

改进的基于语义理解的文本情感分类方法研究

王日宏¹ 崔兴梅¹ 周 炜¹ 王成龙² 李永珺¹

(青岛理工大学计算机工程学院 青岛 266033)¹

(中国联合网络通讯有限公司济南软件研究院 济南 250199)²

摘 要 文本分类在信息检索、Web文档自动分类、数字图书馆、自动文摘、文档的组织和管理等多个领域都有着广泛的应用。提出一种改进的基于语义理解的文本情感分类方法,在情感相似度计算中加入情感义原来重新修正定义,并综合情感短语倾向性的研究,侧重情感词与否定词、程度副词的组合形式分析,重点提出否定词程度副词综合处理模块。结合以连词为划分标准的语句情感倾向性分类处理,提出一种文本倾向度算法,以实现基于语义理解的文本情感分类。实验结果表明,与传统的语义理解算法相比,该方法的分类效果有了一定程度的提升。

关键词 改进语义理解,文本情感分类,否定词,程度副词,句子结构化模板

中图法分类号 TP391 文献标识码 A

Research of Text Sentiment Classification Based on Improved Semantic Comprehension

WANG Ri-hong¹ CUI Xing-mei¹ ZHOU Wei¹ WANG Cheng-long² LI Yong-jun¹

(Computer Engineering Institute, Qingdao University of Technology, Qingdao 266033, China)¹

(China UNICOM Software Research Institute Jinan Branch, Jinan 250199, China)²

Abstract Text classification has a wide range of applications in information retrieval, Web automatic document classification, digital library, automatic abstracting, document organization and management. An improved text sentiment classification method was put forward based on semantic understanding. Emotional sememe was joined to revise the definition in the emotional similarity calculation and the development of emotional phrases' sentiment was combined. Focusing on emotional words and negative words, the degree of adverbs combining form of analysis, and module complex negative word and adverb were put forward. Combining with the use of conjunctions as the standard of the sentence for emotional tendencies classified processing, a text propensity algorithm was given to judge the text sentiment and classify the text. Experimental results show that the classification results with the method improve when comparing with the previous algorithm.

Keywords Improved semantic comprehension, Text sentiment classification, Negative words, Adverb of degree, Sentence structure template

1 引言

第三十八次中国互联网发展状况统计报告显示,中国网民达7.10亿,互联网普及率达51.7%^[1]。随着互联网的普及,网络作为信息传播和获取的主要平台,承载了信息交流的主要工作。人们可以在论坛、博客等社交网站随意发表个人观点和意见,而通过对文本倾向性的判断可以处理这些观点及意见,并分析其情感等,因此文本情感分类得到了广泛的应用。

国内外诸多学者对词语、语句相似度计算做出了一系列研究。从20世纪80年代开始,R. Pada首先基于WordNet对英文词语相似度计算进行了研究,但并未将语义树和语义

距离中概念节点之外的因素纳入计算,因此该词语相似度计算方法仅仅考虑了语义距离^[2]。P. Resnik于90年代引入了信息量这一因子来对R. Rada的研究进行了改进,由于概念节点和祖先节点之间存在信息量差值,因此运用差值进行计算从而得到词语相似度^[3]。随后,Rigau和Agirre等人综合前人的研究结论,提出了一种综合语义距离、语义树深度、概念节点信息量的词语相似度算法^[4]。Lillian^[5]主张如果上下文存在语义相近的词,那么这些语义相近的词也应该是相似的,我们称其为基于相关熵的计算方法,以此得出的词语相似度可以借助词典和统计计算模型来判断文本倾向性,可以利用抽取的短语倾向来判定^[6],也可以人工训练机器学习语料来进行判定^[7],但准确率较高的是支持向量机的方法。在中

本文受国家自然科学基金项目(61502262),山东省研究生教育创新计划项目(SDY16023)资助。

王日宏(1964—),男,教授,主要研究方向为智能信息处理、数据挖掘,E-mail: rihongw@126.com;崔兴梅(1991—),女,硕士生,主要研究方向为智能信息处理;周 炜(1981—),男,博士,讲师,主要研究方向为网络信息安全;王成龙(1990—),男,硕士,工程师,主要研究方向为机器学习、数据挖掘;李永珺(1994—),硕士生,主要研究方向为智能信息处理。

