

文章编号: 1003-0077(2020)03-0088-11

基于句法规则和 HowNet 的商品评论细粒度观点分析

韦婷婷¹, 陈伟生¹, 胡勇军², 骆威¹, 包先雨³

(1. 华南农业大学 数学与信息学院, 广东 广州 510642;

2. 广州大学 工商管理学院, 广东 广州 510006;

3. 深圳市检验检疫科学研究院, 广东 深圳 518001)

摘要: 该文提出一种基于句法规则和 HowNet 词典的商品评论细粒度观点分析方法, 主要包括三个模块: 评价对象抽取、评价对象—评价词对抽取、评价对象总体观点得分计算。具体思路为: 首先, 结合词性标注和频繁项集方法构建一个初始的评价对象词典, 便于重用和修正商品的总体评价维度; 其次, 基于爬取的电商评论文本真实数据设计了评价对象—评价词对抽取规则; 最后, 借助 HowNet 词典分别计算不同评价维度的观点综合得分, 进而对比同一商品不同品牌在各个维度下的总体观点评价, 该方法在商品评论语料集上验证了有效性。

关键词: 商品评论; 细粒度观点分析; 句法规则; HowNet

中图分类号: TP391

文献标识码: A

Fine-grained Opinion Analysis of Product Reviews Based on Syntactic Rules and HowNet

WEI Tingting¹, CHEN Weisheng¹, HU Yongjun², LUO Wei¹, BAO Xianyu³

(1. School of Mathematics and Informatics, South China Agricultural University,

Guangzhou, Guangdong 510642, China;

2. School of Management, Guangzhou University, Guangzhou, Guangdong 510006, China;

3. Shenzhen Academy of Inspection and Quarantine, Shenzhen, Guangdong 518001, China)

Abstract: This paper proposes a fine-grained opinion analysis model based on syntactic rules and HowNet for product reviews. The model consists of three main modules: the target mining, the target-opinion mining, and the overall opinion estimation. Firstly, a target lexicon is constructed in terms of part-of-speech tagging and frequent item set mining, which is convenient to reuse and modify the overall opinion target of products. Secondly, rules are designed for target-opinion extraction based on the real data of e-commerce reviews. Finally, HowNet dictionary is adopted to estimate the overall score of all targets, and then to compare different brands of the same product in each evaluation perspective. The validity of this method is verified on product review corpus.

Keywords: product reviews; fine-grained opinion analysis; syntactic rules; HowNet

0 引言

商品评论细粒度观点分析可以获取关于商品局部性的结构、属性和功能等方面的观点评价情况^[1], 更能满足商家和用户的需求, 已日益成为自然语言处理等领域的热门研究方向。细粒度观点分析又可

称为评价对象—评价词(也有研究称之为“情感词”)对的挖掘, 即抽取评论文本中用户所评价的对象及其观点词的搭配对。目前评价对象与评价词的抽取方法主要分为两大类: ①基于监督学习的方法; ②基于无监督学习的方法。基于监督学习的方法目前较为流行的有利用深度学习与传统机器学习方法相结合的模型, 主要通过各种优化手段自动学到一个

收稿日期: 2019-07-25 定稿日期: 2019-10-15

基金项目: 国家自然科学基金(61702197); 国家重点研发计划课题(2016YFF0202504); 广东省社科“十三五”规划项目(GD17CGL01); 华南农业大学大学生创新创业计划训练项目(201810564284)

有效的抽取模型,但是这些方法依赖于人工标注的训练集,并且语料中隐含的知识通过训练过程记忆在模型的参数中,用户无法直接感知和修正^[2]。无监督学习方法主要借助评价对象与评价词之间的语法或语义关系,结合规则和语言学特征直接抽取评价对象—评价词对。由于无监督方法不需要人工标注数据的参与,具有移植性更强等优点,更加适合当前开放的网络环境,因而得到广泛应用^[3]。

本文主要关注利用句法关系的无监督细粒度观点分析研究。句法关系可以反映出句子各成分之间的语义修饰关系,其可以获得长距离的搭配信息,并与句子成分的物理位置无关。因此,近几年对评价对象—评价词对的抽取研究都融入了句法分析^[4]。但是,目前利用句法分析对商品评论中评价对象—评价词对的挖掘还存在以下几点不足:一是几乎完全依靠修饰关系进行抽取,抽取到很多无意义的评价对象(非商品属性词),导致产生了大量的噪声,而在后续筛选环节中由于修饰错误的原因很有可能丢失了有用的评价对象或者评价词,不利于评价对象或评价词词典的扩充和修正;二是多数研究只分析句子中的主谓结构,仅抽取到名词性的商品特征作为评价对象,然而,用户评论中商品特征还存在不少通过动词表达的现象,例如,“散热很快”,“散热”的词性标注虽为动词,但其也同样反映了商品的功能,这种“动词+程度副词+形容词”的表达句式在评论文本中经常出现,但无法通过主谓结构进行抽取;三是多数观点分析的研究仅基于正面和负面的粗极性分类,难以从整体上解释已购买用户群体对商品属性的偏好程度,因此其他用户无法深入了解商品口碑,商家也无法对比自身产品和竞争对手家的差异,而这却是细粒度观点分析的重要价值。

