文本情感分析综述

王梦圆 汪安平 惠宁

摘 要 由于主观性文本有很多应用价值，情感分析近年来引起了很多研究人员的兴趣。情感分析是对主观性文本进行挖掘与分析，获取有用的知识和信息。本文首先对文本情感分析的研究现状与应用进行了总结；将文本情感分析归纳为3项主要任务，即情感信息抽取，情感信息分类以及情感信息的检索与归纳，并对它们进行了细致的介绍和分析；接下来介绍了文本情感分析的常用方法和近年来前沿的方法；进而对情感分析评测和资源建设情况进行介绍；最后阐述了文本情感分析存在的问题，并结合研究热点对文本情绪分析的前沿进展进行了概括和展望。重在对文本情感分析研究的主流方法和前沿进展进行概括、比较和分析。

关键词 中文文本；情感分析；观点挖掘；语料库；情感词典

成员负责分工 王梦圆：第5,6,8,9章 汪安平：第7章 惠宁：第1,2,3,4,5章

成员工作量 王梦圆：33% 汪安平：33% 惠宁：33%

# 引言

随着Web2.0的蓬勃发展，互联网逐渐倡导以用户为中心，用户参与的开放式构架理念。互联网用户由单纯的“读”网页，开始向“写”网页、“共同建设”互联网发展，并由被动地接收互联网信息向主动创造互联网信息迈进。因此，互联网(如博客和论坛)上产生了大量的用户参与的、对于诸如人物、事件、产品等有价值的评论信息。这些评论信息表达了人们的各种情感色彩和情感倾向性，如喜、怒、哀、乐和批评、赞扬等。基于此，潜在的用户就可以通过浏览这些主观色彩的评论来了解大众舆论对于某一事件或产品的看法。由于越来越多的用户乐于在互联网上分享自己的观点或体验，这类评论信息迅速膨胀，仅靠人工的方法难以应对网上海量信息的收集和处理，因此迫切需要计算机帮助用户快速获取和整理这些相关评价信息。情感分析(sentiment analysis)技术应运而生。

文本情感分析又称意见挖掘，简单而言，是对带有情感色彩的丰观性文本进行分析、处理、归纳和推理的过程。最初的情感分析源自前人对带有情感色彩的词语的分析[l]，如，“美好”是带有褒义色彩的词语，而“丑陋”是带有贬义色彩的词语。随着互联网上大量的带有情感色彩的主观性文本的出现，研究者们逐渐从简单的情感词语的分析研究过渡到更为复杂的情感句研究以及情感篇章的研究。基于此，按照处理文本的粒度不同，情感分析可分为词语级、短语级、句子级、篇章级以及多篇章级等几个研究层次[2]。按照处理文本的类别不同。可分为基于新闻评论的情感分析和基于产品评论的情感分析。其中，前者处理的文本主要是新闻评论，如情感句“他坚定地认为台湾是中国不可分割的一部分”，表明了观点持有者“他”对于事件“台湾归属问题”的立场：后者处理的主要是网络在线的产品评论文本，如“Polo的外观很时尚”。表明了对评价对象“Polo的外观”的评价“时尚”是褒义的。由于基于产品评论的情感分析可以帮助用户了解某一产品在大众心目中的口碑，因此受到很多消费者和商业网站的青睐。而基于新闻评论的情感分析多用于舆情监控和信息预测中，是国内外评测中重要的评测任务。

情感信息抽取是情感分析的最底层的任务，它旨在抽取情感评论文本中有意义的信息单元。其目的在于将无结构化的情感文本转化为计算机容易识别和处理的结构化文本，继而供情感分析上层的研究和应用服务。如将情感句“我觉得Canon的相片质量不错”转化为如图1所示的结构化文本形式。情感信息分类则利用底层情感信息抽取的结果将情感文本单元分为若干类别，供用户查看，如分为褒、贬两类或者其他更细致的情感类别（如喜、怒、哀、乐等）。按照不同的分类目的，可分为主客观分析和褒贬分析；按照不同的分类粒度，可分为词语级、短语级、篇章级等多种情感分类任务。这些分类任务在情感分析初期吸引了大量的研究者。最高层的情感信息的检索与归纳可以看作与用户直接交互的接口，着重强调检索和归纳两项应用。该层次的研究主要在前两项任务即情感信息抽取和分类的结果的基础上进行进一步的加工处理。

# 情感信息抽取

情感信息抽取旨在抽取情感文本中有价值的情感信息，它可以看作情感分析的基础任务。一直以来，学术界对它兴趣不减。纵观目前的研究现状，有价值的情感信息单元主要有评价词语（如优秀、好用）、评价对象（如GPS、屏幕分辨率）、观点持有者（如国家政府、台湾当局）等。在对大量的情感文本进行分析之后，不少研究者发现，某些组合搭配对于情感分析的上层任务如情感信息分类以及情感信息的检索与归纳有更直接的帮助，如评价搭配（评价对象和评价词语的搭配，如屏幕分辨率-高）、评价短语（程度副词及其修饰的评价词语的搭配，如不怎么-好）等。下面本文将一一介绍目前情感信息抽取的具体任务及其主要实现技术。

## 2.1评价词语的抽取和判别

评价词语又称极性词、情感词，特指带有情感倾向性的词语。显然，评价词语在情感文本中处于举足轻重的地位，评价词语的识别和极性判断在情感分析领域创建伊始就引起了人们极大的兴致。基于前人大量的研究工作，评价词语的抽取和判别往往是一个一体化的工作，主要分为基于语料库和基于词典两种方法[6]。

基于语料库的评价词语抽取和判别主要是利用大语料库的统计特性，观察一些现象来挖掘语料库中的评价词语并判断极性。早期的一些学者发现，由连词（如and或but）连接的两个形容词的极性往往存在一定的关联性，如and连接的形容词（如lovelyandbeautiful）极性相同，然而but连接的形容词（如lovelybutunnatural）极性相反。基于这种现象，Hatzivassiloglou和McKeown[1]从大语料库华尔街日报（WallStreetJournal）中发掘出大量的形容词性的评价词语。Wiebe等人[7]沿袭了较为相似的工作，他们使用了一种相似度分布的词聚类方法在大语料库上完成了形容词性的评价词语的获取。然而，以上的两种方法仅将评价词语的词性局限于形容词词性，却忽略了其他词性的评价词语。为了避免评价词语词性的限制，Riloff等人[8]手工制定一些模板并选取种子评价词语，使用迭代的方法获取了名词词性的评价词语。随后，Turney和Littman[9]提出了点互信息（pointmutualinformation）的方法判别某个词语是否是评价词语。这种方法适用于各种词性的评价词语的识别，但是较为依赖种子褒/贬词语集合。鉴于此，基于语料库的方法最大的优点在于简单易行，缺点则在于可利用的评论语料库有限，同时评价词语在大语料库中的分布等现象并不容易归纳。

基于词典的评价词语抽取及判别方法主要是使用词典中的词语之间的词义联系来挖掘评价词语。这里的词典一般是指使用WordNet或HowNet等。很自然地，有学者想到利用词典将手工采集的种子评价词语进行扩展来获取大量的评价词语[10−12]。这种方法简单易行，但是较依赖于种子评价词语的个数和质量，并且容易由于一些词语的多义性而引入噪声。为了避免词语的多义性，一部分学者使用词典中词语的注释信息来完成评价词语的识别与极性判断[13−16]。此外，一些学者[17]沿用了Turney等人的点互信息的方法[9]，通过计算WordNet中的所有形容词与种子褒义词代表good和贬义词bad之间的关联度值来识别出评价词语。然而，并非所有语种的情感资源都像英文一样丰富，对于某些词典资源非常稀缺的语种，有学者将词典资源丰富的语种的情感词典翻译到资源较少的语种中[18]，如将英文的情感词典翻译成中文，供中文情感分析应用。但是实验显示，不少评价词语在经过翻译之后极性发生了改变。这也印证了Wiebe在文献[19]中所指出的“词语的词义和其极性有一定的关系，但是相同的词义并不一定有相同的极性”。鉴于此，基于词典的方法的优点在于获取的评价词语的规模非常可观，但是由于很多词存在一词多义现象，构建的情感词典往往含有较多的歧义词，如词语“好”在大多数情况下表现为“优秀”的意思，但在某些情况下扮演修饰成分（如“他跑得好快啊!”）。

此外，还有一部分学者采用基于图的方法来识别评价词语的极性[6，20]。具体来说，该方法将要分类的词语作为图上的点，利用词语之间的联系形成边来构建图，继而采用各种基于图的迭代算法（propagationalgorithm）来完成词语的分类。如，有学者考察图中两个词语的注释信息而构建图[20]，继而使用Spin模型对图中的点迭代地进行概率计算，得出每个词语的极性。还有一些学者尝试使用多种图模型[6]，如最小切分模型（mincuts）、随机最小切分模型（randomizedmincuts）、标签迭代模型（labelpropagation）等完成评价词语的褒贬分类。实验证实了基于图的方法的有效性。基于图的方法是一种新颖的方法，它可以灵活地将词语间的各种联系作为特征融入图中，继而进行迭代计算。然而，寻找更为有效的词语间特征以及如何选取图算法是值得深入研究的问题。

## 2.2评价对象的抽取

评价对象是指某段评论中所讨论的主题，具体表现为评论文本中评价词语所修饰的对象，如新闻评论中的某个事件/话题或者产品评论中某种产品的属性（如“屏幕”）等。现有的研究大部分集中于产品领域的评价对象的抽取，他们大多将评价对象限定在名词或名词短语（候选评价对象）的范畴内，进而对它们进行进一步的识别。

一部分学者使用基于规则/模板的方法抽取评价对象。规则的制定通常要基于一系列的语言分析与预处理过程，如词性标注、命名实体识别、句法分析等。相应地，制定的规则也包括词序列规则、词性规则以及句法规则等形式。Yi[21]使用3条限制等级逐渐递进的词性规则从候选评价对象中抽取出真正的评价对象。还有的学者[22，23]使用关联规则挖掘的方法或是基于句法分析的结果[24]找出频繁出现的候选评价对象，继而使用两种剪枝方法去除错误样例。然而，这些方法仅能找出频繁的评价对象。为了发掘出非频繁的评价对象，有学者尝试使用含有评价词语和评价对象槽（slot）的词序列模板[22]。此类方法最主要的优点在于针对性强，可以直接针对待解决的问题或特定的语言现象制定规则/模板；而其缺点则在于规则/模板的可扩展性差，人工编写的工作量大，成本较高。

有学者[24]从另一个角度诠释了评价对象的抽取。他们将评价对象看作产品属性的一种表现形式（如对数码相机领域而言，“相机的大小”是数码相机的一个属性，而“相机滑盖”是数码相机的一个组成部分），继而考察候选评价对象与领域指示词（如“整体-部分”关系指示词“scannerhas”）之间的关联度来获取真正的评价对象。实验结果表明，这种方法取得了较好的实验效果，超过了基于规则/模板的方法，但难点在于领域指示词的获取。

近年来，随着话题模型（topicmodel）[26，27]的逐渐兴起，很多学者将其应用到情感分析领域。由于评价对象是蕴涵于情感文本中的某些话题，因此可以使用话题模型来评价对象的识别。有学者[28]采用多粒度的话题模型挖掘产品领域情感文本中的评价对象，并将相似的评价对象进行聚类。这种方法理论上能够提高评价对象抽取的召回率。但遗憾的是，还没有实验将这种方法与上述传统的基于名词短语的方法进行对比。

此外，还有一部分学者从事新闻评论文本中的话题评价对象的抽取[29，30]。如，对于情感句“所有人都认为政府应该加强改善医疗卫生条件”，抽取话题评价对象“政府应该加强改善医疗卫生条件”。

## 2.3观点持有者抽取

观点持有者的抽取在基于新闻评论的情感分析中显得尤为重要，它是观点/评论的隶属者，如新闻评论句“我国政府坚定不移的认为台湾是中国领土不可分割的一部分”中的“我国政府”。很自然地，人们会想到评论中的观点持有者一般是由命名实体（如人名或机构名）组成，因此可以借助于命名实体识别技术来获取观点持有者[31]。此外，还有学者曾尝试借助语义角色标注来完成观点持有者的抽取[29]。但是这些方法较为依赖自然语言处理的基础技术，有较低的语言覆盖现象和较差的领域适应性。

还有人将观点持有者的抽取定义为分类任务，这种方法的关键在于分类器和特征的选取。如Choi将其看作一个序列标注问题[32]，并使用CRF（conditionalrandomfield）模型融合各种特征来完成观点持有者的抽取。相似地，Kim[11]将所有名词短语都视为候选观点持有者，使用ME（maximumentropy）模型来进行计算。

以上的方法将观点持有者的抽取当作一个独立的任务。通过观察许多研究者发现，观点持有者一般是与观点同时出现的，所以可以将观点和观点持有者的识别作为一个任务同时解决。Bethard[33]在抽取出情感句中的观点单元（多是由一些短语组成）之后，分析句中观点和动词的句法关系，即可同步获取观点持有者。

由于产品评论中一般默认观点持有者是用户本身，因此鲜有研究者在产品评论领域研究这一任务。

## 2.4组合评价单元的抽取

评价词语在情感分析中的作用是不言而喻的。然而在某些情况下，单独的评价词语存在一定的歧义性，如评价词语“高”在以下3个句子中的使用：

•Sen1：凯越的油耗真高。

•Sen2：捷达的性价比相当高。

•Sen3：这辆车有1米多高。

Sen1和Sen2是情感句，但是评价词语“高”在修饰不同的评价对象时表现出不同的极性。如，“高”在Sen1中表示贬义，而在Sen2中则表示褒义。此外，评价词语往往也会出现在非情感句中，如Sen3。因此，仅考虑单独的评价词语在情感分析中的应用是远远不够的。研究者们发现，有些包含评价词语的“组合评价单元”（如组合“油耗-高”、“相当-高”）对于处理情感分析的上层任务更有帮助。下面将具体来介绍各种形式的组合评价单元。

# 情感信息分类

情感信息的分类任务可大致分为两种：一种是主、客观信息的二元分类；另一种是主观信息的情感分类，包括最常见的褒贬二元分类以及更细致的多元分类[44]。

## 3.1主客观信息分类

在对情感文本进行情感分析时，往往由于情感文本中夹杂着少量的客观信息而影响了情感分析的质量[44]，因此将情感文本中的主观信息和客观信息进行分离变得非常必要。由于情感文本单元表现格式比较自由，且区分主、客观文本单元的特征并不明显，在很多情况下，情感文本的主客观识别比主观文本的情感分类更有难度。

一部分学者通过考察文本内部是否含有情感知识（具体表现为第1节情感信息抽取的结果）来完成主客观信息分类[10，46]。然而我们发现，许多客观句中也可能会包含评价词语，如客观句“这位英雄名叫张三丰”同样含有评价词语“英雄”。为了在更大程度上消除歧义性，很多学者挖掘并使用情感文本中的组合评价单元。此外，还有学者[8]构建情感模板识别情感文本的主客观性（如贬义模板“〈x〉drives〈y〉upthewall”）。以上这些基于情感知识的主客观分类方法的工作重心在于情感文本中情感知识的挖掘以及各种情感知识融合的方法研究。

还有一部分学者将情感文本单元的主客观分类定义为一种二元分类任务，即对任意给定的情感文本单元，由分类器协助判断其主客观性。这种方法的关键在于分类器和分类特征的选取。具体来说，Hatzivassiloglou[47]使用了词语作为特征，并采用了NB（NaïveBayes）分类器完成篇章级情感文本的主客观分类。Yao[48]着重从一些特殊的特征角度考察了主客观文本，如标点符号角度、人称代词角度、数字角度等。Pang[49]则采用基于图的分类算法完成句子级的主客观分类。基于特征分类的方法目前还是主客观信息分类的主流方法。这种方法定义明确，其根本问题在于特征的选取。因此，尝试使用更深层、更复杂的分类特征也许是这类方法的突破方向之所在。

## 3.2主观信息情感分类

主观信息情感任务按不同的文本粒度可分为词语级、短语级、句子级和篇章级等。其中，第1节已经对词语级和短语级的情感分类方法进行了总结，因此本节将着重介绍句子级和篇章级的主观信息情感分类方法。

一般而言，研究者将主观本文的极性分为褒义和贬义两类（thumbsup?thumbsdown?）。纵观目前的研究工作，与主客观信息分类类似，可分为两种研究思路：基于情感知识的方法以及基于特征分类的方法。相似地，前者主要是依靠一些已有的情感词典或领域词典以及主观文本中带有情感极性的组合评价单元进行计算，来获取主观文本的极性。后者主要是使用机器学习的方法，选取大量有意义的特征来完成分类任务。这两种研究思路有很多代表性的研究工作。文献[10，47，40，41]首先分析句子/篇章中的评价词语或组合评价单元的极性，然后进行极性加权求和。这种方法的重点一般都放在评价词语或组合评价单元的抽取和极性判断方法的研究上。在基于特征分类的方法中，Pang[42]首次将机器学习的方法应用于篇章级的情感分类任务中。他们尝试使用了n-gram词语特征和词性特征，并对比了NB，ME和SVM（supportvectormachine）这3种分类模型，发现unigram特征效果最好。然而，Cui[43]通过实验证明，当训练语料较少时，unigram的效果较优；但随着训练语料的增多，n-gram（n>3）发挥了越来越重要的作用。Kim[44]除了考察传统的n-gram模型之外，还引入了位置特征和评价词特征来完成句子级的褒贬分类。Zhao[44]则将句子级情感分类任务提炼为一个三层分类任务，利用各层之间类别标签的相互作用，并考虑上下句之间情感的互相影响，使用CRF模型将这些特征进行融合。类似于主客观信息分类任务，基于特征的方法的研究重点在于有效特征的发现以及特征选择和特征融合等问题的研究。

除了对主观文本信息的褒贬二元分类之外，还有一些研究工作进行更细致的情感分类任务。Pang[46]将褒贬等级分为3类，并使用了one-vs-all多元分类算法和回归分类算法完成情感分类。Goldberg[47]则使用了一种基于图的半指导的分类算法，完成评论的褒贬包括4个等级的分类。

## 3.3观点分类与挖掘

情感分类还可以体现在对某些事件的观点分类上。Lin等人[48]主要使用3种分类模型识别有关“巴以冲突”主题的评论文本所表达的观点，即是“支持巴方”还是“支持以方”。而Kim等人[49]主要对美国大选时涌现出来的大量评论文章进行分类汇总，从而推断大部分选民是支持“共和党”还是“民主党”。该文献同样也是使用分类器和分类特征相结合的算法，其中对分类特征进行了泛化，取得了较好的效果。和主观信息情感分类不同的是，“观点分类与挖掘”任务除了需要使用情感知识之外，还需要发掘一部分与“观点”相关的知识。

# 情感信息的检索与归纳

情感分析是一项以应用为导向的研究课题，然而，情感信息抽取和分类后呈现的结果并不是用户所能直接使用的。经过大量的调研我们发现，情感分析技术与用户的交互主要集中于情感信息检索和情感信息归纳两项任务上。其中，情感信息检索旨在为用户检索出主题相关，且包含情感信息的文档；情感信息归纳则针对大量主题相关的情感文档，自动分析和归纳整理出情感分析结果提供给用户参考，以节省用户翻阅相关文档的时间。因此，情感信息归纳可以看作是情感信息检索结果的一个深入加工。

