年: ProQuest.
5. Andreevskaia, Alina和Sabine Bergler. 从WordNet的定义中挖掘模糊情感: 从WordNet的解释中提取情感标签. 在欧洲计算语言学协会(*EACL-06*)会议的论文中. 2006年.
6. Andreevskaia, Alina和Sabine Bergler. 当专家和通才共同工作: 克服情感标记中的领域依赖. 在计算语言学协会(*ACL-2008*)年会议的论文中. 2008年.
7. Andrzejewski, David和Xiaojin Zhu. 具有主题集知识的潜在狄利克雷分配. 在*NAACL HLT*会议的论文中. 2009年.
8. Andrzejewski, David, Xiaojin Zhu和Mark Craven. 通过狄利克雷森林先验将领域知识融入主题建模. 在*ICML*会议的论文中. 2009年.
9. Archak, Nikolay, Anindya Ghose, and Panagiotis G. Ipeirotis. 给我看看钱! : 通过挖掘消费者评论来推导产品特性的定价能力。在*ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-2007)*的论文集中。2007年。
10. Asher, Nicholas, Farah Benamara, and Yvette Yannick Mathieu. 在话语中提炼观点: 一项初步研究。在*International Conference on Computational Linguistics (COLING-2008): Companion volume: Posters and Demonstrations*的论文集中。2008年。
11. Asur, Sitaram and Bernardo A. Huberman. 用社交媒体预测未来。Arxiv预印本arXiv:1003.5699, 2010年。
12. Aue, Anthony and Michael Gamon. 将情感分类器定制到新领域: 一个案例研究。在*Recent Advances in Natural Language Processing (RANLP-2005)*的论文集中。2005年。
13. Banea, Carmen, Rada Mihalcea, 和 Janyce Wiebe. 多语言主观性: 更多语言是否更好? 在国际计算语言学会议 (*COLING-2010*) 的论文中. 2010.
14. Banea, Carmen, Rada Mihalcea, Janyce Wiebe, 和 Samer Hassan. 使用机器翻译的多语言主观性分析. 在自然语言处理经验方法会议 (*EMNLP-2008*) 中. 2008年.
15. Bar-Haim, Roy, Elad Dinur, Ronen Feldman, Moshe Fresko, 和 Guy Goldstein. 在股票微博中识别和跟踪专家投资者. 在自然语言处理经验方法会议 (*EMNLP-2011*) 中. 2011年.

情感分析与观点挖掘

16. Barbosa, Luciano 和 Junlan Feng. 从有偏差和嘈杂数据中进行强大的推特情感检测. 在国际计算语言学会议 (COLING-2010) 中. 2010年.
17. Bautin, Mikhail, Lohit Vijayarenu, 和 Steven Skiena. 国际新闻和博客的情感分析. 在国际AAA/博客和社交媒体会议(ICWSM-2008). 2008年.
18. Becker, Israela 和 Vered Aharonson. 最后但绝对不是最不重要的: 关于自动极性分类中最后一句的作用. 在ACL 2010会议短文集中. 2010年.
19. Beineke, Philip, Trevor Hastie, Christopher Manning, 和 ShivakumarVaithyanathan. 情感总结的探索. 在AAA/春季研讨会上探索文本中的态度和情感: 理论和应用. 2003年.
20. Benamara, Farah, Baptiste Chardon, Yannick Mathieu, 和 VladimirPopescu. 基于上下文的主观性分析. 在第5届国际自然语言处理联合会议(IJCNLP-2011). 2011年.
21. Bessalov, Dmitriy, Bing Bai, Yanjun Qi, and Ali Shokoufandeh. 基于监督潜在n-gram分析的情感分类. 在信息与知识管理 (CIKM-2011) ACM会议论文集. 2011.
22. Bethard, Steven, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, 和 Dan Jurafsky. 自动提取观点命题及其持有者. 在AAA/春季研讨会论文集上探索文本中的态度和情感. 2004.
23. Bickerstaffe, A. 和 I. Zukerman. 应用于多类情感检测的分层分类器. 在第23届国际计算语言学会议 (Coling 2010) 论文集上. 2010.
24. Bilgic, Mustafa, Galileo Mark Namata, 和 Lise Getoor. 结合集体分类和链接预测在挖掘图和复杂结构研讨会论文集上. 2007.
25. Bishop, C. M. 模式识别与机器学习. 第4卷. 2006: springer纽约.
26. Blair-Goldensohn, Sasha, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A. Reis, and Jeff Reynar. 为本地服务评论构建情感总结器. 在WWW-2008工作坊上的NLP in the Information Explosion Era. 2008.
27. Blei, David M. and Jon D. McAuliffe. 监督主题模型. 在NIPS会议上. 2007.
28. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 潜在狄利克雷分配. 机器学习研究杂志, 2003. 3: p. 993-1022.
29. Blitzer, John, Mark Dredze, and Fernando Pereira. 传记, 宝莱坞, 音箱和搅拌机: 领域自适应情感分类. 在计算语言学协会年会论文集 (ACL-2007). 2007年.
30. Blitzer, John, Ryan McDonald, 和 Fernando Pereira. 结构对应学习的领域适应. 在自然语言处理实证方法会议论文集 (EMNLP-2006). 2006.
31. Blum, Avrim 和 Shuchi Chawla. 使用图最小割学习标记和未标记数据. 在国际机器学习会议论文集 (ICML-2001). 2001年.
32. Blum, Avrim, John Lafferty, Mugizi R. Rwebangira, 和 Rajashekar Reddy. 使用随机最小割的半监督学习. 在

情感分析与观点挖掘

- 国际机器学习会议论文集 (*ICML-2004*). 2004年.
33. Boiy, Erik 和 Marie-Francine Moens.多语言网络文本情感分析的机器学习方法信息检索, 2009年.
12(5): p. 526-558.
34. Bollegala, Danushka, David Weir, and John Carroll. 使用多个来源构建情感敏感的词库, 用于跨领域情感分类. 在第49届计算语言学年会 (*ACL-2011*) 的论文集中. 2011.
35. Bollen, Johan, Huina Mao, and Xiao-Jun Zeng. *Twitter*情绪预测股市. 计算科学杂志, 2011.
36. Boyd-Graber, Jordan and Philip Resnik.跨语言的整体情感分析: 多语言监督潜在狄利克雷分配. 在第十届自然语言处理会议 (*EMNLP-2010*) 的论文集中. 2010.
37. Branavan, S. R. K., Harr Chen, Jacob Eisenstein, and Regina Barzilay.从自由文本注释中学习文档级语义属性. 在第一届计算语言学年会 (*ACL-2008*) 的论文集中. 2008.
38. Breck, Eric, Yejin Choi, and Claire Cardie. 在上下文中识别观点表达. 在国际人工智能联合会议(*IJCAI-2007*). 2007年.
39. Brody, Samuel 和 Nicholas Diakopoulos. Cooooooooooooooooo!!!!!!!!!!!!!!使用单词延长来检测微博中的情感. 在自然语言处理实证方法会议(*EMNLP-2011*).
2011.
40. Brody, Samuel and Noemie Elhadad.一种无监督的方面-情感模型用于在线评论. 在北美*ACL*年年会会议中的论文. 2010年.
41. Brooke, Julian, Milan Tofiloski, and Maite Taboada. 跨语言情感分析: 从英语到西班牙语. 在*RANLP*会议上.
2009年.
42. Burfoot, Clinton, Steven Bird, 和 Timothy Baldwin. 国会辩论记录的集体分类. 在计算语言学协会第49届年会(*ACL-2011*). 2011年.
43. Carenini, Giuseppe, Raymond Ng, 和 Adam Pauls. 评价文本的多文档摘要. 在欧洲协会计算语言学分会 (*EACL-2006*). 2006年.
44. Carenini, Giuseppe, Raymond Ng, 和 Ed Zwart.从评价文本中提取知识. 在知识捕获第三届国际会议 (*K-CAP-05*). 2005年.
45. Carvalho, Paula, Luís Sarmento, Jorge Teixeira, 和 Mário J. Silva.政治辩论评论的情感标注语料库中的说谎者和救世主. 在计算语言学协会第49届年会:短论文. 2011年.
46. Castellanos, Malu, Umeshwar Dayal, Meichun Hsu, Riddhiman Ghosh, Mohamed Dekhil, Yue Lu, Lei Zhang, and Mark Schreiman. *LCT*: 一种用于实时客户智能的社交渠道分析平台. 在2011年国际数据管理会议论文集(*SIGMOD-2011*). 2011年.
47. Castillo, Carlos 和 Brian D. Davison. 对抗性网络搜索 . Foundations and Trends in Information Retrieval, 2010. 4(5): p. 377-486.

