

网络舆情分析中共性知识挖掘方法研究^{*}

段建勇 程利伟 张梅 高振安

(北方工业大学信息工程学院 北京 100144)

【摘要】 共性知识挖掘是网络舆情中实现领域可移植的有效途径,提出从共性情感元素、共性语言模式两方面建立共性舆情知识库。共性情感元素挖掘主要通过半自动方法识别并从训练库中学习量化权值实现动态扩展知识库;共性语言模式挖掘主要从语法、语义角度弥补句法分析引入的错误,提出三类修正模型,包括主语转移模型、极端情感动词模型与情感修饰短距离依赖模型。最后从宗教、酒店两个领域进行验证,证实共性知识挖掘在系统可移植性方面具有一定效果。

【关键词】 舆情分析 共性知识挖掘 情感元素 语言结构

【分类号】 TP391

The Common Knowledge Mining for the Internet Public Opinion Analysis

Duan Jianyong Cheng Liwei Zhang Mei Gao Zhen'an

(College of Information Engineering, North China University of Technology, Beijing 100144, China)

【Abstract】 The common knowledge mining is an effective way for the Internet public opinion analysis. This paper builds the common knowledge base for the common sentimental elements and the common language patterns. The common sentimental knowledge is mined by the semi-supervised method from the training corpus. This knowledge base is also quantified and dynamically expanded. The common language pattern knowledge includes three kinds of fixed models, such as transform model, extreme sentimental verb model and distance dependency model. Finally the common knowledge bases are testified in the domains of religions and hotels, and proved the effectiveness in the system implant performance.

【Keywords】 Public opinion analysis Common knowledge mining Sentimental element Language structure

1 引言

网络舆情分析是将带有情感倾向的用户评论运用相关技术进行分析与综合,给出倾向性分类与标注,帮助决策者及时把握相关舆论动态发展。例如产品评论的情感分析有助于公司制定正确的战略赢得最大利润;社会舆论的情感分析有助于政府把握民情与社会热点以及对政策的意见,维护国家的长治久安。

国际上多家相关研究团体纷纷组织舆情研究^[1]。国际文本检索会议(Text REtrieval Conference, TREC)从2006年开始,每年都会组织情感倾向性的评测任务,从文本数据集上检索出带有倾向性观点的文档。NTCIR(NACSIS Test Collections for IR)是由日本情报信息研究所主办的多语言处理国际评测会议,也于2006年开始了多语言观点倾向性评测,每年举行一次,拥有中、日、英三种语言的标准语料库,评测目的是从新闻报道中提取带有主观性信息的内容。国内2008年发起的COAE(Chinese Opinion Analysis Evaluation)是第一个在中文情感分析

收稿日期:2013-07-01

收修改稿日期:2013-08-28

* 本文系教育部人文社会科学基金项目“基于多层次情感分析的网络文本舆情监测方法研究”(项目编号:10YJC870003)、北京市哲学社会科学规划基金项目“北京市公共危机事件在网络传播中的演化机制与模型研究”(项目编号:13SHC031)和国家自然科学基金项目“面向维基百科的多粒度一体化信息抽取方法研究”(项目编号:61103112)的研究成果之一。

领域的评测,近年来在不断推动中文情感倾向性分析理论研究和应用的同时,逐步建立了舆情分析研究的基础数据集。

2 相关工作

网络舆情分析在政府、研究团体等共同推动下,其研究领域与传播载体向纵深立体化发展^[2]。舆情分析领域涵盖了产品评论、新闻评论、股市评论等,舆情分析的媒介也由传统的网页扩展到BBS、微博等社交媒体,载体呈多样化、立体化。舆情分析在技术上呈精细化方向发展^[3],覆盖了词语、句子以及篇章等不同粒度的情感计算,特别是研究重心集中在句子级情感计算,发展了一系列句子级的情感倾向性计算方法。除了句法分析为核心技术外,特征分类为核心技术在情感倾向性计算方法上也较为常见。