## 4.1情感信息检索

传统的搜索技术往往仅关注于检索事实性的相关文档，然而随着人们参与互联网建设的增多，尤其是博客、论坛的蓬勃发展，情感相关的文档逐渐成为用户检索需求的一部分。因此，很多研究机构和组织结合传统的搜索技术，掀起了情感信息检索（也称观点检索）研究的热潮。

情感信息检索任务最早出现在Hurst和Nigam的工作中[60]。2006年，TREC（TextRetrievalEvaluationConference）首次引入了博客检索任务（Blogtrack）[61]，更多的研究者致力于该任务的研究。情感信息检索要求检索回的文档同时满足两项准则：（1）主题相关；（2）具有情感倾向性。一般而言，该任务主要包含3个步骤：

①结合传统的信息检索模型进行主题相关的文档检索。即给定某一主题，检索出所有与其相关的文档。在这一步骤中，研究者一般使用传统的检索模型以及一些较为成熟的查询扩展技术[62]。也有研究者在查询扩展中融入了情感知识[63]，以提高检索的性能。

②相关文档的主客观识别。即针对某一主题的所有相关文档，判别它们的主客观性，并获取带有情感的主观性文档。在这一步骤中，研究者们一般借鉴情感信息抽取和情感信息分类的技术，如第1节和第2节所述。

③主题相关的情感（主观性）文档排序。此时的排序策略需要同时兼顾文档的情感打分以及相关性打分，不少学者对该步骤进行了深入的研究。一种最直接的方式是使用线性加权函数来融合两部分的打分[62]。这种方法首先将两部分的打分剥离开来，分别进行打分的设计和计算，继而对这两个分数进行加权求和。然而，该方法缺乏理论基础和细致的分析。清华大学的张敏提出了一种新颖的基于概率生成模型的情感文档排序方法[64]。该模型从生成文档的角度考察文档的情感打分和相关性打分，取得了不错的效果。

通过以上3个步骤的分析我们发现，情感信息检索融合了传统的信息检索技术和新型的情感分析技术。其中，如何使二者进行更好的融合是研究者近期和未来所需关注的重点。

## 4.2情感信息归纳

就目前的研究现状而言，情感信息的归纳往往以情感文摘的形式存在。传统的基于事实性新闻语料的文摘旨在提取重要的事实性信息，并去除冗余信息。相比而言，情感文摘的处理对象为某一产品或某一事件的大量用户评论，因此这种文摘融入了更多的情感信息。它主要侧重于提取具有明显情感倾向性的主观信息，是对某一产品或某一事件的评论信息的归纳和汇总。针对产品类评论信息，情感文摘共有两种呈现方式：一种是基于产品属性的情感文摘，另一种是基于情感标签的情感文摘。

### 4.2.1基于产品属性的情感文摘

到目前为止，大部分研究者致力于基于产品属性的情感文摘的研究[25，50，65，66]，其中的产品属性特指在产品评论中的评价对象，如“相片质量”等。如图2所示，这种文摘按照产品的属性进行分门别类的整理，一般都标引出表示“支持”和“反对”该产品属性的具体句子供用户参考，以适应不同用户个性化的需求。纵览目前学者们的研究工作，标准的基于产品属性的情感文摘共包含3个重要的步骤：

（1）识别出评论信息（句子或篇章）中的产品属性，即评价对象的识别。如前面所述，很多学者从事这一任务的研究，并取得了不错的效果[21−25]。但需要关注的是，许多研究工作忽视了产品属性的别称现象，如“胶卷”和“胶片”。如果能将其进行别名消解，则能更好地将产品的情感文摘进行归纳。

（2）抽取出描述产品属性的情感句，即针对产品的每一种属性，收集与其相关的所有的情感句[46−49]。这是因为含有产品属性的句子不一定是情感句，如句子“这个相机不需要用胶片。”就不含有任何情感倾向性。这个步骤类似于句子级的主客观识别任务，然而也有很多学者省略了这一步骤，直接进行步骤（3）。

（3）针对产品属性的每一个情感句，判断其情感倾向性。这个步骤类似于句子级的情感信息分类任务[50−55]。基于此，针对产品的每一项属性列出用户的各种情感信息（如表示褒义和贬义的句子）形成情感文摘供用户参考。还有一些情感文摘将一些统计信息也罗列出来，方便用户对多种产品进行对比。如有的情感文摘将用户针对某一项属性的“支持”和“反对”情感句的个数列举出来[65]；此外，还有情感文摘为每种产品计算出推荐等级[67]等。

通过以上3个步骤的分析我们发现，基于产品属性的情感文摘技术是情感信息抽取和情感信息分类技术的融合。因此，只有情感分析各项基础技术（如上面3个步骤）指标的提高，才能增强这类情感文摘的质量。

### 4.2.2基于情感标签的情感文摘

由于基于产品属性的情感文摘较为依赖情感分析的底层技术，而且用户若想了解某一产品属性的具体情况仍需阅读大量标有“褒/贬”等情感类别的情感句，为了更便捷地为用户提供参考意见，基于情感标签的情感文摘应运而生。出现这类情感文摘的灵感来自于某些用户自由建设的在线评论网站，如国外著名的网站epinions（http：//www.epinions.com/）。用户在这类网站上发表对某一产品的看法时，要求填写对自己观点的简要概括，一般用若干个词语或短语的标签形式表现，如smallsize（pros），shortbatterylife（cons）等。这些标签能够很好地概括评论的主要内容，并以简短精悍的方式吸引了大量的用户群，可以看作是一种新形式的基于情感标签的文摘。然而，这类在线评论网站比较少，因此大多数的在线网络评论并没有人工标注的标签。此外，依赖用户提供标签也存在一些问题：一来用户写评论时非常随意，很有可能概括不全自己的评论；二来用户书写的评论用词较为丰富，不方便计算机自动对比两个相似产品。如在评价一个餐馆的饭菜时，有人使用gooddiet，而有的人则使用healthy这个标签。因此，近年来（2008年至今），不少学者[68，69]开始研究为网络评论自动生成标签，方便用户快速阅读评论以及对比产品。

Titov[68]的工作主要是为一组相关产品评论集的产品属性打标签。具体来说，该方法将产品属性当作文档集中潜在的话题（topic），继而使用一个改进的结合产品文档和属性等级的话题模型（topicmodel）[26，27]对产品的属性进行潜在的标签词语生成。如对于“宾馆”评论的属性“房间（rooms）”，该方法可以自动发掘出相关文档集中的相关词语，如small，clean等。然而，该方法是建立在已知产品属性的基础上的，也就是说，已知这组产品相关的文档集中所有的产品属性（如room，service，location等），继而为每种产品属性找出潜在的情感标签。然而在现实的评论语料中，一般不会事先预知产品的属性以及属性等级，因此需要自动地发掘产品属性及其情感标签。Branavan[69]较为巧妙地解决了这个问题，他将标签定义为“评价搭配”的形式（如smallsize，longbatterylife等），并为单篇的网络产品评论打上合适的标签。具体地，该方法首先在一些带有标签的网站（如epinion）上获取某一产品的大量的人工标注的标签（如smallsize等）作为这一产品的标签库；接着，使用相似度聚类的方法对这标签库进行聚类，每一类被视为一个潜在的话题（即产品属性），并且每一类话题包含若干个相似的情感标签（如smallsize和smartsize）；继而，使用话题模型分析一篇评论中潜在话题的分布情况（用概率值表示）；最终通过对概率值的分析获取最显著的几个潜在话题，并取其代表性情感标签作为这篇评论的标签。该种方法主要针对单篇网络评论（单文档）进行分析，优点在于标签更加清晰，有完整的意义；此外，由于同一类产品共享一个标签库，方便产品之间的对比。当然，该方法还可以推广到多文档情感文摘中。

### 4.2.3基于新闻评论的文摘

还有部分学者从事基于新闻评论领域的情感文摘研究。该种文摘和普通的新闻文摘比较类似，除了抽取重要的、信息含量大的核心句子之外，还要重视该句子中的情感信息。有学者[70]提出，这种情感文摘的核心句需要包含两方面的词语：概念词（conceptwords）和情感词（sentimentwords）。因此，他们将新闻评论信息中的词语进行了细致的分析，通过分析每个词与话题的联系找出概念词，并通过拆分字的方法找出情感词，进而挑选出核心情感句形成情感文摘。由于基于新闻评论的文摘应用面并不是很广，而且与普通的新闻文摘较为相似，所以目前从事这方面研究的学者并不多。

# 情感分析方法

文本情感分类任务主要指通过提取文本内容中的情感要素，将文本划分到一个或多个预定义的情感类别中，通过判定文本中所表达的情感类别，实现对文本发布者情感的监控、预测和管理。目前，情感分类方法主要包括：基于词典和规则的方法、基于机器学习的方法、复合方法以及其他方法。

## 5.1基于词典和规则的情感分类方法

基于词典和规则的方法能体现文本的非结构化特征，易于理解和解释。此外，该方法处理速度快且精度较高，在相对短的时间内能够对大型数据源做出可行且效果良好的结果。

### 5.1.1基于词典的情感分类方法

基于词典的方法主要利用情感词典资源，将语料库中的情感表达关键字提取出来，并藉此对语料进行情感分类。在早期的研究中，Ma等人[71]提出了一种基于词典的情感分类方法，并将其应用到即时通讯系统上。该方法首先利用关键字识别出文本中情感相关的内容，再利用句法特征检测其中的情感意义，并通过文本消息中的情感对系统中的语音合成、手势等功能进行调整，帮助用户更好地与远距离用户沟通。在此基础上，Aman等人[72]提出一种基于情感强度知识的分类方法，该方法除使用情感词典之外，还增加了情感强度知识库。该方法在博客语料的情感分类任务中，其准确率可达66％以上。

由于情感词典中的情感词有较大程度的领域依赖性、时间依赖性和语言依赖性，同一词汇在不同的领域、时间和语言环境中可能会表达完全不同的情绪，然而传统方法在构建情感词典时并未考虑词典的应用环境因素，甚至无法应用到其他语种。因此，在跨领域、跨时间、跨语言的文本情感分类任务中效果并不理想。

为解决领域依赖性的问题，Yang等人[73]提出一种基于特定领域情感词典分类的方法。该方法利用情感感知LDA(emotion-aware LDA，EaLDA)模型，为预定义的情感构建特定领域的情感词典。EaLDA模型使用一组领域无关的最小种子词作为先验知识，来发现特定领域的词典。在SemEval一2007数据集的综合实验中，该方法可以有效辅助文本情感分类任务，其中对最难辨别的Disgust类别，其F1值可达10.52％，其他分类如Sadness可达36.85％。

为挖掘时间依赖性对情感分类的影响，Golder等人[17]利用情感词典对全球不同地域、不同文化背景博主所发表的Twitter进行统计，分析了数百万篇公开Twitter中所表达的情感，并明确地识别出人们的情感会随着季节、星期、昼夜呈现出周期性变化的模式。

解决情感词典语言依赖性问题最常用的方法是构建本语言的情感词典，因此，很多研究者们开展了构建中文情感词典的研究工作。

对于中文及其他语料资源匮乏的语言，难以获得用于构建本语言情感词典的语料素材。为解决该问题，Xu等人[7]提出一种基于英文情感词典WordNet-Affect自动构建中文情感词典的方法。该方法首先将英文情感词典中所有的英文单词都翻译成中文；再借助中文同义词词典《同义词词林》将每个情感类别构建一个双语无向图，并提出一个图算法用以过滤翻译过程中引入的非情感词，以获得种子情感词集；最终通过同义词扩大表示类似情感的词汇量，从而获得数量大、质量高的中文情感词典。

文献[74]主要解决语料库资源严重不足条件下构建情感词典的问题，但由于不同语言不同文化中词汇所表达的情感存在差异，对情感分类的准确率存在影响，为此一些研究者开始研究利用少量种子词构建情感词典的方法。Song等人[75]提出一种基于异构图的情感词典分类方法，该方法利用种子词和表情符号构建情感词典，并使用随机游走算法强化对情感分布的评估效果。在新浪微博真实数据的实验中，利用该方法构建的情感词典对7种情感anger，disgust，fear，happiness，like，sadness，surprise分类的准确率可达54.1％。

此外，传统的情感词典方法还存在词典中情感词固定，难以及时捕捉新词、变形词的缺陷。为此，Wu等人[76]提出一种基于数据驱动的微博专用情感词典分类方法。该方法设计了一个包含3种词典的情感知识统一框架。为了提高情感词的覆盖率，该方法还支持将检测到的情感新词加入到词典中，不断扩展情感词典的样本集。在新浪微博数据集的实验中，该方法准确率可达58.04％。

影响情感词典方法分类准确率的主要因素包括情感词典的覆盖率和标注的准确率，目前的情感词典在这2个方面仍有不足，一些研究者利用互联网的便利，通过网民的帮助构建了一个高质量情感词典。Staiano等人[77]提出一种基于“众包”的情感分类方法，该方法利用大规模“众包”方式建立情感注释与新闻文章之间的联系。使用分布语义自动构建高质量、高精度的情感词典。在rappler。com新闻消息数据集的实验中，该方法较好地完成了fear，anger，surprise，joy，sadness这5类情感的分类任务，其中对fear的分类效果最好，准确率可达56％，surprise效果最差，准确率为25％。

总体上，基于情感词典的分类方法能体现出文本的非结构化特征，在词典中情感词覆盖率和标注准确率较高的情况下，分类效果比较理想。然而，该类方法依赖领域、时间、语言等方面的背景知识，且难以及时捕捉新词、变形词，使如何构造高质量的情绪词典成为其难点。

### 5.1.2基于规则的情感分类方法

除了情感词典，还有一类基于规则的情感分类方法，可以快速实现对情感语料的分类。在早期的工作中，Strapparava等人[78]提出了一种基于语义规则的情感分类方法，该方法利用隐形语义算法(1atentsemanticanalysis，LSA)计算通用语义词和情感词的语义相似度，再根据语义相似度对新闻标题进行分类。该方法在Times，BBC，CNN等新闻语料的情感分类任务中，准确率可达38.28％。

由于网络中的非正式文本比较多，文献[78]在对不规范文本的情感分类任务中表现并不理想。为解决该问题，Neviarouskaya等人[79]首先对网络文本中非正式缩写、情感图标以及语法错误等不规范文本进行了预处理，再利用基于语义规则的方法分阶段处理每个句子，最终将目标语料中的情感分为9类interest，joy，surprise，anger，sadness，fear，disgust，guilt，shame，其架构如图4所示。在包括日记博客、童话和新闻标题等不同领域的数据集中，该方法情感分类的准确率可达72.6％。此外，该方法在具有复杂句子的环境中也具有较好的分类效果。

对于句子级情感分析，现有的基于情感词典的方法表现并不理想，因为其并未考虑文本顺序和篇章结构。Wen等人[80]提出一种利用类序列规则情感分类的方法，将指定文本中的情感分为7类anger，disgust，fear，happiness，like，sadness，surprise。该方法首先分别利用情感词典和机器学习方法获得句子的2个潜在情感标签，并将每个微博文本看作一个数据序列，再从数据集中挖掘文本的类规则序列，最后根据规则的特性对微博进行情感分类。在2013年新浪微博的情感分析评测任务中，该方法对7种情感的分类准确率均达到41。33％以上。

通过分析社会媒体中公众情感的成因，可以利用情感与起因事件的关联关系来提高情感分类的准确率。传统的情感分类方法主要基于统计手段，没有考虑到引起情感的触发事件。Lee等人[81]提出一种基于文本驱动的情感成因检测方法。该方法通过对中文语料库数据的分析，确定了7种语言线索，包括原因事件的位置、体验者情感关键词的位置、使役动词、动作词、认知标记、连接词和介词，并据此归纳出语言规则来检测情感的成因。

在此基础上，部分研究者将基于规则的方法与事件成因相结合，从而实现对媒体文本的情感分类。Li等人[82]提出一种结合语义规则和事件触发理论的情感分类方法。该方法根据社会学以及其他领域的知识和理论，从推断和提取情感的原因人手进行分析。通过构造一个基于语义规则的系统，自动检测并抽取每个博客语料中情感的起因事件。再利用起因事件训练分类器对语料库进行情感分类。在新浪微博的实验中，该方法在加入了原因事件后，对6个情感类别anger，disgust，fear，happiness，sadness，surprise的分类准确率均有所提高。例如，Happiness从85.41％提升至87.36％，Surprise从71.71％提升至72.52％等。

此后，Gao等人[83]结合情感起因事件，提出一种基于规则的情感分析系统。该系统从事件的结果、代理的行为和对象的性质中分析产生情感的原因事件，并根据这些事件挖掘其中的基本情感、复合情感(满意、感激、悔恨和愤怒)和扩展情感(信任、失望、怜悯等)的关联规则，通过抽取的情感规则对中文微博进行分类。在对新浪微博语料中实验，分类准确率可达82.50％。

综上，虽然基于规则的情感方法可以在较短时间内获得分类结果，且可以加入事前起因等其他规则来提高情感分类的准确率，但在数据量较大时，规则的维护比较复杂，且不易扩展。

## 5.2基于机器学习的情感分类方法

在大数据时代的背景下，捕获文本数据并从中萃取有价值的情感信息是基于机器学习情感分类方法的主要任务，该方法主要包括有监督和半监督方法。

### 5.2.1有监督学习情感分类方法

在有监督学习的过程中，只需要给定输入情感样本集，即可推演出目标情感分类的可能结果。有监督学习相对比较简单，针对不同的情感分类任务可分为单标签和多标签情感分类方法。

特征选取是否合适是影响有监督学习分类效果的一个主要因素，现有方法中用于情感分类的特征主要包括词级、句子级和篇章级特征。其中，词级特征主要包括词频(例如词袋特征)、词性(例如名词、动词、连接词)、语义(例如词向量的相似度)、表情符号及其组合等。

在早期的研究中，为了使得语音合成技术TTS(text-to-speech)朗读时语调更自然，研究者们需要挖掘出童话故事中所蕴含的情感。Alm等人[84]提出一种将SVM与SNoW(sparse network of winnows)架构结合的文本情感预测方法。该方法通过选取故事首句、特定的连接词等30种特征将22篇格林童话划分为happy，sad，fearful，angry，disgusted，surprised，non-emotion这7类，其中surprised又细化为正向Su+和负向Su一。实验通过对比SVM、朴素贝叶斯等基准分类器，结果显示将SVM与SNoW架构结合的方法效果最好，其准确率可达到69.37％。

相较于童话故事、博客等长文本语料，社交网络信息通常是简短的，如微博、即时消息、新闻标题等。短文本受字数的限制，呈现出特征稀疏、内容简短、表述直接等特点，这使得以往的情感分类方法在面向短文本语料时，难以保证其分类效果。面向短文本的情感分析[8妇是近几年最热门的研究方向之一。

在微博环境中，表情符号被广泛用来表达不同的情感，社交网络运营商也为用户提供了丰富的情绪图标，方便用户表达对事物的情感。因此，表情符号也被视为情感分类的重要信号。此外，表情符号具有其他词语所不具备的独立性，在大多数话题、领域、时间段中，表情符号所代表的情感基本保持不变，因此，很多研究者都将表情符号作为其特征中的一个重要组成部分。

Read[85]提出一种基于表情符号的情感分类方法。该方法从语料库中抽取指定情感符号的文本集合，把所得的样本集作为训练数据来实现情感分类。

该方法在包含金融主题、并购主题以及2个主题混合的数据集上进行测试，证明了表情符号特征的主题独立性，其分类准确率可达70％。

此后，Ouyang等人[43]提出了一种基于卷积神经网络的情感分类架构。该架构使用Googleword2vec方法从文本中提取词向量，并将其作为卷积神经网络的输入。通过一个基于3对卷积层和池化层的卷积神经网络架构对影评语料进行情感分类。在电影评论语料库(rottentomatoes.com)的实验中，将该架构与递归神经网络和矩阵向量递归神经网络等其他的神经网络模型对比，该方案在5类情感分类任务中表现良好，准确率达到45.4％。