针对上述不足,本文的目标是在中文商品评论中首先抽取出初步的评价对象词典,进而通过句法关系识别评价对象对应的评价词,最后计算评价对象的总体观点得分。本文主要贡献为:

(1) 提出一种基于频繁项集挖掘的评价对象抽取方法,构建初始的评价对象词典,约束后续评价对象—评价词对的抽取范围,从而减少无效的干扰词;

(2) 通过分析评价对象和评价词的可能语义构成,设计评价对象—评价词对的抽取句法规则,并根据词对抽取结果反过来进一步修正和扩充初始的评价对象,获取语义上更完整的评价对象—评价词对知识库;

(3) 基于 HowNet 词典设计了评价对象的观点

得分计算方法,综合考虑了否定副词和程度副词的影响,以精准刻画用户对商品属性的偏好程度。

本文组织结构安排如下:第1节介绍相关的工作;第2节介绍评价对象—评价词的抽取方法,这一节共包括两个部分:评价对象的抽取和评价对象—评价词对的抽取;第3节计算每个评价对象的观点得分,给出相应的计算得分方法;第4节介绍实验步骤和结果;第5节进行总结与展望。

1 相关工作

目前的细粒度观点分析研究主要集中在评价对象—评价词对的抽取方法上,大致可以分为有监督和无监督方法两类。有监督学习方法^[5-9]主要是利用机器学习的思想,将评价对象和评价词当作一个序列标注任务来识别,需要大量的训练集。例如,郁圣卫等^[9]提出一种基于领域情感词典特征表示的细粒度意见挖掘方法,使用领域情感词典在观点文本上构建特征表示并将其加入序列标注模型的输入部分。然而,无论是领域词典还是训练集,都需要不少人工干预,具有一定的领域局限性。

近年来,融入句法分析的无监督方法在细粒度观点分析研究中备受关注,这是由于句法分析可以反映出句子各成分之间的语义修饰关系,因而获得较高的准确率。彭云等^[1]从句法分析、词义理解和语境相关等多角度获取词语间的语义关系,然后将其作为约束知识嵌入到主题模型,提出语义关系约束的主题模型,提升了主题模型细粒度情感极性分类的准确性。由于商品评论具有短文本所表现的特征稀疏性问题,Yan等^[10]提出了一种新的基于句中词对的主题模型 BTM,相对其他已有主题模型体现出了更优越的无监督主题发现性能。李勇敢等^[11]通过构建多级词库、制定成词规则、开展串频统计等给出一种基于规则和统计的新词识别方法,在评价词和评价对象的依存模式的基础上给出基于词语特征的观点句识别算法,实现中文微博情感倾向性自动分类。Xu等^[12]根据评价对象和评价词的依存句法关系建立情感图,接着利用随机游走算法迭代计算评价对象和评价词的置信度,通过自学习策略精炼提取评价对象,最后利用提取的评价对象进一步过滤评价词,该方法取得了较好的准确率和召回率。陈放等^[2]首先利用斯坦福句法分析器获取评论文本的词性标签和依存句法关系,再按照一定的语法规则抽取评价对象—情感词二元搭配,在情感倾向分

类实验中取得较高的准确率。

上述研究较好地推动了商品评论细粒度观点分析的研究进展。然而,大部分相关研究忽略了评价对象中动词的成分,且随着评论数据的增长,将会导致评价对象—评价词词典越发不完备。此外,相关研究在观点倾向性分析中仅考虑评价对象的极性分类(如正面或负面情感),并未计算具体的得分,无法准确描述商品属性的细腻观点表达。为此,本文在抽取更为全面的评价对象—评价词对的基础上,提出一种针对评价对象的观点得分计算方法。

2 评价对象—评价词对抽取

本文以商品评论为研究对象,因此提到的评价对象指的是商品属性。本节中我们首先利用频繁项集挖掘方法构建初始的评价对象词典,而后针对每个评价对象抽取其对应的评价词,再根据词对的抽取结果对初始评价对象进行扩充和修正,最终生成评价对象—评价词对知识库。

