

基于购物领域词典扩建的评论情感研究

吴 潇¹, 王 磊^{1,2}

(1. 南京邮电大学 宽带无线通信与传感网技术教育部重点实验室, 江苏 南京 210003;
2. 东南大学 移动通信国家重点实验室, 江苏 南京 210096)

摘 要:针对购物评论中如何高效提取有用的情感信息,提出了构建领域情感词典进行评论情感分类方法。对购物评论语料进行分词去重,就各领域评论文本进行词性标注,选择词性为名词、形容词及部分其他词性的词语,通过计算该部分词语的 PTF-IDF 进行排序,设置阈值筛选后得到购物评论语料的领域情感词,从而构建领域情感词典。将该词典作为情感特征应用于购物评论情感分类实验中,并与基于普通情感词典分类方法的性能进行了分析比较。实验结果表明,利用提出方法进行购物评论情感分类的效果,尤其是在分类准确率方面要明显高于基于普通情感词典的情感分类方法,且所提出的方法可适用于各领域的购物评论,有效降低了情感特征空间的维度,具有普适性和可扩展性等优点。

关键词:购物评论情感研究;情感分类;领域情感词典;情感特征

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2017)07-0194-06

doi:10.3969/j.issn.1673-629X.2017.07.043

Investigation on Sentiment of Reviews with Shopping Field Dictionary Construction

WU Xiao¹, WANG Lei^{1,2}

(1. Key Lab of Broadband Wireless Communication and Sensor Network Technology of MoE, Nanjing University of Posts & Telecommunications, Nanjing 210003, China;
2. National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China)

Abstract: Aiming at effectively extracting sentimental information from the shopping reviews, a sentiment classification method has been proposed for online shopping reviews based on the construction of domain emotional dictionary, which segments words from the reviews corpus, nouns, adjectives and words of some other lexical category in the reviews after part of speech tagging. Sorting by calculating the PTF-IDF of words and setting threshold, selection of domain emotional words has been obtained to construct the domain emotional dictionary which has been applied to the sentiment classification of shopping reviews compared with the performance of the classification method based on the general sentiment dictionary. The experimental results show that the proposed method of shopping review sentiment classification is better than one based on general sentiment dictionary, and it can be applied to various fields for effectively reducing the dimension of the feature space, with universality and extensibility.

Key words: research on sentiment of shopping reviews; sentiment classification; domain emotional dictionary; emotional characteristics

1 概述

随着互联网的蓬勃发展,电子商务的兴起吸引了越来越多的用户开始在网上购物,体验网上购物所带来的“足不出户”和“物美价廉”。与此同时,用户也通过对购买的商品进行评论,表达对产品的主观看法和意见。然而,由于网上购物没有地域的限制,在为用户

带来便利的同时也导致用户无法直接触摸和具体了解商品的质量,可能造成网上商城对商品的描述与实际有区别,给用户带来不便。用户只有通过了解已购顾客对商品的评论,做出相对可靠的决策。因此,为了使用户能更加方便快捷地找出富价值的评论,浏览自己想要的信息,为了促进商家提高用户对网上购物服

收稿日期:2016-05-23

修回日期:2016-08-30

网络出版时间:2017-06-05

基金项目:国家自然科学基金资助项目(61201164, 61322104, 61201165, 61271240);南京邮电大学宽带无线通信与传感网技术教育部重点实验室开放研究基金(NYKL201509);东南大学移动通信国家重点实验室开放研究基金(2016D01)

作者简介:吴 潇(1992-),女,硕士研究生,研究方向为文本情感分析、机器学习等;王 磊,博士,副教授,硕士研究生导师,研究方向为现代通信中的智能信号与信息处理、随机矩阵理论与无线通信、认知无线电等。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20170605.1506.022.html>

务的满意度,对购物评论进行情感分类就显得十分重要。

对购物评论进行情感分类,就是按照评论文本所表达的情感倾向进行分析、处理、归纳和推理^[1],判别评论中用户想要表达的观点^[2]、喜好、感受以及对商品或者商家服务的态度^[3],进而为用户提供更有效和更可靠的商品信息,辅助用户做出合理的决策,提高网上购物效率和服务质量。目前,针对评论文本情感分类的研究主要是将购物评论分为两类,即正向情感评论和负向情感评论。也有部分研究是将其分为三类,即正向情感评论、负向情感评论和中性情感评论。购物评论的情感分类研究属于文本情感分类研究的一个分支。主要针对网上商城的购物评论进行情感分类研究。