词级特征可以将大多数信息表示成词向量形式，并且可以较方便地衡量2个词之间的相似度，在情感分类任务中具有难以替代的作用。然而，句子中的词语并非词汇的堆砌，不同的句法会带来完全不同的情感表达，词级特征缺乏对文本语料整体上的考虑，在复杂句式中对情感的分类并不理想。随着深度学习方法的兴起，许多研究者开始将其应用于文本情感分析工作中。通过构建多隐层的模型，深度学习可以提取更深层的句子级特征，从而提高文本分类的准确率。Santos等人[86]提出一种基于字符到句子的卷积神经网络情感分类方法(character to sentence convolutional neural network，CharSCNN)，该方法利用一个含有双卷积层的神经网络，从字符、词和句子级别的信息中分别抽取特征。该方法在斯坦福影评情感树库(ssTb)的5种情绪分类任务中，平均准确率可达48.3％。

此外，部分研究者将递归神经网络也引入到情绪分类的工作中。由于普通的递归神经网络缺乏层级表示能力，Irsoy等人[90]提出了一种基于深度递归神经网络的情感分类方法。该方法将3个递归层叠加，利用非线性递归信息构成一个树型结构，递归计算每个节点的贡献值。该方法在斯坦福情感树库(SSTb)上的5种情感分类任务中，分类效果略优于文献[89]，平均准确率可达49.8％。

句子级特征在表达整体情感时的优秀表现，引发了研究者们向更高级情感特征的思考，一些研究者们开始着手研究篇章级的情感特征。Kang等人[91]从情感空间的角度对情感表达的作用进行了分析。该方法利用在中文情感语料(Ren-CECPs)中抽取的8种情感标记：exception，joy，love，surprise，anxiety，sorrow，anger，hate构成8维情感空间。根据这8种情感所构成的矩阵空间描述情感成分(词，词汇)，将这些情感成分通过内积的形式构成更高级(句子，篇章)的情感矩阵，利用SVM对中文语料(Ren-CECPs)进行情感分类。在中文博客9类情感分类的实验中，该方法的F值可达39.24％。

此后，Rao等人[92]提出一种基于主题的篇章级情感分析方法。该方法通过潜在主题建模、多种情感标签和众多读者共同标记来生成主题的特征。其中最大熵的过度拟合问题也通过将特征映射到概念空间得到缓解。在实际数据集(包括BBC论坛博客、顶客网的博客、MySpace评论、RunnersWorld论坛的博客、Twitter微博以及YouTube的评论)的实验中，验证了该方法在长文本和短文本情感分类上同样有效，准确率可达86.06％。

另一种情感分析问题是因为文本语料往往会涉及多个属性，文本情感分类可以仅仅看作是多标签分类任务中的一个属性，结合情感属性和其他相关属性，可以有效提高情感分类的准确率。Huang等人[93]提出了一个基于多任务的情感分类方法。该方法在按照情感分类的同时也进行基于主题的分类，对于每个任务用多个标签进行训练，有助于解决类歧义的问题。对真实的Twitter数据试验中，该方法准确率要高于朴素贝叶斯、SVM和最大熵模型，可达到74.4％。

此后，Zhang等人[94]结合了情感与社会领域知识2个重要指标，提出了一种基于因子图算法(factorgraph)的情感分类模型。该方法通过观察带注释的Twitter数据集，归纳出影响用户情感的2个主要因素：情感相关性和社会相关性。并将这2个因素作为特征，相较于决策树、SVM和逻辑回归等基准方法，该方法使用因子图算法取得了较好的分类效果，准确率可达72％。

总体上，基于有监督学习的方法在准确率上优于基于词典和规则的方法，但对样本数据的质量要求较高，需要花费巨大的时间成本和人力成本对语料进行标注，影响了该方法的推广。

### 5.2.2半监督学习分类方法

随着大数据时代的到来，数据的采集变得比以往任何时候都容易，标记数据却成为了有监督学习方法的瓶颈，半监督学习方法可以充分利用大量的未标记样本改善分类器性能，在情感分类任务中扮演着重要的角色，研究者们对此开展了大量研究。

半监督学习方法主要利用少量标记数据对训练样本进行初始化，Sun等人[95]将表情符号与一元特征、二元特征结合起来，提出了一种面向中文微博的半监督情感分类方法。该方法利用表情符号对未标记数据进行初始化，将语句中所含表情符号最多的一类标记为该语句的情感标签；再通过提取语句中词语的一元特征与二元特征，用SVM与朴素贝叶斯分类器将微博中表达的情感分为7种类别：乐、喜、悲、怒、恐、恶、惊。实验表明情感自动标记的准确率可达到88.7％，情感分类中朴素贝叶斯分类器要优于SVM，其精准率和召回率都超过71％。

此后，Sintsova等人[96]提出一种基于多项贝叶斯的半监督情感分类方法。该方法首先将根据情感词典将未标记的数据分为36个情感类别，并根据每个文本中选出的最突出标签对标注进行改进；然后从文本中抽取挖一grams特征并过滤无关信息；最后利用重新平衡的伪标记数据和多项贝叶斯分类器对微博语料进行分类。该方法在Twitter语料库的实验中F1值可达到20.2％。

另一部分研究者将半监督学习与Ekman，Plutchik等情感分类体系相结合，利用心理学情感分类知识对训练样本进行初始化。Purver等人[97]提出一种半监督学习与Ekman分类体系相结合的情绪分类方法。该方法采用少量人工选取的标签(hashtag)和情感符(emoticon)来自动标注微博情绪，以省去大量手工标注语料的过程。在Twitter语料的实验中，利用hashtag分类的准确率可达67.4％以上，利用emoticon分类的准确率可达75.2％以上。然而，该方法对恐惧、惊讶和憎恶(fear，surprise，disgust)3类情感的区分度不高，因为训练语料中标签和情感符的意思含糊，对区分情绪起到了干扰。

在此基础上，Suttles等人[98]也提出了一种基于离散二进制半监督学习的情感分类方法。与文献[97]的研究不同，该方法根据Plutchik的情感轮进行情感分类，把固有的多层次情感分类问题转换成一个含有4组对立情感的二进制问题，选取情感符号(emoticon)、标签以及表情符(emoji)作为参考进行人工标记。该方法首先提取不同类型的特定标签，并将表示相同情感类别的标签进行归类，然后对比每个独立的二元分类器的准确性。在Twitter微博数据测试中，该方法的分类准确率最高可达91％。

此后，Jiang等人[99]提出一种基于表情符号空间模型(emoticon space model)的半监督情感分类方法，该方法利用表情符号从未标记的数据中构建词向量，通过将词和微博映射到表情符号空间来确定微博的主观性、极性和情感。在中文微博基准语料库(NLP&&CC2013)实验中，该方法的情感分类准确率优于SVM及朴素贝叶斯，可达61.7％。

在文本情感分类任务中，有很多情况下需要对读者评论中的情感进行分析，而读者评论又与源文本之间存在紧密的联系。针对该现象，Li等人[100]提出一种基于双视图标签传播的半监督方法，对读者评论中的情感进行分类。该方法先通过词袋、二元特征提取文本中的情感信息，再通过双视图提取文本之间的对应关系。双视图依赖于2个图关系：包括源

文本之间的关系以及评论文本之间的关系。此外还将源文本与相应评论文本之间的依赖关系集成到这2个图中。最后在源文本和评论文本之间配置一个权重以处理评论文本中信息的不足。该方法在Yahoo Kimo News语料的情感分类实验中，准确率可达74.5％。

该类方法的优点在于可以较方便地获得大量的标记数据用以训练样本集，解决了有标记数据集稀缺的问题。然而，该类在第1次分类过程中分错的样本，会影响到第2次分类的准确率。

### 5.2.3复合情感分类方法

由于前2种分类方法的优缺点都很明显，一些研究者开始考虑综合2种方法，吸取各自方法的优点。这些复合分类方法主要分为2类，其中一类是将情感分类任务分解成有无情感、正负情感、细粒度情绪等子任务，再分别针对不同子任务设计不同算法的层次情感分类方法。

情感分类中，类别之间不是互相独立的，它们之间有一定层次关系。基于层次结构的复合方法就是利用这种层次关系，提高情感分类的准确率。Ghazi等人[101]提出一种基于层次情感分类的方法，该方法包含3层结构：1)第1层定义文本是否包含情感；2)第2层对第1层中有情感的文本进行正负划分；3)第3层将第2层的正类划为happiness，负类细化为sadness，fear，anger，disgust，surprise。该方法有助于粗粒度到细粒度的分类，在格林童话上分类的准确率可达60％以上。

在此基础上，Esmin等人[102]提出一种面向短文本的层次情感分类方法。该方法仍将第1层用来确定文本是否含有情感；第2层对上一层有情感的文本做极性分类，其中仅happiness是正类；第3层将负类分为sadness，fear，surprise，disgust，anger，最终利用MulticlassSVM分类器对微博语料进行分析。在Twitter语料库的情感分类实验中，该方法的平均准确率达63.2％。

此后，Xu等人[103]也用层级分类法对中文微博进行了情感分类，同时还将主成分分析法引入到情绪分类中，计算微博中主要情感的比例。该方法采用4层结构，将情感细分为19种类别。在新浪微博数据上进行了4层实验，其中第1层只采用平面型文本分类；第2层与第1层不同，采取了层级分类；第3层在第2层的基础上加人了词性特征；第4层在第3层的基础上还加入了心理学情感词典。通过第1层对文本是否有情感信息的分类和第2层正、负情感的分类，将无关的文本剔除，使后续分类工作更加容易。在层级分类的实验中，第3层的分类准确率可以达到90％左右，

由于微博语料通常是隐式的、不平衡的，为解决该问题，Zhang等人[104]提出了一种基于主题模型的层级情感分类法。该方法先对微博语料进行去除无关信息的预处理；然后根据主题模型进行特征选择，用选出的特征词和情感词典构成(情感，情感指示)关联，识别隐含的情感；最后构造一个树结构的层级分类器对微博进行分类。该方法在新浪微博语料库的情感分类实验中，F值可达70％。

该类通过将情感分类任务分解成较为有无情感检测、情感分类、细粒度情感等子任务，利用更为成熟的主客观检测、情感分类的技术对样本进行预分类，从而降低情感分类的难度。

另一类基于子类的复合方法是将语料库先分为更细致的子类，再利用分好的子类对样本进行分析，从而获得最终的情感分类。Keshtkar等人[105]也提出了一种基于层次的心情分类方法。该方法对博客的心情进行分类，总共分为132类，但是情感与心情有所区别，情感持续的时间比心情的要短，二者之间的层次结构和分类任务并不相同。此后，刘宝芹等人[21]在文献[105]的基础上，根据Ekman的6类情绪理论中情感极性与情感间的相互关系，为6类情感建立了3层树状结构，并利用该结构对不同话题微博的情感进行自动分析。在新浪微博的情感分类实验中，该方法比传统的贝叶斯方法情感识别精度更高，同时还降低了情感数据分布不均衡对结果的影响，该方法在6种情感分类任务中，平均精准率可达70.6％。

此后，欧阳纯萍等人[107]提出一种基于情感词汇本体的多种有监督学习复合方法。该方法使用朴素贝叶斯算法对微博是否有情感进行预分类，并根据分类结果对情感进行精确分析。该方法首先将情感词汇本体库的7种类别细分为21小类(快乐、安心、尊敬等)，并把这2l小类作为每条微博的最终特征，分别采用SVM和kNN(最一nearestneighbors)算法对预分类后的新浪微博数据集进行细粒度情感分类。在2013年CCF自然语言处理与中文计算机会议的中文微博情感分析评测任务中，该方法相较于单纯使用SVM和kNN分类器，其F值提高近11％，其对情感判别的准确率可达72.71％，表现优于单一分类算法。

归纳上述3类方法，主要是针对极性或单一情绪标签分类，忽略了情感标签在实例中多情感共存的情况。因此，情感分析与传统的情感分析相比，从另一个维度还可以看作是一个多标签情感分类问题。

## 5.3其他情感分类方法

传统的情感分析方法很少认为一个文本可以同时表达多种情感，而事实中一条语料可能出现有多种情感共存的情况。为了更准确地把握文本中所表达的情感信息，研究者们从另一个新的维度出发，开展了基于多标签情感分类的研究。

Yang等人[108]提出了一种多标签情感分类方法。该方法利用表情符号、标点符号和小型词典对数据进行标记，再用多标签情感分类(multi-labelemotionclassification，MEC)算法对微博进行分类。MEC算法同时考虑文本级和词级信息，先用kNN收集特定微博的情感信息，再利用朴素贝叶斯计算微博在词语层面属于任何一个情感类别的概率，最后设置一个阈值抽取微博的情感标签。准确率最高83.6％。

此后，Buitinck等人[109]提出了一种面向影评的多标签情感检测系统。该方法先通过词袋和篇章特征将句子标记为预设情感标签集的一个子集，然后分别用one-VS.-rest SVM和RAKEI。方法对文本进行分类。在IMDB影评数据集的实验中，RAKEL分类器的表现最好，因为多标签分类相较于单标签分类器，分类规则更加复杂。

在此基础上，Liu等人[110]提出了一种基于多标签的情感分类方法。该方法利用DUTSD，NTUSD，Hownet这3个情感词典提取微博语料中的情感特征和原始分割词特征，通过与MIkNN，BRkNN)等使用kNN算法的基准方法做对比，发现CIR分类效果最好，其平均准确率可达65.5％，HL为16.7％，AP为76.6％，OE为37.3％。

因为一个句子可能包含多种不同强度的情感，有些情感可以共存，而有些则不会同时出现。根据这一特性，一些研究者在多标签情感分析工作中加入了情感强度分布的要素。Wang等人[111]提出了一种多标签情感分类方法。该方法利用skip-gram语言模型训练词汇情感的分布式表达，将微博语句降维成词向量，并将二者作为卷积神经网络的输入，最终使用基于CI。R的多标签学习方法获得每条微博的最终情感标签排序。

针对现有语料库中情感类别不平衡的特征，Li等人[112]提出了一种基于最大熵(multi-label maxi-mum entropy，MME)短文本情感分布检测方法。该方法利用最大熵原理估计词与社会情感之间的关系，为了提高在新闻、微博等多种规模语料库中的预测能力，引入了L-BFGS算法来缓解约束。该方法在Semeval，SSTweet，ISEAR这3个语料库上的F1值分别为36.96％，90.30％，54.86％。

在此基础上，Zhou等人[113]提出一种基于情感分布学习的分类方法。该方法设计了一个从句子到情感分布的映射函数，用以描述多重情感和它们各自的强度，并引入Plutehik情感轮理论以提高情感检测的准确率。该方法在中文博客Ren-CECPs语料库的实验中，其平均准确率可达66.4％，HL为17.72％，AP为64.10％，OE为52.39％。

由于篇章级语料上下文信息对情感分布有着很大的影响，Xu等人[114]提出一种迭代多标记的情感分布检测方法，该方法利用句子内部特征(intra-sentence features)对句子进行初始分类，结合上下文信息，根据初始分类结果与整篇微博情感类别的转移概率，综合考虑转移概率、ML-kNN和随机多标签RAKEL这3种多标签分类器的分类结果，最终获得该微博的情感分布。

虽然基于多情感标签的分类方法的复杂性和难度都比较高，且该研究方向才刚刚起步，但其应用前景较好，将会成为一个新的研究热点。

文本情感分类任务主要指通过提取文本内容中的情感要素，将文本划分到一个或多个预定义的情感类别中，通过判定文本中所表达的情感类别，实现对文本发布者情感的监控、预测和管理。目前，情感分类方法主要包括：基于词典和规则的方法、基于机器学习的方法、复合方法以及其他方法。

# 情感分析评测及资源

## 6.1情感分析的评测

随着互联网的发展和带有情感色彩的主观性文本的增多，情感分析得到了越来越多的学者和研究机构的关注。近年来，为了推动情感分析技术的发展，国内外的很多研究机构纷纷组织了一些公共评测，为情感分析的方法研究提供统一的平台。

情感分析首先引起了国际文本检索会议TREC的关注，并从2006年开始每年都有情感分析相关的评测任务出现。由于TREC长年专注于检索方面任务的评测，因此TREC首次关注的情感分析任务是博客检索任务。对于给定的查询（话题），该任务要求在博客数据集上（近30GB，320万篇，2006年规模）检索带有观点的文档，并且这些文档必须含有主观性信息，而不能是纯客观的叙述[115]。除了观点检索任务之外，还有一个篇章情感分类的子任务，即为检索返回的文档进行情感分类，分为褒义、贬义和混合（positive，negative，mixed）3类。BlogTrack任务发展到TREC2009，有更多的情感分析的元素加入。如：判断返回的文档是主观评论还是客观事实、是深入的剖析还是浅显的总结；判断返回文档的博主是男士还是女士，以及是否是专家；判断返回的文档是个人博文还是公司博文等非常有意思的情感分析任务。

NTCIR（NIItestcollectionforIRsystems）的情感分析评测（multilingualopinionanalysistask，简称MOAT）同样出现在2006年，每年举行一次，并拥有中、英、日3种语言的标准语料库。不同于TREC所关注的观点检索，NTCIR评测的主要任务是从新闻报道中提取主观性信息。给定各个语种的句子，要求参加评测的系统判断句子是否与篇章的主题相关，并从句子中提取出观点持有者、评价词极性、评价对象等信息。分析NTCIR观点分析的路线可以看出，其目标是进行多语种、多信息源、多粒度、深层次的主观性信息提取[2]。MOAT任务发展到NTCIR-8，也融入了一些新的内容，如情感问答任务。给定某一情感问题，如“猪流感有哪些负面影响?”，从相关文本中找出正确的情感评价；又如，跨语言情感分析，即给定一个英文的查询，从4种不同语言的文档池中返回相关文档。

在国内，尤其是针对汉语的情感分析问题的研究才刚刚开始。COAE（Chineseopinionanalysisevaluation）始办于2008年，是国内第一个情感分析方面的评测。它致力于推动中文情感分析理论和技术的研究和应用，同时建立中文情感分析研究的基础数据集。COAE共设置6个任务[116]，可分为3个方面：一是中文评价词语的识别和分析，侧重于词语级的倾向性评测；二是中文文本倾向性相关要素的抽取，主要是抽取句子中的评价对象，侧重于有关倾向性的相关信息的抽取；三是中文文本倾向性的判别，侧重于篇章级的倾向性评测。COAE是首个提供产品类评价语料的评测，为中文情感分析的发展提供了很好的施展平台。

## 6.2情感分析的资源建设

### 6.2.1情感分析的语料

除了第5.1节中3个国际/国内评测所提供的语料以外，不少研究单位和个人也提供了一定规模的语料：

（1）康奈尔（Cornell）大学提供的影评数据集（http：//www.cs.cornell.edu/people/pabo/movie-review-data/）：由电影评论组成，其中持肯定和否定态度的各1000篇；另外，还有标注了褒贬极性的句子各5331句，标注了主客观标签的句子各5000句。目前，影评库被广泛应用于各种粒度如词语、句子和篇章级的情感分析研究中。

