
文本情感分析综述*

赵妍妍⁺, 秦兵, 刘挺

(哈尔滨工业大学 计算机科学与技术学院 信息检索研究中心, 黑龙江 哈尔滨 150001)

A Survey of Sentiment Analysis *

ZHAO Yan-Yan⁺, QIN Bing, LIU Ting

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: Phn: +86-451-86413683 ext 800, E-mail: zyy@ir.hit.edu.cn

Abstract: Sentiment analysis is a novel research topic with the quick development of online reviews, which has drawn interesting attention due to its research value and extensive applications. This paper surveys the state-of-the-art research on sentiment analysis. First, three important tasks of sentiment analysis are summarized and analyzed in detail, including sentiment extraction, sentiment classification, sentiment retrieval and summarization; then the evaluation and corpus for sentiment analysis are introduced; finally the applications of sentiment analysis are concluded. This paper aims to take a deep insight into the mainstream methods and recent progress in this field, making detailed comparison and analysis. It is expected to be helpful to the future research.

Key words: sentiment analysis; sentiment extraction; sentiment classification; sentiment retrieval and summarization; evaluation; corpus

摘 要: 文本情感分析是随着网络评论的海量增长而迅速兴起的一个新兴研究课题,其研究价值和应用价值受到人们越来越多的重视.本文对文本情感分析的研究现状与进展进行了总结.首先将文本情感分析归纳为三项主要任务,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;进而介绍了文本情感分析的国内外评测和资源建设情况;最后介绍了文本情感分析的应用.文本重在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,以期对后续研究有所助益.

关键词: 文本情感分析;情感信息抽取;情感信息分类;情感信息的检索与归纳;评测;资源建设

中图法分类号: TP391 文献标识码: A

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

* Supported by the National Natural Science Foundation of China under Grant Nos. 60803093, 60975055 (国家自然科学基金) and the “863” National High-Tech Research and Development of China via grant 2008AA01Z144(863 计划探索类专题项目)

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(本文中提及的情感分析,都是指文本情感分析).

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主观性文本进行分析、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^[1],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.

情感分析涉及多项非常有挑战性的研究任务.本文综合已有的研究成果,将情感分析归纳为三项层层递进的研究任务,即情感信息的抽取、情感信息的分类以及情感信息的检索与归纳,如图 1 所示.

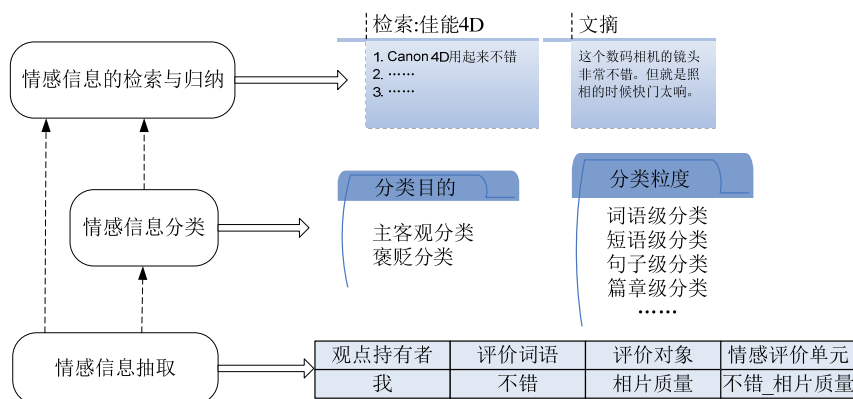


Fig.1 The framework of sentiment analysis

图 1 情感分析的研究框架

其中,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.

情感分析是一个新兴的研究课题,具有很大的研究价值和应用价值^[3-5].鉴于此,该研究课题受到国内外越来越多的研究机构的重视.本文在接下来的部分首先分别详细阐述情感分析的三个主要研究任务,重点针对各任务的主流方法和前沿进展进行对比分析;接着介绍国内外主流的评测会议以及现有的资源建设情况;然后,本文介绍情感分析几个重要的应用点;最后,展望情感分析技术的发展趋势.

1 情感信息抽取

情感信息抽取旨在抽取情感文本中的有价值的情感信息.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.

1.1 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.很显然,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^[10].

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“*lovely and beautiful*”)极性相同,然而“*but*”连接的形容词(如:“*lovely but unnatural*”)极性相反.基于这种现象, Hatzivassiloglou 和 McKeown^[11]从大语料库华尔街日报(Wall Street Journal)中发掘出大量的形容词性的评价词语. Wiebe 等人^[11]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,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^[12]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^[13]提出了点互信息(Point Mutual Information)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.很自然的,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^[14-16].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^[17-20].此外,一些学者^[21]沿用了 Turney 等人的点互信息的方法^[13],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,并非所有的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^[22],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[23]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“*他跑的好快啊!*”).