目前,在对评论文本进行情感倾向性分析方面,国内外研究采用的技术主要分为两大类:基于机器学习的方法和基于情感词典或者语义知识的方法。其中,使用情感词典的方法^[4]是通过借助情感词典查找并统计待分类评论文本中正面情感词和负面情感词的数量,计算其数量的差值来达到分类评论文本情感倾向性的目的,若差值为正则判定为正面情感,反之若差值为负即为负面情感,差值若等于零则视为中性情感。基于情感词典的方法能取得较好的分类准确性,但其局限性在于无法识别未登录词语,例如领域情感词语和网络流行词语等。

张成功等^[5]通过构建一个包含基础词典、网络用语词典、领域词词典和修饰词词典在内的较为全面高效的情感极性词典,并深入分析了修饰词对情感词的影响,将修饰词与情感极性词组合成情感极性短语,作为计算文本情感极性的基本单元。实验结果表明,这种基于情感极性词典的文本分类方法效果不错,但其分析文本感情时过多地依赖于现有词典,一旦文本中不存在情感词典中的词语,就无法判断该文本的感情倾向与情感强度。利用知网提供的情感词典作为基础词典^[6],并构建3类模型用以扩展该基础词典。但在此过程中没有考虑到领域情感词的作用以及修饰词对情感词的影响。通过整合知网情感词典和 NTUSD 词典,扩展为一个大型情感词典用以对文本进行分析^[7],并构造了修饰词词典,但并未深入研究在待分析文本中修饰词对情感词的影响。研究了如何用领域情感词典扩展构建通用情感词典^[8],但其忽视了基础情感词典对文本分类的重要性。

采用机器学习的方法则是将语料标注为训练语料和测试语料,通过使用最大熵、支持向量机、朴素贝叶斯等分类器对评论文本进行情感倾向性分类。Pang 等^[9]主要是采用对比实验对电影评论进行情感倾向性

分类,分为正面情感和负面情感。其研究工作首先对待分析的评论文本进行预处理,提取相关特征,然后使用这些特征训练包括最大熵、支持向量机和朴素贝叶斯在内的三种机器学习模型。其实验结果表明,使用支持向量机能取得较为理想的分类效果。Yu 等^[10]选择具有情感倾向性的词语作为特征,并采用朴素贝叶斯分类器对篇章级的文本进行情感分类以及主客观分类。樊康新^[11]采用基础情感词典提取情感特征,代入支持向量机算法,对网络文本分类进行分析和研究。蒋婉婷等^[12]使用 MapReduce 分布式朴素贝叶斯算法对中文文本分类器进行训练和测试。

针对购物评论中如何高效提取有用的情感信息,提出了构建领域情感词典进行评论情感分类的方法。对现有的情感词典资源进行整合,构建了基础情感词典。在分析网络流行词语对情感分析的影响并进一步扩建基础情感词典。通过研究领域情感词的特殊性,即词性不同的词语在各领域购物评论中所表达的情感与目的存在差异的特性,使用 PTF-IDF 方法构建购物评论领域情感词典。以此为基础,设计了两组实验对上述方法进行验证。结果表明,提出的情感分类方法能够很好地提取情感特征,并取得了较高的分类准确率,可适用于各领域的购物评论情感分类。

2 基础情感词典构建

通常,能够显示情感的词语会被选作分类特征用于对评论的情感倾向进行分类。因此,构建一个普遍适用的基础情感词典十分重要,为提取特征打下坚实的基础^[13]。目前,用于研究评论情感倾向的词典主要有:知网情感词典,包含 836 个正向情感词语和 900 个负向情感词语;由台湾大学自然语言处理实验室建立的简体中文情感词典:NTUSD (National Taiwan University Sentiment Dictionary)。由于知网情感词典所提供的正向词语和负向词语的数目相对均衡,而台湾大学提出的情感词典中褒义词语的数量远低于贬义词语的数量,存在非均衡性,因此采用知网情感词典作为基础情感词典的候选词典并进行相应改善。具体过程如图 1 所示。



图1 基础情感词典的构建方法

基础情感词典是分析评论情感倾向性的基础,采用的方法是在知网情感词典的基础上,通过机器和手工结合的方式,选取日常生活中用户使用频率较多的、具有相对明显情感倾向性的词语。首先,把词典中包括“的”“地”“得”后缀的词与前面的词语合并。例