（2）伊利诺伊大学芝加哥分校（UniversityofIllinoisatChicago，简称UIC）的Hu和Liu提供的产品领域的评论语料：主要包括从亚马逊和Cnet下载的5种电子产品的网络评论（包括两个品牌的数码相机、手机、MP3和DVD播放器）。其中，他们将这些语料按句子为单元详细标注了评价对象、情感句的极性及强度等信息。因此，该语料适合于评价对象抽取和句子级主客观识别，以及情感分类方法的研究。此外，Liu还贡献了比较句研究[117]方面的语料。

（3）Wiebe等人所开发的MPQA（multiple-perspectiveQA）库：包含535篇不同视角的新闻评论，是一个进行了深度标注的语料库。其中，标注者为每个子句手工标注出一些情感信息，如观点持有者、评价对象、主观表达式以及其极性与强度。文献[118]描述了整个的标注流程。MPQA语料适合于新闻评论领域任务的研究。

（4）麻省理工学院（MassachusettsInstituteofTechnology，简称MIT）的Barzilay等人[119]构建的多角度餐馆评论语料：共4488篇，每篇语料分别按照5个角度（饭菜、环境、服务、价钱、整体体验）分别标注上1~5个等级。这组语料为单文档的基于产品属性的情感文摘提供了研究平台。

（5）中国科学院计算技术研究所的谭松波博士提供的较大规模的中文酒店评论语料：约有10000篇，并标注了褒贬类别，可以为中文的篇章级的情感分类提供一定的平台。

### 6.2.2情感分析的词典资源

情感分析发展到现在，有不少前人总结出来的情感资源，大多数表现为评价词词典资源：

（1）GI（generalinquirer）评价词词典（英文，http：//www.wjh.harvard.edu/~inquirer/）。该词典收集了1914个褒义词和2293个贬义词，并为每个词语按照极性、强度、词性等打上不同的标签，便于情感分析任务中的灵活应用。

（2）NTU评价词词典（繁体中文）。该词典由台湾大学收集，含有2812个褒义词与8276个贬义词[75]。

（3）主观词词典（英文，http：//www.cs.pitt.edu/mpqa/）。该词典的主观词语来自OpinionFinder系统。该词典含有8221个主观词，并为每个词语标注了词性、词性还原以及情感极性。

（4）HowNet评价词词典（简体中文、英文，http：//www.keenage.com/html/e\_index。html）。该词典包含9193个中文评价词语/短语，9142个英文评价词语/短语，并被分为褒贬两类。其中，该词典提供了评价短语，为情感分析提供了更丰富的情感资源。

# 情感分析最新方法

## 7.1 基于深度混合文本群体模型的情感分析

### 7.1.1概述：

众包技术提供了一个有效的平台来应用人才分析，由于语境，写作风格，观点等的大变化，这对于自动语言模型来说是一项艰巨的任务。然而，当每个工人的人群标签数量不足以训练参数时，或者当不可能为大数据集中的每个样本收集标签时，标准众包聚合模型是不称职的。在本文中，我们提出了一种新的混合模型，用于利用人群和文本数据进行情感分析，包括生成众包聚合模型和深度感性自动编码器。这两个子模型的组合是基于概率框架而不是启发式方法获得的。我们引入了统一的目标函数来结合两个子模型的目标，并推导出一种有效的优化算法来共同解决相应的问题。实验结果表明，与最先进的模型相比，我们的模型取得了优异的成果，特别是当人群标签稀缺时。

社交媒体的快速增长使用提供了关于不同主题的公众意见的巨大来源。有效挖掘这些意见对各行各业都非常有价值。例如，酒店，航空公司，贷方，银行甚至政客都利用这些数据来寻找新的客户，定位新产品，分析客户的个性并做出更好的决策。然而，由于文本的不同变化，例如不同的语境，作者的性别，写作风格和不同的观点，探索公众意见的情绪对于自动语言模型来说是一项非常具有挑战性的任务。

### 7.1.2 混合情感分析模型

首先介绍混合模型，展示其架构并解释其背后的直觉。 然后，基于概率框架制定其统一的目标函数，并导出用于更新参数和估计真值的优化算法。

#### 7.1.2.1 CrowdDeepAE架构

所提出的混合模型，由CrowdDeepAE表示，由两个主要部分组成，一个用于文本数据的深度去噪自动编码器和一个用于人群标签的聚合模型。图2演示了CrowdDeepAE的体系结构，其中深度去噪自动编码器有两个任务：通过噪声重建损坏的文本数据并从文本数据中估计真值。众包聚合模型也应该用来估计嘈杂的人群标签的真相。因此，真相是通过人群和文本数据的贡献获得的。在CrowdDeepAE中耦合深度自动编码器和人群聚合模型有几个优点：1）Crowd- DeepAE利用两个信息源来更准确地估计事实，通过自动编码器的编码器路径文本数据和通过聚合模型的人群数据; 2）多层自动编码器为文本样本提供了强大的判别功能，在学习现实世界文本数据的非线性嵌入空间方面具有比浅模型更多的能力; 3）去噪自动编码器中的重构损失函数起着数据依赖正则化项的作用，间接阻止了众包聚合模型过度拟合; 4）CrowdDeepAE能够注释整个数据集，甚至是没有任何人群标签的样本，因为自动编码器可以使用有限数量的人群标签和无人监督的重建任务的监控进行有效训练; 5）聚合模型利用群众工作者的语义知识协助训练自动编码器，由于文本数据的变化很大，这是非常有益的。 6）与天真的非联合学习方法相比，用于CrowdDeepAE的联合学习框架导致更优化的结果，其中文本和人群子模型被单独训练。

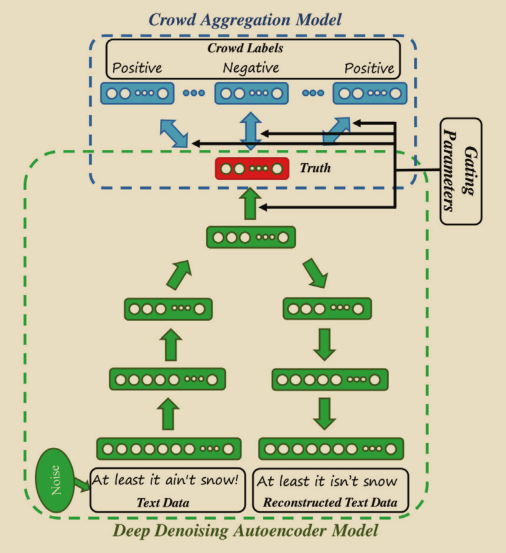
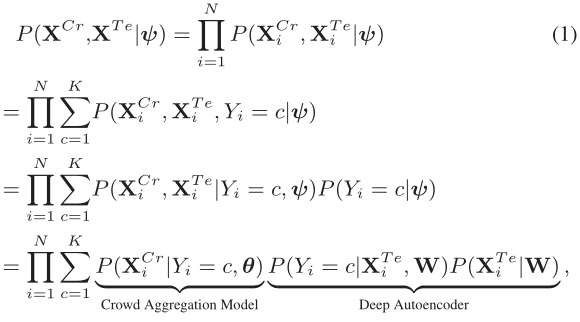


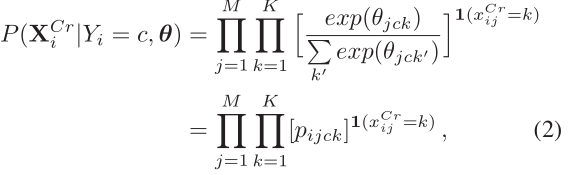
图1 CrowdDeepAE架构，由深度去噪自动编码器和人群聚合模型组成

#### 7.1.2.2 CrowdDeepAE目标函数

让我们考虑众包任务包括N个问题，每个问题都有K个可能的选项。 人群和文本数据分别由X = {X Cr，X Te}表示，Y表示未知的真实标签。 我们提供了一个概率框架来组合我们的自动编码器和聚合子模型，从而为我们的混合模型定义一个统一的目标函数。 给出观察结果（X Cr，X Te）的CrowdDeepAE参数（ψ）的一般似然函数是：

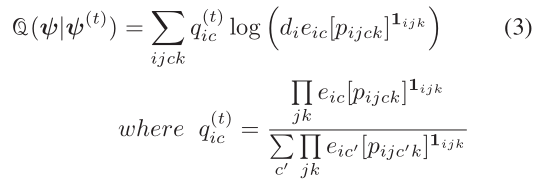


其中i和c是问题和选项的索引，W和θ分别代表自动编码器和人群聚合子模型的参数。 注意，假设样本是独立且相同分布的（i.i.d），并且假定X Cr i和X Te i在条件上是独立的，给定真实标签。 我们现在能够分解方程式中的似然函数。 （1）进入人群聚合和深度自动编码器目标。 考虑到M群众工作人员在众包任务中受雇，我们的生成人群聚合模型具有以下形式。



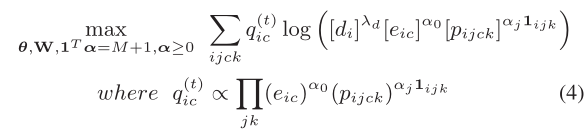
其中X Cr i = {x Cr i1，...，x Cr iM}是第i个问题的人群标签集。另外，当i是真实标签时，p ijck显示人群标签的概率，使得第j个工作者为第i个问题选择第k个选项。因此，每个问题的人群数据的联合概率是基于每个有条件独立的人群标签的概率。聚合模型将一个混淆矩阵θj视为每个工人的可靠性参数，其中较高的对角元素θjkk表示工人的可靠性更高。此外，指数非线性增加了我们的众包聚合模型在处理嘈杂的人群标签时的灵活性。对数似然函数L（ψ| X）= logP（X Cr，X Te |ψ）的直接优化是困难的，因此我们使用期望最大化（EM）学习方法来解决该问题。接下来，我们提出命题1以减轻优化问题，并在附录A中提供其证明。为了更简单的符号，此后我们表示P（Y i = c | X Te i，W）= e ic和P（X Te i | W）= di作为编码器和解码器路径的概率，分别为1（x Cr ij = k）= 1 ijk。

命题1：迭代地改进以下辅助函数Q足以使对数似然函数L（ψ| X）最大化。

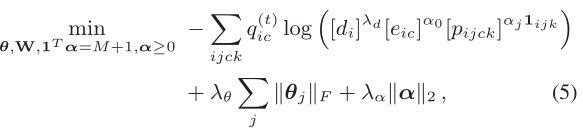


其中q（t）ic表示事实的概率分布。 从技术上讲，相对于当前参数，期望未知的真实标签。 因此，我们可以迭代地改进辅助函数Q而不是对数似然函数L.

在Q函数中，自动编码器和人群工作者在目标函数和计算真值方面具有相似的效果。 然而，由于所有问题的学习知识，预计深度自动编码器比不熟练的人群工作者具有更准确的预测。 因此，我们使用可调整的权重来控制每个因素在我们的目标函数中的影响。 接下来，我们提出更新的目标函数，它可以类似于命题1得出。



其中α和λd分别是可调整权重和自动编码器重建损失的超参数。 注意，α可以看作是门控参数（见图2），它可以调整每个工人和自动编码器在估算真值时的贡献。 换句话说，α为我们的混合模型提供了关于群体工作者和自动编码器可信度的更多自由度。 例如，当有几个（非专家）人群工作者用（非常嘈杂的）人群标签标记问题时，（判别性）自动编码器的高权重可以帮助准确地估计真相。 注意，我们通过λd定义解码器路径概率的权重，因为d i不影响真值，并且仅调节自动编码器目标函数。 此外，我们为参数添加了两个更多的正则化惩罚项，以避免过度拟合。



其中λθ和λα是正则化项的超参数。 另外，为我们的目标函数添加两个约束（在最小操作下）是有益的，因为它具有竞争性学习并避免了平凡的解α= 0。

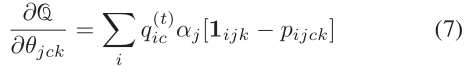
#### 7.1.2.3 CrowdDeepAE优化算法

为了有效地解决问题（5），我们采用交替学习策略来更新参数并估计真相。 特别是，参数ψ= {θ，α，W}中的每一个都被更新，而其他参数和真值是固定的，而真值的概率Q = {q 1，...，q N} 当假定参数已知时估计。

更新θ：更新人群聚合模型参数的问题减少为：



有几种一阶优化算法可用于解决此问题。 使用参数θ的目标函数的以下梯度，我们使用L-BFGS算法迭代地更新参数。

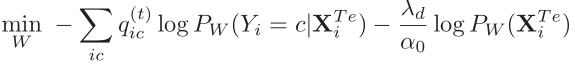


更新α：更新门控参数的问题归结为：



我们使用拉格朗日乘法器方法有效地解决了这个问题。

更新W：更新深度去噪自动编码器参数的问题有以下形式。

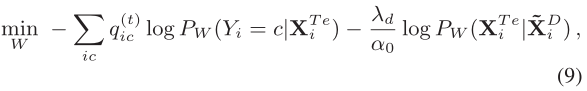


其中第一项是分类问题的标准交叉熵损失函数。 但是对于第二个概率项，我们使用（Bengio et al.2013）中的一个定理，以便在标准去噪自动编码器中将术语更改为重建损失函数。

一般的想法是，如果观测变量X被具有条件分布C（~X | X）的噪声破坏成~X，则训练去噪自动编码器实际上估计反向条件分布P（X | ~X）。 已经证明，P（X）的一致估计可以使用马尔可夫链来估计，马尔可夫链在P（X | ~X）的采样和C（~X | X）的采样之间交替进行，如下所示



该定理证明了传统去噪自动编码器的P W（X | ~X）（Vincent等人2008; Bengio等人2013; Ghasedi Dizaji等人2017）是真实条件分布的一致估计。 同样，当样本数N→∞时，去噪自动编码器生成的样本的渐近分布收敛于原始数据生成分布。 因此，我们将文本模型的目标函数重新表述如下：



其中~X D i是被随机噪声破坏的样本。现在很清楚我们如何使用去噪自动编码器作为基于文本的子模型。

该学习方法在处理非常大的数据集时没有内存耗尽问题。对于群体工作者来说，我们可以将人群数据分成几个小批量，每个小批量只包括一些工人的人群标签。处理大量文本样本，我们能够将文本样本分发到一组小批量中，并使用随机优化算法训练自动编码器参数。因此，可以使用随机和并行学习方法来管理计算和空间复杂性。

算法1显示了CrowdDeepAE算法，其中真实性首先通过多数表决来初始化。然后，它在更新模型参数和估计真值之间交替，直到收敛。值得一提的是，我们使用E-step中的干净文本样本来计算真值。但是方程中带有噪声文本输入的分类丢失函数。 （9）在训练参数W时具有正则化效应，并导致更加鲁棒和通用的自编码器模型。

### 7.1.3 实验

首先评估我们的混合模型在人群聚合任务中的表现，然后检查学习语言模型的质量。 为了将所提出的模型与最先进的聚合模型进行比较，我们使用两个大规模的众包数据集，其中包含文本数据以及用于情绪分析的人群标签。

#### 7.1.3.1 数据集

CrowdFlower（CF）数据集是2013年Crowdsourcing at Scale共享任务挑战的一部分，由CrowdFlower 2收集，作为有关天气的推文的情感分析的丰富资源。该数据集包括569,375个人群标签，共98,980条推文。但黄金标准（真实）标签仅提供300条推文，相当于从761名工人收集的1720个人群标签。在人群任务中，要求工作人员使用以下选项标记与天气相关的推文的情绪，负数（0），中性（1），正数（2）与天气无关（7）。人群工作者也能够通过不能告诉（5）选项跳过问题。

情感极性（SP）数据集包括人群工作者关于两个类别的电影评论的情绪分析，“新鲜”（正面）和“腐烂”（负面）。该数据集由来自RottenTomatoes网站3的电影评论中的5,000个句子组成，其由（Pang和Lee 2007）提取。一名任务请求者雇用了203名人群工作人员来标记数据集，总共产生了27,777个人群标签。所有问题的黄金标准标签都可以在SP数据集中找到。

#### 7.1.3.2 实验设计

第一步，对于CF和SP数据集，我们首先使用词干分析方法来解析文本（Porter 1980），然后删除常用的英语停用词，最后根据术语频率 - 逆文档频率（tf-idf）提取前1000个词。 得分（Baeza-Yates，Ribeiro-Neto等1999）。

第二步，对于深度自动编码器，我们考虑编码器和解码器路径的三个完全连接的层，其中512,256和128个神经元作为特征映射，然后在编码器路径的顶部添加softmax层。 泄漏整流激活（泄漏RELU）用作自动编码器层的激活功能，除了解码器路径末端的重建层，其已经整流激活（RELU）以重建文本样本。 此外，我们将学习率设置为10 -7并采用Adam（Kinga和Adam 2015）作为我们的优化方法。 所有层的权重也通过Xavier或GlorotUniform初始化方法（Glorot和Bengio 2010）初始化。

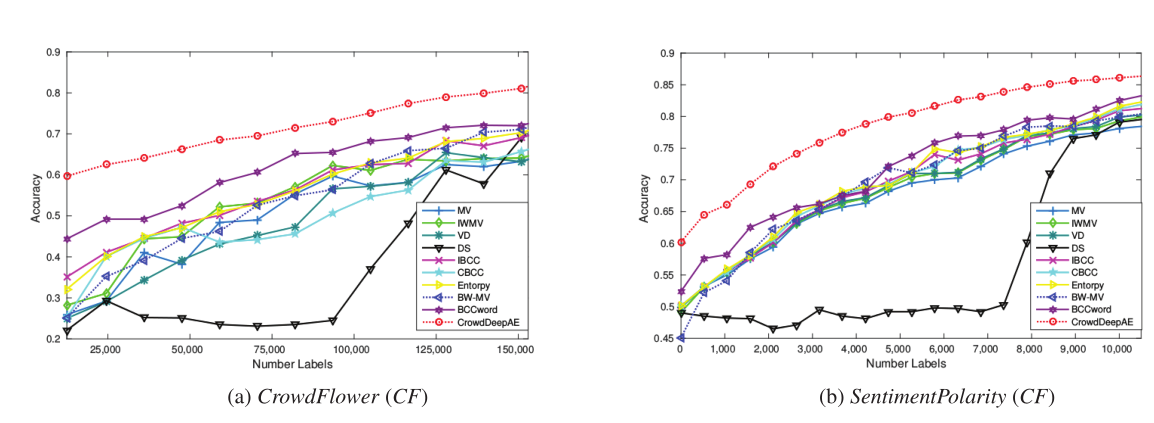
第三步，设置参数，由于众包任务是一个无监督的问题，我们没有使用任何真正的标签来设置超参数{λθ，λα，λd}和辍学噪声值。 我们使用（Tian和Zhu 2015）中的技巧，其使用非相关可能性来选择超参数。 特别地，我们利用似然函数p（X Cr | Y，θ）来选择λα，λd和λ集中的丢失α= {0.01,0.1,1}，λ集d = {0.01,0.1， 1}和dropout set = {0.1,0.2,0.3}，并采用p（Y | X Te，W）作为从λ集θ= {0.01,0.1,1}中选择λθ的标准。 因此，使用这种方法，我们确保在没有真正标签知识的情况下选择超参数。