48. Chaudhuri, Arjun. 消费者行为中的情感和理性2006: Elsevier Butterworth-Heinemann.
49. Chen, Bi, Leilei Zhu, Daniel Kifer, and Dongwon Lee. 关于什么是观点的探索? 使用观点评分模型探索政治立场. 在 人工智能AAAI会议论文集(AAAI-2010)中. 2010.
50. 陈宇波和谢锦红. 在线消费者评论: 口碑作为营销传播组合的新要素. 管理科学, 2008. **54**(3): p. 477-491.
51. 崔叶进, 埃里克·布雷克和克莱尔·卡迪. 联合提取实体和关系以识别观点. 在会议论文集中 经验方法在自然语言处理中的应用 (EMNLP-2006) 。 2006.
52. 崔叶进和克莱尔·卡迪. 使用整数线性规划调整极性词典进行领域特定情感分类. 在 2009年经验方法在自然语言处理中的会议论文集 (EMNLP-2009) 。 2009年。
53. 崔叶进和克莱尔·卡迪. 分层顺序学习提取观点及其属性. 在计算语言学协会年会论文集 (ACL-2010) 。 2010年。
54. Choi, Yejin和Claire Cardie. 学习组合语义作为子句情感分析的结构推理. 在自然语言处理中的经验方法会议论文集 (EMNLP-2008) 。 2008年。
55. Choi, Yejin, Claire Cardie, Ellen Riloff和Siddharth Patwardhan. 使用条件随机场和提取模式识别观点来源. 在人类语言技术会议和经验方法会议论文集 (HLT/EMNLP-2005) 。 2005年。
56. Cilibrasi, Rudi L.和Paul M. B. Vitanyi. 谷歌相似度距离. IEEE知识与数据工程交易, 2007年. **19** (3): p. 370-383.
57. Cui, Hang, Vibhu Mittal和Mayur Datar. 在线产品评论的情感分类比较实验. 在AAAI-2006会议论文集中. 2006年。
58. Das, Dipanjan. 关于自动文本摘要的调查-单一文档摘要语言, 2007. **4**: p. 1-31.
59. Das, Sanjiv and Mike Chen. 雅虎! 亚马逊: 从股票留言板中提取市场情绪. 在 APFA-2001会议记录. 2001.
60. Das, Sanjiv and Mike Chen. 雅虎! 亚马逊: 从网络闲聊中提取情感管理科学, 2007. **53**(9): p. 1375-1388.
61. Dasgupta, Sajib and Vincent Ng. 挖掘简单的, 分类困难的: 半监督方法用于自动情感分类. 在ACL和AFNLP的第47届年会和第4届IJCNLP会议记录(ACL-2009). 2009.
62. 戴夫, 库沙尔, 史蒂夫·劳伦斯和大卫·M·佩诺克. 挖掘花生画廊: 产品评论的观点提取和语义分类. 在国际万维网会议论文集 (WWW-2003). 2003.
63. 大卫多夫, 德米特里, 奥伦·苏尔和阿里·拉波波特. 使用Twitter标签和表情符号进行增强情感学习. 在Coling-2010会议论文集. 2010.
64. Dellarocas, C., X.M. Zhang, and N.F. Awad. 探索在线产品评论在销售预测中的价值: 以电影为例. 互动营销杂志, 2007. **21**(4): p. 23-45.

情感分析与观点挖掘

65. 迪, 利皮卡和S K Mirajul Haque。从嘈杂文本数据中进行观点挖掘。在第二届嘈杂非结构化文本数据分析研讨会论文集 (AND-2008)。2008。
66. 丁晓文和刘冰。解决观点挖掘中的对象和属性共指问题。在国际计算语言学会议 (COLING-2010) 的论文集中。2010年。
67. 丁晓文, 刘冰和Philip S. Yu。一种基于整体词典的观点挖掘方法。在网络搜索和网络数据挖掘会议 (WSDM-2008) 的论文集中。2008年。
68. 丁晓文, 刘冰和张磊。用于观点挖掘应用的实体发现和分配。在 ACM SIGKDD国际知识发现与数据挖掘会议 (KDD-2009) 的论文集中。2009年。
69. Dowty, David R., Robert E. Wall和Stanley Peters。蒙塔古语义学导论。第11卷。1981年: 斯普林格。
70. Dragut, Eduard C., Clement Yu, Prasad Sistla, and Weiyi Meng. 构建情感词典。在ACM国际信息与知识管理会议(CIKM-2010)。2010年。
71. 杜伟富和谭松波。通过改进的信息瓶颈方法构建面向领域的情感词典。在ACM信息与知识管理会议 (CIKM-2009)。2009年。ACM.72. 杜伟富, 谭松波, 程学琪和云晓春。自适应信息瓶颈方法用于自动构建面向领域的情感词典。在ACM国际网络搜索与数据挖掘会议 (WSDM-2010)。2010年。
73. Duh, Kevin, Akinori Fujino, and Masaaki Nagata。机器翻译是否适合跨语言情感分类? 在第49届年度会议 的 the 关联 为了 计算语言学: 短论文 (ACL-2011)。2011年。
74. 江口浩司和维克多·拉夫伦科。使用生成模型进行情感检索。在会议论文集的经验方法会议 (EMNLP-2006)。2006年。
75. Esuli, Andrea和Fabrizio Sebastiani。确定观点挖掘的术语主观性和术语方向。在欧洲计算语言学协会会议(EACL-2006)的会议论文中。2006年。
76. Esuli, Andrea和Fabrizio Sebastiani。通过词汇分类确定术语的语义方向。在 ACM国际信息与知识管理会议(CIKM-2005)的会议论文中。2005年。
77. Esuli, Andrea和Fabrizio Sebastiani。SentiWordNet: 一种公开可用的用于观点挖掘的词汇资源。在语言资源和评估会议 (LREC-2006) 中。2006年。
78. Feldman, Ronen, Benjamin Rosenfeld, Roy Bar-Haim和Moshe Fresko。基于混合方法的股票情感分析 - 基于股票的情感分析。在第23届人工智能应用国际会议 (IAAI-2011) 中。2011年。
79. Feng, Song, Ritwik Bose和Yejin Choi。使用基于图的算法学习词语的一般内涵。在自然语言处理经验方法会议 (EMNLP-2011) 中。2011。
80. Fiszman, Marcelo, Dina Demner-Fushman, Francois M. Lang, Philip Goetz和Thomas C. Rindfleisch。解释生物医学文本中的比较结构。在BioNLP会议中。2007年。

情感分析与观点挖掘

81. Frantzi, Katerina, Sophia Ananiadou, and Hideki Mima. 自动识别多词术语 : *C-value/NC-value*方法. 国际数字图书馆杂志, 2000. 3(2): p. 115-130.
82. Gamon, Michael. 客户反馈数据的情感分类: 嘈杂数据, 大特征向量, 和语言分析的作用. 在国际计算语言学会议论文集 (*COLING-2004*). 2004.
83. Gamon, Michael, Anthony Aue, Simon Corston-Oliver, and Eric Ringger. *Pulse*: 从自由文本中挖掘客户观点智能数据分析进展 VI, 2005: p. 121-132.
84. Ganapathibhotla, Murthy and Bing Liu. 在比较句中挖掘观点. 在国际计算语言学会议论文集 (*COLING-2008*). 2008.
85. Ganesan, Kavita, ChengXiang Zhai, 和 Jiawei Han. *Opinosis*: 一种基于图的方法来抽象总结高度冗余的观点. 在第23届国际计算语言学会议 (*COLING-2010*) . 2010年.
86. Ganter, Viola 和 Michael Strube. 通过追逐黄鼠狼来寻找避讳词: 使用维基百科标签和浅层语言特征进行避讳词检测. 在 *ACL-IJCNLP 2009*会议论文集, 短论文. 2009年.
87. Gao, Sheng 和 Haizhou Li. 一种用概率潜在分析进行情感分类的跨领域适应方法. 在ACM信息与知识管理会议 (*CIKM-2011*) . 2011年.
88. Ghahramani, Zoubin 和 Katherine A. Heller. 贝叶斯集合. 在 神经信息处理系统进展, 2006年. 18: p. 435.
89. Ghani, Rayid, Katharina Probst, Yan Liu, Marko Krema, 和 Andrew Fano. 用于产品属性提取的文本挖掘. *ACM SIGKDD Explorations Newsletter*, 2006. 8(1): p. 41-48.
90. Ghose, Anindya 和 Panagiotis G. Ipeirotis. 设计新颖的评论排名系统: 预测评论的有用性和影响力. 在国际电子商务会议论文集. 2007.
91. Ghose, Anindya 和 Panagiotis G. Ipeirotis. 估计产品评论的有用性和经济影响: 挖掘文本和评论者特征. *IEEE Transactions on Knowledge and Data Engineering*, 2010.
92. Ghose, Anindya, Panagiotis G. Ipeirotis, 和 Arun Sundararajan. 使用计量经济学进行观点挖掘: 声誉系统的案例研究. 在计算语言学协会会议论文集 (*ACL-2007*). 2007.
93. Gibbs, Raymond W和Herbert L. Colston. 语言和思维中的讽刺: 认知科学读本, 2007年: 劳伦斯埃尔蒙. 94. Gibbs, Raymond W. 讽刺的心理语言学. 《实验心理学杂志: 普通版》, 1986年. 115(1): 第3页.
95. Goldberg, Andrew B.和Xiaojin Zhu. 当没有很多明星时看到明星: 基于图的半监督学习用于情感分类. 在 *HLT-NAACL 2006*研讨会上的文本图形: 基于图的自然语言处理算法. 2006年.
96. González-Ibáñez, Roberto, Smaranda Muresan和Nina Wacholder. 在Twitter中识别讽刺: 仔细观察. 在第49届年会上的论文会议 的 *the* 关联 为了 计算语言学 : 短论文 (*ACL-2011*) . 2011年.