如:原本的知网情感词典中“感激”和“感激的”“感激地”“感激得”是 4 个词,在中文语法中这 4 个词都表示感激的意思,只是在句子中的词性不同。现在经处理统一删除其后缀含“的”“地”和“得”的词语,只保留“感激”一词。然后,对整理后知网情感词典中的情感词语进行筛选,剔除不常用的冷僻词汇。

Google 浏览器在进行搜索时会把出现在一个网页信息中的一系列关键字存储进命中列表 Hit Lists 中,这其中还包括这些关键字在网页中出现的位置,关键字的字体大小以及字母大小写等内容^[14]。因此,对每个词语依照 Google 返回的 Hits 值从大到小排序,去除 Hits 值低于阈值即应用效率相对较低的词语,从而得到初始情感词典。

3 添加网络词扩展词典

知网情感词典具有时效性,其包含的情感词语一般适用于对传统的情感词语分析,然而由于购物评论中出现的用词用语不规范,以及近年来大量涌现的网络流行用语,这种基于传统情感词语的评论情感分析常常不能取得较好的效果。为了能够较为合理地完善对基础情感词典的扩展,并充分考虑当下流行的网络用语对分析评论情感倾向性的影响^[15],通过搜狐、网易、新浪及腾讯这四个较为流行的大型中文微博服务网站中人工抽取约 200 个使用频率较高并且含有较明显情感倾向性的网络用语,例如:“腹黑”、“尼玛”、“伤不起”、“稀饭”、“何弃疗”等等。再经过人工标注和多次仔细校对后,将这些网络流行用语分为正向情感词语和负向情感词语,分别加入基础情感词典中。

4 构建购物评论领域情感词典

在实验中,首先从三个主流的网上购物商城(携程网、当当网、亚马逊网)分别下载了关于电脑、书本和酒店的评论语料,经过预处理整理后得到 6 000 条评论语料集,将使用该语料集作为实验的研究对象。

4.1 购物评论候选情感词集的产生

该购物评论语料集在经过分词处理后,去除重复的词语,产生一个评论词集。然后,对每条购物评论进行词性标注处理,可以得到共 28 种词性的词语,例如:i 指成语,取自英文单词 idiom(其中文意思为成语)的第 1 个字母。具体词性标注内容如表 1 所示。

在研究过程中可以发现,诸如:“我们”、“这个”、“那个”和“了们”等等不存在实际意义的词语对情感分类的影响甚微,因此对这些词语进行滤除。由文献[16]可知,在对中文文本的语法研究中,词性为形容词和动词的词语通常用于描述主语或宾语,带有一定的情感倾向。因此,在实验中,抽取评论语料中词性为

习用词、副词和形容词的词语进行保留。同时,名词、数词、介词、助词等词性的词语基本不具备情感倾向,因此对这些词语也进行去除。最终得到一个购物评论候选情感词集。

表 1 中文词性标注

标记	词性	标记	词性
a:	形容词	ni:	机构名
b:	区别词	nl:	处所名词
c:	连词	ns:	地名
d:	副词	nt:	时间词
e:	叹词	nz:	其他专名
g:	语素字	o:	拟声词
h:	前接成分	p:	介词
i:	习用语	q:	量词
j:	简称	r:	代词
k:	后接成分	u:	助词
m:	数词	v:	动词
n:	普通名词	wp:	标点符号
nd:	方位名词	ws:	字符串
nh:	人名	x:	非语素字

例如:评论文本“定了豪华大床房,可惜到了的时候,由于都满只有 3 层的,虽然视野不够开阔,还好景色很美。服务也很好,服务员都会主动问好。打电话要风油精也很快送到。中餐厅太小很吵,还要跑去点菜,汤味道不错,其他很一般。花园风景非常漂亮,游泳池很晚都有人在游泳。网上介绍有烧烤,结果根本没有。晕。总的来说非常满意,出去旅游了三次,这次是最满意的,兴奋不已,说下次还想来”。经过词性标注后可知“问好”、“出去”、“满意”等词语的词性为动词,而“好”、“漂亮”、“豪华”等词语的词性为形容词。显然,“了”、“中餐厅”、“风油精”等词性为助词、名词以及标点符号“,”不具有情感倾向,因此进行去除。而“还好”,“开阔”,“可惜”等词语极有可能带有情感倾向,从而保留至购物评论候选情感词集中。