#### 7.1.3.3 实验结果分析

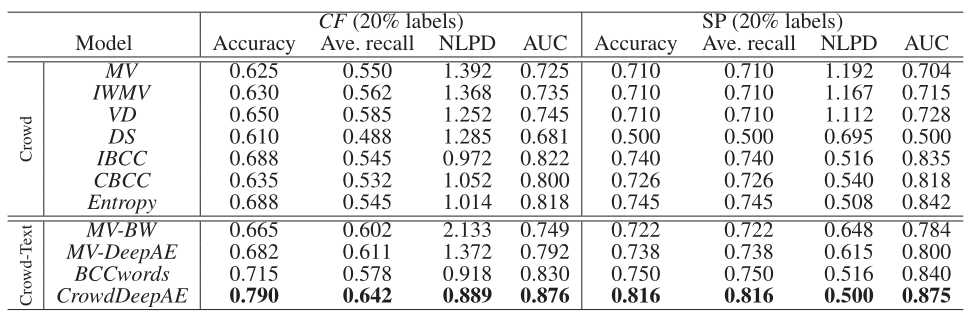
为了评估我们模型的性能，我们使用CF和SP数据集进行了几次实验，以使用人群标签和文本数据来估计真相。 为了便于比较，我们使用以下替代模型和比较度量。

实验一：替代模型

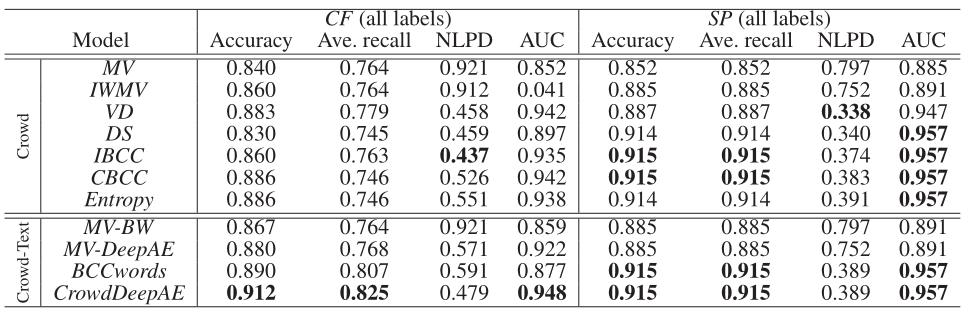
将模型Crowd-DeepAE与几种基线方法进行比较，包括多数投票（MV），迭代加权多数投票（IWMV）（Li和Yu 2017），投票分配（VD），Dawid和Skene模型（DS）（Dawid and Skene 1979），独立贝叶斯分类器组合模型（IBCC）（Simpsonetal.2013），基于社区的贝叶斯分类器组合模型（CBCC）（Venanzi等人2017），多类极小极大熵模型（熵） ）（Zhou et al.2017），主要投票聚合模型和词袋文本分类器（MV-BW）的组合，多数投票聚合模型的组合和类似于我们的自动编码器的深感情自动编码器（MV-DeepAE），贝叶斯分类器与单词模型（BCCwords）相结合（Simpson等，2015）。应该注意的是，VD可以被认为是MV的概率版本，因为它计算每个选项的概率，同时假设所有工人具有相同的可靠性。此外，MV-BW模型训练一个经典的词袋使用由多数表决聚合模型引发的目标标签的文本数据分类器。同样，MV-DeepAE使用多数投票聚合模型的预测标签来训练文本数据的深度自动编码器模型。替代模型的结果来自参考文献，除了MV-DeepAE之外，我们使用与CrowdDeepAE类似的自动编码器网络实现。



当增加人群标签的数量时，CrowdFlower（CF）和SentimentPolarity（SP）数据集上的众包聚合模型的准确性



CrowdFlower（CF）和SentimentPolarity（SP）数据集上的众包聚合模型的比较，当20％的人群标签可用时。 比较指标是准确性，ave。 回想一下，AUC（越高越好）和NLPD（越低越好）



当所有人群标签可用时，CrowdFlower（CF）和SentimentPolarity（SP）数据集上的众包聚合模型的比较。 比较指标是准确性，ave。 回想一下，AUC（越高越好）和NLPD（越低越好）

为了检验上述聚合模型的有效性，我们使用CF和SP数据集的不同子集（人群标签数量）进行了多次实验。以下（Simpson等人2013），当只有2％随机选择的人群标签可用时，我们使用聚合模型估计真相。然后，我们通过添加额外2％随机选择的人群标签来增加人群标签的数量，并重新运行所有模型。重复该过程，直到所有人群标签用于训练。图3显示了CF和SP数据集上聚合模型的准确性。如图所示，CrowdDeepAE始终优于其他具有显着利润率的模型，特别是当有少量人群标签可用时。有趣的是，在CF数据集中，我们的模型只需要16％的所有其他模型，这些模型使用30％的人群标签。 CrowdDeepAE在CF数据集中与8％的人群标签相比也实现了更高的准确度，而拥有30％人群标签的MV模型。此外，CrowdDeepAE不断提高MV-DeepAE的性能，从而证实了我们的联合学习框架和人群聚合子模型的重要性。请注意，为了清晰可视化，我们仅在图3中显示了有限部分的结果（CF和SP数据集中大约150,000和10,000个人群标签）。

此外，表1和表2分别报告了CF和SP数据集上聚合模型的上述比较指标，当时有20％和100％的人群标签可用。我们将表格中的模型分为两组单一模型和混合模型，其中第一组仅使用人群标签来估计真相，第二组使用人群标签和文本数据进行预测任务。仅使用20％的标记，大约70 ％文本样本至少有一个人群标签。在这种情况下，使用文本数据更为重要，因为没有足够的人群标签来训练人群参数。混合人群文本模型具有比人群模型相对更好的性能，因为混合模型能够使用语言模型来对没有人群标签的样本进行分类。但人群模型的人群标签不足以进行培训，并根据先前的分布为未标记的样本分配默认类别。我们提出的模型CrowdDeepAE受益于由一小部分人群数据训练的深度自动编码器，并且能够在没有人群标签的情况下有效地标记样本。当只有20％的人群标签可用时，我们的模型根据所有指标优于SP和CF数据集上的替代模型。此外，与使用所有人群标签的两个数据集上的最新模型相比，CrowdDeepAE仍然获得了优越或竞争性的结果。这表明CrowdDeepAE利用强大的深语言模型以及有效的人群聚集模型，使用人群和文本数据提供准确的预测。

## 7.2 针对aspect层面情感分类的潜层观点学习

### 7.2.1 实验目标

方面级别情绪分类旨在检测对句子中的特定目标表达的情绪。基于观察到情感极性通常与给定句子中的特定跨度相关，可以利用这些信息来更好地分类。另一方面，这些信息也可以作为与预测相关的理由。我们提出了一种基于分割关注的LSTM模型，该模型可以使用线性链条件随机场（CRF）层有效地捕获目标和传感表达式之间的结构依赖性。该模型模拟人类在阅读时推断情绪信息的过程：当给出目标时，人们倾向于在对基础情绪信息做出明智决定之前在句子中搜索周围相关的文本跨度。我们通过SemEval任务和Twitter的社会评论，对不同语言的在线评论的公开数据集执行情绪分类任务。

### 7.2.2 实验方法

问题可以表述如下。 给出一个带有n个单词的句子{w 1 ... wi，w i + 1，...，wj，w j + 1，...，wn}和一个目标{wi，...，wj}这是 在句子中的跨度，我们需要预测特定目标的情绪极性。 架构如图2所示。接下来，我们将从下到上依次引入所有组件。

#### 7.2.2.1 模型

问题可以表述如下。 给出一个带有n个单词的句子{w 1 ... wi，w i + 1，...，wj，w j + 1，...，wn}和一个目标{wi，...，wj}这是 在句子中的跨度，我们需要预测特定目标的情绪极性。

架构如图2所示。接下来，我们将从下到上依次引入所有组件。

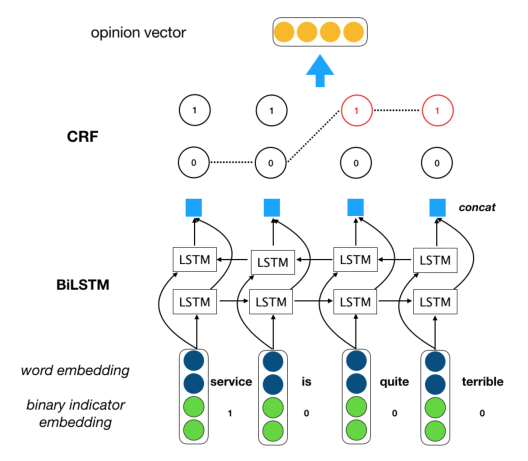


图2 模型架构图

目标术语“服务”通过嵌入以绿色表示的二进制指示符来表示。 该模型使用BiLSTM对每个单词的上下文信息进行编码。 基于其隐藏状态，内部分割关注层旨在选择感兴趣的意见词。 以红色显示的节点意味着它们对应的单词对目标情绪的分类影响最大。

#### 7.2.2.2 输入层

首先，通过查找嵌入表将单词映射到它们的向量表示。 与使用单词嵌入直接编码目标的传统方法（Wang等人2016b; Chen等人2017）不同，我们使用二进制特征来指示每个单词是否是目标的一部分，因为目标的位置信息可以 由此由指示序列编码。 遵循（He et al.2017）中的类似想法，使用随机初始化的嵌入表将每个二元指示符映射到向量表示。

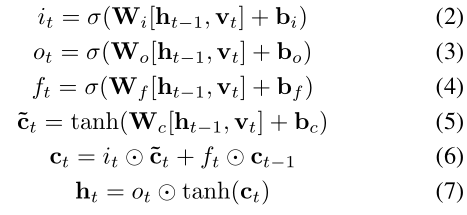
从图2中可以看出，每个单词的表示形式是通过连接单词的嵌入和二进制特征嵌入形成的：



其中t∈[i，j]是指示第t个字是否属于目标跨度[i，j]的二元函数。 Wemb∈Rd 1×| V | Wmask∈Rd 2×2是两个矩阵，其中| V | 是词汇大小，d 1和d 2分别是字嵌入和二进制特征嵌入的维度。 如果给出了方面术语，我们还使用可学习的向量对它们进行编码，这些向量连接到上面的表示。

#### 7.2.2.3 双向LSTM层

然后使用BiLSTM来捕获每个单词的上下文信息。 前向LSTM计算如下：



其中W i，W o，W f，Wc∈Rdh×（d 1 + d 2 + dh）是用于不同门的权重矩阵，bi，bo，bf，bc∈Rdh是偏置矢量和d 1 + d 2和dh分别表示LSTM的输入矢量和隐藏状态的维数。 表示元素乘法和σ代表sig moid函数。

后向LSTM与前向LSTM非常相似，只是输入序列以反向方式馈送。 我们连接前向和后向LSTM的隐藏状态以形成最终表示。 注意，我们将目标信息包括在等式1中作为输入，因此每个r i可以被视为给定句子中位置i处的单词的特定于目标的表示。

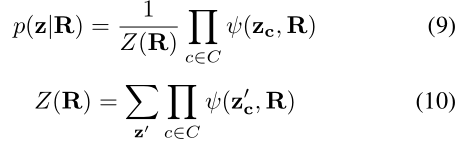


其中[·，·]指的是连接两个列向量以形成单个列向量的操作。 我们使用R = {r 1，r 2，...，r n}来表示为句子生成的所有单词表示。

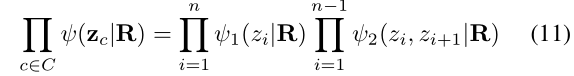
#### 7.2.2.4 注意力机制分割

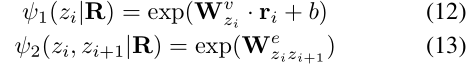
为了直接模拟目标与其意见（以及意见词之间）之间的结构依赖关系，我们引入了一个分段关注层来捕获它们。

如图2所示，我们为每个单词引入一个潜在的二进制变量z∈{0,1}。 此变量指示其对应的单词是否是意见表达的一部分。 具体而言，我们结合线性链CRF来指定这些潜在变量之间的结构依赖性。 形式上，所选序列的分布参数化如下：



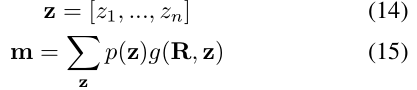
其中z由z c形成，每个都在单个集合c上定义，并且ψ（z c，R）是集合c的潜在函数，Z（R）是分区函数。 通常，我们分别在这些无向图形模型的顶点和边上定义两种电位。



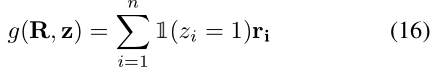


这里Wv∈R2×2d h将上下文表示映射到每个潜状态的特征得分，We∈R2×2是为每对潜状态定义的转移矩阵。

特征函数 我们的目的是根据先前的选择生成潜在意见的表示，我们称之为图2中的意见向量。 方程15给出了计算该值的函数的一般形式。



其中g（R，z）是基于意见选择定义的特征函数，m是该特征函数的期望。 但是，枚举所有可能的选择在计算上是昂贵的。 为了使程序易于处理，我们只定义每个顶点上的特征，以便可以根据边际概率计算此函数，这些边际概率可以使用动态编程通过消息传递或前向 - 后向样式算法有效地计算：



给定等式15和16，意见向量可以简化为以下形式。



我们可以直观地从这个等式17中看出，分段注意层实际上是从有助于目标情感极性的句子中提取信息。 在模型预测期间，我们使用维特比式解码算法明确且准确地提取潜在观点。 这些以不受控制的方式提取的意见使我们有机会进一步分析和评估我们的模型。

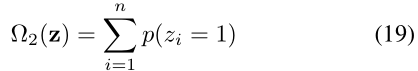
正则化 在最初的实验中，我们观察到这种基于注意力的模型所犯的常见错误是它们倾向于关注情感词，即使这些词在语义上与目标没有关联。 基于这样的假设，即意见表达通常是短而连贯的跨度而不是断开的句子词，我们引入了两个额外的正则化指导模型。

隐藏变量z基本上有两种状态：是意见的一部分。 根据观察结果，只有少数意见跨度实际上会对目标的情绪产生影响，因此不鼓励不同国家之间的频繁转换。 这引起了我们下面的第一个调节器，它试图鼓励国家保持不变：



具体而言，它强制不同状态之间的转换特征值小于同一状态之间的转换特征值。 否则，该模型将受到惩罚。

第二个正规化器试图强制执行模型以处理真正重要的短距离和少量跨度：



从本质上讲，这些正则化因素带来了一些结构偏差，因此模型可以通过细分关注层专注于短暂但有意义的意见跨度。

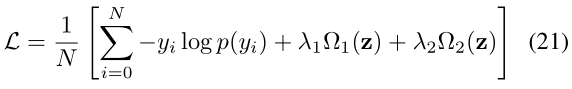
#### 7.2.2.5 目标函数

使用softmax计算情感标签的分布：



其中Wtag∈R2dh将意见向量m映射到每个情感标签的特征分数，b标签是偏差项。

我们主要关注两种模型：基于香草分段注意力的LSTM（SA-LSTM）和用附加正则化器（SA-LSTM-P）增强的版本。 在小批量训练模式中，SA-LSTM-P的损失函数定义如下：

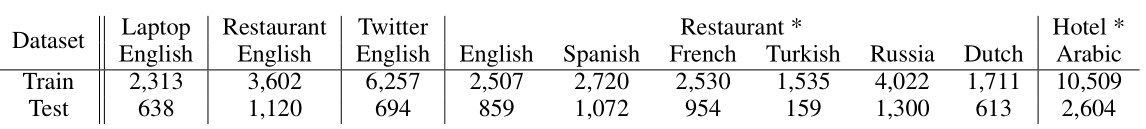


其中λ1和λ2表示每个正则化器的系数。 没有这些惩罚条款，它就成为SA-LSTM的目标函数。

### 7.2.3 实验分析及结论

我们主要使用两组数据集进行两组实验。第一组用于详细分析模型的每个组件。在第二组中，我们专注于检查模型的语言敏感性问题。

第一组由来自SemEval- 2017任务7（Pontiki等人2017）的评论数据集和（Dong等人2017年）收集的Twitter评论组成。 SemEval-2017任务7包含来自两个领域的评论：餐厅和笔记本电脑。在之前的工作（Tang，Qin和Liu 2016; Chen et al.2017）之后，作为预处理的一部分，我们还丢弃了包含“冲突”标签的句子（其中不同的情绪表达了相同的方面）。为了评估模型所学习的潜在观点，我们还使用了来自（Wang等人2016a）的这两个数据集的附加注释，其中包括用于意见术语的手动注释标签。通过使用实体词查询Twitter API来收集Twitter数据集，因此每个句子通常仅与一个目标配对。在第二组中，我们使用来自SemEval-2016任务5的餐厅和酒店评论，其中包括七种不同的语言。所有统计数据如表1所示。



每个域中的h个句子，目标，极性的数量。 带有“\*”的数据集来自SemEval 2016 Task 5。

#### 7.2.3.1 基线方法

•SVM（Kiritchenko等，2017）：SemEval 2017任务7中报告的最佳方法，它采用词典功能，表面特征和SVM的解析功能。

•AdaRNN（Dong et al.2017）：该模型学习基于使用递归NN为目标转换的依赖关系树，适应性地将上下文词的情绪传播到目标节点。

•AT-LSTM（Wang等人，2016b）：目标的表示用于通过在LSTM之上采用的标准注意来捕获相关联的单词。

•MemNet（Tang，Qin和Liu 2016）：该模型在获得正确的情绪之前，会在一定次数（多跳）中进行标准关注。

•RAM（Chen et al.2017）：与MemNet类似，它在多个注意事项之间添加了一个重新呈现的功能，以模拟不同注意步骤之间的内在依赖关系。

请注意，AT-LSTM，MemNet，RAM都是基于注意力的方法。 MemNet和RAM的结果来自（Chen et al.2017）。 多个注意力背后的直觉是，可能存在多个语义成分，这些成分有助于特定目标的情绪。 类似地，分词注意也可以被视为单词级别的“多重注意”。 2使用此设置，模型可以更好地模拟多个注意事项之间的依赖关系。 将模型与它们区分开来的另一点是我们使用二进制指示器功能在输入层对目标信息进行编码。 对于SemEval 2016评论数据集，我们的模型可能是第一种利用注意力机制的方法。 我们将其与竞赛中报告的两个最佳模型和两个基于LSTM的模型进行比较。

•XRCE（Brun，Perez和Roux 2016）：该模型包含许多基于解析器生成的语法特征的手工制作规则。

•IIT-TUDA（Kumar等人，2016年）：它结合了域依赖图功能和每种语言的大量情感词典。

•LSTM，HP-LSTM（Ruder，Ghaffari和Breslin 2016）：BiLSTM用于对句子进行编码，然后将第一个和最后一个隐藏状态组合起来用于情绪预测。 HP-LSTM添加了审查级别的LSTM，以捕获同一评论中的句子之间的信息。

为了了解BiLSTM和分段关注层的工作情况，我们分别实现了基本版本。 我们还实施了标准的基于注意力的模型进行比较。

•A-Softmax：不使用LSTM，该模型直接对每个单词的输入嵌入执行注意机制。 它可以被视为直接从句子中选择情感词而不考虑上下文信息的过程。

•SA-Softmax，SA-Softmax-P：该型号以分段注意取代了A-Softmax的标准注意力。 SA-Softmax-P增加了罚款条款。

•A-LSTM：LSTM用于捕获关注层之前的上下文信息。 除了我们使用不同的目标表示之外，该模型与AT-LSTM非常相似。

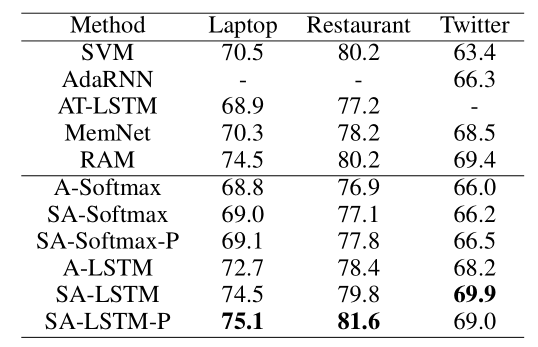
•SA-LSTM，SA-LSTM-P：在LSTM之上使用分段注意层。 这是本文的主要模型。对于SA-LSTM-P中的学习过程，还给出了一个学习过程。