文研究方面,张积家最早提出的词语相似度计算是通过引入字符相似度得出的^[8]。关毅等人通过对实词的研究,将汉语词汇相似度计算与统计数据联系起来,取得了良好的效果^[9]。章成志则对词素进行了改进,从而得到一种相似度算法^[10]。田久乐等通过研究《同义词语林》进而发现了中文词语相似度的计算方法^[11]。闻彬等赋予概念情感语义,进而有效判断文本的情感倾向性^[12]。马力等采用向量空间模型的方法,提出了支持向量机的文本情感倾向分析方法^[13]。徐健锋等结合机器学习的方法,提出了一种混合中文情感分类算法^[14]。中文文本倾向性研究起步较晚,尚存在不足之处。

基于机器学习的方法和基于语义的方法是最常用的两种文本分类方法。基于机器学习的方法利用标注好的训练集通过成熟的机器学习算法来构建分类器,将其作为一个基准用于以后对测试集进行分类,但是其不能很好地处理文本的语义信息;基于语义理解的方法以词语相似度为基础来计算褒义、贬义种子词从而得到词语情感极性值,并整合短语倾向、句子结构以更全面地得到文本的情感倾向,本文基于此提出一种改进的语义理解文本情感分类方法。

2 引入情感义原来计算词语情感相似度

2.1 《知网》语义相似度计算

对于两个词语 W_1 和 W_2 ,假设词语 W_1 有 n 个义项 Y_1, Y_2, \dots, Y_n ,词语 W_2 有 m 个义项 Z_1, Z_2, \dots, Z_m ,则可以通过义原的相似度来计算词语的相似度,其计算公式如下:

$$Sim(W_1, W_2) = \max_{i=1,2,\dots,n, j=1,2,\dots,m} Sim(Y_i, Z_j) \quad (1)$$

2.2 义原相似度计算

义原是概念的基本单位,也是一切计算的基础。语义距离有助于义原的计算。语义距离是一种最短路径,该路径用于连接实体义原树中的节点。语义距离副作用于相似度。假设语义距离为 d ,则其义原相似度为:

$$Sim(p_1, p_2) = \frac{\partial}{d + \partial} \quad (2)$$

其中, P_1 和 P_2 表示义原, d 是路径距离, ∂ 是一个调节参数。式(2)只利用路径距离来计算相似度,但是义原之间的关系众多,若将其都考虑进来则能够使义原相似度的计算更为准确,计算公式如下:

$$Sim(P_1, P_2) = \frac{\partial \theta_1 \theta_2}{d + \partial} \quad (3)$$

其中, θ_1, θ_2 为义原 P_1 和 P_2 在义原树中的深度。

2.3 概念义原相似度计算

2.3.1 虚词概念的相似度计算

实词和虚词是两类毫不相干的词语,因此可以认定他们之间的相似度为零。虚词不表示实在的意义,对于虚词而言,相似度的计算比较简单,只需与关系义原和句法义原对应即可。

2.3.2 实词概念的相似度计算

实词相似度的计算较复杂,其中涉及语句环境的影响。其基本思想与模块化处理的思想基本一致:将整体拆分为几个模块,通过对各个模块之间的相似度计算结果进行汇总从而得到整体相似度。整体相似度可以记为:

$$Sim(S_1, S_2) = \sum_{i=1}^4 \beta_i Sim_i(S_1, S_2) \quad (4)$$

通过观察发现, Sim_1 和 Sim_4 的取值在很大程度上影响了相似度计算的准确性,因此对公式进行了改进,得到:

$$Sim(S_1, S_2) = \sum_{i=1}^4 \beta_i \prod_{j=1}^i Sim_j(S_1, S_2) \quad (5)$$

仅仅考虑第一义原对其他义原之间的相互制约关系,是与《知网》的描述相悖的。根据《知网》的描述,除了第一独立义原相似度之外的其他义原相似度也应该是相互独立的,因此游春晖^[15]提出独立计算义原相似度的计算公式:

$$SimS(s_1, s_2) = \beta_1 SimP_1(p_1, p_2) + \sum_{i=2}^4 \beta_i SimP_1(p_1, p_2) \times \beta_i P_i(p_1, p_2) \quad (6)$$