情感分析与观点挖掘

97. Greene, Stephan和Philip Resnik. 不仅仅是词语: 句法包装和隐含情感。在人类语言技术的论文中: 北美ACL分会的2009年年会 (NAACL-2009)。2009年。
98. Griffiths, Thomas L. 和 Mark Steyvers. 预测和语义关联. 在神经信息处理系统15. 2003年.
99. Griffiths, Thomas L., Mark Steyvers, David M. Blei, 和 Joshua B. Tenenbaum. 整合主题和语法. 神经信息处理系统进展, 2005年 17: p. 537-544.
100. Groh, Georg 和 Jan Hauffa. 通过基于自然语言处理的情感分析来表征社会关系. 在第五届国际AAAI网络日志和社交媒体会议(ICWSM-2011). 2011年.
101. Guo, Honglei, Huijia Zhu, Zhili Guo, Xiaoxun Zhang, 和 Zhong Su. *OpinionIt*: 一个用于跨语言观点分析的文本挖掘系统. 在 ACM信息和知识管理会议 (CIKM-2010). 2010年.
102. 郭洪磊, 朱慧佳, 郭志立, 张晓勋和苏忠. 产品特征分类与多级潜在语义关联. 在ACM国际信息和知识管理会议 (CIKM-2009) 中。2009年。
103. 海, 张琪, 金正宰. 通过共现关联规则挖掘隐含特征。计算语言学和智能文本处理, 2011年: 第393-404页。
104. 汉考克, 杰弗里T., 劳伦E. 库里, Saurabh Goorha和Michael Woodworth. 关于说谎和被骗: 计算机中介通信中的语言分析。话语过程, 2007年。45 (1): 第1-23页。
105. 哈迪斯蒂, 埃里克 A., 乔丹·博伊德-格拉伯和菲利普·雷斯尼克. 使用适配器语法建模透视。在 2010年自然语言处理会议论文集 (EMNLP-2010) 中。2010年。
106. Hassan, Ahmed, Amjad Abu-Jbara, Rahul Jha, and Dragomir Radev. 识别外语词的语义倾向. 在计算语言学协会第49届年会短文集 (ACL-2011). 2011.
107. Hassan, Ahmed, Vahed Qazvinian, and Dragomir Radev. 态度是怎么回事? : 识别在线讨论中带有态度的句子. 在2010年经验方法在自然语言处理中的应用会议论文集 (EMNLP-2010). 2010.
108. Hassan, Ahmed and Dragomir Radev. 使用随机游走识别文本极性. 在计算语言学协会年会论文集 (ACL-2010). 2010.
109. Hatzivassiloglou, Vasileios, Judith L. Klavans, Melissa L. Holcombe, Regina Barzilay, Min-Yen Kan, and Kathleen R. McKeown. *Simfinder*: 一个灵活的聚类工具用于摘要. 在NAACL-01摘要研讨会论文集中. 2001.
110. Hatzivassiloglou, Vasileios和Kathleen R. McKeown. 预测形容词的语义倾向. 在计算语言学协会年会论文集 (ACL-1997). 1997年.
111. Hatzivassiloglou, Vasileios和Janyce Wiebe. 形容词的倾向和可分级性对句子主观性的影响. 在国际计算语言学会议论文集 (COLING-2000). 2000.

情感分析与观点挖掘

112. He, Yulan.从标记特征中学习情感分类模型. 在ACM信息与知识管理会议论文集 (CIKM-2011). 2010年.
113. He, Yulan, Chenchua Lin和Harith Alani. 自动提取跨领域情感分类的极性主题. 在计算语言学协会第49届年会论文集 (ACL-2011). 2011年.
114. Hearst, Marti. 基于方向的文本解释作为信息访问
细化, 在基于文本的智能系统, P. Jacobs, 编辑 1992,
劳伦斯埃尔蒙联合出版社. 第257-274页.
115. Hobbs, Jerry R. and Ellen Riloff. 信息提取, 在自然语言处理手册, 第二版, N. Indurkha 和 F.J.Damerau, 编辑. 2010, Chapman & Hall/CRC Press.
116. Hofmann, Thomas. 概率潜在语义索引. 在不确定性人工智能会议论文集 (UAI-1999). 1999年.
117. 洪廷成和Steven Skiena. 博彩的智慧? 情感分析与NFL点差. 在国际博客和社交媒体会议论文集 (ICWSM-2010). 2010年.
118. 胡敏清和刘冰挖掘和总结客户评论. 在
ACM SIGKDD国际知识
发现和数据挖掘会议论文集 (KDD-2004). 2004年.
119. 胡南, 保罗·A·帕夫洛和张珍妮弗在线评论能揭示一个产品的真实质量吗? : 在线口碑传播的实证研究和分析建模. 在电子商务会议论文集 (EC-2006). 2006年.
120. 黄炫静和W. Bruce Croft. 一个统一的相关性模型用于
观点检索. 在ACM信息和
知识管理会议论文集 (CIKM-2009). 2009年.
121. 池田大辅, 高村宏也, Lev-Arie Ratinov和奥村学学习改变词语的极性用于情感分类. 在第三届国际联合会议论文集自然语言处理 (IJCNLP-2008). 2008年.
122. Indurkha, Nitin和Fred J. Damerau. 自然语言处理手册, 2010年: 第二版, Chapman & Hall.
123. Jakob, Niklas和Iryna Gurevych. 使用条件随机场在单一和跨领域环境中提取观点目标. 在会议论文集的经验方法会议 (EMNLP-2010). 2010年.
124. 贾立峰, Clement Yu和孟维义. 否定对情感分析和检索效果的影响. 在第18届ACM信息与知识管理会议 (CIKM-2009). 2009年.
125. Jiang, Jay J.和David W. Conrath. 基于语料库统计和词汇分类的语义相似性. 在计算语言学研究论文集中. 1997年.
126. 姜, 龙, 莫宇, 周明, 刘晓华和赵铁军. 目标依赖的推特情感分类. 在第49届计算语言学协会年会 (ACL-2011) 中. 2011年.
127. Jijkoun, Valentin, Maarten de Rijke和Wouter Weerkamp. 生成专注于特定主题的情感词典. 在计算语言学协会年会 (ACL-2010) 的论文集中. 2010.

情感分析与观点挖掘

128. 金伟和洪海河。一种新颖的基于词汇化的HMM学习框架用于网络观点挖掘。在国际机器学习会议 (ICML-2009) 的论文集中。2009年。
129. Jindal, Nitin和Bing Liu。在文本文档中识别比较句。在 ACM SIGIR会议上进行的研究和信息检索开发 (SIGIR-2006) 的论文集中。2006年。
130. 金德尔, 尼廷和刘冰。挖掘比较句子和关系。在人工智能国家会议论文集 (AAAI-2006)中。2006b。
131. 金德尔, 尼廷和刘冰。观点垃圾邮件和分析。在Web搜索和Web数据挖掘会议论文集 (WSDM-2008)中。2008。
132. 金德尔, 尼廷和刘冰。评论垃圾邮件检测。在WWW会议论文集 (海报论文)中。2007年。
133. 金德尔, 尼廷, 刘冰和林怡鹏。使用意外规则发现不寻常的评论模式。在ACM国际信息与知识管理会议论文集 (CIKM-2010)中。2010。
134. 乔, 尤汉和爱丽丝奥。在线评论分析的方面和情感统一模型。在ACM Web搜索和数据挖掘会议论文集 (WSDM-2011)中。2011年。
135. Joachims, Thorsten. 使大规模SVM学习变得实用, 在核方法进展 - 支持向量学习, B. Schölkopf, C. Burges, 和 A. Smola, 编辑。1999, MIT出版社。
136. Johansson, Richard 和 Alessandro Moschitti. 在细粒度观点分析中的重新排序模型。在国际计算语言学会议论文集 (COLING-2010)。2010。
137. Joshi, Mahesh, Dipanjan Das, Kevin Gimpel, 和 Noah A. Smith. 电影评论和收入: 一项文本回归实验。在北美计算语言学协会人类语言技术会议论文集 (NAACL 2010)。2010。
138. Kaji, Nobuhiro 和 Masaru Kitsuregawa. 从HTML文档中自动构建极性标记语料库。在 COLING/ACL 2006主会议海报展示 (COLING-ACL-2006)。2006。
139. 加治, 信博和北川胜。从大量HTML文档中构建情感分析词典。在自然语言处理和计算自然语言学习的联合会议 (EMNLP-2007) 中。2007。
140. Kamps, Jaap, Maarten Marx, Robert J. Mokken和Maarten De Rijke. 使用WordNet衡量形容词的语义倾向。在LREC-2004中。2004。
141. 金山浩和名川哲也。面向领域的情感分析的全自动词典扩展。在自然语言处理会议 (EMNLP-2006) 中。2006。
142. 肯尼迪, 阿利斯泰尔和戴安娜·英克彭。使用上下文情感转移词对电影评论进行情感分类。计算智能, 2006. 22(2): p. 110-125。
143. 肯尼迪, 克里斯托弗。比较, 语义学, 在语言和语言学百科全书, 第二版, 2005年, 爱思唯尔。
144. 凯斯勒, 杰森S.和尼古拉斯·尼科洛夫。通过监督排序语言配置来定位情感表达。在会议论文集中。

