

网络情感词自动识别方法研究^{*}

张清亮 徐 健

(中山大学资讯管理学院 广州 510006)

【摘要】针对情感词识别及情感词库构建效率不高的问题,提出一种自动提取基准情感词集的方法,从词频、词的领域性情感倾向和词的情感强度三方面进行基准词筛选,再凭借目标词与正、负基准词集的不同语义相似度进行情感词的识别和情感倾向的判断,使机器能够自动完成大部分工作,提高效率,降低构建不同领域情感词库的成本。以京东商城 71 061 条评论和卓越网 1 736 条评论为数据集进行实验,获得的召回率为 76.36%,准确率为 76.94%,情感倾向判断的准确率为 62.70%。

【关键词】情感分析 情感倾向 PMI-IR 情感词库

【分类号】G250

Research on Automatic Extraction of Web Sentiment Words

Zhang Qingliang Xu Jian

(School of Information Management, Sun Yat-Sen University, Guangzhou 510006, China)

【Abstract】To improve the efficiency of extracting sentiment words and building sentiment lexicon, the authors propose a method to extract a set of basic sentiment words, and then to calculate both the PMI-IR value between candidate word and the positive basic sentiment word set and the PMI-IR value between candidate words and the negative basic sentiment word set, to judge the orientation of a candidate word. Taking account of frequency, orientation, intensity and definiteness of words, computers are able to finish most of the work. It improves the efficiency and reduces cost of building sentiment lexicon. Experiment is processed on the dataset constituted with 71 061 reviews from 360buy and 1 736 reviews from Joyo. With the dataset, the method achieves a recall rate of 76.36%, a precision of 76.94%, and the precision of sentiment orientation is 62.70%.

【Keywords】Sentiment analysis Sentiment orientation PMI-IR Sentiment lexicon

1 引言

情感词库的构建是网络用户情感分析任务的一个重要环节。在丰富、完善的情感词表基础上,研究人员可以进行观点的识别、分类、提炼、整合、追踪,分析调研结果,拓展搜索引擎的功能等更加深入的研究^[1,2]。此外,也可以通过情感词反向发现其描述的实体,从而提取产品特征,构建产品本体,实现将产品特征依据重要度排序^[3]的目的。

目前,已经存在一些通用的情感词库,然而,在不同的领域,词的情感并不一致,甚至完全相反。比如“很硬”在食品行业是个贬义词,而在建筑材料行业却是个褒义词。这就要求在构建情感词库时要考虑词的领域属性,并

收稿日期:2011-07-28

收修稿日期:2011-08-07

^{*} 本文系国家自然科学基金项目“用户评论情感分析及其在竞争情报服务中的应用研究”(项目编号:11CTQ022)的研究成果之一。

对情感词表构建方法的灵活性提出更高的要求。而当前很多情感词表还采用人工方法进行构建和维护(比如《知网》情感词表^[4]),存在效率低、更新不及时(比如网络用语)、没有从领域的角度考虑词的情感倾向等问题。因而迫切需要一种能囊括词的领域属性,并自动构建、动态更新词表的方法。本文提出一种机器自动获取具有领域特点的基准情感词的方法,该方法主要思路是基于某个领域的语料库,采用统计和 PMI-IR 算法,提取语料库内与其他词的语义相似度最高的 N 个词作为基准词集。获得基准词集后,再利用基准词集和 PMI-IR 算法来判断候选情感词的情感倾向性和情感强度。其任务包括了正、负倾向情感基准词集的选取和情感词的自动发现以及情感倾向的判断。

2 情感词识别相关研究

当前已经有一些学者在情感词情感倾向方面进行了研究,主要研究思路可概括为以下两种:

(1) 基于语义计算的方法。语义方法主要是基于 WordNet 和 HowNet 进行。Kamps 等^[5]就是基于 WordNet,计算目标词和基准词“good”和“bad”间的语义距离来分析词汇的褒贬性。中文方面,朱嫣岚等^[6]利用 HowNet 语义相似度和语义相关度计算工具,进行目标词与基准词之间紧密程度的计算;路斌等^[7]提出利用《同义词词林》计算词汇的褒贬。这些方法在通用情感词的倾向判断上获得了一定的效果,但是对于和领域属性有较大相关度的词的情感倾向就无法很好地判断。

(2) 基于统计分析的方法。Dave 等^[8]利用抽取技术获得特征词,再通过贝叶斯网络计算特征词与已标注情感类别文档的语义相似度,判定词汇的情感倾向。Turney 等^[1]和彭学仕等^[9]也采用了统计方法,他们利用点互信息(PMI-IR),通过搜索引擎来统计词汇和种子词之间的 PMI-IR 值,根据 PMI-IR 值确定两个词之间的紧密程度,得出词的情感倾向。