上述公式都是在概念的意义上考虑词语的可替换程度,并未考虑词语的情感意义,因此本文在计算情感相似度时引入了情感义原 $Emo(Y, Z)$,概念 Y, Z 的相似度计算如下:

$$SimS(Y, Z) = \sigma_1 (\beta_1 SimP_1(p_1, p_2) + \sum_{i=2}^4 \beta_i SimP_1(p_1, p_2) \times \beta_i P_i(p_1, p_2)) + \sigma_2 Emo(Y, Z) \quad (7)$$

引入情感义原来对 $Sim_2(S_1, S_2), Sim_3(S_1, S_2), Sim_4(S_1, S_2)$ 进行修正。当情感义原相同时,该值被设定为 1;否则被设定为 0。根据经验并结合研究结果,参数 σ_1 和 σ_2 分别取 0.7 和 0.3 时实验效果最佳。实验数据详见第 3 节。

2.4 词语情感倾向度计算

根据相似度可得情感值,知网中的情感词并不涉及情感强度的计算。经研究发现,使用基于语义相似度的方法对情感词典中词语的情感权重进行计算可以得到较高的准确率,即将情感词与基准词配对后进行计算^[16-17]。假设种子集 $Seedsets = \{PO, NG\}$,其中, PO 指褒义基准词集, NG 为贬义基准词集,褒义贬义基准集是从 HowNet 中人工挑选的,则词 w 的情感权值为:

$$o(w) = \frac{1}{M} \sum_{i=1}^M sim(w, po_i) - \frac{1}{N} \sum_{j=1}^N sim(w, ng_j) \quad (8)$$

其中, $po_i \in PO, ng_j \in NG, M$ 和 N 分别为褒贬基准词的个数。设置数字 0 为阈值, $O(w) > 0$ 表示该词为褒义, $O(w) < 0$ 表示该词为贬义,而 $O(w)$ 的数值则表示该词的褒贬倾向度。

3 情感词句处理

文本情感倾向的判定不能单纯地依靠情感词语,带有感情倾向的短语也是研究的内容之一,因为情感词汇与不同的词汇搭配而出现的情感倾向是截然不同的,所以本文在研究情感词情感倾向的基础上侧重研究情感词与其他词的组合形式,比如否定词、连词、程度副词,它们通过修饰目标词而形成的短语是对句子进行情感分析的基础。

3.1 情感词的分类定义

情感词是能够表达情感的一类词语,但同样的情感词又可以划分为不同的种类,其具体分类如表 1 所列。

表 1 所列为句子处理中常见的两类情感词,只有准确判断动态情感词的情感倾向才能准确地计算句子的相似度。例如:

句 1 “这辆车的悬挂真软”

句 2 “沙发真软啊”

句 1 中“软”表示贬义,句 2 中“软”则表示褒义。难以判

表 3 程度副词表

最高 (2)	最、最为、太、极、极为、极其、极度、极端、至、至为、顶、过、过于、过分、分外、万分
次高 (1.5)	更、更加、更为、更其、越、越发、倍加、愈、愈发、愈加、愈为、愈益、越加、格外、益发、还、很、挺、怪、老、非常、特别、相当、十分、好、不好、甚、甚为、颇、颇为、异常、深为、满、蛮、够、多、多么、殊、特、大、大为、何等、何其、尤其、无比、尤为、不胜
中 (1)	比较、较、较比、较为、不大、不太、不很、不甚
低 (0.5)	稍微、稍、稍稍、稍为、稍许、略、略略、略微、些微、多少、有点、有些