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^[10,24].具体的,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(Propagation Algorithm)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^[24],继而使用 Spin 模型对图中的点迭代的进行概率计算,得出每个词语的极性.还有一些学者尝试使用多种图模型^[10],如:最小切分模型(Mincuts)、随机最小切分模型(Randomized Mincuts)及标签迭代模型(Label Propagation)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,它可以灵活的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.

1.2 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,命名实体识别和句法分析等.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^[25]使用三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.还有学者^[26-27]使用关联规则挖掘的方法或是基于句法分析的结果^[28]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^[26].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^[29]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:“*scanner has*”)之间的关联度来获取真正的评价对象.实验证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(Topic Model)^[30-31]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用于评价对象的识别.有学者^[32]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,但是遗憾的是,还没有实验将这种方法和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^[33-34].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*”.

1.3 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*”.很自然的,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^[35].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^[33].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^[36],并使用CRF (Conditional Random Fields)模型融合各种特征来完成观点持有者的抽取.相似的,Kim^[15]将所有名词短语都视为候选观点持有者,使用ME (Maximum Entropy)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,观点持有者一般是和观点同时出现的,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^[37]在抽取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.

1.4 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*”在以下三个句子中的使用.

- Sen1: 凯越的油耗真高.
- Sen2: 捷达的性价比相当的高.
- Sen3: 这辆车有 1m 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.如在 Sen1 中“*高*”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.下面将具体的介绍各种形式的组合评价单元.

1.4.1 主观表达式的抽取

主观表达式(Subjective Clues)是指表示情感文本单元主观性的词语或词组. 1.1 节的评价词语是主观表达

式的一部分.此外,某些词语的组合(如:“*village idiot*”或“*get out of here*”)也能很明显的标识文本的主观性,虽然它们中的任何一个词语单独可能都并非评价词语.如何获取这些有意义的词组是主观表达式抽取的重点.

Wiebe 和 Wilson 是这项任务的引领者^[38].近几年来,他们挖掘大量的主观表达式形成主观表达式库,并基于此完成文本的主客观分类和褒贬分类.具体的,他们首先从语料中抽取所有的 n 元词语/词组($1 \leq n \leq 4$)作为候选主观表达式;继而通过对比训练语料中的标准的主观表达式,为每个候选主观表达式计算出可能成为主观表达式的概率;最后通过对概率值的分析,获得这些主观表达式.Wiebe 和 Wilson^[39]在随后的工作中又引入了“主观表达式密度”协助判断主观表达式.2004 年,Wiebe 和 Wilson 将他们前期的工作进行了总结^[40],从不同的语料中扩充了大量的主观表达式,主要包括手工收集的一部分主观表达式以及自动从标注/未标注语料中学习而来的一部分主观表达式.此外,他们首次利用句法分析的结果发掘了句法主观表达式^[41].随后,Wiebe 和 Wilson 采用多种特征及机器学习方法对他们获取的大量的主观表达式的情感程度(strong 或 weak)进行了识别.

1.4.2 评价短语的抽取

评价短语表现为一组连续出现的词组,但不同于主观表达式,该词组往往是由程度副词和评价词语组合而成,如:“*very good*”等.因此,这种组合评价单元不仅顾及了主观表达式的情感极性,还考察了其修饰成分.这些修饰成分或加强或减弱或置反了主观表达式的情感极性,使得评价短语成为一种情感色彩丰富的组合评价单元.

有学者采用基于一些情感词典的方法识别这种评价短语.如 Whitelaw^[42]结合 WordNet 使用半自动的方法构建了形容词性的评价词词典以及修饰词词典.对于一个含有评价词语的情感文本,该方法首先查看评价词前面的词语,如果属于修饰词词典,获取这个词组作为评价短语.根据两个词典中的属性值计算出情感极性.这种方法由于基于较为细致的词典,因此准确率较高.然而,由于词典中词语有限而限制了召回率.还有学者使用依存句法结构(如 ADV、ATT 以及 DE 结构),在句法树上获取评价短语^[27].这种方法巧妙的利用了评价短语中所含词语之间的句法修饰关系,但是较为依赖句法分析的结果.

评价短语考察的是连续出现的词组,然而,有些表示修饰关系的词语并非总是和评价词语连续出现.如在情感句“[I did [not] have any [doubt] about it.]”中,修饰词“*not*”和评价词“*doubt*”并非连续出现,但它们共同决定了情感句的最终极性. Moilanen 等人^[43]和 Choi 等人^[44]将其定义为“组合语义单元”(Compositional Semantics),具体表现为一组非连续的词语,通过相互作用来表达出某种情感极性.“组合语义单元”可以看作一种更复杂的评价短语,多使用人工总结或半自动生成的模板来识别.