情感分析与观点挖掘

第三届国际AAAI网络日志和社交媒体会议
(ICWSM-2009)。2009年。

145. 金贤德和翟成祥。在文本中生成相互矛盾观点的比较摘要。在ACM会议信息与知识管理会议(CIKM-2009)中。
2009.
146. 金正基金, 李金基和李钟赫。评估多语言可比性主观性分析系统。在计算语言学协会第48届年会(ACL-2010)中。2010年。
147. 金, 钟, 李金基, 和李钟赫。发现有区别的观点: 为情感分析测量术语权重。在ACL和AFNLP的第47届年会和第4届IJCNLP中。2009年。
148. 金, 秀敏和Eduard Hovy。在线评论中自动识别赞成和反对的原因。在COLING/ACL 2006主会议海报展示中。2006年。
149. 金, 秀敏和Eduard Hovy。Crystal: 分析网络上的预测性观点。在联合会议上的经验方法自然语言处理和计算自然语言学习(EMNLP/CoNLL-2007)。2007年。
150. 金秀敏和Eduard Hovy。确定观点的情感。在计算语言学国际会议论文集(COLING-2004)。2004年。
151. 金秀敏和Eduard Hovy。提取在线新闻媒体文本中的观点、观点持有者和主题。在自然语言处理经验方法会议论文集(EMNLP-2006)。2006年。
152. 金秀敏和Eduard Hovy。识别和分析判断观点。在北美ACL人类语言技术会议论文集。2006年。
153. 金秀敏、Patrick Pantel、Tim Chklovski和Marco Pennacchiotti。自动评估评论的有用性。在自然语言处理经验方法会议论文集(EMNLP-2006)。2006年。
154. Kleinberg, Jon M.在超链接环境中的权威来源。ACM期刊(JACM), 1999. 46(5): p. 604-632.
155. 小林望, 飯田龍, 乾健太郎, 松本裕治。通过提取主题-属性-值关系在网络上进行观点挖掘。在AAAI-CAAW'06会议中。2006。
156. 小林望, 乾健太郎, 松本裕治。在观点挖掘中提取方面-评价和方面-关系。在2007年联合会议上的经验方法和计算自然语言学习。2007。
157. Kouloumpis, Efthymios, Theresa Wilson, 和 Johanna Moore。Twitter情感分析: 好的、坏的和OMG! inWeblogs和Social Media(ICWSM-2011)的第五届国际AAAI会议。2011。
158. Kovelamudi, Sudheer, Sethu Ramalingam, Arpit Sood, and Vasudeva Varm a.使用维基百科的用户评论领域无关模型进行产品属性提取。在第五届国际自然语言处理联合会议(IJCNLP-2010)的论文中。2011年。
159. Kreuz, Roger J and Gina M Caucchi.词汇对讽刺感知的影响。在计算机处理比喻语言研讨会的论文中。2007年。

情感分析与观点挖掘

160. Kreuz, Roger J. and Sam Glucksberg. 如何讽刺: 语音回声理论的口头反讽。《实验心理学杂志》, 1989年。118(4): p. 374。
161. Ku, Lun-Wei, Yu-Ting Liang, and Hsin-Hsi Chen. 新闻和博客语料库中的观点提取、总结和跟踪。在AAAI-CAAW'06的论文中。2006年。
162. Lafferty, John, Andrew McCallum, and Fernando Pereira. 条件随机场: 用于分割和标记序列数据的概率模型。在国际机器学习大会(ICML-2001)。2001。
163. Lakkaraju, Himabindu, Chiranjib Bhattacharyya, Indrajit Bhattacharya, and Srujana Merugu. 利用连贯性同时发现潜在的方面和相关的情感。在SIA M数据挖掘大会 (SDM-2011)。2011。
164. Lappas, Theodoros and Dimitrios Gunopulos. 在大型评论语料库中高效的自信搜索。在ECML-PKDD 2010大会。2010。
165. Lee, Lillian. 分布相似度的度量。在计算语言学协会年会 (ACL-1999)。1999。
166. Lerman, Kevin, Sasha Blair-Goldensohn, and Ryan McDonald. 情感总结: 评估和学习用户偏好。在第12届欧洲计算语言学协会会议(EACL-2009)中。2009年。
167. Lerman, Kevin and Ryan McDonald. 对比总结: 消费者评论实验。在NA ACL HLT 2009会议论文集中: 短论文。2009年。
168. Li, Binyang, Lanjun Zhou, Shi Feng, and Kam-Fai Wong. 基于句子的观点检索的统一图模型。在计算语言学协会年会论文集 (ACL-2010)中。2010年。
169. Li, Fangtao, Chao Han, Minlie Huang, Xiaoyan Zhu, Ying-Ju Xia, ShuZhang, and Hao Yu. 结构感知的评论挖掘和总结。在第23届国际计算语言学会议 (COLING-2010)中。2010年。
170. 李方涛, 黄敏烈, 杨毅和朱晓燕. 学习如何识别评论垃圾。在人工智能国际联合会议(IJCAI-2011)的论文集中。2011年。
171. 李方涛, 黄敏烈和朱晓燕. 情感分析与全局主题和局部依赖关系。在第二十四届人工智能大会(AAAI-2010)的论文集中。2010年。
172. 李俊辉, 周国栋, 王红玲和朱巧明. 通过浅层语义解析学习否定的范围。在第23届国际计算语言学大会(COLING-2010)的论文集中。2010年。
173. 李莎莎, 林金友, 宋永寅和李周军. 从比较性问题中挖掘可比较实体。在计算语言学协会年会(ACL-2010)的论文集中。2010。
174. 李守山, 黄初仁, 周国栋和李一美. 在监督和半监督的情感分类中采用个人/客观观点。在计算语言学协会 (ACL-2010) 年会论文集中。2010年。
175. 李守山, 李一美, 陈颖, 黄初仁和周国栋. 情感分类和极性转移。在

情感分析与观点挖掘

第23届国际计算语言学会议 (COLING-2010) 论文集中。2010年。

176. 李守山, 王中庆, 周国栋和李一美。用于不平衡情感分类的半监督学习。在人工智能国际联合会议 (IJCAI-2011) 论文集中。2011年。
177. 李, 陶, 张毅和Vikas Sindhwani。一种基于非负矩阵三因子分解的情感分类方法, 具有词汇先验知识。在计算语言学协会年会论文集 (ACL-2009) 中。2009年。
178. 李晓丽, 张磊, 刘冰和吴思琼。分布相似性与PU学习用于实体集扩展。在计算语言学协会年会论文集 (ACL-2010) 中。2010年。
179. 林怡鹏, Nguyen Viet-An, Nitin Jindal, 刘冰和Hady W.Lauw。使用评分行为检测产品评论垃圾邮件。在ACM国际信息与知识管理会议论文集 (CIKM-2010) 中。2010年。
180. 林成华和何玉兰。情感/主题模型用于情感分析。在ACM国际信息和知识管理会议 (CIKM-2009) 中。2009年。
181. 林德康。自动检索和聚类相似词。在计算语言学协会第36届年会和第17届国际计算语言学会议 (COLING-ACL-1998) 中。1998年。
182. 林德康。Minipar。 <http://webdocs.cs.ualberta.ca/lindek/minipar.htm>。2007年。
183. 林伟豪, 特蕾莎·威尔逊, 詹妮丝·威贝和亚历山大·豪普特曼。你站在哪一边? : 在文档和句子级别上识别观点。在自然语言学习会议 (CoNLL-2006) 中。2006年。
184. 刘兵。情感分析和主观性, 在自然语言处理手册, 第二版, N. Indurkha和F.J. Damerau, 编辑。2010年。
185. 刘兵。Web数据挖掘: 探索超链接、内容和使用数据, 2006年和2011年: Springer。
186. 刘兵, Wynne Hsu和Yiming Ma。集成分类和关联规则挖掘。在国际会议论文集上知识发现与数据挖掘 (KDD-1998)。1998年。
187. 刘兵, 胡敏清和程俊生。观点观察员: 分析和比较网络上的观点。在国际会议论文集上万维网 (WWW-2005)。2005年。
188. 刘兵, 李伟新, Philip S. Yu和Xiao-Li Li。部分监督文本文档分类。在国际会议论文集上机器学习 (ICML-2002)。2002年。
189. 刘飞凡, 李斌和刘洋。使用统计分类器和词汇特征找到有观点的博客。在第三届国际AAAI网络日志和社交媒体会议 (ICWSM-2009) 中。2009年。
190. 刘飞凡, 王东, 李斌和刘洋。通过主题分析和自适应方法改进博客极性分类。在北美ACL (HLT-NAACL-2010) 的人类语言技术会议的论文集中。2010年。
191. 刘晶晶, 曹云波, 林钦业, 黄亚楼和周明。在观点摘要中检测低质量产品评论。在

情感分析与观点挖掘

- 经验方法和计算自然语言学习的联合会议 (EMNLP-CoNLL-2007) 中。2007年。
192. 刘静静和斯蒂芬妮·塞尼夫通过解析和改写范式进行评论情感评分。在2009年经验方法在自然语言处理中的会议论文集 (EMNLP-2009) 中。2009。
193. 刘阳, 黄祥基, 安爱军和于晓辉通过博客进行销售业绩预测的情感感知模型。在ACM SIGIR信息检索研究与开发会议 (SIGIR-2007) 的论文集中。2007。
194. 刘阳, 黄祥基, 安爱军和于晓辉对在线评论的有用性进行建模和预测。在ICDM-2008会议论文集中。2008。
195. 龙冲, 张杰和朱晓燕的准确特征评级的评论选择方法。在Coling 2010会议论文集的海报卷中。2010。
196. 陆斌。使用依赖解析器在中国新闻文本中识别观点持有者和目标。在人类语言技术会议论文集: 北美ACL分会的2010年年会 (HLT-NAACL-2010)。2010年。
197. 陆斌, 谭晨昊, Claire Cardie和Benjamin K. Tsou。使用无标签平行语料库进行联合双语情感分类。在计算语言学协会第49届年会论文集 (ACL-2011)。2011年。
198. 陆越, Malu Castellanos, Umeshwar Dayal和ChengXiang Zhai。一种上下文感知情感词典的自动构建方法: 一种优化方法。在第20届国际会议论文集 on World wide web (WWW-2011)。2011年。
199. 卢, 岳, 段慧中, 王宏宁和翟成祥。利用结构化本体来组织分散的在线观点。在计算语言学国际会议论文集 (COLING-2010)。2010年。
200. 卢, 岳, Panayiotis Tsaparas, Alexandros Ntoulas和Livia Polanyi。利用社交背景进行评论质量预测。在国际万维网会议论文集 (WWW-2010)。2010年。
201. 卢, 岳和翟成祥。通过半监督主题建模进行观点整合。在世界互联网大会论文集 (WWW-2008)。2008年。
202. 卢, 岳, 翟成祥和Neel Sundaresan。对短评论进行评级方面的总结。在世界互联网大会论文集 (WWW-2009)。2009年。
203. 马腾飞和万晓军。在中国新闻评论中提取观点目标。在Coling 2010海报集 (COLING-2010)中。2010。
204. Maas, Andrew L., Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 学习情感分析的词向量。在第49届计算语言学年会 (ACL-2011)中。2011年。
205. Macdonald, Craig, Iadh Ounis, and Ian Soboroff. TREC 2007博客赛道概述。2007年。
- Manevitz, Larry M. and Malik Yousef. 用于文档分类的单类支持向量机。机器学习研究杂志, 2002年。2: 第139-154页。