第一种主流方法是以句法分析为基础的情感计算方法,主要依赖句法分析技术,但不同于普通的句法分析,它依赖于一定的情感知识作为情感计算基础。例如文献[4]主要采用手工方法建立情感词的极性词典,并引进同义词来提高词典覆盖率,在情感知识库的支持下,结合句法分析技术进行情感分析。情感分析特定任务不需要深度的语法分析,因而浅层的句法分析常常被应用于情感分析中,它不需要全面分析句法的内部结构,只需要辅助分析出各种情感元素之间的关联关系。例如文献[5]采用局部浅层句法分析来研究情感句子的倾向性计算方法,并结合语言极性分析方法研究舆情。文献[6]研究句子级文本情感分析,在系统模型的训练过程中引入浅层句法信息和启发式位置信息,同时在不增加领域词典的情况下,有效提高系统的精确率。由于句法分析的主要目的是获取情感元素之间的关联关系,文献[7]研究句子级情感分析中自动获取句法路径来描述评价对象及其评价词语之间的修饰关系,并通过计算句法路径编辑距离来改进情感评价单元抽取的系统性能。除了在整个句法分析,还有对其中局部的骨架元素进行分析,例如文献[8]研究了深度情感分析中的动词模型,采用句法语义的方法精细描述动词的配价关系。实际上,在情感文本中文字表述经常是口语类型或网络语,具有很大的随意性,而常规的句法分析器是建立在规范文本的基础上的,需要结合局部句法分析方法以及词汇语

义知识进行补充,使其适应非规范文本。

第二种主流思路是将情感倾向性分析看成是特征分类问题,这样所有特征分类技术便可应用于情感倾向性分析问题中。特征分类模型不需要文本具有较好的规范性也能表现出较好的特性。例如文献[9]研究了产品评论信息的不规范问题,采用机器学习的方法研究产品评论信息分析方法。特征分类技术包含特征选择与特征集成等关键技术,例如文献[10]研究情感分析中的特征选择问题,关注主观性表达语句,抽取其中明显的特征。情感分析的特征集成方法主要有异构特征集成以及多策略方法等,例如文献[11]研究了基于Stacking组合分类方法的中文情感分类方法,将4组统计方法集成到情感分类任务中,为异构特征集成提供思路;多策略方法如文献[12]研究博客舆情分析中的表情符号的规则方法、情感词典的规则方法、基于SVM的层次结构的多策略方法。运用特征分类法进行情感倾向性计算主要依赖于语料库的训练,不同的语言特征分布对系统性能有较大的影响。

语言分析技术与特征分类技术是情感倾向性分析的两种主流技术,语言分析技术具有较好的领域通用性,但要求文本具有一定规范性且需要情感元素知识库支撑,而情感元素知识库具有领域性;特征分类技术没有文本规范性要求与情感元素知识库支撑,但其分类性能对训练语料有领域依赖性。两种技术都存在移植性问题,包括跨领域与跨时间,例如文献[13]跨领域知识的基金评论情感分析,通过标注语料库训练来提取语言知识,提升跨领域舆情分析水平;文献[14]研究评论的情感倾向性随时间推移的变化情况,研制了自适应情感分析方法并运用于销售预测。虽然舆情分析领域与载体差异很大,但在舆情分析技术上有较多的共通之处可供借鉴。目前全领域范围的舆情分析技术尚不成熟,舆情分析系统的开发一般都针对特定领域,如何以最小代价将系统移植到新的领域具有现实意义。

本文侧重从语言分析的角度研究情感分析的移植性问题,主要研究宗教类与酒店类舆情语料,获取舆情分析的共性情感元素与语言模式知识,并围绕移植过程中的共性知识挖掘与集成问题,探讨跨领域的可移植性。

3 共性知识挖掘

舆情分析是在简化句法分析的基础上进行情感倾向性判断的过程,包括情感元素的识别、借助句法分析的语言结构分析、情感倾向性计算等主要过程。人类在表达情感的时候所用的语言结构是相似的,不同领域间舆情分析方法的差异在于相似语言结构下的情感元素的替换。共性知识挖掘的目的是在较少的人工干预条件下,尽可能多地发现共性的语言结构与基础性情感元素,既保持特定领域较高的准确率,又方便地扩展舆情分析的适用范围,提高系统的移植性,主要进行共性情感元素挖掘与共性语言结构模式挖掘。