4.2 购物评论领域情感词的选择

选择合理的领域情感词对于构建领域情感词典十分重要。由文献[16]可知,若一条评论表现为某种情感倾向,那么在这条评论中出现的情感词语也会集群出现在具有同样情感倾向的其他评论中。换言之,这些集群出现在具有某种相同情感倾向评论中的词语大都是相似的,其情感倾向性也趋于一致。由文献[17]可知,TF-IDF(Term Frequency-Inverse Document Frequency)算法对评论词集进行计算的过程中,TF 意为词频,表示该评论词集中的每个词语在评论中出现的频率;IDF 意为逆向文档频率,是对该词语普遍重要性的度量,其基本思想是:如果包含该词语的评论数量越少,则说明该词语区分类别的能力越好。其计算方法

是:由总评论数目除以包含该词语的评论数目,然后将得到的商取对数,即可求得该指定词语的 IDF。如果某词语在指定评论中具有高词频,且该词语在整个评论语料集中具有低评论频率,其 TF-IDF 就会产生高权重值。TF-IDF 算法的公式如下:

$$TF(\text{word}) = \frac{\text{word}}{\sum \text{total words}} \quad (1)$$

其中,分子 word 表示该词语;分母 total words 表示该评论词集中所有词语出现的次数之和。

通过式(1)计算后可得到该词语的 TF。

$$IDF(\text{word}) = \log_2 \frac{|\text{Reviews}|}{|\{\text{word} \in \text{Review}\}|} \quad (2)$$

其中,分子 |Reviews| 表示语料集中评论的总数;分母 |{word ∈ Review}| 表示包含词语 word 的评论数目。

通过式(2)计算后可得到该词语的 IDF。最后,通过式(3)计算后可得到该词语的 TF-IDF:

$$TF-IDF(\text{word}) = TF \times IDF \quad (3)$$

设在该过程中,在上述评论候选种子情感词集的基础上,为了从诸多候选情感词中,筛选出具有代表性并且能够明显影响购物评论情感分类的情感词作为购物评论语料集中的领域情感词,减少计算复杂度,提高计算精度,通过以 TF-IDF 算法为基础进行改进,使用 PTF-IDF 算法,计算得到候选情感词的 PTF-IDF 值,采用阈值法根据排序情况依次选择前 20 个候选情感词作为领域情感词。具体方法如下:

在使用 PTF-IDF (Part of speech Tag Frequency-Inverse Document Frequency) 算法对评论候选情感词集进行计算的过程中,需要指出 PTF 定义为经由词性标注后候选情感词语的词频,它表示候选情感词语在所有购物评论中出现的频率;而 IDF 同上。

PTF-IDF 算法是对词频型特征权重算法的修正,旨在针对某一领域内购物评论情感分类以 TF-IDF 算法为基础进行的改进。由于词频型特征权重算法仅考虑了情感特征出现在评论中的次数,容易将许多出现频率高却不能够区分类别或者根本没有意义的词语代入情感分类的计算中,造成了评论情感不能得到准确分类的结果。而 TF-IDF 算法在词频排序的基础上,还考虑了包含该词语的评论数量,认为包含该词语的评论数越多,其区分类别的能力越弱。在计算过程中,保留了对评论情感分类影响不大的助词和名词,而忽视了许多具备明显情感倾向但计算结果低于阈值的情感词语。因此,PTF-IDF 算法能有效改善这一缺陷,筛选出更多有价值的情感词语。

PTF-IDF 算法的公式如下:

$$PTF_{x,j} = \frac{\text{word}_{x,j}}{\sum_k \text{word}_{k,j}} \quad (4)$$

其中,分子 $\text{word}_{x,j}$ 表示 word_x 在评论 Review_j 中出现的次数;分母 $\sum_k \text{word}_{k,j}$ 表示 Review_j 中所有词语出现的次数之和。

$$IDF_x = \log_2 \frac{|\text{Reviews}|}{|\{j: \text{word}_x \in \text{Review}_j\}| + 1} \quad (5)$$