1.4.3 评价搭配的抽取

评价搭配是指评价词语及其所修饰的评价对象二者的搭配,表现为二元对<评价对象,评价词语>,如情感句“*凯越的油耗很高*”中的“*油耗-高*”.前面所介绍的“主观表达式”和“评价短语”主要是考察含有情感极性的一些词和短语,然而情感句中出现的某些“主观表达式”和“评价短语”并非真正的表现出情感极性.如情感句 s_1 “*车跑的好快啊*”中的词语“*好*”并不存在情感极性,需要过滤掉.此外,还有一些“主观表达式”和“评价短语”存在一定的歧义,其极性需要根据上下文而确定.“评价搭配”则可以很好的解决上述两点问题.

针对评价搭配的抽取任务,大部分学者采用了基于模板的获取方法. Kobayashi 等人^[45]考察评价对象和评价词语之间的修饰关系,并用 8 个共现模板来描述.然而,由于模板太过简单且修饰关系仅仅停留在词表面,该方法产生了大量的噪声.为了深入挖掘评价对象和评价词语之间的修饰关系,一部分学者尝试使用句法关系模板. Bloom 等人^[46]利用 Stanford Parser 手工构建了 31 条句法规则.此外, Popescu 等人^[29]利用 MINIPAR Parser 手工构建了 10 条依存句法抽取模板来获取评价搭配.同时,国内的姚天昉等人^[47]基于依存句法分析总结出“上行路径”和“下行路径”的匹配规则;后续总结出 SBV 极性传递规则,用于评价搭配的识别.可以看出,他们的工作融入了更多对评价对象和评价词语之间深层关系的挖掘.然而,由于匹配规则或模板的制定参与了过多的人工,覆盖率较低.因此,在未来工作中我们应该侧重于研究自动生成评价对象和评价词语之间的匹配规则的策略.

2 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^[48].

2.1 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^[49],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部分情感信息抽取的结果)来完成主客观信息分类^[14,50].然而,我们发现许多客观句中也可能包含评价词语,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,如1.4节中提到的“主观表达式”、“评价短语”和“评价搭配”等组合信息.此外,还有学者^[12]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> up the wall”).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.具体的,Hatzivassiloglou^[51]使用了词语作为特征,并采用了NB(Naïve Bayes)分类器完成篇章级情感文本的主客观分类.Yao^[52]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,数字角度等等.Pang^[53]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,根本的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.

2.2 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部分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,研究者将主观文本的极性分为褒义和贬义两类(Thumbs up? Thumbs down?).纵观目前的研究工作,和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.相似的,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,选取大量的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[14,51,54-55]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^[56]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(Support Vector Machine)三种分类模型,发现unigram特征效果最好.然而,Cui^[57]通过实验证明,当训练语料较少的时候,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^[58]除了考察传统的n-gram模型外,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^[59]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^[60]将褒贬等级分为三类,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^[61]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四个等级的分类.

2.3 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面.Lin^[8]主要使用三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^[9]主要对美国大选时涌现出来的大量的评论文章进行分类汇总,来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,其中作者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,还需要发掘一部分和“观点”相关的知识.

3 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的,且包含情感信息的文档;情感信息归纳则针对大量的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情感信息检索结果的一个深加工。

3.1 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^[62]。2006 年 TREC 首次引入了博客检索任务(Blog Track)^[63],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,该任务主要包含三个步骤:

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,研究者们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^[64]。也有研究者在查询扩展中融入了情感知识^[65],用以提高检索的性能。

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 章和第 2 章所述。

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^[64]。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^[66]。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。

通过以上三个步骤的分析,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们近期和未来所需关注的重点。

3.2 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的用户评论,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。一是基于产品属性的情感文摘,一是基于情感标签的情感文摘。

3.2.1 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^[6,29,54,67],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三个重要的步骤:

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^[25-29]。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,将能更好的将产品的情感文摘进行归纳。

2. 抽取描述产品属性的情感句,即针对产品的每一种属性,收集和其相关的所有的情感句^[50-53]。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^[54-59].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,如图 2 所示.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^[6];此外,还有情感文摘为每种产品计算出推荐等级^[68]等.

通过以上三个步骤的分析,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:上面三个步骤)指标的提高,才能增强这类情感文摘的质量.

佳能400D	支持	反对
焦距	佳能400D焦距不错	佳能400D焦距很难进行调整
机身重量	佳能的机身重量还不错	佳能的机身重量太重了

Fig.2 Sentiment summarization based on product features

图 2 基于产品属性的情感文摘

3.2.2 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,为了更便捷的为用户提供参考意见,基于情感标签的情感文摘应运而生,如图 3 所示.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*¹.用户在这类网站上发表对某一产品的看法的时候,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“*small size*”(“Pros”),“*short battery life*”(“Cons”)等.这些标签能够很好的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(国内还没有发现),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:一来用户写评论的时候非常随意,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的时候,有人使用“*good diet*”,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^[69,70]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.