3.1 共性情感元素挖掘

舆情分析中的情感元素主要有评价主体、评价客体、评价词语、评价指标等^[2]。评价主体为句子中观点的持有者,情感的发出者,在大部分情况下会省略或者是隐含的。例如“我国政府始终坚持人民的利益高于一切”,这里的观点持有者就是“我国政府”,评价主体是判定情感倾向的基础,其决定了情感立场。评价客体为情感句中所讨论的对象。评论文本中涉及的新闻事件、产品等都可看成是评价客体,评论内容主要围绕这些客体进行讨论,发表具有倾向性的评论语言。例如“国产手机质量也不错!”,这里的“国产手机”为评价客体,评价客体还常常伴有评价指标等围绕客体某个侧面的细化评价指标,这里“手机”就包含了“质量”评价指标。评价词语指具有情感倾向的修饰性词语,是人类用特定的词汇来表达情感或进行评价,评价词语的性质表达了评价主体对评价客体的情感倾向^[4]。例如“母亲是那么的善良和和蔼”,这里的“善良”和“和蔼”就是评价词语。形容词和副词是最常见的评价词语,在某些情况下动词、名词也可以作为评价词语。

评价主体和评价客体与舆情领域直接相关,一般会有较大的差别,每个领域的知识库需要以人工或者半自动的方式建立。评价词语是表达情感倾向的词语,不同领域间具有一定的共性,构建公共的评价词语库可以实现领域间的共享。

本文提出共性情感元素挖掘方法可以解决不同领域间舆情知识共享问题。先从特定领域挖掘开始,逐步发现不同领域间的重复知识,按照知识的重合度不断动态反馈,建立稳定的共性知识库。共性知识库从

舆情语料库中的情感句中获取,情感句中包含评价客体与评价词语两个基本条件,需要手工建立少量的评价客体库与评价词语库作为自动扩充学习的起点。

评价客体库的扩充可看成实体抽取任务,在分词与词性标注基础上结合错误词性标注规律,发现新的评价客体。构建评价词语库可以从常规词典中抽取,并人工分析词语的褒贬性生成经验型评价词语库。为了实现共享,后期对褒贬义模糊的词语进行逐步强化,在人工介入的基础上分别建立评价客体库与情感词汇库,还应当对各种情感元素进行量化,在此基础上建立灵活的共性情感库。

情感元素量化学习算法通过较少的人工介入半自动来实现,本方法对学习语料库的构成有一定要求,需要人工将语料库设置成正面与负面语料,并结合已经人工建立的评价客体库、评价词汇库将语料库自动细分成以情感句子为单位的 4 个语料库,方便后期训练。例如,对正面评价客体进行正面描述的语料,简称正正语料;对正面评价客体进行负面描述的语料,简称正负语料;对负面评价客体进行正面描述的语料,简称负正语料;对负面评价客体进行负面描述的语料,简称负负语料。在细分语料库的基础上,借助细分语料库的倾向性标注以及已建立的情感知识库进行交叉验证,并结合实体识别技术逐步发现更多的潜在评价实体,加强共性情感元素库的量化权值,步骤如下所示:

算法 1:情感元素量化算法

输入:半自动标注的 4 类情感语料库

输出:系统词典:评价客体与评价词汇的正负面量化特征
计算过程:

- ①语料库预处理,进行分词与词性标注;
- ②全局识别潜在的评价客体与评价词语;
- ③提取包含潜在评价客体与评价词语的情感句;
- ④统计各类特征出现频率并进行量化处理;
- ⑤将新学习的情感元素进行人工甄别。

3.2 共性语言结构模式挖掘

情感分析过程是对情感句进行句法分析与语义计算的过程^[5,6],句法分析过程将情感句在词性标注以及情感元素识别的基础上进行句法树的构造;语义计算过程重点关注情感元素的语义关系,并依据情感元素间的相对语义关系进行情感倾向性评价。情感分析需要借助句法分析获取情感元素间的依存关系,在情感分析过程中充分结合浅层句法分析与命名实体识别方

法进行情感倾向性评价,但是通用的句法分析等工具在句子成分相对完备的情况下表现良好,在缺失重要句法成分或增加特殊情感元素的情况下,需要建立修正模型。为此,本文提出主语转移模型、极端动词模型以及短距离依赖模型。

(1) 主语转移模型

在中文常用句子中经常会出现复句,复句中由于每个单句意义紧密,具有很强的关联关系,主语常常会被略去而不影响句子表达。复句实例如表1所示:

表1 复句实例

原始句子	主语转移后生成各分句
假如产品质量不好,数量再多,也怕没有使用价值。	①假如产品质量不好 ②产品数量再多 ③产品也怕没有使用价值

原始句子中从第一句可以看出主语为产品,但是第二句和第三句的主语就被省略了,由于主语的缺失导致主谓宾结构不完整,识别句子情感倾向时会直接丢掉该复句。如果把第一句的“产品”转移到第二句、第三句做主语,就会得到完整的主谓宾式的句子结构,有利于句子情感倾向性的计算。

如何有效解决复句中单句主语迁移问题是这类情感分析问题的关键。首先识别段落中相关的主语,可以结合名词短语识别技术识别评价客体,将其保留在栈结构中,遍历完相关段落,结合复句类型列出相关主语缺失单句进行主语补充。具体步骤如下:

算法2:主语转移算法

输入:待分析文本段落

输出:主语补充完整后文本段落

计算过程:

- ①依次按单句输入,识别复句结构,抽取名词性短语,识别评价客体;
- ②将评价客体压入栈结构,该栈主要存贮替换主语,并记录其在句中位置;
- ③按顺序列出主语缺省句,结合句子位置补充或替换缺省句;
- ④输出包含主语完整表达句子。

这类复句的主语缺失问题在语言现象中具有一定的普遍性,在复句分析过程中可以结合复句的类型,包括并列、承接、递进、选择、转折、因果、假设、条件、解说、目的等复句,识别复句主要利用具有关联作用的连词、副词以及一部分起关联作用的短语等,然后建立以复句为主要分析对象的主语转移模型作为共性语言结

构模式分析模型。

(2) 极端情感动词模型

极端情感动词模型主要解决在模型分析过程中动词不做褒贬倾向支持的问题。实际上,在语言运用过程中,许多动词具有极强的倾向性。除了中性动词外,可分为褒义动词与贬义动词,例如迫害、侵犯等属于贬义动词,而爱戴、拥护等属于褒义动词。“这家店铺侵害消费者权益。”这句话可以依据其动词褒贬倾向综合计算句子的褒贬义。

另外动词模型中还要结合动词本身的配价问题,这里主要考虑一价动词与二价动词,并为每类动词建立分析配价模型,主要原因是这两类动词建模相对容易且占有较大的比例。一价动词如“苹果的营销计划成功了”;二价动词“这种思想毒害年轻人”。

动词模型是一种词汇级的修正模型,它具有较高的可信度,计算过程中的优先级高于句法树,可以修改原来句法树得到的倾向性值。动词模型能提高倾向性分析结果的正确率,但主要依靠语言知识支持,合理的动词配价分析模型及一定规模的动词库是性能提高的保障,并可以作为倾向性分析的共性知识。

(3) 情感修饰的短距离依赖模型

句法树的处理粒度是句子级别的,有时句子结构分析引入局部错误或由于复句的某些成分的省略,会造成一定的误差,所以系统加入了处理对象为短距离语义依赖的处理方法。情感词修饰短语是一种短距离依赖模型,它由情感词以及其所修饰的评测对象组成,两者缺一不可,否则不能将其定义为情感词修饰短语。由于语言在实际运用中无论语义表达还是句子结构,一般都较为复杂,中文处理中没有确定的模型可以完全正确分析出结果。如果主语的自动转移分析错误,那将会得到情感倾向性相反的结果,而短距离依赖的处理对象有助于修正这种失误所带来的负面影响。

情感倾向性分析通常是在整句分析的基础上得到倾向性结果,但是有时候整句倾向性主要集中在局部的某些短语,这些短语中包含了情感倾向性词汇与评价对象,形成整句的情感倾向性的极性。这类情况主要依据两个重要指标:评价对象具有明显的情感倾向;情感词作为短语成分修饰评价对象,这样整句的情感倾向性集中在局部结构上。这类语言现象具有一定的普遍性,由其计算生成的局部情感倾向正确率相对较

高,在修正模型中优先级高于主语转移模型。

(4) 修正模型间的优先顺序

本文引入的几类语言模式修正模型,主要是解决目前句法分析尚不完全成熟以及舆论文本表述不规范的问题。语言修正模型针对初始句法分析有可能导致的问题及时进行调整,本质上讲是一种规则的方法,在修正规则逐步增加的时候,必然会导致规则间的冲突需要消解,建立合适规则的优先级能在一定程度上避免规则冲突引入的问题。优先顺序一般是局部修正模型优先,例如极端动词模型、情感修饰的短距离依赖模型优先级高;而主语自动转移模型是句子级修正模型,优先级较低。目标是引入新的修正模型后能够提高系统性能,并具有一定的通用性。