情感分析与观点挖掘

207. Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schutze. 信息检索导论. 第1卷. 2008年: 剑桥大学出版社.
208. Manning, Christopher D. and Hinrich Schutze. 统计自然语言处理基础. 第999卷. 1999年: MIT出版社.
209. Martineau, Justin and Tim Finin. *De lta tfidf*: 一种改进的情感分析特征空间. 在第三届国际AAAI网络日志和社交媒体会议 (ICWSM-2009). 2009年.
210. McDonald, Ryan, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. 用于细粒度情感分析的结构化模型. 在年度计算语言学协会会议 (ACL-2007). 2007年.
211. McGlohon, Mary, Natalie Glance, and Zach Reiter. 星级质量: 汇总评论以排名产品和商家. 在国际博客和社交媒体会议 (ICWSM-2010) 中. 2010年.
212. Medlock, Ben and Ted Briscoe. 科学文献中的弱监督学习用于对冲分类. 在计算语言学协会第45届年会上. 2007年.
213. Mei, Qiaozhu, Xu Ling, Matthew Wondra, Hang Su, and ChengXiangZhai. 主题情感混合: 建模网络日志中的方面和观点. 在国际万维网会议 (WWW-2007) 中. 2007年.
214. Mejova, Yelena and Padmini Srinivasan. 探索情感分类器的特征定义和选择. 在第五届国际AAAI博客和社交媒体会议 (ICWSM-2011) 中. 2011年.
215. 孟新凡和侯凤王. 挖掘用户评论: 从规范到总结. 在ACL-IJCNLP 2009年会议短文集中. 2009年.
216. Mihalcea, Rada, Carmen Banea和Janyce Wiebe. 通过跨语言投影学习多语言主观语言. 在计算语言学协会年会 (ACL-2007) 会议论文集中. 2007年.
217. Mihalcea, Rada和Carlo Strapparava. 测谎仪: 自动识别欺骗性语言的探索. 在ACL-IJCNLP 2009年会议短文集中. 2009年.
218. Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, 和 Katherine Miller. *WordNet*: 一个在线词汇数据库 1990年: 牛津大学出版社.
219. 米勒, 马哈利亚, 科纳尔萨蒂, 丹尼尔维森塔尔, 朱尔莱斯科维奇和克里斯托弗波茨. 情感通过超链接网络流动. 在第五届国际AAAI网络日志和社交媒体会议 (ICWSM-2011). 2011年.
220. 闵慧真和朴钟成. 检测和阻止虚假情感传播. 在第5届国际自然语言处理联合会议 (IJCNLP-2010). 2011年.
221. 米切尔, 汤姆. 机器学习 1997年: 麦格劳希尔.
222. 莫哈达姆, 萨玛内和马丁埃斯特. *ILDA*: 依赖 LDA 模型用于从在线产品评论中学习潜在方面及其评级. 在年度 ACM SIGIR 国际研究与开发会议 (SIGIR-2011). 2011年.
223. Moghaddam, Samaneh和Martin Ester. *Opinion digger*: 一种无监督的从非结构化产品评论中挖掘观点的挖掘器. 在 *Proceeding of the*

情感分析与观点挖掘

- ACM信息与知识管理会议 (CIKM-2010)。2010年。
224. Moghaddam, Samaneh, Mohsen Jamali和Martin Ester。ETF: 扩展的张量分解模型用于个性化预测评论的帮助。在ACM国际网络搜索和数据挖掘会议 (WSDM-2012)。2012年。
225. Mohammad, Saif。从从前到幸福结局: 追踪小说和童话中的情绪。在ACL 2011年语言技术与文化遗产、社会科学和人文学研讨会 (LaTeCH)。2011年。
226. Mohammad, Saif和Tony Yang。追踪邮件中的情感: 性别在情感轴上的差异。在ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA-2011)的论文集中。2011年。
227. Mohammad, Saif, Cody Dunne和Bonnie Dorr。从明确标记的词和词典中生成高覆盖的语义倾向词典。在2009年经验方法在自然语言处理中的会议 (EMNLP-2009) 中。2009年。
228. Mohammad, Saif和Graeme Hirst。概念距离的分布式度量: 面向任务的评估。在经验方法在自然语言处理中的会议 (EMNLP-2006) 的论文集中。2006。
- Mohammad, Saif M.和Peter D. Turney。常见词语引发的情绪: 使用机械土耳其人创建情绪词典。在 NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text的论文集中。2010年。
230. 莫伊拉宁, 卡罗和斯蒂芬·普尔曼。情感组合。在最近自然语言处理进展会议论文集 (RANLP 2007) 中。2007年。
231. 蒙塔格, 理查德。形式哲学; 理查德蒙塔格的选集, 1974年: 耶鲁大学出版社。
232. 莫尼, 雷蒙德J.和拉兹万·布内斯库。从文本中挖掘知识使用信息提取。ACM SIGKDD探索通讯, 2005年。7(1): 第3-10页。
233. 莫兰特, 罗瑟, 萨拉·斯克劳文和沃尔特·达勒曼斯。基于语料库的处理否定线索范围的方法: 现有技术的评估。在第九届国际计算语义学会议论文集 (IWCS-2011) 中。2011年。
234. 森永, 聪, 山西健二, 立石健二, 和福岛敏和在网络上挖掘产品声誉。在ACM SIGKDD国际知识发现和数据挖掘会议 (KDD-2002)。2002年。
235. Mukherjee, Arjun 和 Bing Liu。通过半监督建模进行方面提取。在计算语言学协会第50届年会 (ACL-2012) (已接受发表)。2012。
236. Mukherjee, Arjun 和 Bing Liu。建模评论评论。在计算语言学协会第50届年会 (ACL-2012) (已接受发表)。2012年。
237. Mukherjee, Arjun, Bing Liu 和 Natalie Glance。在消费者评论中发现虚假评论者群体。在国际万维网会议 (WWW-2012)。2012年。
238. Mukherjee, Arjun, Bing Liu, Junhui Wang, Natalie Glance, and Nitin Jindal。检测群组评论垃圾。在国际万维网会议论文集 (WWW-2011, 海报论文)。2011。

情感分析与观点挖掘

239. Mukund, Smruthi and Rohini K. Srihari. 一种基于共训练的乌尔都语主观性分类的向量空间模型. 在 *Coling 2010: 海报论文集*. 2010.
240. Mullen, Tony and Nigel Collier. 使用多样信息源的支持向量机进行情感分析. 在 *EMNLP-2004会议论文集*. 2004.
241. Murakami, Akiko and Rudy Raymond. 支持还是反对?: 从回复活动和观点表达中分类在线辩论的立场. 在 *Coling 2010: 海报论文集*. 2010.
242. Na, Seung-Hoon, Yeha Lee, Sang-Hyob Nam, and Jong-Hyeok Lee. 基于查询特定情感词典的观点检索改进
信息检索的进展, 2009年: 第734-738页。
243. Nakagawa, Tetsuji, Kentaro Inui, and Sadao Kurohashi. 基于依赖树的情感分类使用具有隐藏变量的CRFs。在人类语言技术的论文集: 北美ACL分会的2010年年会 (HAACL-2010)。
- 2010.
244. Narayanan, Ramanathan, Bing Liu, and Alok Choudhary. 条件句的情感分析. 在会议论文集中自然语言处理的经验方法 (EMNLP-2009)。
- 2009.
245. Nasukawa, Tetsuya and Jeonghee Yi. 情感分析: 利用自然语言处理捕捉偏好. 在知识捕获的第二届K-CAP-03国际会议论文集。2003年。
246. Neviarouskaya, Alena, Helmut Prendinger, and Mitsuru Ishizuka. 从文本中识别细粒度情绪的组性原则. 在第三届国际博客和社交媒体会议(ICWSM-2009)中. 2009.
247. Neviarouskaya, Alena, Helmut Prendinger, and Mitsuru Ishizuka. 在文本中识别情感、判断和欣赏. 在第23届国际计算语言学会议(COLING-2010)中. 2010.
248. Newman, Matthew L., James W. Pennebaker, Diane S. Berry, and Jane M. Richards. 谎言的言辞: 从语言风格预测欺骗. *Personality and Social Psychology Bulletin*, 2003. **29**(5): p. 665.
249. Ng, Vincent and Claire Cardie. 改进机器学习方法来解决共指问题. 在年度计算语言学协会会议(ACL-2002)中. 2002.
250. Ng, Vincent, Sajib Dasgupta, 和 S. M. Niaz Arifin. 检查语言知识源在自动识别和分类评论中的作用. 在COLING/ACL 2006 MainConference Poster Sessions (COLING/ACL-2006)的会议论文中. 2006.
251. Nigam, Kamal 和 Matthew Hurst. 朝着一个稳健的观点度量标准. 在AAAI Spring Symp. on Exploring Attitude and Affect in Text的会议论文中. 2004.
252. Nigam, Kamal, Andrew K. McCallum, Sebastian Thrun, 和 Tom Mitchell. 使用标记和未标记的文档进行文本分类. *机器学习*, 2000. **39**(2): p. 103-134.
253. Nishikawa, Hitoshi, Takaaki Hasegawa, Yoshihiro Matsuo, 和 Genichiro Kikui. 使用整数线性规划公式进行观点摘要的句子提取和排序. 在Coling 2010: Poster Volume的会议论文中. 2010a.
254. 西川仁, 长谷川孝明, 松尾佳宏和菊井玄一郎. 为情感优化信息量和可读性

情感分析与观点挖掘

总结. 在计算语言学协会年会论文集 (ACL-2010) 中. 2010b.