Tag: 焦距很一般, 机身重量还可以, 成像非常赞,
Review: 昨天刚刚买了佳能400D。赶紧上来跟大家说一说。昨天拍了一天照, 感觉这个相机成像非常好, 像传说中的一样☺但是, 焦距不容易控制, 不是很满意。但是机身重量还好, 不是很重, 起码我用起来还是不错的。.....

Fig.3 Sentiment summarization based on sentiment tag

图 3 基于情感标签的情感文摘

Titov^[69]的工作主要是为一组相关产品评论集的产品属性打标签.具体的,该方法将产品属性当做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(Topic Model)^[30-31]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

¹ <http://www.epinions.com/>

档集中所有的产品属性(如“room”,“service”以及“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,因此需要自动的发掘产品属性及其情感标签. Branavan^[70]较为巧妙的解决了这个问题,他将标签定义为“评价搭配”的形式(如: “small size”, “long battery life”等),并为单篇的网络产品评论打上合适的标签.具体的,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: “small size”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:“small size”和“smart size”);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,优点在于标签更加的清晰,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.

3.2.3 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^[71]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(Concept Words)和情感词(Sentiment Words).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,而且和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.

4 情感分析的评测与资源建设

4.1 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(Text Retrieval Evaluation Conference)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(Blog Track).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^[72].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,贬义和混合三类(positive, negative, mixed). Blog Track 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(NII Test Collection for IR Systems)的情感分析评测(MOAT: multilingual opinion analysis task)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和篇章的主题相关,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,可以看出他们的目标是进行多语种、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,从四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(Chinese Opinion Analysis Evaluation)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^[73],可分为三个方面,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.

4.2 情感分析的资源建设

4.2.1 情感分析的语料

除了 4.1 节中三个国际/国内评测所提供的语料外,不少研究单位和个人也提供了一定规模的语料.

1. 康奈尔大学(Cornell)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 1,000 篇;另外还有标注了褒贬极性的句子各 5,331 句,标注了主客观标签的句子各 5,000 句.目前影评库被广泛应用于各种粒度的,如词语、句子和篇章级情感分析研究中.

2. 伊利诺伊大学芝加哥分校(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 下载的五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 & 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^[74]方面的语料.

3. Janyce Wiebe 等人所开发的 MPQA(Multiple-Perspective QA)库:包含 535 篇不同视角的新闻评论,它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^[75]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(MIT)的 Barzilay 等人构建的多角度餐馆评论语料:共 4,488 篇,每篇语料分别按照五个角度(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. 国内的中科院计算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 10,000 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.

4.2.2 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(General Inquirer)评价词词典(英文,<http://www.wjh.harvard.edu/~inquirer/>).该词典收集了 1,914 个褒义词和 2,293 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 2,812 个褒义词与 8,276 个贬义词^[76].

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 8,221 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 9,193 个中文评价词语/短语, 9,142 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.

5 情感分析的应用

随着互联网上评论文本的爆炸式增长,迫切需要计算机帮助用户加工整理这些情感信息,这使得情感分析研究有着重要的应用.下面,本文就情感分析的应用现状以及应用前景进行概括介绍.

✧ 用户评论分析与决策

这是目前情感分析技术使用最频繁的一个应用点.人们在购买某一产品之前,往往倾向于网上查询该产品的相关评论,并通过与其他产品的对比来做最终的决策.由于用户没有足够的时间和精力浏览全部的评论信息,导致最终的决策带有风险性.情感分析技术则可以很好解决这一难题.该技术首先自动获取大量的相关评论信息,进而挖掘出主要的产品属性(如:“油耗”)和评价词语(如:“高”),最终通过统计归纳推理,给用户该产品各个属性的评价意见,方便用户做最终的决策.

目前国内外有很多研究机构根据现实生活中的具体需求,研发出各个领域的情感分析系统,帮助用户对海量信息进行分析 & 决策.例如,Liu 等人研发的 OpinionObserver 系统可以处理网上在线顾客产品评价^[4],采用可

视化方式对若干种产品评价对象的综合质量进行比较. Wilson 等人研发的 OpinionFinder 系统可以自动识别主观性句子以及抽取句子中情感信息^[77].上海交通大学则开发了一个用于汉语汽车论坛的情感分析系统,挖掘并概括人们对各种汽车品牌的评论和意见^[47].