2.1 评价对象

本文借鉴文献[13]的思想,认为用户在商品评论文本中更多提到的是商品属性,因此商品属性应该是一种频繁出现的词。此外,不相关的内容在不同的评论文本中也会不相同^[14],因此不频繁出现的词要么并非商品属性,要么属于不重要的属性^[13]。然而,并非所有频繁出现的词都是商品属性。基于此,本文首先通过 FP 增长算法获取初始的频繁项集作为候选评价对象(商品属性),再利用 PrefixSpan 算法进行二次过滤,最终得到评价对象词典。

2.1.1 候选评价对象生成

FP 增长算法是一种用于频繁项集挖掘的算法,其通过频繁构造 FP 树来实现频繁模式的增长,这种分而治之的挖掘策略避免了大量候选集的产生,因此本文采用 FP 增长算法作为抽取候选评价对象的方法。

本文将所有评论文本中词性标注后的名词和动词作为初始候选频繁项集。由于评论文本中包含不少主观认识类动词,本文还根据现代汉语主观词的相关研究,整理了一份如表 1 所示的主观认识类动词词典,将这些词从频繁项集中删除,避免进行无效的计算。最后采用与文献[13]中相同的最小支持度 1% 来过滤某些非商品属性的特殊频繁项集,将该结果作为候选评价对象集合。

表 1 主观认识类动词词典

主观认识类动词
认为、以为、觉得、感觉、你看、我看、你说、我说、你想、我想

2.1.2 候选评价对象过滤

由于 FP 增长算法仅依据词性标注的结果和最小支持度挖掘频繁项集,抽取的候选频繁项集中含有大量噪声,有必要进行二次过滤。本文尝试将 PrefixSpan 算法引入初始候选评价对象集的噪声过滤中。PrefixSpan 算法本身是用于挖掘频繁序列模式的关联算法,将由 FP 算法抽取的候选频繁项集进行简单的片段抽取作为 PrefixSpan 算法的输入,期望能够更精准地定位可能的评价对象在文本中的位置,从而排除大量其他无关噪声的干扰。PrefixSpan 算法与 Apriori 和 FP 增长算法都有类似之处,它从序列数据中长度为 1 的前缀开始挖掘序列模式,不断递归挖掘,每次挖掘都会根据所设定的支持度阈值进行过滤筛选,直到无法挖掘时,递归结束。

本文首先在词性标注后的评论数据中抽取符合匹配要求的文本序列片段。根据中文的语言习惯,本文认为对应评价对象的形容词在一句话中的词距不超过 i 个($0 \leq i \leq 3$)。依据此原则,在序列数据中以“先右后左”的顺序策略进行遍历查找,抽取符合要求的片段作为候选序列集,在此基础上再利用 PrefixSpan 算法进行过滤,根据经验设定名词的支持度阈值仍为 1%,动词的支持度阈值为 1.5%,保留最终挖掘出来的序列模式前缀作为最终的评价对象,部分示例如表 2 所示。

表 2 部分评价对象示例

词性	评价对象
名词	性价比、屏幕、包装、颜值、分辨率、价格、外观、速度、做工、机身
动词	散热、开机、运行、送货、配置、反应

2.2 评价对象—评价词对

评论文本中,评价对象通常与评价词共同出现,因此可以认为评价对象和评价词总是最近距离搭配^[15]。但是目前的评论文本中包含的评价词不仅丰富而且还出现不少网络新词,若仅凭其在句子中的位置来关联距离最近的描述对象,很可能导致关联错误。因此,本文首先依据句法规则来分析评价对象及其对应的评价词之间的依存关系,再结合词

性标注抽取评价对象—评价词对。

本文使用的辅助分析工具是哈尔滨工业大学 LTP 的汉语语言处理模块——词性标注和依存句法分析,部分词性标注和依存句法关系的符号标记及解释如表 3、表 4 所示。抽取出的词对以<评价对象,评价词>形式的二元组表示。

表 3 部分词性标注类型及示例

标记	含义	示例
n	名词	屏幕
v	动词	散热
a	形容词	清晰
d	副词	非常

表 4 部分依存关系类型及示例

标记	含义	示例
SBV	主谓关系	外观好看(外观←好看)
ATT	定中关系	快递速度快(快递←速度)
CMP	动补关系	打游戏卡(打游戏←卡)
ADV	状中关系	屏幕很清晰(很←清晰)