4 实验

舆情分析中共性知识挖掘主要实现舆情知识复用,实现舆情系统在不同领域间的移植。在实验阶段考察共性知识的挖掘及其对舆情系统不同领域间移植性支持效果的评价。

4.1 实验数据

实验数据选择宗教与酒店领域的舆情数据,宗教类舆情数据来自哈尔滨工业大学的机器翻译实验室;酒店类舆情语料来自中国科学院计算技术研究所组织的评测语料。在宗教、酒店领域各选择了 400 个样本进行人工分类标注形成正面样本 200 篇、负面样本 200 篇为测试样本,其余作为训练数据备用,主要获取领域相关的评价对象和部分情感词。

4.2 共性情感元素挖掘实验结果

笔者在先期舆情系统的基础上^[4],增加了共性语言知识移植性模块,建立系统词典,包括共性词典与领域词典。共性情感元素挖掘主要为系统移植性,需要具有广泛的适用范围;领域词典则只适用于特定领域的词典。采用领域词典与共性词典相结合的方式能节省地将系统从一个特定领域转移到另一个特定领域进行舆情分析并保持良好性能。

(1) 实验结果

首先将各种基本词语进行整理,按照频率进行初筛,然后人工甄别,形成可以使用的系统词典。例如宗教类训练文本中,从 7 991 个名词中选取排名靠前的名词作为正面和负面评测对象;从 1 312 个情感词

语中选取排名靠前的词语组成正面和负面情感词词典;从 3 644 个情感动词中找出具有褒贬义色彩的情感动词组成正面和负面情感动词表。在此基础上,人工参与进行甄别,大大降低了人工的工作量。同样的方法在酒店领域整理词表,调整后的领域细分词典收录情况如表 2 所示:

表 2 细分系统词典收录情况

词典类别	宗教领域			酒店领域		
	最大词频	最小词频	词条数量	最大词频	最小词频	词条数量
正面的评测对象	1 080	50	31	396	2	315
负面的评测对象	1 765	52	17	4	2	36
正面情感词词典	167	2	115	26	2	88
负面情感词词典	434	2	153	35	2	112
正面情感动词词典	182	2	6	21	2	9
负面情感动词词典	643	2	622	20	2	42

(2) 实验结果分析

共性情感元素挖掘主要是针对以形容词为主的情感词典和情感动词,获取方法是先单独从不同领域训练再进行合并形成共性情感元素库。在两个领域的挖掘中发现,负面的情感动词所占比例最多,大约占总词汇量的 66%,正面的情感动词所占比例极少,其次是正面情感词和负面情感词,约占总词量的 28%。每个领域形成上述几个细分词典后,对情感词词典、情感动词词典进行合并,依据算法由领域词典逐步形成共性情感元素词典,其中情感词词典 103 个,在宗教与酒店领域中重合的情感词语占 40% 左右,并且这些情感词语在两个领域中出现的频次大都超过 3 次,较为稳定;情感动词 11 个,主要是动词使用具有一定领域性。随着领域的增加,每个共性情感词语的权值会动态调整。

由上述共性情感元素库构建过程可知,构建共享的情感元素库具有很强的可操作性。共性情感元素库是系统运行的必要条件,主要优点是无需重新训练便能够获取数量可观的可共享的情感元素库,随着系统领域扩展具有累加式效应。

4.3 共性语言模式实验结果

舆情分析系统中共性情感元素是情感倾向性分析的基础,但是情感倾向性需要考虑到目前句法分析本身的不成熟以及舆论文本存在不规范表述问题等现状,为此引进语言模式修正模型来改善系统运行性能。

(1) 实验结果

舆情分析系统在结合共性情感元素的基础上,采用隶属度刻画语句的情感倾向性,取值范围为 $[-1, 1]$,

其中 -1 表示情感倾向为负面最大值,1 表示情感倾向为正面最大值。为了给出明确的情感倾向分类,在取值范围基础上设置褒贬两类阈值,高于褒义阈值才能判定为正面评论,低于贬义阈值才能判定为负面评论。在保证总体正确率稳定的前提下,滑动阈值窗口观察正确率,分别将阈值设定成符合实际需求的情况。例如宗教类语料的负面样本识别率高于正面样本的识别率,因为负面样本的危害性更大, -0.1 以下即可判定为负面评价,0.3 以上为正面评价;而酒店领域则按照经验值阈值进行设定,0.3 以上为正面评价, -0.3 以下为负面评价。