◆ 舆情监控

互联网具有开放性、虚拟性、隐蔽性、发散性、渗透性和随意性等特点,越来越多的网民乐意通过这种渠道来表达观点,逐渐成为舆情话题产生和传播的主要场所.网络信息和社会信息的交融对社会的直接影响越来越大,甚至关系到国家信息安全和长治久安.因此,社会管理者应及时对这些舆论进行反馈.然而由于互联网上的信息量十分庞大,仅靠人工的方法难以应对网上海量信息的收集和处理,因此需要依靠情感分析技术自动的对舆情信息进行监控.虽然目前这一应用点的研究成果还不是很多,但不影响其成为一个有价值的应点.

◆ 信息预测

随着互联网的蓬勃发展,网络信息对人们生活的影响已经越来越不容忽视.某一个新事件的发生或者网络上对某个事件的热议都在很大程度上左右着人们的思维和行动.如:在金融市场上,网络上对某支股票的热议都在很大程度上左右着金融实践者们的行为,同时进一步影响着股市变化的趋势;又如:国外总统或议员大选的时候,很多参选者希望通过汇总选民的网络言论来预测自己是否能够获选.因此,信息预测变得非常必要.情感分析技术可以帮助用户通过对互联网上的新闻、帖子等信息源进行分析,预测某一事件的未来状况.

Devitt^[7]等人通过对金融评论文本的情感极性识别,对未来的金融走势做出预测. Lin^[8]则构造了一个“巴以战争”评论分析系统,来区分某一评论是“支持巴方”还是“支持以方”.此外, Kim^[9]通过分析大量美国大选时的网络新闻评论,来预测美国大选的结果.

除了上面介绍的三个主要的应用领域外,情感分析在其他一些自然语言处理领域也扮演着重要的角色.例如,在信息抽取领域,抽取对象一般是反映客观事实的文本,情感分析技术可用于将文本中的主观句和客观句进行分离,提高信息抽取的准确率^[49].情感分析技术还可以用于问答系统中,当用户所问问题是情感相关的问题时,该技术可以帮助问答系统提供更真实的答案^[78].此外,情感分析技术还可以用于情感文摘的生成,进而达到汇总归纳的目的^[6,29,54,67].情感分析技术的快速发展很大程度上源于人们改进人机交互现状的愿望.该技术在以上众多研究领域的应用,使其成为一个非常重要的研究方向.

6 结束语

本文在充分调研和深入分析的基础上对情感分析的研究进展进行了综述.其中重点介绍了情感分析研究中的几个关键问题,包括情感信息的抽取,情感信息的分类,情感信息的检索与归纳,以及情感分析的评测与资源建设等.情感分析是一个新兴的研究方向,在自然语言处理领域对其进行广泛研究却只有十年左右的时间,所以情感分析研究中尚存许多值得深入探索的问题.在本文的最后,我们基于大量的调研和近几年来研究经验提出一些值得进一步挖掘的研究点,希望对本领域的其他研究者有所启发.

问题 1: 如前所述,情感信息抽取的研究属于情感分析的基础,有众多的基础研究任务.然而,虽然人们已经投入了很大的精力,但总的来看,还有许多工作需要进一步细致化.如在评价词语的抽取与判别任务中,大量的研究者局限于词本身来识别词的主客观性以及褒贬性,而忽视了词语所在语境的影响(如 1.4 节中的 Sen1, Sen2, Sen3 中的“焉”).因此,在接下来的工作里,迫切需要我们把情感信息抽取的各项任务做细做深.此外,组合评价单元作为一种情感信息完整丰富的情感单元,也非常值得深入且重点研究.

问题 2: 目前来看,情感信息分类的大部分工作都集中在句子级和篇章级的褒贬分类任务中.然而,与普通的分类任务不同,更有意义的情感分类任务是针对评价对象的情感分类.即:在一个情感句或情感篇章中,挖掘出某一具体的评价对象所对应的情感类别,具体表现为 COAE 评测的任务 3.这是因为用户关注的是大众对某一评价对象的情感倾向性,而非对某一句子或篇章的情感倾向性.然而,这方面的研究工作还比较少且并不深入.此外,情感句的主客观分类也非常具有研究价值,为情感信息分类起到了过滤不相关(客观)信息的作用.目前主要使用基于特征分类的方法.然而,由于主客观句的特征并不明确而导致效果并不理想.因此,如何找到一种有效的方法进行主客观句的分类是一个亟待解决的问题.

问题 3: 基于情感标签的情感文摘是一种很简洁有效的评论归纳方式,也引起了国外一些学者的关注.但总的来看,由于研究的比较粗糙,准确率并不是很高,无法真正达到实用.因此,在未来工作中,需要我们进一步将工作细化,如:确定情感标签的形式,研究抽取情感标签的方法以及研究挖掘或生成评论的情感标签的模型等都是值得深入研究的问题.