255. 奥康纳, 布伦丹, 拉姆纳斯·巴拉苏布拉马尼扬, 布莱恩·R·劳特利奇和诺亚·A·史密斯. 从推文到民意调查: 将文本情感与公众舆论时间序列联系起来. 在国际AAAI会议论文集网络日志和社交媒体 (ICWSM 2010) 中. 2010.
256. O'Mahony, Michael P.和Barry Smyth. 学习推荐有用的酒店评论. 在第三届ACM推荐系统会议论文集中. 2009.
257. Ott, Myle, Yejin Choi, Claire Cardie, 和 Jeffrey T. Hancock. 通过任何手段找到欺骗性观点垃圾邮件. 在第49届计算语言学年会 (ACL-2011). 2011年.
258. Ounis, Iadh, Craig Macdonald, Maarten de Rijke, Gilad Mishne, 和 Ian Soboroff. TREC-2006博客赛道概述. 在第十五届文本检索会议 (TREC-2006). 2006年.
259. Ounis, Iadh, Craig Macdonald, 和 Ian Soboroff. TREC-2008博客赛道概述. 在第16届文本检索会议 (TREC-2008). 2008年.
260. Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. PageRank引用排名: 为网络带来秩序. 1999.
261. Paltoglou, Georgios 和 Mike Thelwall. 情感分析的信息检索加权方案研究. 在第48届计算语言学协会年会 (ACL-2010) 中. 2010年.
262. Pan, Sinno Jialin, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang 和 Zheng Chen. 通过谱特征对齐进行跨领域情感分类. 在国际万维网会议 (WWW-2010) 中. 2010年.
263. Pang, Bo 和 Lillian Lee. 观点挖掘和情感分析. 信息检索的基础与趋势, 2008年. 2 (1-2): p. 1-135.
264. 庞博和李莉安·李. 看星星: 利用类别关系进行情感分类与评分尺度相关的挖掘. 在计算语言学协会会议 (ACL-2005) 的论文集中. 2005.
265. 庞博和李莉安·李. 情感教育: 基于最小割的主观性总结的情感分析. 在计算语言学协会会议 (ACL-2004) 的论文集中. 2004.
266. 庞博和李莉安·李. 使用非常简单的统计方法进行评论搜索: 一次探索. 在国际计算语言学会议上的海报论文 (COLING-2008) 中. 2008年.
267. 庞博, 李莉安·李和Shivakumar Vaithyanathan. 竖起大拇指?: 使用机器学习技术进行情感分类. 在自然语言处理实证方法会议 (EMNLP-2002) 的论文集中. 2002年.
268. Pantel, Patrick, Eric Crestan, Arkady Borkovsky, Ana-Maria Popescu, and Vishnu Vyas. Web规模的分布相似性和实体集扩展. 在自然语言处理中的经验方法会议论文集 (EMNLP-2009). 2009.
269. Park, Do-Hyung, Jumin Lee, and Ingo Han. 在线消费者评论对消费者购买意向的影响: 参与度的调节作用. 国际电子商务杂志, 2007. 11(4): p. 125-148.

情感分析与观点挖掘

270. Park, Souneil, KyungSoon Lee, and Junehwa Song. 对有争议问题的新闻文章进行对比. 在计算语言学协会第49届年会 (*ACL-2011*). 2011.
271. Parrott, W. Gerrod. 社会心理学中的情绪: 必读2001: 心理学 Pr.
272. Paul, Michael J., ChengXiang Zhai, and Roxana Girju. 总结观点性文本中的对比观点. 在会议论文集中的经验方法自然语言处理 (*EMNLP-2010*). 2010.
273. Peng, Wei and Dae Hoon Park. 使用约束非负矩阵分解为社交媒体情感分析生成形容词情感词典. 在 第五届国际AAA/网络日志和社交媒体会议 (*ICWSM-2011*). 2011.
274. Pennebaker, James W., Cindy K. Chung, Molly Ireland, Amy Gonzales, 和 Roger J. Booth. *LIWC2007*的开发和心理测量特性. www.LIWC.Net, 2007.
275. Polanyi, Livia和Annie Zaenen. 上下文情感转移. 在 AAA/春季研讨会上探索文本中的态度和情感. 2004年.
276. Popescu, Ana-Maria和Oren Etzioni. 从评论中提取产品特征和观点. 在 自然语言处理中的经验方法会议论文集 (*EMNLP-2005*). 2005年.
277. Qiu, Guang, Bing Liu, Jiajun Bu和Chun Chen. 通过双向传播扩展领域情感词典. 在 国际人工智能联合会议论文集 (*IJCAI-2009*). 2009.
278. Qiu, Guang, Bing Liu, Jiajun Bu和Chun Chen. 通过双向传播进行观点词扩展和目标提取. 计算语言学, 第37卷, 第1期: 9.27, 2011年.
279. 邱立坤, 张伟, 胡长建和赵凯. *Selc*: 一种用于情感分类的自监督模型. 在第18届ACM信息与知识管理会议 (*CIKM-2009*) 中. 2009年.
280. Qu, Lizhen, Georgiana Ifrim和Gerhard Weikum. 基于稀疏文本模式的评论评分预测的袋装观点方法. 在国际计算语言学会议论文集 (*COLING-2010*) 中. 2010年.
281. Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech和Jan Svartvik. 英语语言的综合语法. 第397卷. 1985年: 剑桥大学出版社.
282. Raaijmakers, Stephan和Wessel Kraaij. 主观性分类的浅层方法, 在 *ICWSM-2008*会议论文集中, 2008年. 第216-217页.
283. Raaijmakers, Stephan, Khiet Truong, 和 Theresa Wilson. 多模态主观性分析的多方对话. 在 *Empirical Methods in Natural Language Processing (EMNLP-2008)*会议论文集中. 2008.
284. Rabiner, Lawrence R. 关于隐马尔可夫模型和语音识别中的一些应用的教程IEEE会议论文集, 1989. 77(2): 第257-286页.
285. Radev, Dragomir R., Simone Teufel, Horacio Saggion, Wai Lam, John Blitzer, Hong Qi, Arda Celebi, Danyu Liu, 和 Elliott Drabek. 大规模文档摘要评估挑战. 在 Association for Computational Linguistics (*ACL-2003*)年会上. 2003.

- Annual Meeting of the Association for Computational Linguistics (ACL-2003)*. 2003.
286. Rao, Delip和Deepak Ravichandran. 半监督极性词典归纳。在第12届欧洲分会议的*ACL(EACL-2009)*。2009年。
 287. Ravichandran, Deepak和Eduard Hovy. 学习表面文本模式用于问答系统。在年度会议上计算语言学协会(*ACL-2002*)。2002年。
 288. Riloff, Ellen. 自动构建信息的字典提取任务。在*AAAI-2003*的处理中。1993年。
 289. Riloff, Ellen. 从未标记的文本中自动生成提取模式。在*AAAI-1996*的会议上。1996年。
 290. Riloff, Ellen, Siddharth Patwardhan和 Janyce Wiebe. 特征包容性的观点分析。在会议上自然语言处理中的经验方法(*EMNLP-2006*)。2006。
 291. Riloff, Ellen和Janyce Wiebe. 学习主观表达的提取模式。在会议论文集集中的经验方法会议(*EMNLP-2003*)。2003年。
 292. Ruppenhofer, Josef, Swapna Somasundaran和Janyce Wiebe. 找到主观表达的来源和目标。在*LREC*会议论文集中。2008年。
 293. Sadikov, Eldar, Aditya Parameswaran和Petros Venetis. 博客作为电影成功的预测器。在第三届国际网络日志和社交媒体会议(*ICWSM-2009*)。2009年。
 294. Sakunkoo, Patty和Nathan Sakunkoo. 在线书评中的社交影响分析。在第三届国际AAA/网络日志和社交媒体会议(*ICWSM-2009*)。2009年。
 295. Santorini, Beatrice. Penn树库项目的词性标注指南, 1990年: 宾夕法尼亚大学, 工程与应用科学学院, 计算机与信息科学系。
 296. Sarawagi, Sunita. 信息抽取。数据库的基础与趋势, 2008年。1(3): 第261-377页。
 297. Sauper, Christina, Aria Haghighi和Regina Barzilay. 带有态度的内容模型。在计算语言学协会第49届年会论文集(*ACL-2011*)中。2011年。
 298. Scaffidi, Christopher, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng和Chun Jin. *Red Opal*: 从评论中进行产品特征评分。在第十二届ACM电子商务会议论文集(*EC-2007*)中。2007年。
 299. Schapire, Robert E. 和 Yoram Singer. *BoosTexter*: 一种基于提升的文本分类系统。机器学习, 2000。39(2): p. 135-168.
 300. Seki, Yohei, Koji Eguchi, Noriko Kando, 和 Masaki Aono. 面向观点的摘要和其在*DUC 2006*的分析。在文档理解会议(*DUC*)。2006年。
 301. Shanahan, James G., Yan Qu, 和 Janyce Wiebe. 在文本中计算态度和情感的理论和应用. Vol. 20. 2006: Springer-Verlag New York Inc.
 302. Shawe-Taylor, John 和 Nello Cristianini. 支持向量机, 2000, 剑桥大学出版社。
 303. Snyder, Benjamin 和 Regina Barzilay. 使用good grief算法进行多方面排名。在北美计算语言学协会会议(*NAACL/HLT-2007*)的论文集中。2007年。