通过依存关系抽取副词,对其机制进行分析并在后文提出该算法的相关改进,采用哈尔滨工业大学的技术平台来分析依存关系,如图 1 所示。依存关系的种类如表 4 所列。

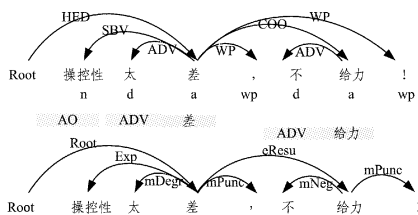


图 1 句子成分的依存关系示意图

表 4 Deparser 词性和依存关系标记注释表

依存关系	标记	依存关系	标记
定中关系	ATT	连动结构	VV
数量关系	QUN	同位关系	APP
并列关系	COO	前附加关系	LAD
后附加关系	RAD	动宾关系	VOB
介宾关系	POB	主谓关系	SBV
比拟关系	SIM	核心	HED
关联结构	CNJ	语态结构	MT
独立关系	IS	状中关系	ADV
动补关系	CMP	“的”字结构	DE
“地”字结构	DI	“得”字结构	DEI
“把”字结构	BA	“被”字结构	BEI
独立分句	IC	依存分句	DC

“不是很好”这个例子反映了程度副词与否定词一般会配对出现,这增加了处理的难度,本文提出了否定词程度副词混合处理模块来处理句子中否定词和程度副词共同出现的情况。

否定词和情感词组合(“不好”)带来的情感强度应该比单纯情感词(“差”)的强度弱,因此可以将否定词看作强度是 $-θ$ ($0 \leq \theta \leq 1$)的副词。结合否定词提出的窗口理论来处理与情感词的搭配问题。

本文从以上 24 种关系中选择比较有代表性的几种进行分析,如表 5 所列。

表 5 关系模式分析

模式	举例说明	原倾向	修正后倾向
ADV 状中关系	操作性太差	-1	-2
CMP 动补关系	差的很	-1	-1.5
VOB 动宾关系	不是很好	1	1.5
ATT 定中关系	非常好	1	1.5

采用窗口理论来处理与情感词搭配问题的中心思想为:以情感词为中心向外分散搜索,检索在窗口大小范围内的否定词与程度副词,遇到“,”“;”“。”等标点时停止查找。拟解决的问题如下:

(1)否定词、程度副词修饰情感词的顺序不同,比如“不很漂亮”“很不漂亮”等。

(2)特殊修饰词的处理。“偏”“太过”等通常会改变句子的倾向,比如“室内的温度太过微暖”“刹车踏板灵敏度偏高”等。其处理方法如下:若否定词与情感词的距离比与程度副词的距离近,则依据式(15)进行计算。

$$S = S_{Emotional\ Words} \times (-\theta) / L_{adv} \tag{15}$$

若程度副词与情感词的距离比与否定词的距离近,则依据式(16)进行计算。

$$S = S_{Emotional\ Words} \times (-\theta) / L_{adv} \tag{16}$$

其中, L_{adv} 为程度副词的强度, $S_{Emotional\ Words}$ 为情感词汇的倾向度。对于“偏”“太过”这类词汇,需要人工总结来汇成字典,进而加入判断过程。

3.5 语句情感倾向的研究

采用模块化的思想将语句拆分为情感词模块和短语模块,使用模块化的思想求解,具体的计算公式如式(17)所示。

$$EmoSent(sentence) = \sum_{i=1}^n [SO(phrase_i) \times W(phrase_i)] \tag{17}$$

其中, $SO(phrase_i)$ 为短语 $phrase_i$ 的情感倾向值; $W(phrase_i)$ 为短语 $phrase_i$ 在句子 $sentence$ 中的权值,满足 $\sum_{i=1}^n W(phrase_i) = 1$,且 $W(phrase_i)$ 取值不均等。这里未考虑并列句这类的复合句,仅是对简单句的计算。因此应对含有转折等连词的复合句建立句子模板(见表 6),并计算情感词的权重 $W(phrase_i)$ 。语句情感倾向的计算流程如图 2 所示。

表 6 句子结构化模板

模板编号	句子结构	权重确定比例	说明
1	、	各情感词均等划分权重	表示并列、递进
2	和	各情感词均等划分权重	
3	并且	各情感词均等划分权重	
4	也	各情感词均等划分权重	
5	而且	各情感词均等划分权重	
6	既...又...	各情感词均等划分权重	表示转折
7	不但...而且...	各情感词均等划分权重	
8	...,相反的...	1:2	
9	却	1:2	
10	但是	1:2	
11	不过	1:2	
12	虽然...但是...	1:2	
13	...另一方面...	各情感词均等划分权重	