#### 7.2.3.2 训练过程

我们使用来自GloVe（Pennington，Socher和Manning 2017）的300维词嵌入用于英语数据集。 其他语言的预训练嵌入来自（Ruder，Ghaffari和Breslin 2016）3，它们也被我们比较的方法所使用。 目标二进制指示器嵌入的维数为30.我们在所有实验中修复了嵌入字，因为我们发现如果我们不断微调它们就很容易过度拟合。 Dropout也在输入图层之后使用，并针对每个数据集进行调整。 λ1在0和1之间调谐，λ2从[0,0.2]中选择，步长为0.07。 对于LSTM，我们将隐藏的尺寸大小设置为300.将六分之一的训练数据作为调整超参数和进行模型选择的验证集。 使用随机梯度下降训练模型，使用Adam的更新规则（Kingma和Ba 2015）。

#### 7.2.3.3 实验结果

表2：来自SemEval 2017任务7的评论结果以及来自Twitter的评论准确度（％）



基于第一组数据集的主要结果可以在表2中找到。与SVM和AdaRNN相比，基于注意力的模型在不需要手工制作的特征或外部解析器的情况下是有效的。 LSTM层和分段关注层都不断提高所有域的性能。与MemNet和RAM相比，它们通过多跳扩展了标准注意力，分段注意模型获得了更好的性能，因为它可以捕获更多有利于此任务的结构信息。我们可以凭经验得出结论，在这个任务中，同时参与情感跨度比一次多于一个单词更自然有效。惩罚术语通常对分割注意层有积极影响，除了来自Twitter的社交评论。我们发现Twitter文本相对嘈杂且结构较少，因此，正规化程序的引入往往导致包含错误的观点。此外，从结果我们可以观察到，当与LSTM组件配对时，分段关注层的表现更好。

表3：SemEval 2016任务5审核数据集的准确性（％）结果

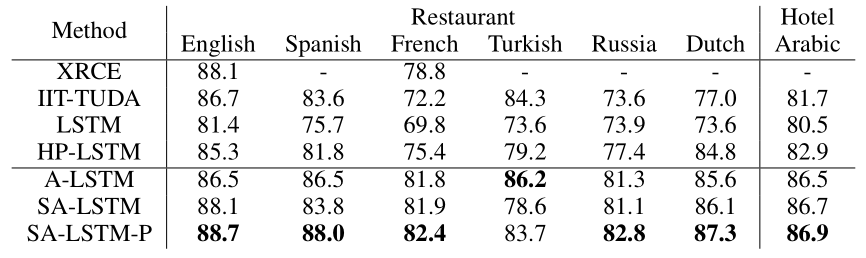
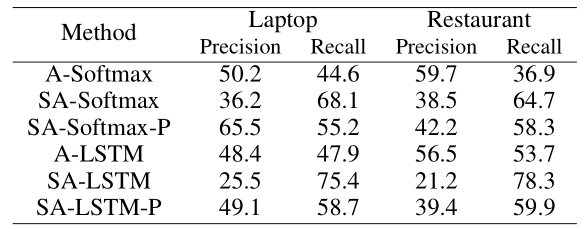


表3显示了来自SemEval 2016任务5的七种语言的评论数据集的结果。根据上述发现，关注层提高了所有数据集的性能。 与前两个基于昂贵的手动特征工程的模型相比，这种改进尤其适用于手工制作的功能不足的低资源语言。 由于注意机制的有效性，我们的模型优于传统的LSTM。 此外，除了土耳其评论7之外，基于细分注意力的模型通常比标准的基于注意力的模型获得更好的结果，这在很大程度上证明了我们模型的语言不敏感性。 此外，分段注意力一直受益于额外的正则化器，我们可以得出结论，我们对跨度的假设适用于不同的语言。

我们已经观察到了我们提出的注意机制的有效性。但是，注意力层是否根据需要关注真实的意见表达。为了理解这一点，我们再采取进一步措施来评估预测期间的潜在意见。我们提取情感词或跨度，即对最终预测具有相对显着影响的意见。对于标准的基于注意力的模型，我们通过在注意层上执行最大操作来提取单个意见词。对于分段注意，我们使用维特比解码算法提取最佳意见跨度，这导致一个或多个意见跨度。两者都使用（Wang et al.2016a）在单词级别提供的注释意见进行评估。

表7：在单词级别无监督提取意见的表现（％）



主要结果显示在表7中。与不使用BiLSTM的模型相比，我们发现BiLSTM通常有助于回忆更多的意见表达，因为一些句子词是依赖于上下文的。例如，当“技术”与“技术”配对时，“高”是积极的，但当与“价格”配对时，它变为否定。 BiLSTM为模型提供了对句子的基本理解。通过直接模拟意见词之间的依赖关系，分段注意力以类似的方式工作。由于这些数据集相对较小，BiLSTM可能在捕获上下文信息方面受到限制。分段注意为模型带来的结构偏差有助于识别更多的意见表达。此外，当与BiLSTM结合时，分段注意通常表现得更好，因为如果在每个表示中编码上下文信息，则更容易对交互进行建模。

与标准注意相比，分割注意力可以自然地回忆更多的意见表达，因为模型捕获的是一系列单词而不是单个单词。 直观地，更多识别的情感词或跨度可以揭示模型的更多情绪信息以做出正确的决定。 然而，当插入分段注意力时，高召回率也导致低精度。引入两个正则化器来平衡它们。 他们帮助该模型基于假设意见应该是短的和连贯的跨度来关注正确的意见。

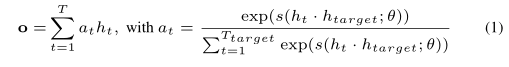
## 7.3基于注意的LSTM用于目标依赖的情感分类

我们提出了一种基于注意力的双向LSTM方法来改善依赖于目标的情绪分类。 我们的方法学习目标实体和最显着特征之间的对齐。 我们对现实数据集进行了广泛的实验。 实验结果表明，我们的模型实现了最先进的结果。

### 7.3.1 实验模型

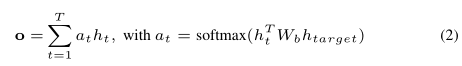
我们提出了两种基于注入的双向LSTM（AB-LSTM）方法来改善目标依赖情绪分类的性能。 AB-LSTM学习根据与任务的相关性将注意力分数分配到不同的单词位置。 我们通过根据每个单词的注意力得分计算每个单词的隐藏状态的加权和来将输入序列压缩为固定长度的向量。

AB-LSTM1我们的第一种注意方法是计算序列中每个单词的权重，作为h t和h target的点积。 这里，h t是基于当前输入x t通过双向LSTM获得的时间步长t的隐藏状态，h target是目标串的隐藏状态。 在数学上，最终句子表示采用以下形式：



其中T是序列的长度，s（h t\*h target ;θ）是关注网络（θ是关注网络的参数集），它将向量映射到实值得分。 为了在上下文中形成适当的概率分布，我们使用softmax函数对整个上下文中的分数进行归一化并获得注意分数a t。 注意分数a t表示相应时间步长t的重要性。 通过这种方式，通过适当训练的注意力网络，它将更高的注意力分数分配给与特定任务更相关的单词。 在我们的模型中，s（h t\*h target ;θ）被实现为另一个神经网络，其参数与B-LSTM的参数共同训练。

AB-LSTM2受（Chen，Bolton和Manning 2016）的启发，我们的第二种方法通过使用biliner项来计算注意力得分，而不是h t和h target的点积。 使用双线性项，我们可以比使用点产品更灵活地计算h t和h目标之间的相似性。 输入序列的输出向量是



其中Wb∈Rh×h用于双线性项。

我们将outpur向量o馈送到softmax分类器以预测目标字符串的情绪分布：



这里，V和b是要学习的参数。 我们通过最小化预测分布y和地面实况分布y之间的交叉熵来训练整个模型。 使用L 2 - 正则化惩罚。 给定训练样本x i，其真实标签yi∈{1,2，...，k}其中k是类别的数量，并且预测概率yi∈[0，1]



其中Θ代表模型的参数。 I {·}是一个指标，I {true} = 1且I {false} = 0.我们使用小批量随机梯度下降（SGD）算法来训练模型。

### 7.3.2 实验分析及结论

#### 7.3.2.1 数据集

Twitter对话（Twitter）：原始数据集1是Twitter推文的集合（Dong et al.2017）。 培训数据包括6,278条推文，测试数据包含692条推文。 训练和测试集中的正，负和中性推文的百分比分别为25％，25％，50％。

  对于数据集，执行数据预处理。 我们从文本中删除非字母字符，数字，代词，标点和停用词。

#### 7.3.2.2 基线方法

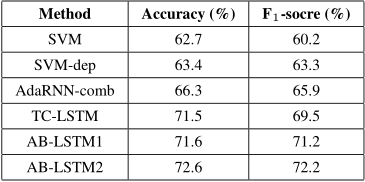
我们将我们的方法与几种基线方法进行比较，包括SVM（Pang，Lee和Vaithyanathan 2002），具有目标依赖特征的SVM（SVM-dep）（Jiangetal.2011），AdaRNN-comb（Dong et al.2017） ，TC-LSTM（Tang et al.2015）。

#### 7.3.2.3 实施细节

在我们的工作中，我们使用具有100维度的GloVe向量2来初始化单词嵌入。 前向和后向LSTM各有500个单位，深输出中softmax隐藏层的大小为500.我们使用minibatch SGD算法训练模型。 每个小批量包含20个输入样本。 我们将循环权重矩阵初始化为随机正交矩阵。 所有偏置矢量都初始化为零。 任何其他权重矩阵通过高斯分布的采样来初始化，均值为0，变量为0.01 2。 超参数λ= 0.8。

#### 7.3.2.4 实验结果

表1：评估结果



在我们的实验中，使用分类精度和F 1 -score评估结果。 我们总结了表1中的实验结果。与以前的方法相比，我们的方法在实验数据集上取得了更好的结果。 例如，AB-LSTM2的F 1 -score比现有技术结果（Target-dep +）高2.7％。 为了准确起见，我们的两个模型都比Target-dep +和TC-LSTM表现更好。 这验证了所提出的方法的有效性。

## 7.4 用于目标情感分析的门控神经网络

概述：针对性的情绪分析将给定文本文档中提到的每个目标实体的情绪极性分类。精通方法从自动语法分析树中提取手动离散特征，以便捕获关于目标实体提及的封闭句子的语义信息。最近，已经表明，可以在不使用语法分析器的情况下实现竞争准确性，语法分析器在诸如推文之类的噪声文本上可能非常不准确。这是通过根据目标实体提及在简单和直观的推文分段上应用分布式词表示和丰富的神经池功能来实现的。在本文中，我们通过提出一个句子级神经模型来扩展这个想法，以解决池功能的限制，这些功能没有明确地模拟推文级语义。首先，使用双向门控神经网络来连接推文中的单词，以便可以在隐藏层上应用池化功能而不是单词，以更好地表示目标及其上下文。其次，使用三向门控神经网络结构来模拟目标提及与其周围环境之间的相互作用。实验表明，与目前最佳的目标传感分析方法相比，我们提出的模型具有更高的精度。

## 7.5 利用L1正则化优化模型识别情感词

概述：情感词识别是情感分析和意见挖掘的众多应用中的基础工作，例如审阅挖掘，意见持有者查找和推特分类。 在本文中，我们提出了一种具有L 1正则化的优化模型，称为ISOMER，用于从语料库中识别情感词。 我们的模型可以使用具有情感标签的种子词和文档，这与仅采用种子词的大多数现有研究不同。 目标函数中的L 1惩罚产生稀疏解，因为大多数候选词没有情绪。 对真实数据集的实验表明，ISOMER优于传统方法，ISOMER学习的词汇可以有效地适应文档级情感分析。

## 7.6 使用神经网络的上下文敏感Twitter情感分类

概述：Twitter上的情感分类近年来吸引了越来越多的研究。 根据推文内容本身，大多数现有工作都侧重于特征工程。 在本文中，我们提出了一种基于上下文的Twitter情感分析神经网络模型，将相关推文的语境化特征以单词嵌入向量的形式结合到模型中。 平衡和非平衡数据集的实验表明，我们提出的模型优于当前最先进的模型。

## 7.7 利用上下文嵌入发现情感分析的概念本源

概述：随着最近深度学习的发展，人工智能的研究获得了新的活力和突出。 虽然机器学习已经成功地振兴了许多研究领域，例如计算机视觉，语音识别和医学诊断，但我们还是看到了自然语言理解方面令人印象深刻的进步。 这种不匹配的期望背后的原因之一是，尽管自下而上的方法对于模式识别是可行的，但推理和理解通常需要采用自上而下的方法。 在这项工作中，我们将子符号和符号AI结合起来，自动发现文本中的概念原语，并将它们与常识概念和命名实体联系起来，形成新的三级知识表示，用于情感分析。 特别是，我们采用循环神经网络通过词汇替换来推断原语，并通过多维缩放将它们用于基础知识和常识知识的基础。

## 7.8 用于情感分析的半监督自动编码器

概述：在本文中，我们研究了自动编码器在建模文本数据中的用法。传统的自动编码器至少受到两个方面的影响：具有词汇量大的高维度的可扩展性和处理与任务无关的词。我们通过自动编码器的丢失功能引入监督来解决这个问题。特别地，我们首先在标记数据上训练线性分类器，然后使用从线性分类器获知的权重来定义自动编码器的损失。为了减少单个分类器带来的偏差，我们对分类器的权重进行了后验概率分布，并利用拉普拉斯近似推导出自动编码器的边缘化损失。我们表明，我们可以从Bregman Divergence的角度合理化我们对损失函数的选择，这证明了我们模型的合理性。我们评估了我们的模型对六个情绪分析数据集的有效性，并表明我们的模型在分类准确性方面显着优于所有竞争方法。我们还表明，我们的模型能够利用未标记的数据集并提高性能。我们进一步表明，我们的模型成功地学习了高度特定的特征图，这解释了其卓越的性能。

## 7.9 使用主题丰富的多原型Word嵌入改进Twitter情感分类

概述：已经表明，学习分布式单词表示对于Twitter情绪分类非常有用。 大多数现有模型依赖于每个单词的单个分布式表示。 这对于情绪分类是有问题的，因为词语通常是多义的，并且每个词在不同的主题下可以包含不同的情感极性。 我们通过学习主题丰富的多原型字嵌入（TMWE）来解决这个问题。 特别是，我们开发了两个神经网络，1）学习通过结合主题信息更好地捕获推文上下文的单词嵌入，以及2）为每个单词学习主题丰富的多个原型嵌入。 SemEval 2013中Twitter情绪基准数据集的实验表明，TMWE优于顶级系统，具有手工制作的功能，以及当前最好的神经网络模型。

## 7.10 利用语义相似度和方面关联改进意见aspect提取

概述：方面提取是细粒度意见挖掘的关键任务。 虽然许多研究人员对它进行了研究，但仍然具有很大的挑战性。 本文提出了一种新的无监督方法来进行重大改进。 该方法基于终身学习框架，并以两种形式的建议实施，这些建议分别基于语义相似性和方面关联。 使用8个评论数据集的实验结果显示了所提出方法的有效性。

## 7.11 9篇最新 AAAI 情感分析相关的论文

[1]Ghasedi K, Huang H. Sentiment Analysis via Deep Hybrid Textual-Crowd Learning Model[C]//Thirty-Second AAAI Conference on Artificial Intelligence (AAAI 2018). 2018.

[2]Cambria E, Poria S, Hazarika D, et al. SenticNet 5: discovering conceptual primitives for sentiment analysis by means of context embeddings[C]//Proceedings of AAAI. 2018.

[3]Wang B, Lu W. Learning Latent Opinions for Aspect-level Sentiment Classification[J]. 2018.

[4]Yang M, Tu W, Wang J, et al. Attention Based LSTM for Target Dependent Sentiment Classification[C]//AAAI. 2017: 5013-5017.

[5]Liu Q, Liu B, Zhang Y, et al. Improving Opinion Aspect Extraction Using Semantic Similarity and Aspect Associations[C]//AAAI. 2016: 2986-2992.

[6]Liu Q, Liu B, Zhang Y, et al. Improving Opinion Aspect Extraction Using Semantic Similarity and Aspect Associations[C]//AAAI. 2016: 2986-2992.

[7]Zhang M, Zhang Y, Vo D T. Gated Neural Networks for Targeted Sentiment Analysis[C]//AAAI. 2016: 3087-3093.

[8]Deng Z H, Yu H, Yang Y. Identifying Sentiment Words Using an Optimization Model with L1 Regularization[C]//AAAI. 2016: 115-121.

[9]Zhai S, Zhongfei (Mark) Zhang. Semisupervised Autoencoder for Sentiment Analysis[C]//AAAI. 2016: 1397-1700.