其中,分子 |Reviews| 表示语料集中的评论总数;分母 |{j: word_x ∈ Review_j}| 表示包含词语 word_x 的评论数目。

如果 word_x 不存在于语料集中,则会导致该公式的分母为零。为了防止影响计算出现差错,在研究中采用分母+1 的方法。最后:

$$PTF-IDF(\text{word}_x) = PTF(\text{word}_x) \times IDF(\text{word}_x) \quad (6)$$

通过上述方法处理后得到领域情感词,从而构建领域情感词典。

5 实验结果与分析

实验采用了 Python 编程开发环境,使用目前针对中文文本处理较好的 Jieba 分词系统对评论语料进行分词,该系统还支持添加自定义词典、关键词提取等功能。分别使用不同语料进行实验。实验一,语料采用了中国科学院谭松波博士收集的关于酒店的网络评论数据集,该数据集名称分别是 ChnSentiCorp_hlt_ba_1000, ChnSentiCorp_hlt_ba_2000, ChnSentiCorp_hlt_ba_4000,在这三个数据集分别为 1 000 条、2 000 条、4 000 条,每个数据集中正负类语料均衡,各 500 篇、1 000 篇、2 000 篇。实验二,语料一部分采用了携程网中有关酒店领域的评论语料集,总共 2 000 篇,其中正负评论语料各为 1 000 篇。另一部分采用从当当网书籍领域采集的评论语料共 2 000 篇,以及从亚马逊网站采集的笔记本电脑领域评论语料共计 2 000 篇,两者中正负评论语料各为 1 000 篇。

对比实验方法 M₁ 采用基础情感词典选取评论中的情感特征,若基础情感词典中的词语出现在待分类评论中,则将该词语作为情感特征,对该情感特征赋予布尔型特征权重。其中,布尔型 (Presence) 特征权重的原理是:如果情感特征出现在待分类评论中,则其权重为 1,反之其权重为 0。

对比实验方法 M₂ 使用基础情感词典筛选情感词语作为情感特征,同时采用绝对词频算法计算各领域购物评论的分词去重后得到的所有候选词语的绝对词频,按照阈值筛选排序后数值较高的词作为情感特征,加入基础情感词典,并带入对该领域评论的情感分类

计算中。

对比实验方法 M_3 采用了提出的构建领域词典的 PTF-IDF 算法,使用基础情感词典筛选购物评论中的情感词语作为情感特征,同时对进行过分词和去重处理后的评论词集词性标注,选择词性为形容词和习用语的词语作为候选情感词,并以该候选情感词集为基础筛选领域情感词,从而构建领域情感词典,并对该领域的购物评论进行情感分类。

实验一和实验二的分类结果如表 2 和表 3 所示。

从实验结果可以看出,方法 M_1 中只使用基础情感词典对评论分类的准确率很低,主要原因是基础情感词典包含的词语数量有限,并且对于指定领域的情感词不能识别,从而忽视了这些领域情感词对于评论情感分类的重要性。方法 M_2 的准确率比方法 M_1 略高,却仍然相对较低,其原因是方法 M_2 能够筛选出现频次高且对评论情感分类有影响的词语,在基础情感词典的基础上有所提高和扩展,但是这些词汇中仍包含了一些没有情感特征或情感特征不明显的词语被选作用于分类的特征代入计算,从而降低了评论分类的准确率。由图 2 和图 3 也可以看出,方法 M_3 的分类准确率较其他两种方法在各个领域的评论情感分类中都有了明显提高。在图书领域,经关键词提取后可发现,该领

域评论存在大量专业词汇,例如书名、人名、诗句等等,在该领域还可以进行更细致的特征提取,以达到进一步提高情感分类准确率的目的。

表 2 实验一分类结果

数据集	选择方法		
	M_1 准确	M_2 准确	M_3 准确
	率/%	率/%	率/%
ChnSentiCorp_htl_ba_1000_pos500	67.6	70.2	88.0
ChnSentiCorp_htl_ba_1000_neg500	60.0	63.8	82.4
ChnSentiCorp_htl_ba_2000_pos1000	63.3	73.6	85.9
ChnSentiCorp_htl_ba_2000_neg1000	60.4	70.1	83.4
ChnSentiCorp_htl_ba_4000_pos2000	61.6	71.9	84.0
ChnSentiCorp_htl_ba_4000_neg2000	61.0	72.2	84.2