问题 4: 情感分析领域在国外已经有了十几年的发展史,积累了一些情感资源和研究成果.然而,国内的研究却刚刚起步.由于语言的差异性,一些国外的研究技术和情感资源无法直接移植到中文处理中.因此,对于我们这些中文处理领域的研究者来讲,如何结合中文处理的特点,将一些成熟的技术和资源应用到中文情感分析领域,是一个值得我们积极探索的任务.

致谢 在此,我们向对本研究工作提供帮助的老师和同学表示感谢.

References:

- [1] Hatzivassiloglou V, McKeown KR. Predicting the semantic orientation of adjectives. In: Proceedings of EACL-1997. Morristown: ACL, 1997. 174-181.
- [2] Huang XJ, Zhao J. Sentiment Analysis for Chinese Text. Communications of CCF, 2008, 4(2).
- [3] Yao TF, Cheng XW, Xu FY. A Survey of Opinion Mining for Texts. Journal of Chinese information processing, 2008, 22(3):71-80 (in Chinese with English abstract).
- [4] Pang B, Lee L. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2008,2(1-2): 1-135.
- [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).
- [6] Liu B, Hu MQ, Cheng J. Opinion Observer: Analyzing and Comparing Opinions on the Web. In: Proceedings of WWW-2005. 2005. 342-351.
- [7] Devitt A, Ahmad K. Sentiment Polarity Identification in Financial News: A Cohesionbased Approach. In: Proceedings of the Association for Computational Linguistics (ACL). 2007. 984-991.
- [8] Lin WH, Wilson T, Wiebe J. Which Side Are You On? Identifying Perspectives at the Document and Sentence Levels. In: Proceedings of the Conference on Natural Language Learning (CoNLL). New York: ACL, 2006. 109-116.
- [9] Kim SM, Hovy E. Crystal: Analyzing Predictive Opinions on the Web. In: Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). 2007. 1056-1064.
- [10] Rao D, Ravichandran D. Semi-supervised Polarity Lexicon Induction. In: Proceedings of EACL-2009. 2009. 675-682.
- [11] Wiebe J. Learning Subjective Adjectives from Corpora. In: Proceedings of AAAI. 2000.
- [12] Riloff E, Wiebe J. Learning Extraction Patterns for Subjective Expressions. In: Proceedings of EMNLP-2003. 2003. 105-112.
- [13] Turney P, Littman ML. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. ACM Transactions on Information Systems (TOIS), 2003, 21(4):315-346.
- [14] Kim SM, Hovy E. Automatic Detection of Opinion Bearing Words and Sentences. In: Proceedings of IJCNLP-2005. 2005. 61-66.
- [15] Kim SM, Hovy E. Identifying and Analyzing Judgment Opinions. In: Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL). 2006. 200-207.
- [16] 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).
- [17] Andreevskaia A, Bergler S. Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses. In: Proceedings of the European Chapter of the Association for Computational Linguistics (EACL). 2006. 209-216.
- [18] Su F, Markert K. Subjectivity Recognition on Word Senses via Semi-supervised Mincuts. In: Proceedings of NAACL-2009. 2009. 1-9.
- [19] Esuli A, Sebastiani F. Determining the Semantic Orientation of Terms Through Gloss Analysis. In: Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM). 2005.
- [20] Esuli A, Sebastiani F. Determining Term Subjectivity and Term Orientation for Opinion Mining. In: Proceedings of the European Chapter of the Association for Computational Linguistics (EACL). 2006. 193-200.