统计方法基本都涉及基准词选择的问题。Turney 等^[1]通过实验证明对上下文敏感度低的词作为基准词能得到更高的准确度,于是他们通过人工成对地挑选对上下文敏感度低的 7 对正负性基准词对。朱嫣岚等^[6]提出基准词应该是褒贬态度明显、强烈,具有代表性的词,他们通过对词频进行降序排列,挑选了三组基

准词,然后进行实验。王素格等^[10]采用基于词频概率估计的 Fisher 准则函数挑选出类别区分能力强的 M 个词,并与自建的情感词词表做交集,然后按照词频筛选出 N 个基准词。

上述研究在情感词提取上取得了一定的效果,但是这些方法在进行文档标注或基准词选取时不仅需要大量的人工干预,而且存在较大的随意性和主观性。同时,因为需要大量的人工判断,所以不能由机器自动迁移到其他领域,无法考虑词的领域属性,具有一定的局限性。最近,彭学仕^[9]提出基于词聚类的基准词选择方法。他们构建目标领域本体,通过领域本体获取初始种子词进行扩展,利用迭代聚类的方法,获得最优的聚类种子词,并作为最终选取的基准词。这种方法虽然将基准词提取过程中的人工干预降低,但同时又引入本体构建,需要大量人工干预。

针对上述研究的不足,本文提出利用某个领域的语料库进行基准词的提取,再通过网络评论文本和搜索引擎进行领域情感词的自动识别,从而达到自动构建和扩展领域相关情感词库的目的。本文提出的方法是一种基于统计的方法,同时为了实现情感词的领域相关和提高情感词自动识别的效率,所以本方法对基准情感词具有更强的依赖性。基准词的优劣将直接决定能否有效体现领域相关性和准确识别情感词。

3 基准词提取及情感词库构建

情感基准词是指具有非常明显褒贬义倾向的词汇。在对当前研究进行分析的基础上,笔者认为领域相关情感基准词还应具备如下特点:

(1) 情感基准词在领域内有明确的褒贬倾向。词的褒贬倾向并不是在所有领域都是一致的,一个词表也不可能涵盖所有领域,所以笔者认为,词在领域内的褒贬倾向鲜明更为重要。

(2) 情感基准词应该是较为常用的词。有些词的情感倾向非常明确,但与领域的相关度很低,被使用的概率比较低,那么这些词的敏感度就会降低,可能与绝大多数词的语义相似度都较小。比如,“惨绝人寰”是个很负面的词,但是几乎不会有人在手机产品评论中使用这个词。

笔者在考虑这些特点后,提出如下基准词集的提取方法。

3.1 基准词集生成过程

基于以上分析,本文提出了基准词集的生成流程,如图1所示:

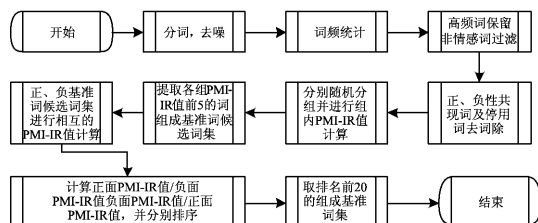


图1 基准词集生成流程

图1中,第4步保留下来的高频词就可能是领域内特有的情感词。第8步正、负基准词集相互进行 PMI-IR 计算是为了获得情感倾向更为鲜明的词作为基准词。获得基准词后,即可通过基准词自动识别情感词。

3.2 情感词识别与领域相关的情感词库的自动构建

领域相关的情感词库的自动构建,首先需要构建领域内的情感词基准词集,再利用基准词集自动识别情感词,实现情感词的自动扩展。本文情感词识别及词库构建整体思路如图2所示:

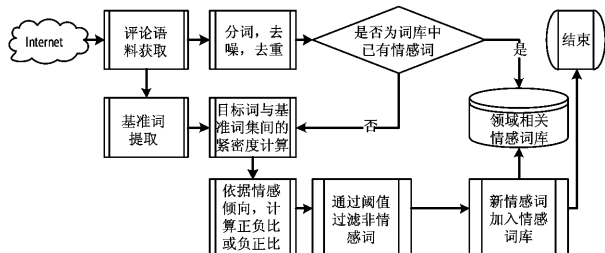


图2 情感词自动识别和情感词库

自动构建与自动扩展流程

图2中,情感词库构建的方法类似基准词提取的过程,只是在计算完正负 PMI-IR 值后,不是直接根据该值大小进行提取,而是通过语义相似度阈值过滤非情感词,语义相似度阈值将由实验给出。