# 8文本情感分析应用

文本情绪分析已经在各个领域得到应用并具有良好的应用前景。

## 8.1舆情监控

舆情分析，就是通过收集和整理民众态度，发现相关的意见倾向，从而客观反映出舆情状态。从古代的“防民之口甚于防川”，到现在的网络时代，“每个人都有了自己的麦克风”。互联网为社情民意的表达提供了平台，体现用户意愿、评论和态度的网络舆情也愈发受到重视。所谓网络舆情，就是对社会热门问题持有不同看法的网络舆论，是社会舆论的一种表现形式，也是公众通过互联网对现实生活中某些热点、焦点问题发表具有较强影响力、倾向性的言论和观点。网络舆情的2个重要特点就是网络非理性情绪和群体极化。

许多非理性的情绪，如仇富、仇官、反权力、反市场等，借助暴力性和娱乐化的网络表达强化，使人们变得更加情绪化和极端化。网民的非理性情绪，对社会存在潜在威胁，值得警醒。另一种特征“群体极化”是由美国教授Robert提出的，就是“起初团队成员拥有某种方面的潜在倾向，在讨论之后，人们朝着所倾向的方向继续移动，最后形成极端的观点”。例如最初群体中成员的意见都比较保守，在经过了群体的商议后，决策就会更加保守；相反，若个体成员意见倾向于冒险化，则经商议后的群体决策就可能会更趋向于冒险。

社会的安全管理需要不断关注网络舆情动向，并及时正确引导网络舆论方向，保证社会的长治久安。然而，各种渠道得到的信息庞杂，只靠人工方法进行甄别无法应对海量信息。因此，研发精确有效的情绪分析系统，实现对舆情信息的自动处理，对维持社会稳定有着非常重要的意义。

## 8.2商业决策

随着互联网的发展，网购在生活中愈发普及，人们通过C2C(如淘宝网、易趣网等)和B2C(如京东网、亚马逊等)形式的电子商务购买商品后，写下对商品的评论。其他消费者通过这些评论可以了解商品质量、售前售后服务，并且直接影响他们的购买决定。同时，生产商通过分析在线评论信息和情绪特征，可以获得消费者的行为特点，预测消费者偏好的变化趋势。此外，销售商还可以通过分析消费者的对商品或售前售后服务的心理状态，获得促销对消费者情绪的影响，为改善营销行为提供决策基础，从而获得竞争优势。

## 8.3观点搜索

随着信息时代的到来，网络数据呈现出爆炸式的增长，激发了用户从互联网海量信息中搜索有效信息的需求。在搜索过程中同时考虑搜索关键字和用户的情感诉求，可以使搜索变得更加便捷、准确和智能。情绪检索技术是解决该问题的重要方法之一，其任务是从海量文本信息中查询文本所蕴含的观点，并根据主题相关度和观点倾向性对结果进行排序。情绪检索返回的结果需要同时满足主题相关性和情感倾向性，情感倾向性既可以是文本带有的情感倾向，也可以是指定情感倾向的文本。因此，情绪信息检索是比情绪分类更加复杂的任务。为满足互联网用户日益增长的搜索需求，2006年国际文本检索会议(TREC)首次引入博客检索(blog track)任务。Mishne叫在LiveJournal blogs上标注了37种情绪类别，并利用频率统计、博客长度、语义特征等方法实现了对博客的情绪分类，为情绪检索提供了基础。此外，在图书、随笔等长文本观点搜索任务中，Mohammad提出一种基于情绪词密度的观点搜索方法。该方法利用谷歌书库定义的情绪实体与共现词之间的关联，发现了童话和小说之间情绪词密度的区别，并可按照情绪类别组织文本集，提高了长文本的搜索性能。此外，该方法还支持对文本中所含情绪的可视化展示与追踪。

## 8.4信息预测

随着越来越多的人热衷于参与到微博等网络互动中，微博对人们的生活也带来了巨大的影响。情绪分析技术可以通过对微博上的新闻、评论等信息进行分析，预测事件的发展趋势，其主要的应用方向包括3个方面

8.4.1金融预测

情绪分析在金融中的巨大应用潜力引起了研究者们的兴趣。美国印第安纳大学和英国曼彻斯特大学的学者发现了一个有趣的现象：Twitter可以从一定程度上预测3～4天后的股市变化。他们通过OpinionFinder方法将人的情绪分为正面和负面2种模式，再利用GPOMS将情绪分为更加细致的类别，包括：冷静(calm)、警惕(alert)、确信(sure)、活力(vital)、友善(kind)和幸福(happy)6类。若将其中的“冷静”情绪指数后移3天，竟与道琼斯工业平均指数(DJIA)惊人的相似。研究者们推测：在股票市场中，微博上对某支股票的议论可以影响投资者的行为，从而进一步影响股市变化的趋势。Devitt等人胡通过对金融文本所表达的情感极性判断，也实现了在一定程度上预测市场交易、股票价格和公司收益波动性的未来走势。

万方数据

8.4.2选情预测

情绪分析在选情预测中也扮演着重要的角色。在美国大选期间，Tumasjan等人通过挖掘和分析民众在Twitter上对各竞选团队的评论，制定针对摇摆州(美国大选中的一个专有名词，指竞选双方势均力敌，都无明显优势的州的特定宣传政策，从而提高己方的民意支持率。此后，Kim等人通过对网络新闻的分析，以81。68％的准确率成功预测出美国大选花落谁家。此外，在2011年意大利议会选举和2012年法国总统大选过程中，Ceron等人用情绪分析计算出政治领导候选人的Twitter支持率。

8.4.3其他预测

情绪分析还可用于对政策性事件的民意预测，如延迟退休的年龄等，为国家相关政策的制定提供辅助支撑。此外，情绪分析还可以应用到疫情、地震等自然灾害的判断和预测。随着信息预测的应用内容越来越丰富，情绪分析技术愈发受到重视。情绪分析技术通过分析互联网新闻、博客等信息源，可以较为准确地预测某一事件的未来走势，无论是政治经济领域还是日常生活中都具有重大意义。

## 8.5情绪管理

用户在微博、社区和论坛中的社交活动都是现实生活对网络社会的映射，这些社交网站中储存了大量的用户个人言论。由于用户的情绪与其所关注的话题通常具有较强的连续性，分析用户发布的言论可以较为准确地获得人们的生活状态和性格特点。Golder等人通过研究Twitter用户在昼夜和不同季节所展现的情绪节奏，包括用户在工作、睡觉等不同时间段内表现的情绪，绘制出心情曲线，从而了解人们的精神状态。此后，Kim等人通过研究也发现人们的情绪在6点、11点、16点和20点达到了高峰，并总结了用户一天中的情绪总体走向。利用这些研究成果，公司可以了解员工的工作状态，从而更有效地制定工作计划。此外，Zhou等人对不同行业名人的微博进行分析，统计名人发布微博中各类情绪的比例，可以分析出不同名人的性格、关注点和个人喜好。随着时代的进步和社会的发展，人们对自我关注的需求不断提高。通过对用户进行情绪分析可以让用户更加了解自我，从而找到更加适合自己的方式去学习、工作和生活，情绪管理领域也将拥有更广阔的应用市场。

随着应用的发展与需求的变化，文本情感分析研究任务更加繁重，基于正负二分类的情感分析作为多分类情绪分析的前期准备和一种特例，其成熟的研究框架和流程值得研究与借鉴。

# 9研究展望

目前针对文本情感分析的研究已经取得了一定的成果，但该研究领域还处于一个相对年轻的阶段。文本情感分析技术在理论和应用上都还存在一些挑战以及新的方向需要进一步研究探讨。

## 9.1面临的挑战

当前文本情感分析研究面临来自多方面的挑战，主要包括：数据稀缺性，无论是情感训练语料还是情感词典资源，都处于比较匮乏的阶段；类别不平衡，收集到的样本中情感各类别的数量明显存在差异；领域依赖性，情感词在不同领域的表达存在差异；语言不平衡，当前大多数工作都基于英文语料，语言迁移存在困难。

### 9.1.1数据稀缺性

文本情感分析主要包括基于情感词典和规则的方法、基于机器学习的方法。然而，无论是哪种方法，数据都很稀缺。在基于词典的方法中，情感词典很难获取资源，目前尚无公开的情感词典可用。此外，即使有开源的情感词典，由于网络新词层出不穷，需要不断对情感词典进行扩充和更新；在基于机器学习的方法中，需要借助有情感标注的语料库来提取特征并训练情感分类器。然而情感标注语料本身也是稀缺资源，由于不同领域的情感表达有不同特点，通用的情感训练语料无法满足不同领域研究的需求。

### 9.1.2类别不平衡

情感分析的工作已经开展很多年，目前大多数工作都假设正负样本是均衡的。情感分析是在情感分析的基础上进行更细粒度的分类。然而，不同情感的数据集规模往往不均衡，在实际收集的微博语料中，一些情感类别的语料数量明显多于另一些类别，例如表达喜欢的语料明显多于表达害怕的。所以，适用于均衡分类的方法在面对不均衡数据时效果往往并不理想。样本数据的不平衡分布会使机器学习方法在进行分类时严重偏向于样本多的类别，从而影响到分类的性能。

### 9.1.3领域依赖性

同一个词在不同的领域背景下表达着不同的情绪，例如“不可预测”在电影评论领域是褒义的，在汽车评论领域则是贬义的。因此，在进行情感分析时，应该充分考虑情感词的领域依赖性。跨领域情感分析是文本情感分析的一个重要研究课题，跨领域情感分析有很多问题需要解决。例如，在一个领域的意见表达，在另一个领域可能反转。此外，还应该考虑不同领域情感词汇的差异。

### 9.1.4语言不平衡

现有情感分析工作大多基于英文[120]，虽然近些年对中文的情感分析也有了一定的研究成果[83I，但是基于情感词典或语义知识库的工作都需依赖特定语种的外部资源，基于英文的情感分析研究很难迁移到其他语言。此外，由于非英语的情感分析训练集和测试集也相对匮乏，极大限制了非英语语种的情感分析研究。

## 9.2未来研究方向

当不同媒体、不同形态的情感信息“融合”在一起，会随之产生“质变”。与此同时，领域自适应、社交网络分析和深度学习等技术的发展，也给文本情感分析研究指出了新的研究方向。从技术的发展趋势分析看，未来文本情感分析的研究还需要关注如下4个方面。

### 9.2.1基于多媒体融合的情感分析

传统的情感分析主要关注文本，然而图片等多媒体通常可以比文本表达更明显的情感效果，即所谓的“一图胜千言”。此外，另一种情感信息表达的主要载体--语音，也可以很好地反映用户的当前情绪状态。因此，随着图像、音频等不同类型社交网络数据的不断增长，各种类型的用户数据相结合的研究将具有更好的应用前景。

### 9.2.2基于领域自适应的情感分析

领域自适应技术可以利用信息丰富的源域资源提升目标或模型的性能。传统情感分析方法为了克服情感词本身具有的领域依赖性，刻意选择领域无关的特征，如表情符号。而领域自适应的方法可以充分利用情感词在不同领域所表达的不同情感，准确、快速地识别文本情感。因此，随着不同领域情感语料资源的积累，基于领域自适应的情感分析将逐渐成为一个新的研究热点。

### 9.2.3基于社交网络分析的情感分析

社交网络的迅猛发展，产生了大量的用户交互数据，这些数据反应了用户的思想、情感及社交关系。通过结合社交网络的关系分析技术，可以了解不同的社会群体是如何表达情感以及情感倾向。因此，研究基于社交网络分析的情感分析技术可以更好地

掌握大众情感走向，为舆情分析、情感管理等应用提供支撑。

### 9.2.4基于深层语义的情感分析

深度学习作为机器学习研究中的一个新领域，取得了很大的进展。在自然语言处理的各项任务中，深度学习也有着许多可喜的成果。随着计算能力不断提高、数据量不断增加，可以预测未来将涌现出更多优秀的神经网络模型。该类方法将在自动抽取情绪特征、减少人工标记工作等方面做出巨大贡献。另外，在深度学习算法中加入一定的策略，可以更好地学习词汇和句子的语义表达，从而实现理解句子以及整个文档的任务。

在此基础上，随着互联网+时代的到来，涌现出大数据分析、特定主题挖掘、用户画像构建和多语言协同等众多新的应用需求，也给文本情感分析带来了新的机遇。

### 9.2.5面向大数据的文本情感分析

大数据技术的发展使数据的收集变得非常容易且成本低廉，对海量的信息数据进行挖掘，可以获得巨大的产品或服务价值。然而收集的数据大多以非结构化文本形式存储，在对文本数据进行情感分析时，传统的概率潜在语义分析方法的时间复杂度和空间复杂度较高，难以满足训练大规模数据的需求，需要提出面向大数据的文本情感分析方法。

### 9.2.6面向特定主题的情感分析

由于情感表达在不同主题下有所不同，无论是采用有监督还是无监督的学习方法，情感分析的准确率在一定程度上受主题的影响。同样的短语在不同的主题下，其语言规则、词库判断标准都存在不同：现有针对主题差异的研究工作，在实际应用中仍存在不少问题。例如，当主题差别过大时，分析性能会明显下降，需要对跨主题情感分析的算法和相关问题开展进一步研究。

### 9.2.7面向个性化的情感分析

随着互联网+时代的到来，用户对各类应用都提出了个性化的需求，而情感是一种高度主观的用户行为特征，同样的情感词汇根据不同的用户历史情感变化也会产生不同的情感含义及强度。因此，通过观察用户情感变化曲线，为不同的用户构建情感画像，从而实现利用有限的信息对用户情感进行个性化精准分析。

### 9.2.8面向多语言的情感分析

随着文化交流的增加，多种语言的网络信息相互影响与融合。现有工作主要针对单一语言，而在单一语言情感分析中所收集到的语料资源与成果，无法在多语言的环境中直接使用；此外，不同语言情绪分析的语料资源也存在不均衡性，难以在跨语言的环境中直接使用。在解决情感分析任务的基本问题外，还需要考虑机器翻译、多语言文本处理等工作，这都对多语言情感分析提出了新的需求。

参 考 文 献

[1] Hatzivassiloglou V, McKeown KR. Predicting the semantic orientation of adjectives. In: Proc. of the EACL’97. Morristown: ACL,1997. 174−181.

[2] Huang XJ, Zhao J. Sentiment analysis for Chinese text. Communications of CCF, 2008,4(2) (in Chinese with English abstract).

(黄萱菁,赵军.中文文本情感分析.中国计算机学会通讯,2008,4(2).)

[3] Yao TF, Cheng XW, Xu FY, Uszkoreit H, Wang R. A survey of opinion mining for texts. Journal of Chinese Information Processing, 2008,22(3):71−80 (in Chinese with English abstract).

(姚天昉,程希文,徐飞玉,汉思•乌思克尔特,王睿.文本意见挖掘综述.中文信息学报,2008,22(3):71−80)

[4] Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2008,2(1-2):1−135. [doi:10.1561/1500000011]

[5] Zhou LZ, He YK, Wang JY. Survey on research of sentiment analysis. Journal of Computer Applications, 2008,28(11):2725−2728(in Chinese with English abstract).

(周立柱,贺宇凯,王建勇.情感分析研究综述.计算机应用,2008,28(11):2725−2728)

[6] Rao D, Ravichandran D. Semi-Supervised polarity lexicon induction. In: Lascarides A, ed. Proc. of the EACL 2009. Morristown: ACL, 2009. 675−682.

[7] Wiebe J. Learning subjective adjectives from corpora. In: Schultz AC, ed. Proc. of the AAAI. Menlo Park: AAAI Press, 2000.735−740.

[8] Riloff E, Wiebe J. Learning extraction patterns for subjective expressions. In: Collins M, Steedman M, eds. Proc. of the EMNLP2003. Morristown: ACL, 2003. 105−112.

[9] Turney P, Littman ML. Measuring praise and criticism: Inference of semantic orientation from association. ACM Trans. onInformation Systems, 2003,21(4):315−346. [doi: 10.1145/944012.944013]

[10] Kim SM, Hovy E. Automatic detection of opinion bearing words and sentences. In: Carbonell JG, Siekmann J, eds. Proc. of the IJCNLP 2005. Morristown: ACL, 2005. 61−66.

[11] Kim SM, Hovy E. Identifying and analyzing judgment opinions. In: Bilmes J, et al., eds. Proc. of the Joint Human Language Technology/North American Chapter of the ACL Conf. (HLT-NAACL). Morristown: ACL, 2006. 200−207.

[12] Zhu YL, Min J, Zhou YQ, Huang XJ, Wu LD. Semantic orientation computing based on HowNet. Journal of Chinese Information Processing, 2006,20(1):14−20 (in Chinese with English abstract).

(朱嫣岚,闵锦,周雅倩,黄萱菁,吴立德.基于HowNet 的词汇语义倾向计算.中文信息学报,2006,20(1):14−20)

[13] Andreevskaia A, Bergler S. Mining WordNet for a fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In: McCarthy D, Wintner S, eds. Proc. of the European Chapter of the Association for Computational Linguistics (EACL). Morristown: ACL, 2006. 209−216.

[14] Su F, Markert K. Subjectivity recognition on word senses via semi-supervised mincuts. In: Ostendorf M, ed. Proc. of the NAACL 2009. Morristown: ACL, 2009. 1−9.

[15] Esuli A, Sebastiani F. Determining the semantic orientation of terms through gloss analysis. In: Herzog O, ed. Proc. of the ACM SIGIR Conf. on Information and Knowledge Management (CIKM). New York: ACM Press, 2005. 617−624.

[16] Esuli A, Sebastiani F. Determining term subjectivity and term orientation for opinion mining. In: McCarthy D, Wintner S, eds. Proc. of the European Chapter of the Association for Computational Linguistics (EACL). Morristown: ACL, 2006. 193−200.

[17] Kamps J, Marx M, Mokken RJ. Using WordNet to measure semantic orientation of adjectives. In: Calzolari N, et al., eds. Proc. Of the LREC. 2004. 1115−1118.

[18] Mihalcea R, Banea C, Wiebe J. Learning multilingual subjective language via cross-lingual projections. In: Carroll J, ed. Proc. Of the Association for Computational Linguistics (ACL). Morristown: ACL, 2007. 976−983.

[19] Wiebe J, Mihalcea R. Word sense and subjectivity. In: Dale R, Paris C, eds. Proc. of the Conf. on Computational Linguistics/Association for Computational Linguistics (COLING/ACL). Morristown: ACL, 2006. 1065−1072.

[20] Takamura H, Inui T, Okumura M. Extracting semantic orientation of words using spin model. In: Knight K, ed. Proc. of the Association for Computational Linguistics (ACL). Morristown: ACL, 2005. 133−140.

[21] Yi J, Nasukawa T, Bunescu R. Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. In: Wu XD, Tuzhilin A, eds. Proc. of the IEEE Int’l Conf. on Data Mining (ICDM). 2003. 427−434.

[22] Hu M, Liu B. Mining opinion features in customer reviews. In: Hendler JA, ed. Proc. of the AAAI 2004. Menlo Park: AAAI Press, 2004. 755−760.

[23] Ni MS, Lin HF. Mining product reviews based on association rule and polar analysis. In: Zhu QM, et al., eds. Proc. of the NCIRCS 2007. 2007. 628−634 (in Chinese with English abstract).

(倪茂树,林鸿飞.基于关联规则和极性分析的商品评论挖掘.见:第3 届全国信息检索与内容安全学术会议论文集.2007. 628−634.)

[24] Liu HY, Zhao YY, Qin B, Liu T. Target extraction and sentiment classification. Journal of Chinese Information Processing, 2010, 24(1):84−88 (in Chinese with English abstract).

(刘鸿宇,赵妍妍,秦兵,刘挺.评价对象抽取及其倾向性分析.中文信息学报,2010,24(1):84−88)

[25] Popescu AM, Etzioni O. Extracting product features and opinions from reviews. In: Mooney RJ, ed. Proc. of the HLT/EMNLP 2005. Morristown: ACL, 2005. 339−346.

[26] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research, 2003,3:993−1022. [doi: 10.1162/ jmlr.2003.3.4-5.993]

[27] Blei DM, Ng AY, Jordan MI. Correlated topic models. In: Schölkopf B, ed. Advances in NIPS. Hyatt Regency: MIT Press, 2006. 147−154.

[28] Titov I, McDonald R. Modeling online reviews with multi-grain topic models. In: Huai JP, Chen R, eds. Proc. of the WWW 2008. New York: ACM Press, 2008. 111−120.

[29] Kim SM, Hovy E. Extracting opinions, opinion holders, and topics expressed in online news media text. In: Dale R, Paris C, eds. Proc. of the ACL Workshop on Sentiment and Subjectivity in Text. 2006. 1−8.

[30] Stoyanov V, Cardie C. Topic identification for fine-grained opinion analysis. In: McKeown K, ed. Proc. of the Conf. on Computational Linguistics. Morristown: ACL, 2008. 817−824.

[31] Kim SM, Hovy E. Determining the sentiment of opinions. In: Nirenburg S, ed. Proc. of the Coling 2004. Morristown: ACL, 2004. 1367−1373.

[32] Choi Y, Cardie C, Riloff E. Identifying sources of opinions with conditional random fields and extraction patterns. In: Mooney RJ, ed. Proc. of the HLT/EMNLP 2005. Morristown: ACL, 2005. 355−362.

[33] Bethard S, Yu H, Thornton A. Automatic extraction of opinion propositions and their holders. In: Proc. of the AAAI Spring Symp. on Exploring Attitude and Affect in Text. 2004. 22−24.

[34] Wiebe J, Wilson T, Bell M. Identifying collocations for recognizing opinions. In: Webber BL, ed. Proc. of the ACL/EACL Workshop on Collocation: Computational Extraction, Analysis, and Exploitation. Morristown: ACL, 2001. 24−31.