-
- [21] Kamps J, Marx M, Mokken RJ. Using WordNet to Measure Semantic Orientation of Adjectives. In: Proceedings of the LREC. 2004.
- [22] Mihalcea R, Banea C, Wiebe J. Learning Multilingual Subjective Language via Cross-lingual Projections. In: Proceedings of the Association for Computational Linguistics (ACL). 2007. 976–983.
- [23] Wiebe J, Mihalcea R. Word Sense and Subjectivity. In: Proceedings of the Conference on Computational Linguistics / Association for Computational Linguistics (COLING/ACL). 2006. 1065-1072.
- [24] Takamura H, Inui T, Okumura M. Extracting Semantic Orientation of Words Using Spin Model. In: Proceedings of the Association for Computational Linguistics (ACL). 2005. 133–140.
- [25] Yi J, Nasukawa T, Bunescu R. Sentiment Analyzer: Extracting Sentiments about a Given Topic Using Natural Language Processing Techniques. In: Proceedings of the IEEE International Conference on Data Mining (ICDM). 2003.
- [26] Hu M, Liu B. Mining Opinion Features in Customer Reviews. In: Proceedings of AAAI-2004. 2004. 755–760.
- [27] Ni MS, Lin HF. Mining Product Reviews Based on Association Rule and Polar Analysis. In: Proceedings of the NCIRCS-2007. (2007) 628 – 634 (in Chinese with English abstract).
- [28] Liu HY, Zhao YY, Qin B, Liu T. Target Extraction and Sentiment Classification. In: Proceedings of the 10th Chinese National Conference on Computational Linguistics. 2009
- [29] Popescu AM, Etzioni O. Extracting Product Features and Opinions from Reviews. In: Proceedings of hltemnlp2005. 2005.339–346.
- [30] Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 2003(3):993–1022.
- [31] Blei DM, Ng AY, Jordan MI. Correlated Topic Models. In *Advances in NIPS*. 2006:147 – 154.
- [32] Titov I, McDonald R. Modeling Online Reviews with Multi-grain Topic Models. In: Proceedings of WWW-2008. 2008. 111–120.
- [33] Kim SM, Hovy E. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In: Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text. 2006. 1–8.
- [34] Stoyanov V, Cardie C. Topic Identification for Fine-grained Opinion Analysis. In: Proceedings of the Conference on Computational Linguistics. 2008. 817–824.
- [35] Kim SM, Hovy E. Determining the Sentiment of Opinions. In: Proceedings of Coling-2004. 2004. 1367–1373.
- [36] Choi Y, Cardie C, Riloff E. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. In: Proceedings of HLT/EMNLP-2005. 2005. 355–362.
- [37] Bethard S, Yu H, Thornton A. Automatic Extraction of Opinion Propositions and Their Holders. In: Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text. 2004. 22–24.
- [38] Wiebe J, Wilson T, Bell M. Identifying Collocations for Recognizing Opinions. In: Proceedings of the ACL/EACL Workshop on Collocation: Computational Extraction, Analysis, and Exploitation. 2001.
- [39] Wiebe J, Wilson T. Learning to Disambiguate Potentially Subjective Expressions. In: Proceedings of the Conference on Natural Language Learning (CoNLL). 2002. 112–118.
- [40] Wilson T, Wiebe J, Hwa R. Just How Mad Are You? Finding Strong and Weak Opinion Clauses. In: Proceedings of AAAI-2004. 2004. 761–769.
- [41] Wilson T, Wiebe J, Hwa R. Recognizing Strong and Weak Opinion Clauses. *Computational Intelligence*. 2006, 22(2):73–99.
- [42] Whitelaw C, Garg N, Argamon S. Using Appraisal Groups for Sentiment Analysis. In: Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM). 2005. 625–631.
- [43] Moilanen K, Pulman S. Sentiment Composition. In: Proceedings of the Recent Advances in Natural Language Processing International Conference (RANLP-2007). Borovets, Bulgaria, 2007. 378–382.
- [44] Choi Y, Cardie C. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In: Proceedings of EMNLP-2008. 2008. 793–801.
- [45] Kobayashi N, Inui K, Matsumoto Y. Collecting Evaluative Expressions for Opinion Extraction. In: Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP). 2004.
- [46] Bloom K, Garg N, Argamon S. Extracting Appraisal Expressions. In: Proceedings of HLT-NAACL 2007. 2007. 308–315.
- [47] Yao TF, Nie QY, Li JC. An Opinion Mining System for Chinese Automobile Reviews. *Frontiers of Chinese Information Processing*, 2006, 260–281 (in Chinese with English abstract).