表 3 实验二分类结果

数据集	选择方法		
	M_1 准确	M_2 准确	M_3 准确
	率/%	率/%	率/%
酒店-正面 1 000 条	63.4	78.2	91.1
酒店-负面 1 000 条	62.8	76.5	87.7
图书-正面 1 000 条	67.1	77.7	90.5
图书-负面 1 000 条	61.1	76.0	79.4
电脑-正面 1 000 条	66.7	76.9	87.2
电脑-负面 1 000 条	64.3	75.0	83.2

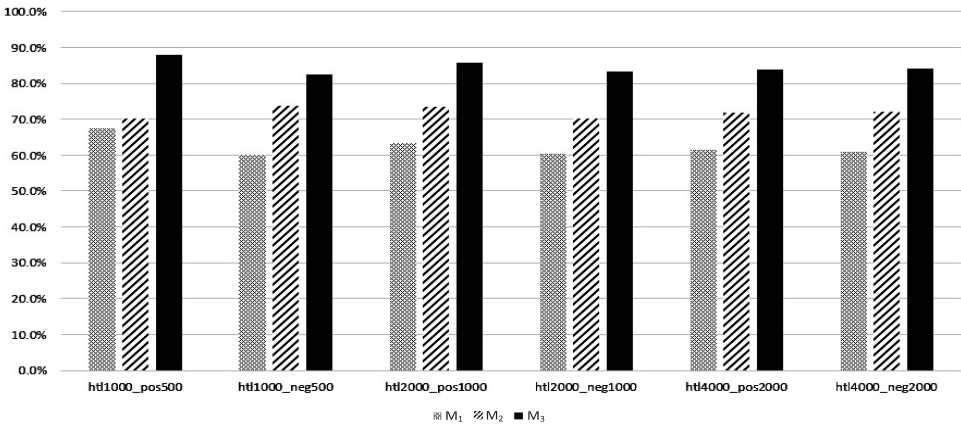


图 2 实验一分类结果图

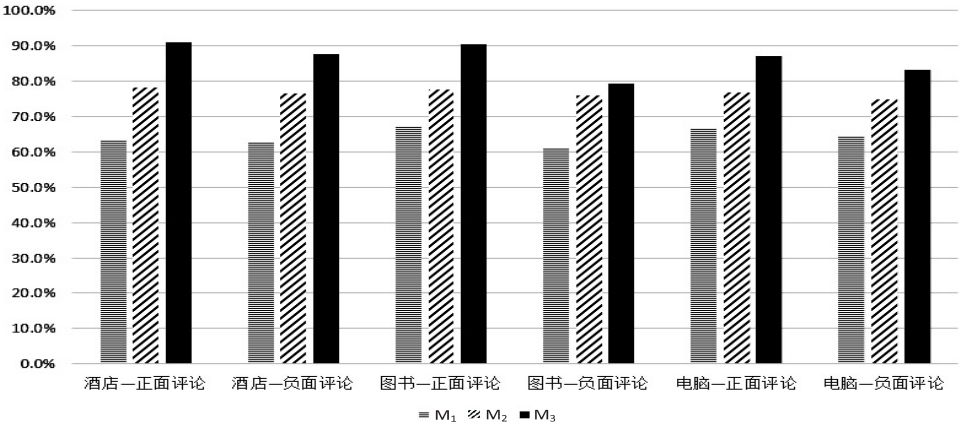


图 3 实验二分类结果图

上述实验可以说明,通过提出的 PTF-IDF 算法得到的领域情感词进而构建领域情感词典,并代入各领域购物评论进行情感分类,能够有效地提升分类准确率,从而验证了该算法的有效性。

6 结束语

面对如今网上购物商城中,用户对消费购物体验过程给出的海量评论,如何根据评论内容较为迅速地提炼出有价值的信息,并对评论的情感倾向做出正确分类,已成为目前迫切需要解决的问题。针对不同领域之间词语相似度存在不同,在文本中表达意义存在差异的特点,通过扩展基础情感词典,利用采集的语料中不同词性词语对评论文本情感倾向性的影响程度筛选合适的词语作为候选情感词,结合评论候选情感词与不同领域的文本之间关联程度信息,构建购物评论领域情感词典,代入购物评论的情感分类研究中。实验结果表明,该方法能有效提高分类准确度,并降低情感特征空间维数,灵活适应情感词在各个领域间存在的语义差异问题。