系统引进语言模式修正模型非常谨慎,主语转移模型只在可识别的复句中推广;极端动词模型依赖独立的动词库;短距离依赖的窗口只限于三个词汇长度以内。因为语言模式的优先级比较高,类似于规则的方式来修正情感分析结果,不合适的修正会降低系统整体性能。共性语言结构分析主要以浅层句法分析为工具,结合情感分析特点进行结果修正,使用前后一定程度上提高了系统识别性能。对两类样本进行正确性测试的结果如表 3 所示:

表 3 采用共性语言模式前后识别结果对比

样本数量	识别结果	宗教领域		酒店领域	
		识别数量(个)		识别数量(个)	
		修正前	修正后	修正前	修正后
正面样本 200 篇	正面	127	122	183	191
	负面	32	44	1	0
	未知	41	34	16	9
负面样本 200 篇	正面	23	11	17	11
	负面	167	188	157	158
	未知	10	1	26	31
正确率		73.50%	77.50%	85%	87.25%

(2) 实验结果分析

从表 3 可知,使用语言模式修正模型后识别结果得到了一定程度提高,宗教领域的提高幅度为 4%,酒店领域为 2%。主要原因是修正规则侧重于对负面评论的发现,而宗教领域中语句的情感倾向要强于酒店领域,特别是极端动词、情感修饰短距离依赖现象尤为明显,因而改进效果较好。

①主语转移修正模型。主语转移修正模型解决复句中主语缺失情况,如果句子分析使用了主语转移模型,那么在句子分析结果的后面会有[SubjectTurn]标识,如表 4 所示。

在这个句群中由于出现了主语缺失或者代词替换情况,在运用主语转移分析后,结果就得到进一步修正。其中涉及

表 4 主语转移修正结果

原始句群	分析结果	修正后结果
①这家酒店客房多, ②但不好住。	①(酒店客房/Nz 多/Vv1)1,第二句 缺乏主语,分析失败	①(酒店客房/Nz 多/Vv1) 1,②(但酒店客房 /Nz 不 好住 Vv - 1) - 1 [Subject- Turn]

的符号意义为:用/Nz 表示名词构成的主语,包括修饰名词的形容词和副词;/Vv 表示动词组成的谓语,最外面的“()”外是句子情感倾向值,1 表示正面, -1 表示负面,0 表示中性,只表示陈述事实。

②极端情感动词模型句法分析。在通常情况下,动词一般不参与情感语义计算过程,但是某些动词具有强烈的感情色彩,例如“报复、污蔑”等表达了负面的情感,而“赞美、嘉奖”等则表达了正面的情感,对情感分析具有重要的影响,极端动词模型通过引入对主要情感动词的情感计算值,修正句法树分析的问题。假设与极端情感动词最近的主语和宾语语义联系最紧密,此模型的优先级要高于句法树分析的结果。运用极端情感动词模型进行修正,得到正确的情感倾向性值。

③情感修饰短距离依赖句法分析。句法分析对于复合句的处理能力较差,特别是嵌套句式可能会导致句法树判断失误,可以结合局部的情感修饰进行情感倾向值修正,局部修正模型由于是短距离依赖关系,其在情感分析过程中正确率较高,在整句分析过程中应该结合局部分析结果得到最终的情感倾向值。

5 结 语

舆情分析在网络时代的经济与政治生活中扮演重要角色,目前虽然特定领域的舆情分析具有较高的准确率,但通用领域的舆情分析系统还很难达到实用水平,研究跨领域的舆情共性知识挖掘具有重要意义,它不但能节省舆情分析系统的开发成本,而且对开发下一代通用舆情分析系统提供基础支持。未来将重点研究情感元素库、语言模式修正模型的扩充与量化评价。

参考文献:

[1] 崔薇,曾润喜,王国华. 中国网络舆情研究文献计量分析[J]. 情报科学, 2011, 29(1): 131 - 135. (Cui Wei, Zeng Runxi, Wang Guohua. Bibliometric Analysis of Internet Public Opinion in China[J]. Information Science, 2011, 29(1): 131 - 135.)