-
- [48] Xu LH, Lin HF, Zhao J. Construction and Analysis of Emotional Corpus. *Journal of Chinese information processing*, 2008: 116–122 (in Chinese with English abstract).
- [49] Riloff E, Wiebe J, Phillips W. Exploiting Subjectivity Classification to Improve Information Extraction. In: *Proceedings of AAAI-2005*. 2005. 1106–1111.
- [50] Hatzivassiloglou V, Wiebe J. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. In: *Proceedings of the International Conference on Computational Linguistics (COLING)*. 2000. 299–305.
- [51] Yu H, Hatzivassiloglou V. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. In: *Proceedings of EMNLP-2003*. 2003. 129–136.
- [52] Yao TF, Peng SW. A study of the classification approach for Chinese subjective and objective texts. In: *Proceedings of the NCIRCS-2007*. (2007) 117–123 (in Chinese with English abstract).
- [53] Pang B, Lee L. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In: *Proceedings of ACL-2004*. 2004. 271–278.
- [54] Hu MQ, Liu B. Mining and Summarizing Customer Reviews. In: *Proceedings of KDD-2004*. 2004. 168–177.
- [55] Turney P. Thumbs up Or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: *Proceedings of ACL-2002*. 2002. 417–424.
- [56] Pang B, Lee L, Vaithyanathan S. Thumbs Up? Sentiment Classification Using Machine Learning Techniques. In: *Proceedings of EMNLP-2002*. 2002. 79–86.
- [57] Cui H, Mittal VO, Datar M. Comparative Experiments on Sentiment Classification for Online Product Reviews. In: *Proceedings of AAAI-2006*. 2006. 1265–1270.
- [58] Kim SM, Hovy E. Automatic Identification of Pro and Con Reasons in Online Reviews. In: *Proceedings of the COLING/ACL-2006*. 2006. 483–490.
- [59] Zhao J, Liu K, Wang G. Adding Redundant Features for Crfs-based Sentence Sentiment Classification. In: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, (EMNLP)*. 2008. 117–126.
- [60] Pang B, Lee L. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. In: *Proceedings of the Association for Computational Linguistics (ACL)*. 2005. 115–124.
- [61] Goldberg AB, Zhu X. Seeing Stars When There Aren't many Stars: Graph-based Semi-supervised Learning for Sentiment Categorization. In: *Proceedings of HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing*. 2006.
- [62] Hurst M, Nigam K. Retrieving Topical Sentiments from Online Document Collections. In: *Proceedings of Document Recognition and Retrieval XI*. 2004. 27–34.
- [63] Ounis I, Rijke MD, Macdonald C, Mishne G, Soboroff I. Overview of the TREC-2006 Blog Track. In: *Proceedings of TREC*. 2006.
- [64] Zhang W, Yu C, Meng WY. Opinion Retrieval from Blogs. In: *Proceedings of CIKM*. 2007.
- [65] Zhang W, Yu C. UIC at TREC 2007 Blog Track. In: *Proceedings of 16th TREC*. 2007.
- [66] Zhang M, Ye XY. A Generation Model to Unify Topic Relevance and Lexicon-based Sentiment for Opinion Retrieval. In: *Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR)*. 2008. 411–418.
- [67] Carenini G, Ng R, Pauls A. Multi-document Summarization of Evaluative Text. In: *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*. 2006. 305–312.
- [68] Qin B, Zhao YY, Gao LL, et al. Recommended Or Not? Give Advice on Online Products. In: *Proceedings of Fifth International Conference on Fuzzy Systems and Knowledge Discovery*. 2008. 208–212.
- [69] Titov I, McDonald R. A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In: *Proceedings of ACL-2008*. 2008. 308–316.
- [70] Branavan S, Chen H, Eisenstein J. Learning Document-level Semantic Properties from Free-text Annotations. In: *Proceedings of ACL-08: HLT*. 2008:263–271.
- [71] Ku LW, Liang YT, Chen HH. Opinion Extraction, Summarization and Tracking in News and Blog Corpora. In: *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*. 2006.
- [72] Ounis I, Rijke MD, Macdonald C. Overview of the Trec-2006 Blog Track. In: *Proceedings of the 15th Text Retrieval Conference(TREC)*. 2006.

-
- [73] Zhao J, Xu HB, Huang XJ, Tan SB, Liu K, Zhang Q. Overview of Chinese Opinion Analysis Evaluation 2008. 2008
- [74] Jindal N, Liu B. Identifying Comparative Sentences in Text Documents. In: Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR). 2006. 244-251.
- [75] Wiebe J, Wilson T, Cardie C. Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation (formerly Computers and the Humanities). 2005, 39(2/3):164-210.
- [76] Ku LW, Lo YS, Chen HH. Using Polarity Scores of Words for Sentence-level Opinion Extraction. In: Proceedings of NTCIR-6 Workshop Meeting. 2007. 316-322.
- [77] Wilson T, Hoffmann P, Somasundaran S. Opinionfinder : A System for Subjectivity Analysis. In: Proceedings of HLT/EMNLP2005 Demonstration Abstracts. 2005. 34-35.
- [78] Lita LV, Schlaikjer AH, Hong W. Qualitative Dimensions in Question Answering: Extending the Definitional QA Task. In: Proceedings of AAAI. 2005. 1616-1617.

附中文参考文献:

- [2] 黄萱菁,赵军.中文文本情感分析.中国计算机学会通讯.2008:第4卷第2期.
- [3] 姚天昉,程希文,徐飞玉等.文本意见挖掘综述.中文信息学报. 2008, 22(3): 71-80.
- [5] 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
- [16] 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 年第 1 期
- [27] 倪茂树,林鸿飞.基于关联规则和极性分析的商品评论挖掘.第三届全国信息检索与内容安全学术会议.2007:628 - 634
- [28] 刘鸿宇,赵妍妍,秦兵,刘挺.评价对象抽取及其倾向性分析.全国第十届计算语言学学术会议. 2009
- [47] 姚天昉,聂青阳,李建超等.一个用于汉语汽车评论的意见挖掘系统.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.
- [48] 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. 2008:116 - 122
- [52] 姚天昉,彭思崴.汉语主客观文本分类方法的研究.第三届全国信息检索与内容安全学术会议. 2007:117 - 123
- [73] 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