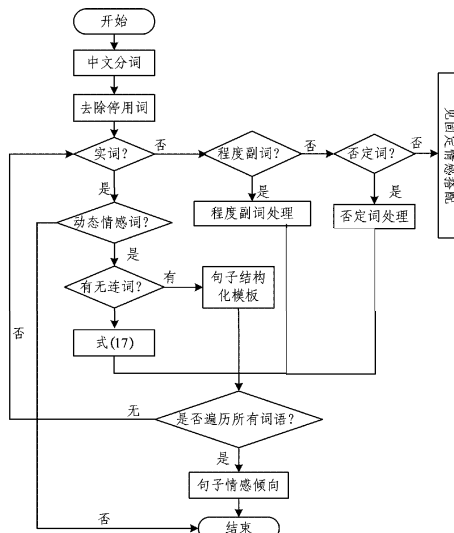


图 2 语句情感倾向的计算流程图

3.6 文本倾向度算法

Step1 判断句子结构,以连词为划分复合句与简单句的标准,判断句子中是否存在连词。若有,则查询连词词典表,去除连词并将其压入连词堆栈,参照句子结构化模板计算语句倾向度;若无,则表示该句子为简单句,按式(17)进行处理。

Step2 若句子是复合句,则按照模块化思想将其拆分为简单句后再进行处理。分词后直接按照 3.4 节给出的计算各类词汇的方法计算情感倾向值。

Step3 结合情感词、否定词、程度副词、句子结构化模板计算语句情感倾向值。

上述步骤的具体算法如下。

算法 1 文本倾向度算法 tendency_Appraise

Input: 语句分词后的词语集合

Output: 语句情感倾向度值

Step1 对分词处理后的所有词语进行编号处理。

Step2 遍历词语集合中的所有词语,若有连词,则记录,将复合句以连词为划分标准划分为 n 个简单句,并对简单句进行处理;若无连词,则直接转到 Step3。

Step3 遍历所有词语,若是名词、代词、量词、数词,则不必判断,直接获得下一个词语;若是情感词,则查询情感词词典,将情感词存储到情感词堆栈,按照否定词程度副词对情感词的处理方法,再查找情感词附近的否定词和程度副词并作相应处理后,压入否定词程度副词堆栈。

Step4 分别弹出情感词、否定词程度副词堆栈,计算语句情感倾向度值。

Step5 重复 Step3,直到所有标记的词汇都已经处理完成。

Step6 查看连词堆栈,若为空,则处理完毕;否则查询连词词典和句子结构化模板修正语句的情感倾向度值。最终得到 EmoSim (sentence)语句的倾向度值。

4 实验数据及分析

实验 1、实验 2 中的语料取自于互联网评论,主题涉及体育、汽车、经济等,包含 1569 个句子,3212 个情感词,3008 个静态情感词,去重后得到 318 个情感词,204 个动态情感词。

实验 1 证明当参数值 l_1, l_2 设定为 0.7 和 0.3 时能获得比较理想的效果。表 7 和表 8 的实验数据证明了引入情感义原并设定参数值的可行性。

表 7 l_1, l_2 的取值对准确率的影响

l_1	l_2	褒义词			贬义词		
		查全率	查准率	F-测量值	查全率	查准率	F-测量值
0.25	0.75	0.861	0.719	0.783	0.877	0.708	0.783
0.30	0.70	0.828	0.741	0.782	0.861	0.726	0.787
0.40	0.60	0.843	0.744	0.791	0.881	0.759	0.811
0.50	0.50	0.833	0.804	0.819	0.831	0.750	0.787
0.60	0.40	0.822	0.877	0.849	0.846	0.810	0.828
0.70	0.30	0.812	0.887	0.869	0.869	0.792	0.836
0.75	0.25	0.822	0.877	0.849	0.846	0.810	0.828

表 8 实验数据集

l_1	l_2	体育			汽车			经济		
		贬义	褒义	平均值	贬义	褒义	平均值	贬义	褒义	平均值
0.25	0.75	0.854	0.812	0.833	0.752	0.712	0.732	0.934	0.752	0.843
0.30	0.70	0.856	0.812	0.834	0.762	0.717	0.739	0.939	0.761	0.850
0.40	0.60	0.861	0.821	0.841	0.767	0.726	0.747	0.941	0.771	0.856
0.50	0.50	0.866	0.824	0.845	0.782	0.731	0.757	0.948	0.779	0.864
0.60	0.40	0.874	0.831	0.853	0.789	0.745	0.767	0.951	0.782	0.867
0.70	0.30	0.882	0.841	0.862	0.803	0.752	0.778	0.959	0.793	0.876
0.75	0.25	0.874	0.831	0.853	0.793	0.742	0.768	0.949	0.781	0.865