[2] 魏韡,向阳,陈千. 中文文本情感分析综述[J]. 计算机应用, 2011, 31(12): 3321 - 3323. (Wei Wei, Xiang Yang, Chen Qian.

- Survey on Chinese Text Sentiment Analysis[J]. *Journal of Computer Applications*, 2011, 31(12): 3321 - 3323.)
- [3] 李纲,程洋洋,寇广增. 句子情感分析及其关键问题[J]. *图书情报工作*, 2010, 54(11): 104 - 107. (Li Gang, Cheng Yangyang, Kou Guangzeng. Key Problems of Sentence Level Sentiment Analysis[J]. *Library and Information Service*, 2010, 54(11): 104 - 107.)
- [4] 曹树金,张学莲,陈忆金. 网络舆情意见挖掘中性词典构建和极性识别方法研究[J]. *图书情报知识*, 2012(1): 60 - 65. (Cao Shujin, Zhang Xuelian, Chen Yijin. A Study on the Polar Word Identification Method and Polarity Dictionary Construction of Online Public Opinion Mining [J]. *Document, Information & Knowledge*, 2012(1): 60 - 65.)
- [5] 段建勇,谢宇超,张梅. 基于句法语义的网络舆论情感倾向性评价技术研究[J]. *情报杂志*, 2012, 31(1): 147 - 150. (Duan Jianyong, Xie Yuchao, Zhang Mei. The Sentimental Analysis Technology for the Internet Public Opinion Using Syntax and Semantic Information[J]. *Journal of Intelligence*, 2012, 31(1): 147 - 150.)
- [6] 徐冰,赵铁军,王山雨,等. 基于浅层句法特征的评价对象抽取研究[J]. *自动化学报*, 2011, 37(10): 1241 - 1247. (Xu Bing, Zhao Tiejun, Wang Shanyu, et al. Extraction of Opinion Targets Based on Shallow Parsing Features [J]. *Acta Automatica Sinica*, 2011, 37(10): 1241 - 1247.)
- [7] 赵妍妍,秦兵,车万翔,等. 基于句法路径的情感评价单元识别[J]. *软件学报*, 2011, 22(5): 887 - 898. (Zhao Yanyan, Qin Bing, Che Wanxiang, et al. Appraisal Expression Recognition Based on Syntactic Path[J]. *Journal of Software*, 2011, 22(5): 887 - 898.)
- [8] Maki I, Vossen P. A Verb Lexicon Model for Deep Sentiment Analysis and Opinion Mining Applications [C]. In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA'11)*. Stroudsburg: Association for Computational Linguistics, 2011: 10 - 18.
- [9] Yang C C, Wong Y C, Wei C. Classifying Web Review Opinions for Consumer Product Analysis [C]. In: *Proceedings of the 11th International Conference on Electronic Commerce (ICEC'09)*. New York: ACM, 2009: 57 - 63.
- [10] Duric A, Song F. Feature Selection for Sentiment Analysis Based on Content and Syntax Models [C]. In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. 2011: 96 - 103.
- [11] 李寿山,黄居仁. 基于 Stacking 组合分类方法的中文情感分类研究[J]. *中文信息学报*, 2010, 24(5): 56 - 61. (Li Shoushan, Huang Churen. Chinese Sentiment Classification Based on Stacking Combination Method [J]. *Journal of Chinese Information Processing*, 2010, 24(5): 56 - 61.)
- [12] 谢丽星,周明,孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取[J]. *中文信息学报*, 2012, 26(1): 73 - 83. (Xie Lixing, Zhou Ming, Sun Maosong. Hierarchical Structure Based Hybrid Approach to Sentiment Analysis of Chinese Micro Blog and Its Feature Extraction [J]. *Journal of Chinese Information Processing*, 2012, 26(1): 73 - 83.)
- [13] 王波. 基于跨领域知识的基金评论情感分析[J]. *情报杂志*, 2011, 30(2): 44 - 47. (Wang Bo. Sentiment Analysis for Fund Reviews Based on Cross Domain Knowledge [J]. *Journal of Intelligence*, 2011, 30(2): 44 - 47.)
- [14] Liu Y, Yu X, An A, et al. Riding the Tide of Sentiment Change: Sentiment Analysis with Evolving Online Reviews [J]. *World Wide Web*, 2013, 16(4): 477 - 496.

(作者 E-mail: duanjy@hotmail.com)