情感分析与观点挖掘

304. Socher, R., J. Pennington, E. H. Huang, A.Y. Ng, and C.D. Manning. 半监督递归自编码器用于情感分布预测. 在 自然语言处理中的经验方法会议论文集 (*EMNLP-2011*). 2011.
305. Somasundaran, S., J. Ruppenhofer, and J. Wiebe. 言论层面观点关系：一项注释研究. 在 第9届*SIGdial*讨论与对话研讨会论文集. 2008.
306. Somasundaran, Swapna, Galileo Namata, Lise Getoor, and Janyce Wiebe. 用于极性和言论分类的观点图. 在基于图的自然语言处理方法研讨会论文集中. 2009.
307. Somasundaran, Swapna和Janyce Wiebe. 在在线辩论中识别立场. 在*ACL*和*AFNLP*的第47届年会和第4届*IJCNLP (ACL-IJCNLP-2009)*中. 2009年.
308. Steyvers, Mark和Thomas L. Griffiths. 概率主题模型. 潜在语义分析手册, 2007年. 427(7): p. 424-440.
309. Stone, Philip. 普通查询器：一种计算机内容分析方法. 区域科学杂志, 1968年. 8(1).
310. Stoyanov, Veselin和Claire Cardie. 通过结构化规则学习进行部分监督的共指消解以进行观点摘要. 在自然语言处理中的经验方法会议论文集 (*EMNLP-2006*)中. 2006年.
311. Stoyanov, Veselin和Claire Cardie. 细粒度观点分析的主题识别. 在计算语言学国际会议 (*COLING-2008*) . 2008年.
312. Strapparava, Carlo和Alessandro Valitutti. WordNet-Affect: WordNet的情感扩展. 在语言资源和评估国际会议 (*LREC*) . 2004年.
313. Su, Fangzhong和Katja Markert. 从词语到意义：主观性识别的案例研究. 在第22届国际计算语言学会议 (*COLING-2008*) . 2008年.
314. Su, Fangzhong和Katja Markert. 用于跨语言词汇替换的词义主观性. 在人类语言技术：北美*ACL*年会 (*HLT-NAACL-2010*) . 2010年.
315. 苏琦, 徐欣颖, 郭宏磊, 郭志立, 吴贤, 张晓讯, 斯文, 苏忠. 隐藏的情感关联在中国网络观点挖掘中. 在国际万维网会议论文集 (*WWW-2008*) . 2008年.
316. Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 基于词典的情感分析方法. 计算语言学, 2011年. 37(2): p. 267-307.
317. Täckström, Oscar and Ryan McDonald. 使用潜在变量结构化预测模型发现细粒度情感. 信息检索进展, 2011年: p. 368-374.
318. Täckström, Oscar and Ryan McDonald. 半监督潜在变量模型用于句子级情感分析. 在第49届年会论文集中
会议 的 the 关联 为了 计算语言学
：短论文 (*ACL-2011*) . 2011年.
319. 高村, 博也, 久保田, 和久保田, 学. 从字典中提取短语的语义倾向. 在联合人类语言技术/北美*ACL*分会会议论文集 (*HLT-NAACL-2007*) . 2007年.
320. 高村, 博也, 久保田, 和
久保田, 学. 使用旋转模型提取词语的语义倾向. 在会议论文集中

情感分析与观点挖掘

计算语言学协会年会 (ACL-2005) 。2005年。

321. 高村宏也、乾健太郎、奥村学。短语语义倾向的潜在变量模型。在计算语言学协会欧洲分会会议 (EACL-2006) 中。2006年。
322. 谭、宋波、吴高伟、唐辉峰和程学琪。情感分析背景下领域转移问题的一种新方案。在信息与知识管理ACM会议论文集(CIKM-2007)中。2007年。
323. 塔塔、斯瓦蒂和巴巴拉·迪·尤金尼奥。从专辑评论中生成细粒度的歌曲评论。在计算语言学年会论文集(ACL-2010)中。2010年。
324. 特斯尼埃尔, L. 结构语法要素: *Jean Fourquet*的前言1959年: C. Klincksieck。
325. 提托夫、伊万和瑞安·麦克唐纳。文本和方面评级的联合模型用于情感摘要。在计算语言学年会论文集(ACL-2008)中。2008年。
326. Titov, Ivan和Ryan McDonald. 使用多粒度主题模型建模在线评论。在国际万维网会议 (WWW-2008) 的论文中。2008年。
327. Tokuhisa, Ryoko, Kentaro Inui和Yuji Matsumoto. 使用从网络中提取的大量示例进行情感分类。在第22届国际计算语言学会议 (COLING-2008) 的论文中。2008年。
328. Tong, Richard M. 用于检测和跟踪在线讨论中观点的操作系统。在 *SI GIR Workshop on Operational Text Classification* 的论文中。2001年。
329. Toprak, Cigdem, Niklas Jakob和Iryna Gurevych. 用户生成的话语中的句子和表达级别的观点注释。在第48届计算语言学协会年会 (ACL-2010) 的论文中。2010年。
330. Tsaparas, Panayiotis, Alexandros Ntoulas, and Evimaria Terzi. 选择一组全面的评论。在 *ACM SIGKDD 知识发现与数据挖掘会议论文集 (KDD-2011)*。2011年。
331. Tsur, Oren, Dmitry Davidov, and Ari Rappoport. 一个很有吸引力的名字: 在线产品评论中的半监督讽刺句子识别。在第四届国际AAA/博客和社交媒体会议论文集 (ICWSM-2010)。2010年。
332. Tsur, Oren and Ari Rappoport. *Revrack*: 一种完全无监督的算法, 用于选择最有帮助的书评。在国际AAA/博客和社交媒体会议论文集 (ICWSM-2009)。2009年。
333. Tumasjan, Andranik, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welp. 用 *Twitter* 预测选举: 140个字符揭示的政治情绪。在国际会议 *Weblogs and Social Media (ICWSM-2010)* 的论文中。2010年。
334. Turney, Peter D. 赞或踩?: 应用于无监督分类的语义倾向。在计算语言学年会 (ACL-2002) 的论文中。
2002.
335. Turney, Peter D. and Micharell L. Littman. 从关联中推断语义倾向: 衡量赞扬和批评 *ACM Transactions on Information Systems*, 2003.

情感分析与观点挖掘

336. Utsumi, Akira. 言语讽刺作为隐含的讽刺环境展示：区分讽刺话语和非讽刺话语。《语用学杂志》2000年 32(12): p. 1777-1806.
337. Valitutti, Alessandro, Carlo Strapparava, and Oliviero Stock. 开发情感词汇资源。 *PsychNology Journal*, 2004. 2(1): p. 61-83.
338. Velikovich, Leonid, Sasha Blair-Goldensohn, Kerry Hannan, and Ryan McDonald. 基于网络的极性词典的可行性。在 *Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics (HAACL-2010)*. 2010.
339. Vrij, Aldert. 检测谎言和欺骗：陷阱和机遇 2008: Wiley-Interscience.
340. Wan, Xiaojun. 跨语言情感分类的共同训练。在 *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP (ACL-IJCNLP-2009)*. 2009.
341. 万晓军. 使用双语知识和集成技术进行无监督的中文情感分析。在会议论文集集中的经验方法中的自然语言处理 (*EMNLP-2008*)。2008.
342. 王东和刘洋. 关于对话中观点总结的初步研究。在计算语言学协会第49届年会 (*ACL-2011*). 2011年。
343. 王冠、谢思宏、刘冰和于飞. 通过社交评论图识别在线商店评论垃圾邮件发送者。ACM智能系统与技术交易, 已接受发表, 2011年。
344. 王宏宁、陆越和翟成祥. 在评论文本数据上的潜在方面评分分析：一种评分回归方法。在ACM SIGKDD国际知识发现和数据挖掘会议论文集 (*KDD-2010*). 2010年。
345. 王, 童和格雷姆·赫斯特. 在基于WordNet的语义相似度度量中细化深度和密度的概念。在自然语言处理中的经验方法会议 (*EMNLP-2011*)。2011年。
346. 王晓龙, 魏福如, 刘晓华, 周明和张明. *Twitter*中的主题情感分析：基于图的标签情感分类方法。在ACM会议的信息和知识管理 (*CIKM-2011*) 中。2011年。
347. 魏斌和克里斯托弗·帕尔. 跨语言适应：情感分类实验。在ACL 2010会议短文集中 (*ACL-2010*) 中。2010年。
348. 魏伟和Jon Atle Gulla. 通过情感本体树对产品评论进行情感学习。在计算语言学协会年会论文集 (*ACL-2010*) 中。2010年。
349. 温, 苗苗和云芳吴. 使用自举方法挖掘名词的情感期望。在第五届国际自然语言处理联合会议(*IJCNLP-2010*). 2011年。
350. Wiebe, Janyce. 识别叙述中的主观角色。在计算语言学国际会议(*COLING-1990*). 1990年。
351. Wiebe, Janyce. 从语料库中学习主观形容词。在人工智能国家会议(*AAAI-2000*). 2000.
352. Wiebe, Janyce. 追踪叙述中的观点。计算语言学, 1994年。20: 第233-287页。