实验 2 情感词实验

204 个动态情感词占总情感词的 6%,但是动态情感词对分析句子倾向具有重大的意义,实验数据如表 9 所列,它含有否定词短语的准确率分析。

表 9 动态情感词判断准确率

动态情感词数目	正确判断个数	准确率/%
204	194	95.1

对上述语料进行筛选,得到“情感词+否定词”组合共 1023 个,占情感词总数的 34%,可见否定词的处理对句子倾向的影响较大,实验数据如表 10 所列。短语综合倾向性准确率比较(如综合情感词、副词、否定词等)的实验数据如图 3 所示。

表 10 否定词+情感词的准确率分析

否定词+情感词数目	正确判断个数	准确率/%
1023	982	96.0

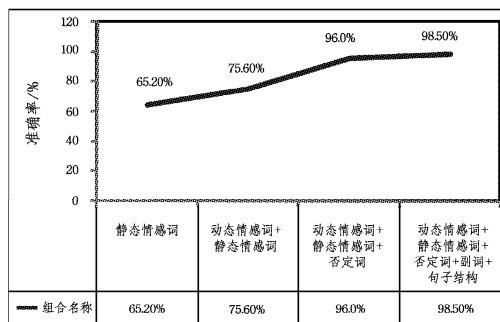


图 3 各类短语组合的倾向性比较

实验 3 改进的语义相似度准确率分析

选取文献[21]的测试语料(谭松波团队整理的关于酒店评论的语料)进行实验。从已经标注的酒店评论文本中选取 900 个文本,褒义、贬义平均分配,在不交叉使用文本的前提

下提取 3 组共 900 个文本进行实验,如表 11 所列。

表 11 实验数据集

组别	文本总数	褒义文本数	贬义文本数
1	300	150	150
2	300	150	150
3	300	150	150

本文方法主要与传统语义理解、原始的语义相似度计算公式、文献[20]以及文献[21]中的基于《知网》改进的倾向性算法进行比较,比较的参数主要是查全率、查准率、F-测量值,实验数据如表 12 所列。实验表明,本文方法比传统的语义理解方法在 3 个指标上均有提高,准确性也更高。

表 12 语义理解对比实验/%

方法	评估标准	第一组	第二组	第三组	平均值
传统语义理解	查准率	69.48	67.68	74.65	70.60
	查全率	57.00	56.50	59.00	57.50
	F-测量值	62.62	61.59	65.91	63.37
原始语义理解	查准率	75.35	69.60	75.32	73.42
	查全率	72.50	67.50	73.50	71.17
	F-测量值	73.90	68.53	74.40	72.28
文献[20]改进算法	查准率	76.58	70.34	75.87	74.26
	查全率	74.21	68.70	72.45	71.79
	F-测量值	75.38	69.51	75.65	73.51
文献[21]改进算法	查准率	78.95	71.67	77.07	75.90
	查全率	76.00	69.50	75.50	73.67
	F-测量值	77.45	70.57	76.29	74.77
本文方法	查准率	79.36	72.63	79.36	76.32
	查全率	78.10	69.89	76.98	74.00
	F-测量值	78.96	70.23	77.63	74.87

结束语 本文改进了以词语相似度为基础的语义相似度计算方法,提出了一种改进的基于语义理解的文本情感分类方法来判定文本的情感倾向性。实验分析证明了该方法的有效性。否定词处理算法和文本倾向度算法没有对复杂并列句进行进一步的研究,这将是今后的重点。