商品属性一般由名词或者名词短语构成,但是由于口语表达随意,会出现省略现象,如“散热性能”在口语中会省略为“散热”,这使得在一般意义上具有动词词性的“散热”成了评价对象的指代词,因此在评论文本中常出现评价对象以动词表达的现象。名词性的评价对象在句子中大多数充当主语成分,与其对应的评价词以形容词为主,在句子中充当谓语成分,因此通过主谓结构可以抽取这部分的评价对象—评价词对。但是动词性的评价对象,其评价词一般以动补成分出现,因此主谓结构的句法规则不适用,需设计动补结构的规则。另外,评价对象和评价词有时还会包含修饰的成分。评价对象前面如果有修饰和限制性的定语成分,那么将该成分与评价对象进行组合或许就构成一个名词性短语,修正原有紧靠主成分抽取的评价对象。评价词前面如果有修饰语,那么该修饰语一般是代表程度偏好的副词,对于计算评价对象的观点倾向得分有一定意义,因此本文考虑将程度副词与评价词一并抽取。

根据上述对评论文本的语法和语义分析,本文提出以下几个抽取规则。

规则 1 在主谓结构中,若评价对象为主语,则抽出其对应的形容词谓语作为评价词。即两者构成主谓关系(SBV),以(n,a,SBV)或(v,a,SBV)表示。

例 1 “包装精美”的词性标注和依存句法分析

如图 1 所示。

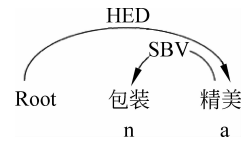


图 1 例 1 的词性标注、依存句法分析

规则 2 当评价对象为动词时,评价词一般表现为动词补足语。寻找存在于动补结构中的评价对象,抽出其动词补足语作为评价词。即两者构成动补结构(CMP),用(v,a,CMP)表示。

例 2 “开机快”的词性标注和依存句法分析如图 2 所示。

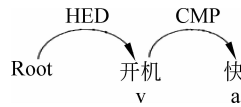


图 2 例 2 的词性标注和依存句法分析

然而,规则 1 和规则 2 抽取出的评价词无法全面准确地表达用户的观点语义,例如,“屏幕非常清晰”,按规则 1 则抽出构成主谓结构的“屏幕—清晰”词对,忽略了副词修饰词,而这正是用户表达观点倾向的程度体现。为此,本文采用 HowNet 的中文程度级别词语词典和现代汉语否定副词词典作为依据,将程度副词和否定副词作为评价词的一部分一并提取,因此补充了以下的规则 3。在词性标注中,为了区分开程度副词和否定副词,本文保留程度副词的 d 标注,而将否定副词另外用 neg 表示。

规则 3 对于符合规则 1 和规则 2 的评价词,若评价词同时有状语 ADV 修饰,需将 ADV 的状语修饰词组合进评价词,用[d(neg),a,ADV]表示。

例 3 “外观非常漂亮”的词性标注和依存句法分析如图 3 所示。

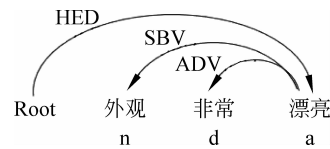


图 3 例 3 的词性标注和依存句法分析

以上的规则 1、2、3 只是针对初始构造的评价对象词典里的商品属性,因而只发现了核心词(名词或动词),却遗漏了评价对象的修饰成分,这可能会造成语义上的缺失。为了使抽取的评价对象更加完整,语义更详实,需要在评价对象词典的基础上,寻找评价对象短语,同时也进一步扩充评价对象词典。

借鉴文献[4]思路,本文提出规则 4。

规则 4 若评价对象有定语 ATT 修饰时,需将 ATT 的非评价词修饰词组合进评价对象,用 (n, n_0, ATT) 、 (v, n_0, ATT) 和 (n, v_0, ATT) 表示。

例 4 “风扇声音大”的词性标注和依存句法分析如图 4 所示。

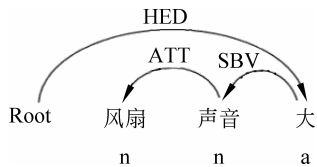


图 4 例 4 的词性标注和依存句法分析

综上,本文设计的总体规则如表 5 所示。

表 5 本文设计的总体规则

规则			示例
依存关系	序号	词性标注	
SBV	(1)	$n+a$ $v+a$	屏幕/n 清晰/a, 散热/v 快/a
	(2)	$n+n_0+a$ $v+n_0+a$ $n+v_0+a$	屏幕/n 分辨率/n 高/a, 运行/v 速度/n 快/a
	(3)	$n+d(neg)+a$ $v+d(neg)+a$	颜值/n 很/d 高/a, 键盘/n 不/neg 顺手/a
	(4)	$n+n_0+d(neg)+a$ $v+n_0+d(neg)+a$ $n+v_0+d(neg)+a$	快递/n 速度/n 太/d 慢/a
CMP	(5)	$v+a$	配置/v 高/a
	(6)	$v+d(neg)+a$	配置/v 不太/d 好/a