情感分析与观点挖掘

353. Wiebe, Janyce和Rada Mihalcea. 词义和主观性. 在计算语言学国际会议和第44届ACL年会(COLING/ACL-2006). 2006年。
354. Wiebe, Janyce, Rebecca F. Bruce, and Thomas P. O'Hara. 开发和使用一个黄金标准数据集进行主观分类. 在计算语言学协会会议论文集 (ACL-1999). 1999.
355. Wiebe, Janyce and Ellen Riloff. 从未标注的文本中创建主观和客观句子分类器. 计算语言学和智能文本处理, 2005: p. 486-497.
356. Wiebe, Janyce, Theresa Wilson, Rebecca F. Bruce, Matthew Bell, and Melanie Martin. 学习主观语言. 计算语言学, 2004. **30**(3): p. 277-308.
357. Wiebe, Janyce, Theresa Wilson, and Claire Cardie. 在语言中注释观点和情感表达. 语言资源和评估, 2005. **39**(2): p. 165-210.
358. Wiegand, M. and D. Klakow. 用于观点持有者提取的卷积核. 在人类语言技术会议: 北美ACL分会的2010年年度会议(HAACL-2010). 2010.
359. Williams, Gbolahan K. and Sarabjot Singh Anand. 使用WordNet预测形容词的极性强度. 在 第三届国际AAAI网络日志和社交媒体会议(ICWSM-2009). 2009.
360. Wilson, Theresa and Stephan Raaijmakers. 比较词、字符和音素 n -gram在主观话语识别中的效果. 在Interspeech会议. 2008.
361. Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 在短语级情感分析中识别上下文极性. 在人类语言技术会议和自然语言处理会议(HLT/EMNLP-2005). 2005.
362. 威尔逊, 特蕾莎, 詹妮丝·威贝和丽贝卡·华. 你有多生气? 寻找强烈和弱的观点子句. 在人工智能国家会议论文集 (AAAI-2004). 2004年.
363. 威尔逊, 特蕾莎, 詹妮丝·威贝和丽贝卡·华. 识别强烈和弱的观点子句. 计算智能, 2006年. **22**(2): p. 73-99.
364. 吴光宇, 德里克·格林, 巴里·史密斯和帕德里格·坎宁安. 扭曲作为鉴定可疑评论的验证标准. 在社交媒体分析论文集. 2010年.
365. 吴琮, 谭松波和程学琪. 情感转移的图排名. 在 ACL-IJCNLP 2009会议短论文集 (ACL-IJCNLP-2009). 2009年.
366. 吴元斌, 张琪, 黄炫靖和吴立德. 短语依赖解析用于观点挖掘. 在会议论文集中经验方法在自然语言处理中的应用 (EMNLP-2009) 。 2009.
367. 吴元斌, 张琪, 黄炫靖和吴立德. 基于图的情感表示的结构观点挖掘. 在2011年经验方法会议自然语言处理中的应用 (EMNLP-2011) 。 2011年.
368. 吴云芳和温苗苗. 消除动态情感模棱两可的形容词. 在第23届国际计算语言学会议 (Coling 2010) 。 2010年.

情感分析与观点挖掘

369. 夏锐和宗成庆。探索词关系特征的使用情感分类。在 *Coling 2010*会议论文集：海报卷。2010年。
370. 夏瑞和宗成庆。基于词性的集成模型用于跨领域情感分类。在第五届国际自然语言处理联合会议(IJCNLP-2010)的论文集。2011年。
371. 徐刚, 孟晓, 王华。使用基于图的算法和多种资源构建中文情感词典。在第23届国际计算语言学大会(Coling 2010)的论文集。2010。
372. 杨辉, 罗思, Jamie Callan。在TREC2006博客赛道中的知识转移和观点检测。在TREC的论文集。2006年。
373. 杨迅和高永中。使用比较类型分类从文本中提取比较实体和谓词。在第49届计算语言学协会年会(ACL-2011)的论文集。2011年。
374. Yano, Tae和Noah A. Smith.值得评论的是什么? 内容和政治博客中的评论数量。在国际 AAAI网络日志和社交媒体会议(ICWSM 2010)。2010年。
375. Yatani, Koji, Michael Novati, Andrew Trusty和Khai N. Truong。在线评论摘要的形容词-名词词对提取方法分析。在国际人工智能联合会议(IJCAI-2011)。2011年。
376. Yessenalina, Ainur和Claire Cardie。情感分析的组合矩阵空间模型。在经验自然语言处理会议论文集(EMNLP-2011)。2011年。
377. Yessenalina, Ainur, Yejin Choi和Claire Cardie。自动生成注释者理由以改进情感分类。在 ACL 2010会议短文集中。2010年。
378. Yessenalina, Ainur, Yison Yue, 和 Claire Cardie.多层次结构化模型用于文档级情感分类。在 *Empirical Methods in Natural Language Processing(EMNLP-2010)*会议论文集。2010。
379. Yi, Jeonghee, Tetsuya Nasukawa, Razvan Bunescu, 和 Wayne Niblack。情感分析器：使用自然语言处理技术提取关于给定主题的情感。在 *IE EE International Conference on Data Mining (ICDM-2003)*会议论文集。2003。
380. Yoshida, Yasuhisa, Tsutomu Hirao, Tomoharu Iwata, Masaaki Nagata, 和Y uji Matsumoto。多领域情感分析的迁移学习-识别领域相关/无关词汇极性。在Twenty-Fifth AAAI Conference on Artificial Intelligence(AAAI-2011)会议论文集。2011。
381. 于洪和Vasileios Hatzivassiloglou。朝着回答观点问题：将事实与观点分开并确定观点句子的极性。在自然语言处理会议论文集 (EMNLP-2003) 中。2003年。
382. 于建兴, 郑军扎, 王萌和蔡达成。方面排名：从在线消费者评论中识别重要的产品方面。在计算语言学协会第49届年会论文集中。2011年。
383. 于建兴, 郑军扎, 王萌, 王凯和蔡达成。领域辅助产品方面层次生成：朝着非结构化消费者评论的层次化组织。在自然语言处理会议论文集 (EMNLP-2011) 中。2011年。

情感分析与观点挖掘

384. 翟忠武, 刘冰, 徐华和贾培发。聚类产品特征进行观点挖掘。在 *ACM 国际网络搜索和数据挖掘会议 (WSDM-2011)* 中。2011年。
385. 翟忠武, 刘冰, 徐华和贾培发。有约束的LDA用于在观点挖掘中分组产品特征。在 *PAKDD-2011* 会议上。2011年。
386. 翟忠武, 刘冰, 徐华和贾培发。使用半监督学习和软约束进行分组产品特征。在 *国际计算语言学会议 (COLING-2010)* 中。2010年。
387. 翟忠武, 刘冰, 张磊, 徐华和贾培发。在在线讨论中识别评价观点。在 *AAAI* 会议上。2011年。
388. 张磊和刘冰。提取情感分析的资源术语。在 *IJCNLP-2011* 会议论文集中。2011a年。
389. 张磊和刘冰。识别暗示观点的名词产品特征。在 *计算语言学协会年会论文集 (短文) (ACL-2011)* 中。2011b年。
390. 张磊, 刘冰, Suk Hwan Lim和Eamonn O'Brien-Strain。在观点文档中提取和排名产品特征。在 *计算语言学国际会议论文集 (COLING-2010)* 中。2010年。
391. 张敏和叶兴耀。一个生成模型, 统一主题相关性和基于词典的情感用于观点检索。在 *年度ACM SIGIR国际会议论文集中, 研究和信息检索开发 (SIGIR-2008)*。2008年。
392. 张, 魏, 李锋佳, 克莱门特于, 和魏义孟。提高观点检索和观点极性分类的有效性。在 *ACM国际信息和知识管理会议 (CIKM-2008)*。2008年。
393. 张, 魏和克莱门特于。*UIC*在 *TREC 2007* 博客报告中, 2007年。
394. 张, 文斌和史蒂文·斯基纳。利用博客和新闻情感的交易策略。在 *国际会议上 Weblogs and Social Media (ICWSM-2010)*。2010年。
395. 张, 朱和巴拉吉·瓦拉达拉简。产品评论的效用评分。在 *ACM国际信息和知识管理会议 (CIKM-2006)*。2006年。
396. 赵, 韦恩·辛, 江静, 严洪飞和李晓明。联合建模方面和意见与 *MaxEnt-LDA* 混合。在 *会议的经验方法自然语言处理 (EMNLP-2010)*。2010年。
397. 周, 兰军, 李斌阳, 高伟, 魏忠宇和黄金辉。无监督发现话语关系以消除句内极性歧义。在 *会议论文集的经验方法自然语言处理 (EMNLP-2011)*。2011。
398. 周, 丽娜, 石永梅和张东松。一种统计语言建模方法用于在线欺骗检测。IEEE知识和数据工程交易, 2008年: 第1077-1081页。
399. 周树森, 陈庆才和王晓龙。主动深度网络用于半监督情感分类。在 *Coling* 会议 2010年: 海报卷。2010年。
400. 朱静波, 王慧珍, Benjamin K. Tsou和朱木华。多方面观点调查从文本评论中。在 *ACM会议国际信息和知识管理会议 (CIKM-2009)*。2009年。

情感分析与观点挖掘

- 401. 朱晓进和Zoubin Ghahramani。从标记和未标记数据中学习标签传播。卡内基梅隆大学计算机科学学院，匹兹堡，宾夕法尼亚州，技术报告CMU-CALD-02-107，2002年。
- 402. 庄丽，景峰和朱晓燕。电影评论挖掘和总结。在 ACM国际信息与知识管理会议 (CIKM-2006) 中。2006年。
- 403. Zirn, Căcilia, Mathias Niepert, Heiner Stuckenschmidt和Michael Strube。基于结构特征的细粒度情感分析。在第5届国际自然语言处理联合会议 (IJCNLP-2011) 中。2011年。