[35] Wiebe J, Wilson T. Learning to disambiguate potentially subjective expressions. In: Roth D, van den Bosch A, eds. Proc. of the Conf. on Natural Language Learning (CoNLL). Morristown: ACL, 2002. 112−118.

[36] Wilson T, Wiebe J, Hwa R. Just how mad are you? Finding strong and weak opinion clauses. In: Hendler JA, ed. Proc. of the AAAI 2004. Menlo Park: AAAI Press, 2004. 761−769.

[37] Wilson T, Wiebe J, Hwa R. Recognizing strong and weak opinion clauses. Computational Intelligence, 2006,22(2):73−99.

[38] Whitelaw C, Garg N, Argamon S. Using appraisal groups for sentiment analysis. In: Fuhr N, ed. Proc. of the ACM SIGIR Conf. on Information and Knowledge Management (CIKM). New York: ACM Press, 2005. 625−631.

[39] Moilanen K, Pulman S. Sentiment composition. In: Mitkov R, ed. Proc. of the Recent Advances in Natural Language Processing Int’l Conf. (RANLP 2007). 2007. 378−382.

[40] Choi Y, Cardie C. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In: Lapata M, Ng HT, eds. Proc. of the EMNLP 2008. Morristown: ACL, 2008. 793−801.

[41] Kobayashi N, Inui K, Matsumoto Y. Collecting evaluative expressions for opinion extraction. In: Nagao M, ed. Proc. of the Int’l Joint Conf. on Natural Language Processing (IJCNLP). Morristown: ACL, 2004. 584−589.

[42] Bloom K, Garg N, Argamon S. Extracting appraisal expressions. In: Sidner C, ed. Proc. of the HLT-NAACL 2007. Morristown: ACL, 2007. 308−315.

[43] Yao TF, Nie QY, Li JC, Li LL, Lou DC, Chen K, Fu Y. An opinion mining system for Chinese automobile reviews. In: Cao YQ, et al., eds. Proc. of the Frontiers of Chinese Information Processing. Beijing: Tsinghua University Press, 2006. 260−281 (in Chinese with English abstract).

(姚天昉,聂青阳,李建超,李林琳,娄德成,陈珂,付宇.一个用于汉语汽车评论的意见挖掘系统.中文信息处理前沿进展——中国中文信息学会成立二十五周年学术年会论文集.2006.260−281.)

[44] Xu LH, Lin HF, Zhao J. Construction and analysis of emotional corpus. Journal of Chinese Information Processing, 2008,22(1): 116−122 (in Chinese with English abstract).

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

[45] Riloff E, Wiebe J, Phillips W. Exploiting subjectivity classification to improve information extraction. In: Yanco H, ed. Proc. Of the AAAI 2005. Menlo Park: AAAI Press, 2005. 1106−1111.

[46] Hatzivassiloglou V, Wiebe J. Effects of adjective orientation and gradability on sentence subjectivity. In: Kay M, ed. Proc. of the Int’l Conf. on Computational Linguistics (COLING). Morristown: ACL, 2000. 299−305.

[47] Yu H, Hatzivassiloglou V. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: Collins M, Steedman M, eds. Proc. of the EMNLP 2003. Morristown: ACL, 2003. 129−136.

[48] Yao TF, Peng SW. A study of the classification approach for Chinese subjective and objective texts. In: Zhu QM, et al., eds. Proc. of the NCIRCS 2007. 2007. 117−123 (in Chinese with English abstract).

(姚天昉,彭思崴.汉语主客观文本分类方法的研究.见:第3 届全国信息检索与内容安全学术会议论文集.2007.117−123)

[49] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Scott D, ed. Proc. of the ACL 2004. Morristown: ACL, 2004. 271−278.

[50] Hu MQ, Liu B. Mining and summarizing customer reviews. In: Kohavi R, ed. Proc. of the KDD 2004. New York: ACM Press, 2004. 168−177.

[51] Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Isabelle P, ed. Proc. of the ACL 2002. Morristown: ACL, 2002. 417−424.

[52] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. In: Isabelle P, ed. Proc. of the EMNLP 2002. Morristown: ACL, 2002. 79−86.

[53] Cui H, Mittal VO, Datar M. Comparative experiments on sentiment classification for online product reviews. In: Gil Y, Mooney RJ, eds. Proc. of the AAAI 2006. Menlo Park: AAAI Press, 2006. 1265−1270.

[54] Kim SM, Hovy E. Automatic identification of pro and con reasons in online reviews. In: Dale R, Paris C, eds. Proc. of the COLING/ACL 2006. Morristown: ACL, 2006. 483−490.

[55] Zhao J, Liu K, Wang G. Adding redundant features for CRFs-based sentence sentiment classification. In: Lapata M, Ng HT, eds. Proc. of the Conf. on Empirical Methods in Natural Language Processing (EMNLP 2008). Morristown: ACL, 2008. 117−126.

[56] Pang B, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: Knight K, ed. Proc. of the Association for Computational Linguistics (ACL). Morristown: ACL, 2005. 115−124.

[57] Goldberg AB, Zhu X. Seeing stars when there aren’t many stars: Graph-Based semi-supervised learning for sentiment categorization. In: Bilmes J, et al., eds. Proc. of the HLT-NAACL 2006 Workshop on Textgraphs: Graph-Based Algorithms for Natural Language Processing. Morristown: ACL, 2006. 45−52.

[58] Lin WH, Wilson T, Wiebe J. Which side are you on? Identifying perspectives at the document and sentence levels. In: Bilmes J, et al., eds. Proc. of the Conf. on Natural Language Learning (CoNLL). Morristown: ACL, 2006. 109−116.

[59] Kim SM, Hovy E. Crystal: Analyzing predictive opinions on the Web. In: Eisner J, ed. Proc. of the Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). Morristown: ACL, 2007. 1056−1064.

[60] Hurst M, Nigam K. Retrieving topical sentiments from online document collections. In: Proc. of the Document Recognition and Retrieval XI. 2004. 27−34.

[61] Ounis I, Rijke MD, Macdonald C, Mishne G, Soboroff I. Overview of the TREC-2006 Blog track. In: Proc. of the TREC. 2006.

[62] Zhang W, Yu C, Meng WY. Opinion retrieval from Blogs. In: Laender A, et al., eds. Proc. of the CIKM. New York: ACM Press, 2007. 831−840.

[63] Zhang W, Yu C. UIC at TREC 2007 Blog track. In: Proc. of the 16th TREC. 2007.

[64] Zhang M, Ye XY. A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In: Chua TS, Leong MK, eds. Proc. of the ACM Special Interest Group on Information Retrieval (SIGIR). New York: ACM Press, 2008. 411−418.

[65] Liu B, Hu MQ, Cheng J. Opinion observer: Analyzing and comparing opinions on the Web. In: Ellis A, ed. Proc. of the WWW 2005. New York: ACM Press, 2005. 342−351.

[66] Carenini G, Ng R, Pauls A. Multi-Document summarization of evaluative text. In: McCarthy D, Wintner S, eds. Proc. of the European Chapter of the Association for Computational Linguistics (EACL). Morristown: ACL, 2006. 305−312.

[67] Qin B, Zhao YY, Gao LL, Liu T. Recommended or not? Give advice on online products. In: Ma J, et al., eds. Proc. of the 5th Int’l Conf. on Fuzzy Systems and Knowledge Discovery. IEEE Computer Society Press, 2008. 208−212.

[68] Titov I, McDonald R. A joint model of text and aspect ratings for sentiment summarization. In: McKeown K, ed. Proc. of the ACL 2008. Morristown: ACL, 2008. 308−316.

[69] Branavan S, Chen H, Eisenstein J. Learning document-level semantic properties from free-text annotations. In: McKeown K, ed. Proc. of the ACL 08: HLT. Morristown: ACL, 2008. 263−271.

[70] Ku LW, Liang YT, Chen HH. Opinion extraction, summarization and tracking in news and Blog corpora. In: Gil Y, Mooney RJ, eds. Proc. of the AAAI 2006 Spring Symp. on Computational Approaches to Analyzing Weblogs. Menlo Park: AAAI Press, 2006.

[71] Ma Chunling, Osherenko A, Prendinger H,et a1．A chat system based on emotion estimation from text and embodied conversational messengers[c]／Proc of the 4th Int Conf on Active Media Technology．Piscataway,NJ：IEEE,2005：546-548

[72] Aman S, Szpakowiez S．Identifying expressions of emotion in text[c]／Proc of the 10th Int Conf on Text,Speech and Dialogue．Berlin：Springer,2007：196—205

[73] Yang Mi. Peng Baolin, Chen Zheng.et a1．A topic model for building fine-grained domain—specific emotion lexicon[C] ／／Proc of the 52nd Annual Meeting of the Association for Computational Linguistics．Stroudsburg.PA：ACI.,2014：421-426

[74] Xu Jun, Xu Ruifeng, Zheng Yanzhen,et a1．Chinese emotion lexicon developing via multi—lingual lexical resources integration[c]／]Proc of the 14th Int Conf on Computational Linguistics and Intelligent Text Processing． Berlin：Springer,2013：174-182

[75]Song Kaisong, Feng Shi,Gao Wei,et a1．Build emotion lexicon from microblogs by combining effects of seed words and emoticons in a heterogeneous graph[c]／Proc of the 26th ACM Conf on Hypertext and Social Media．New York：ACM. 2015：283—292

[76]Wu Fangzhao,Huang Yongfeng,Song Yangqiu,et a1．Towards building a high-quality microblog—specific Chinese sentiment lexicon[J]．Decision Support Systems,2016,87：39-49

[77]Staiano J,Guerini M．DepecheMood：A lexicon for emotion analysis from crowd—annotated news[c3／Proc of the 52nd Annual Meeting of the Association for Computational Linguistics．Stroudsburg,PA：ACL,2014：427—433

[78]Strapparava C.Mihalcea R．Learning to identify emotions in text[c3／Proc of the 23rd ACM Symp on Applied Computing．New York：ACM,2008：1556一1560

[79]Neviarouskaya A,Prendinger H,Ishizuka M． Affect analysis model：Novel rule—based approach to affect sensing from text[J]．Natural Language Engineering,2011,17(1)：95-135

[80]Wen Shiyang,Wan Xiaojun． Emotion classification in mieroblog texts using class sequential rules[C]／Proc of the 28th AAAI Conf on Artificial Intelligence．Menlo Park,CA：AAAI,2014：187-193

[81]Lee S Y M,Chen Ying,Huang Churen．A text—driven rule—based system for emotion cause detection[c]／／Proc of the 10th NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text．Stroudsburg,PA：ACL,2010：45—53

[82]Li Weiyuan,Xu Hua．Text—based emotion classification using emotion cause extraction[J]．Expert Systems with Applications,2014,41(4)：1742—1749

[83]Gao Kai,Xu Hua,Wang Jiushuo．A rule—based approach to emotion cause detection for Chinese micro-blogs[J]．Expert Systems with Applications,2015,42(9)：4517-4528

[84]Alm C,Roth D,Sproat R．Emotions from text：Machine learning for text—based emotion prediction[c]／Proc of the 2nd Conf on Human Language Technology and on Empirical Methods in Natural Language Processing．New York：ACM,2005：579-586

[85]Cheng Xueqi,Yan Xiaohui,Lan Yanyan,et a1．BTM：Topic modeling over short texts．IEEE Trans on Knowledge and Data Engineering,2014,26(12)：2928—2941

[86]Read J．Using Emoticons to reduce dependency in machine learning techniques for sentiment classification[c]／Proc of the 43rd ACL Student Research Workshop．Stroudsburg,PA：ACL,2005：43-48

[87]Zhao Jichang,Dong Li,Wu Junjie,et a1．Moodlens：An emoticon-based sentiment analysis system for Chinese tweets [c]／／Proc of the 18th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining．New York：ACM,2012：1528—1 531

[88]Ouyang Xi,Zhou Pan,Li Chenghua,et a1．Sentiment analysis using convolutional neural network[C]／Proc of the 1 5th Int Conf on Computer and Information Technology；the 14th Int Conf on Ubiquitous Computing and Communications；the 13th Int Conf on Dependable,Autonomic and Secure Computing；the 13th Int Conf on Pervasive Intelligence and Computing．Piscataway,NJ：IEEE,2015：2359-2364

[89] Santos C,Gatti M．Deep Convolutional neural networks for sentiment analysis of short texts It]／／Proc of the 25th Int Conf on Computational Linguistics．New York：ACM,2014：69-78

[90] Irsoy O,Cardie C．Deep recursive neural networks for compositionality in language[c]／Proc of the 28th Conf on Neural Information Processing Systems．Cambridge,MA：MIT Press.2014：2096-2104

[91] Kang Xin,Ren Fuji,Wu Yunong．Bottom up：Exploring word emotions for Chinese sentence chief sentiment classification[C]／／Proc of the 6th Natural Language Processing and Knowledge Engineering．Piscataway,NJ：IEEE,2010：1-5

[92]Rao Yanghui,Xie Haoran,Li Jun,et a1．Social emotion classification of short text via topic-level maximum entropy model[J]．Information and Management,2016,53(8)：978-986

[93]Huang Shu,Peng Wei,Li Jingxuan,et a1．Sentiment and topic analysis on social media：A multi—task multi—label classification approach[c]／Proc of the 5th Annual ACM Web Science Conf．New York：ACM,2013：172—181

[94]Zhang Xiao,Li Wenzhong,Lu Sanglu．Emotion detection in online social network based on multi—label learning[c]／Proc of the 22nd Int Conf on Database Systems for Advanced Applications．Berlin：Springer,2017：659—674

[95]Sun Xiao,Li Chengcheng,Ye Jiaqi．Chinese mieroblogging emotion classification based on support vector machine[c]／／ Proc of the 5th Int Conf on Computing,Communication and Networking Technologies．Piscataway,NJ：IEEE,2014：1-5

[96]Sintsova V,Musat C,Pu Pearl．Semi—supervised method for multi—category emotion recognition in tweets[c]／Proc of the 1 7th Int Conf on Data Mining Workshop．Piscataway.NJ：IEEE,2014：393—402

[97] Purver M, Battersby S． Experimenting with distant supervision for emotion classification[c]／Proc of the 13th Conf of the European Chapter of the Association for Computational Linguistics．New York：ACM,2012：482—491

[98]Suttles J, Ide N．Distant supervision for emotion classification with discrete binary values[c]／／Proc of the 14th Int Conf on Intelligent Text Processing and Computational Linguistics．Berlin：Springer,2013：121—136

[99]Jiang Fei,Liu Yiqun,Luan Huanbo,et a1．Microblog sentiment analysis with emoticon space model[J]．Journal of Computer Science and Technology,2015．30(5)：1120一1129

[100]Li Shoushan,Xu Jian,Zhang Dong,et a1．Two—view label propagation to semi supervised reader emotion classification Ec]／Proc of the 26th Int Conf on Computational Linguistics．New York：ACM,2016：2647-2655

[101]Ghazi D,Inkpen D,Szpakowicz S．Hierarchical versus flat classification of emotions in text[c3／[Proc of the 9th NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text．Stroudsburg,PA：ACI.,2010：140—146

[102]Esmin A A A,De Oliveira R L．Matwin S．Hierarchical classification approach to emotion recognition in Twitter[c] ／Proc of the 1 I th Int Conf on Machine Learning and Applications．Piscataway,NJ：IEEE．2012：381—385

[103]Xu Hua,Yang Weiwei,Wang Jiushuo．Hierarchical emotion classification and emotion component analysis on Chinese micro—blog posts[J]． Expert Systems with Applications,201 5,42(22)：8745—8752

[104]Zhang Fan,Xu Hua．Wang Jiushuo,et a1．Grasp the implicit features：Hierarchical emotion classification based on topic model and SVM[c]／／Proc of the 29th Int Joint Conf on Neural Networks．Piscataway,NJ：IEEE,20 1 6：3592-3599

[105]Keshtkar F．Inkpen D．A hierarchical approach to mood classification in blogs[J]．Natural Language Engineering,2011．18(18)1 6l一81

[106]Liu Baoqin,Niu Yun．Multi—hierarchy emotion analysis of Chinese microblog[J]． Computer Technology and Development,2015,25(11)：23—26(in Chinese)

(刘宝芹,牛耘．多层次中文微博情绪分析[J]．计算机技术与发展,2015,25(11)：23-26)

[107]Ouyang Chunping,Yang Xiaohua,Lei Longyan．et a1． Multi——strategy approach for fine——grained sentiment analysis of Chinese microblog[J]．Acta Scientiarum Naturalium Universitatis Pekinensis,2014．50(1)：67-72(in Chinese)

(欧纯萍．阳小华．雷龙艳,等．多策略中文微博细粒度情绪分析研究[J]．北京大学学报：自然科学版,2014,50(1)：67-72)

[108]Yang Jun,jiang Lan.Wang Chongjun,et a1．Multi—label emotion classification for tweets in weibo：Method and application[c]／Proc of the 26th Int Conf on Tools with Artificial Intelligence．Piscataway,NJ：IEEE,2014：424—428

[109]Buitinck I.,Van Amerongen J,Tan E,et a1．Multi emotion detection in user—generated reviews[C]／Proc of the 37th European Conf on Information Retrieval．Berlin：Springer,2015：43-48

[110]Liu Shuhua,Chen Jiunhung．A multi—label classification based approach for sentiment classification[J]．Expert Systems with Applications．2015．42(3)：1083—1093

[111]Wang Yaqi,Feng Shi,Wang Daling.et a1．Multi—label Chinese microblog emotion classification via convolutionaI neural network[c]／／Proc of the 18th Asia—Pacific Web Conf．Berlin：Springer．2016：567—580

[112]Li Jun．Rao Yanghui,Jin Fengmei,et a1．Multi—label maximum entropy model for social emotion classification over short text[J]．Neurocomputing．2016．210：247-256

[113]Zhou Deyu,Zhang Xuan,Zhou Yin,et a1．Emotion distribution learning from texts[c]／Proc of the 21st Conf on Empirical Methods in Natural Language Processing．Stroudsburg,PA：ACI,,2016：638—647

[114]Xu Ruifeng,Wang Zhaoyu,Xu Jun,et a1．An iterative emotion classification approach for microblogs[c]／Proc of the 16th Int Conf on Intelligent Text Processing and Computational Linguistics．Berlin：Springer,2015：104一ll3

[115] Ounis I, Rijke MD, Macdonald C. Overview of the TREC-2006 Blog track. In: Proc. of the 15th Text Retrieval Conf. (TREC). 2006.

[116] Zhao J, Xu HB, Huang XJ, Tan SB, Liu K, Zhang Q. Overview of Chinese opinion analysis evaluation 2008. 2008 (in Chinese with English abstract). <http://nlpr-web.ia.ac.cn/2008papers/gnhy/nh10.pdf>

(赵军,许洪波,黄萱菁,谭松波,刘康,张奇.中文倾向性分析评测技术报告.2008)

[117] Jindal N, Liu B. Identifying comparative sentences in text documents. In: Efthimiadis EN, ed. Proc. of the ACM Special Interest Group on Information Retrieval (SIGIR). New York: ACM Press, 2006. 244−251.

[118] Wiebe J, Wilson T, Cardie C. Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, 2005,39(2-3):164−210.

[119] Ku LW, Lo YS, Chen HH. Using polarity scores of words for sentence-level opinion extraction. In: Proc. of the NTCIR-6 Workshop Meeting. 2007. 316−322.

[120]Janssens O,Slembrouck M,Verstockt S．et a1．Real—time emotion classification of tweets[c]／／Proc of the 5th IEEE／ACM Int Conf on Advances in Social Networks Analysis and Mining．New York：ACM,2013：1430—1431