下一步的研究工作主要是利用现有语料资源和中文语法规则动态地修正词典,达到规范完善领域词典的目的;同时,综合考虑主题词语、评论对象对情感倾向计算的影响,结合上下文相关信息,提供一种更加合理全面的情感倾向计算方式。

参考文献:

- [1] 薛益定. 中文情感分析研究综述[J]. 电脑编程技巧与维护,2016(5):22-24.
- [2] Neviarouskaya A, Prendinger H, Ishizuka M. SentiFul: a lexicon for sentiment analysis[J]. IEEE Transactions on Affective Computing, 2011, 2(1):22-36.
- [3] 周俊臣, 赵 晖. WEB 文本情感倾向性分析研究综述[J]. 计算机光盘软件与应用, 2014, 17(1):303-304.

(上接第 193 页)

- [1] :Springer Science & Business Media,2013.
- [4] Joachims T. Making large-scale SVM learning practical[R]. [s.l.]:[s.n.],1998.
- [5] Hsu C W, Chang C C, Lin C J. A practical guide to support vector classification[EB/OL]. 2010-04-15. <https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf>.
- [6] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报,2000,26(1):32-42.
- [7] Zadeh L A. Fuzzy sets[J]. Information & Control, 1965, 8(3):338-353.
- [8] Lin C F, Wang S D. Fuzzy support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2):464-471.

- [4] 饶洋辉,李 青,刘文印,等. 公众文本之情感词典研究进展[J]. 中国科学:信息科学,2014,44(7):825-835.
- [5] 张成功,刘培玉,朱振方,等. 一种基于极性词典的情感分析方法[J]. 山东大学学报:理学版,2012,47(3):47-50.
- [6] 王宝成,何新宇. 基于改进情感词域识别的舆情情感分析研究[J]. 电子技术与软件工程,2016(3):167.
- [7] 刘清松,张仰森. 基于多词典融合的词汇语义倾向判别[J]. 计算机技术与发展,2015,25(5):104-109.
- [8] 李爱萍,邸 鹏,段利国. 基于句子情感加权算法的篇章情感分析[J]. 小型微型计算机系统,2015,36(10):2252-2256.
- [9] 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. Philadelphia:ACL Press,2002:79-86.
- [10] Yu H, Hatzivassiloglou V. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences[C]//Proceedings of the EMNLP 2003. Morristown:ACL Press,2003:129-136.
- [11] 樊康新. 基于 SVM 的网络文本情感分类系统的设计[J]. 计算机时代,2015(12):34-37.
- [12] 蒋婉婷,孙 蕾,钱 江. 基于 Hadoop 的朴素贝叶斯算法在中文微博情感分类中的研究与应用[J]. 计算机应用与软件,2015,32(7):60-62.
- [13] 张 森,曹 晖. 基于《知网》概念定义的情感词典构建研究[J]. 计算机工程与应用,2015,51(17):118-123.
- [14] Lawrence P, Sergey B. The anatomy of a large scale hyper textual web search engine[J]. Computer Networks & ISDN Systems, 2012, 56(18):3825-3833.
- [15] 顾益军,刘小明. 融合多种情感资源的微博情感分类研究[J]. 计算机科学,2015,42(4):209-212.
- [16] 陈 明. 多极性形容词的倾向判别及其应用研究[D]. 郑州:郑州大学,2012.
- [17] 李宪毅,刘培玉,朱振方,等. 基于改进情感关键词抽取的自动情感摘要技术[J]. 计算机工程与设计,2016,37(3):778-782.

- [9] Inoue T, Abe S. Fuzzy support vector machines for pattern classification[C]//International joint conference on neural networks. [s.l.]:IEEE,2001:1449-1454.
- [10] 陈雪梅,田 赓,苗一松. 面向智能驾驶行为的机器学习[J]. 道路交通与安全,2014,14(6):60-64.
- [11] 王伟钰,曲仕茹. 基于模糊神经网络的自动驾驶决策系统研究[J]. 计算机测量与控制,2009,17(9):1711-1713.
- [12] Smola A J. Learning with kernels I support vector machines[J]. Lecture Notes in Computer Science, 2008, 42(4):1-28.
- [13] 奉国和. SVM 分类核函数及参数选择比较[J]. 计算机工程与应用,2011,47(3):123-124.
- [14] 李盼池,许少华. 支持向量机在模式识别中的核函数特性分析[J]. 计算机工程与设计,2005,26(2):302-304.