3.3 目标词的情感倾向判断

本文计算目标词与基准词集间的语义相似度,是通过目标词与基准词集中的每个词进行 PMI-IR 值计算,依据 PMI-IR 值进行排序,去掉最大值和最小值,然后求和平均,以均值作为目标词的 PMI-IR 值。其公式^[11]如下所示:

$$PMI-IR(\text{word}, \text{wordSet}) = \frac{\sum_{i=1}^{\text{count}(\text{wordSet})-1} PMI-IR(\text{word}, \text{word}_i)}{\text{count}(\text{wordSet})-2} \quad (1)$$

式(1)中 $PMI-IR(\text{word}, \text{word}_i)$ 是目标词 word 与第 i 个基准词 word_i 间的 PMI-IR 值, $\text{count}(\text{wordSet})$ 代表基准词集合 wordSet 含有的元素个数。最后得到的 $PMI-IR(\text{word}, \text{wordSet})$ 值就是词 word 与基准词集 wordSet 之间的语义相似度。

通过式(1)将每个目标词分别与正、负向基准词集进行 PMI-IR 值计算后,进行如下处理:

$$PMI-IR(\text{word}) = \begin{cases} PMI-IR(\text{word}, \text{word}_{\text{positive}}), & PMI-IR(\text{word}, \text{word}_{\text{positive}}) \geq PMI-IR(\text{word}, \text{word}_{\text{negative}}) \\ -PMI-IR(\text{word}, \text{word}_{\text{negative}}), & PMI-IR(\text{word}, \text{word}_{\text{positive}}) < PMI-IR(\text{word}, \text{word}_{\text{negative}}) \end{cases} \quad (2)$$

如果目标词与正向基准词集的语义相似度大于其与负向基准词集的语义相似度,则初步认定该词为正向词;反之,为负向词。

对判断为正向词的目标词,进行如下计算:

$$P = \frac{PMI-IR(\text{word}, \text{word}_{\text{positive}})}{PMI-IR(\text{word}, \text{word}_{\text{negative}})} \quad (3)$$

对判断为负向词的目标词,进行如下计算:

$$P = \frac{PMI-IR(\text{word}, \text{word}_{\text{negative}})}{PMI-IR(\text{word}, \text{word}_{\text{positive}})} \quad (4)$$

计算出 P 值后,将 P 值与阈值进行比较, P 值大于阈值的词确定为情感词。

4 实验及结果分析

为了验证方法的有效性,本文进行了实验。首先确定实验数据集;其次进行基准词的提取;最后通过基准词识别新情感词,并将新情感词加入情感词表。

4.1 实验数据集获取

目前中文语料库并不十分开放,同时中文评论方面的语料库还很少,因此,笔者采取自行构建的方式。在分析了卓越^[12]、当当^[13]、京东^[14]、淘宝^[15]等中文购物平台上的评论后,笔者发现各个网站都提供了打分的机制,大多也给出了评论是否有用的判定按钮(除了淘宝)。但是这些数据并没有很直接地反映评论内容,也无法很直观地表明评论的情感倾向。只有京东在顾客填写评论时就要求顾客将产品的优点和缺点分开写。这样能比较明确地获得评论的情感倾向,有利于实验。于是,笔者采用京东商城上的评论数据作为研究数据。

本研究使用的数据如下:

(1)低频或非情感词过滤词集——HowNet 负面词集:由《知网》情感分析用词语集(Beta 版)中的负面评价词和负面情感词组成,去重后共 4 320 词。

(2)低频或非情感词过滤词集——HowNet 正面词集:由《知网》情感分析用词语集(Beta 版)中的正面评价词和正面情感词组成,共 4 528 词。

(3)负面语料库:由京东商城上关于手机的 71 061 条负面评价组成。

(4)正面语料库:由与负面语料库对应的 71 061 条正面评价组成。

(5)情感词库自动构建测试语料集:由卓越网上关于电子产品的 1 736 条评论组成。

4.2 基准词提取

情感词的提取依照图 1 所示流程进行。在进行分词、去重并统计词频后,获得正面语料词 7 062 个、负面语料词 7 637 个。通过低频或非情感词过滤词表进行筛选并去除停用词以及在正负面词中都出现的词后,得到正面词 724 个、负面词 479 个。实验最后得到基准词表如表 1 和表 2 所示:

表 1 正面基准词

词	词性	词	词性	词	词性	词	词性
稀罕	动词	期盼	动词	紧凑	形容词	痴迷	动词
赏心悦目	动词	皮实	形容词	乐意	动词	好心	名词
细致	形容词	欣慰	形容词	刚好	副词	灵敏	形容词
钟爱	动词	欣喜	形容词	详尽	形容词	礼貌	形容词
恰到好处	形容词	看中	动词	一如既往	动词	舒适	形容词

表 2 负面基准词

词	词性	词	词性	词	词性	词	词性
费劲	形容词	焦急	形容词	疏忽	动词	乱七八糟	形容词
用不着	区别词	邪恶	形容词	劣质	区别词	骗人	动词
万恶	名词	谴责	动词	担忧	动词	僵硬	形容词
低档	区别词	遗憾	形容词	迷糊	动词	可恶	形容词
差劲	形容词	焦虑	形容词	该死	动词	别扭	形容词

算法将常见的“赏心悦目”、“细致”、“劣质”、“差劲”等词选为基准词,同时也将“礼貌”、“焦急”等具有电子商务特色的词抽取出来,这说明算法不仅能够辨别出情感词,同时也具有了一定的领域特征词抽取功能。此外算法还将“用不着”这样不常见的情感词选取为基准词。“用不着”这个词一般出现在“这些功能我用不着”这样的句子中。这个句子只是一个陈述,并不带有太多的情感,但是当这样的陈述普遍存在时,说明这些功能没用,却引起了人们的注意,给人们带来了不便,这样“用不着”就成了一个具有贬义情感的词

汇。这样的情感词是通过人工也不容易发现的,而本文提出的方法能够发现这样的词,进一步证实该方法能够初步识别出领域相关的情感词。这同时也说明:在领域外为非情感词,而在领域内出现频次特别高的词值得关注。

笔者在实验过程中还发现,有些词情感倾向明确,但却是很常用的词,加上中文分词不十分准确的因素,可能会和绝大多数词有很强的语义相似度,用这些词作为基准词将带来更大的噪音。例如,“好”是褒义词,但是当用“好”进行检索时,会将大多数包含“不好”的记录找出来,这就可能导致“好”与许多贬义词之间的 PMI-IR 值也很高,影响结果的准确性。因此,在提取基准词时还应该对超高频词进行处理。

4.3 情感词识别与词库的自动构建

提取基准词后,便可通过基准词实现情感词的自动识别和情感词库的自动扩展。本文对京东商城评论数据进行分词、去重、去停用词、去低频或非情感词等处理后,得到 4 027 个词。将得到的词与提取的基准词进行 PMI-IR 值计算,并计算正 PMI-IR 与负 PMI-IR 的比值或负 PMI-IR 与正 PMI-IR 的比值。同时,为了进行评价,对这 4 027 个词进行人工情感标注,得到正面词 62 个,负面词 103 个,其余为中性词。最后,提取正 PMI-IR 与负 PMI-IR 的比值或负 PMI-IR 与正 PMI-IR 的比值大于阈值的目标词,作为新情感词,并根据各自的情感倾向归入到正负情感词表中。

4.4 评价指标

本实验效果的评价主要在于观察提取的基准词是否能够准确识别情感词。本文采用被同类实验普遍采用的准确率和召回率^[16,17]进行评价。评价指标主要由以下几部分组成:

(1)情感词的召回率:

$$\text{Recall} = \frac{\text{正确判断为情感词的词语个数}}{\text{人工判断为情感词的词语个数}} \quad (5)$$

(2)获得的情感词的准确率:

$$\text{Precision}_1 = \frac{\text{正确判断为情感词的词语个数}}{\text{所有被算法判断为情感词的词语个数}} \quad (6)$$

(3)召回的情感词的情感倾向判断的准确率:

$$\text{Precision}_2 = \frac{\text{正确判断情感倾向的情感词个数}}{\text{所有被算法判断为情感词的词语个数}} \quad (7)$$

4.5 结果分析

在上述实验测试环境及实验数据的情况下,各评价指标随语义相似度阈值选取的变化产生的变化如图

3 所示:

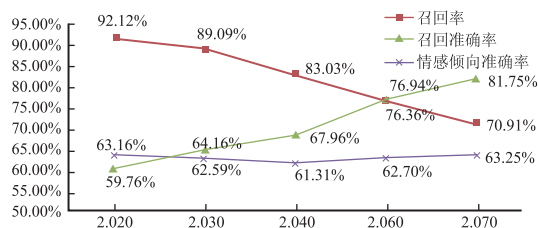


图3 评价指标随阈值变化的趋势

实验数据显示本文提出的方法的情感倾向准确率维持在63%左右。而召回率和召回准确率则在语义相似度阈值为2.06时达到最优。在该阈值下召回率为76.36%,召回词准确率为76.94%。同样,算法识别出了“开心”、“华丽”、“高品质”、“物廉价美”、“窝火”、“倒霉”、“伤心”、“扫兴”等常见情感词,也将“黑屏”、“操作不便”、“半死”、“划痕”、“烂”、“杂音”、“超快”、“高音质”、“人性化”、“推荐”等具有电子产品或电子商务行业特色的词列为了情感词。这些都说明,本文提出的方法不仅在情感词识别上具有一定的效果,在领域情感词识别上也有较好的效果,也证明了前一步骤提取的基准词符合用于发现领域情感词的要求。

5 结 语

本文通过提出一种能够获取具有领域特征的基准情感词自动抽取方法,实现领域相关情感词的识别和词库的构建。该方法从词频、词的领域性情感倾向、词的情感鲜明度三个主要方面,利用 PMI-IR 算法进行基准词提取,再通过计算目标词与基准词集间的语义相似度识别情感词,有效地提高了构建情感词典的效率,并实现了将情感词与领域相关联的构想。数据和抽取结果说明本文提出的方法具有一定的有效性,能初步达到自动构建领域相关情感词典的要求。

通过对实验数据的观察发现,该方法仍有提升的空间。如果将数据稀疏因素考虑进来,并进行同义词合并,可以进一步提高算法的准确率;另外,今后的研究中也考虑使用具有邻近检索功能的搜索引擎或其他相关方法,以更准确地判断两个词间的共现关系。该方法将为构建情感词典提供一种有效的工具,为后续的情感分析提供良好的基础。

参考文献:

- [1] Turney P D, Littman M L. Measuring Praise and Criticism: Inference of Semantic Orientation from Association[J]. *ACM Transactions on Information Systems*, 2003, 21(4): 315-346.
- [2] 王素格. 基于 Web 的评论文本情感分类问题研究[D]. 上海: 上海大学, 2008.
- [3] Tian P, Liu Y, Liu M. Research of Product Ranking Technology Based on Opinion Mining[C]. In: *Proceedings of the 2nd International Conference on Intelligent Computation Technology and Automation*. 2009: 239-243.
- [4] 董振东, 董强. 知网[EB/OL]. [2011-06-20]. <http://www.keenage.com/>.
- [5] Kamps J, Marx M, Mokken R J. Words with Attitude[C]. In: *Proceedings of the 1st International Conference on Global WordNet*. 2002: 332-341.
- [6] 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算[J]. *中文信息学报*, 2006, 20(1): 14-20.
- [7] 路斌, 万小军, 杨建武. 基于同义词词林的词汇褒贬计算[C]. 见: *中国计算技术与语言问题研究——第七届中文信息处理国际会议论文集*. 中国, 武汉: 中国中文信息学会, 2007.
- [8] Dave K, Lawrence S, Pennock D M. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews[C]. In: *Proceedings of the 12th International Conference on World Wide Web*. 2003: 519-528.
- [9] 彭学仕, 孙春华. 面向倾向性分析的基于词聚类的基准词选择方法[J]. *计算机应用研究*, 2011, 28(1): 114-116.
- [10] 王素格, 李德玉, 魏英杰, 等. 基于同义词的词汇情感倾向判别方法[J]. *中文信息学报*, 2009, 23(5): 68-74.
- [11] Turney P D. Mining the Web for Synonyms: PMI-IR Versus LSA on TOEFL[C]. In: *Proceedings of the 12th European Conference on Machine Learning*, Freiburg, Germany. 2001: 491-502.
- [12] 亚马逊公司. 卓越网[EB/OL]. [2011-07-05]. <http://www.amazon.cn/>.
- [13] 北京当当网信息技术有限公司. 当当网[EB/OL]. [2011-07-05]. <http://www.dangdang.com/>.
- [14] 北京京东世纪商贸有限公司. 京东商城[EB/OL]. [2011-07-05]. <http://www.360buy.com/>.
- [15] 阿里巴巴集团. 淘宝网[EB/OL]. [2011-07-05]. <http://www.taobao.com/>.
- [16] Abulaish M, Jahiruddin, Doja M N, et al. Feature and Opinion Mining for Customer Review Summarization[C]. In: *Proceeding of the 3rd International Conference on Pattern Recognition and Machine Intelligence*. 2009: 219-224.
- [17] Hu M, Liu B. Mining Opinion Features in Customer Reviews[C]. In: *Proceedings of the 19th National Conference on Artificial Intelligence*. 2004: 755-760.

(作者 E-mail: issxj@mail.sysu.edu.cn)