参 考 文 献

- [1] 中国互联网信息中心. 第 38 次中国互联网络发展状况统计报告 [EB/OL]. [2016-12-30]. http://www.cnnic.net.cn/hlwfzjy/hlwxbzg/hlwtjbg/201608/t20160803_54392.htm.
- [2] RADA R, MILI H, BICKNELL E, et al. Development and Application of a Metric on Semantic Nets[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1989, 19(1): 17-30.
- [3] RESNIK P. Using information content to evaluate semantic similarity in a taxonomy[C]//Proceeding of the 14th International Joint Conference on Artificial Intelligence. 1995: 448-453.
- [4] RIGAU G, AGIRRE E. A proposal for word sense disambiguation using conceptual distance[C]//International Conference/Recent Advances in Natural Language Processing. 1995: 35-43.
- [5] LEE L J. Similarity-Based Approaches to Natural Language Processing[J/OL]. arXiv: com-lg/9708011.
- [6] TURNEY P D. Thumbs up or thumbs down? Sem-antic orientation applied to unsupervised classification of reviews[C]//Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL). Philadelphia, 2002: 417-424.
- [7] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing(EMNLP). 2002: 79-86.
- [8] 张积家. 测量语义相似性的方法[J]. 心理科学, 1992(3): 53-55, 62.
- [9] 关毅, 王晓龙. 基于统计的汉语词汇间语义相似度计算[C]//语言计算与基于内容的文本处理—全国第七届计算语言学联合学术会议论文集. 2003: 221-227.
- [10] 章成志. 一种基于语义体系的同义词识别研究[J]. 淮阴工学院学报, 2004, 13(1): 59-62, 67.
- [11] 田久乐, 赵蔚. 基于同义词词林的词语相似度计算方法[J]. 吉林大学学报(信息科学版), 2010, 28(6): 602-608.
- [12] 闻彬, 何婷婷, 罗乐, 等. 基于语义理解的文本情感分类方法研究[J]. 计算机科学, 2010, 37(6): 261-264.
- [13] 马力, 刘笑, 宫玉龙. 基于语义的微博短文本倾向性分析研究[J]. 计算机应用研究, 2016, 33(10): 2914-2918.
- [14] 徐健锋, 许园, 许元辰, 等. 基于语义理解和机器学习的混合的中文文本情感分类算法框架[J]. 计算机科学, 2015, 42(6): 61-66.
- [15] 游春晖. 基于语义情感倾向的文本相似度计算[D]. 成都: 电子科技大学, 2008.
- [16] 吴全娥. 汉语句子相似度计算及其在自动问答系统中的应用[D]. 重庆: 西南大学, 2011.
- [17] 闻彬, 何婷婷, 罗乐, 等. 基于语义理解的文本情感分类方法研究[J]. 计算机科学, 2010, 37(6): 261-264.
- [18] HATZIVASSILOGLOU V, MCKEOWN K. Predicting the semantic orientation of adjectives[C]//Proceedings of the 35th Annual Meeting of Association for Computational Linguistics (ACL-97). Madrid, Spain, 1997: 174-181.
- [19] 蔺璜, 郭妹慧. 程度副词的特点范围与分类[J]. 山西大学学报(哲学社会科学版), 2003, 26(2): 71-74.
- [20] 党蕾, 张蕾. 一种基于知网的中文句子情感倾向判别方法[J]. 计算机应用研究, 2010, 27(4): 1370-1372.
- [21] 许元辰. 基于优化的语义理解与 SVM 相结合的文本情感分类研究[D]. 南昌: 南昌大学, 2014.
- [12] 陈玲燕. 多重共线性下的线性回归方法综述[J]. 现代农业, 2008(5): 67-69.
- [13] HOERL A E, KENNARD R W. Ridge regression: biased estimation for nonorthogonal problems[J]. Technometrics, 1970, 42(1): 80-86.
- [14] 张丹平. 基于岭回归方法的我国能源消费影响因素研究[J]. 统计与决策, 2012(21): 146-148.

(上接第 64 页)

- [9] LIU H H, JI W L, ZHANG P, et al. The Research of Wine Quality Evaluation Based on Multiple Linear Regression[J]. Advanced Materials Research, 2013, 756-759: 2489-2493.
- [10] 崔江龙. 多元回归分析在能源利用中的应用[J]. 商, 2015(49): 68.
- [11] 付倩晓. 基于多元线性回归的雾霾预测方法研究[J]. 计算机科学, 2016, 43(s1): 526-528.