另外,评论文本有时不止符合其中的某一条规则,它满足多条规则。如例 5 所示,可能就会抽取符合表 5 中规则(1)的<速度,快>、规则(2)的<快递速度,快>、规则(3)的<速度,挺快>和规则(4)的<快递速度,挺快>。

例 5 “快递速度挺快的”的词性标注和依存句法分析如图 5 所示。

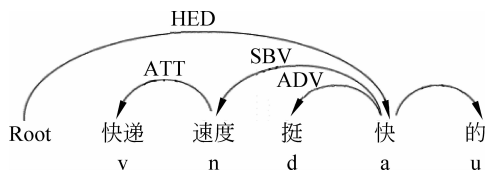


图 5 例 5 的词性标注和依存句法分析

为了避免抽取重复的词对,本文制定了抽取规则的优先级。由于规则(4)包含了扩展评价对象

和评价短语,是最能表达完整观点倾向的词对,所以设置规则(4)为最高优先级;若没有符合规则(4)的词对,再去寻找包含扩展评价对象或评价短语的规则(2)、(3)、(6);最低级别的规则是代表核心规则的规则(1)、(5),即只包含核心词的词对。因此,总体的规则优先级排序为规则(1)、(5)<(2)、(3)、(6)<(4)。部分抽取样例如表 6 所示。

表 6 抽取样例

原始评论文本	抽取结果			
	评价对象	否定词	程度副词	评价词
说实话这个商品不怎么好,画质也没有那么清晰。	商品	不	怎么	好
	画质	没有	那么	清晰
电脑挺好的,就是散热有点慢。	电脑	—	挺	好
	散热	—	有点	慢
很精致的笔记本,性能不错,价格也合适。	性能	—	—	不错
	价格	—	—	合适

3 观点得分计算

本节将利用 HowNet 的义原层次体系结构计算每个评价对象的观点得分。

3.1 评分方法

由于抽取的评价对象—评价词对形如“<‘屏幕’: [‘略/d’, ‘小/a’]>”的表示方式,为了综合评价对象的所有观点,本文首先对所有的评价对象进行预处理,将含有相同评价对象的词对进行合并,合并后的结果形如“{‘内存’: [[‘略/d’, ‘小/a’], [‘很/d’, ‘给力/a’], [‘没有/a’], [‘高/a’], [‘有点/d’, ‘少/a’]……]}”,并将合并后的结果称为组合标签,其由一个评价对象和多个评价词组成。本文针对每一个组合标签,计算所有评价词得分,以此作为评价对象的得分,即商品属性观点得分,该得分可正可负,分别表示正向或负向态度倾向,如式(1)所示。

$$\text{Score}(\text{Target}) = \frac{\sum_{k=1}^n \text{Score}(\text{positive}_k)}{n} + \frac{\sum_{h=1}^m \text{Score}(\text{negative}_h)}{m} \quad (1)$$

其中, $\text{Score}(\text{Target})$ 表示评价对象 Target 的得分, $\text{Score}(\text{positive}_k)$ 和 $\text{Score}(\text{negative}_h)$ 分别表示

第 k 个得分为正的评价词和第 h 个得分为负的评价词, n 是评价词得分为正的个数, m 是评价词得分为负的个数。

由于评价词中包含了否定副词、程度副词和形容词, 于是本文提出的观点得分计算方法如式(2)所示。

$$\text{Score}(\text{opinion}) = \text{Sign}(\text{neg}) \times \text{Score}(d) \times \text{Score}(a) \quad (2)$$

其中, $\text{Sign}(\text{neg})$ 为否定词标记, $\text{Score}(d)$ 为程度副词得分, $\text{Score}(a)$ 为形容词得分。

3.2 形容词极性值计算

本文借鉴刘群等人提出的方法^[16], 基于 HowNet 计算评价词中形容词的极性值。HowNet 使用一系列的“义原”来对每一个“概念”进行描述。所有义原之间组成的是一个复杂的网状结构。义原关系中常用的是上下位关系, 根据上下位关系可构成一个如图 6 所示的树状结构义原层次体系。

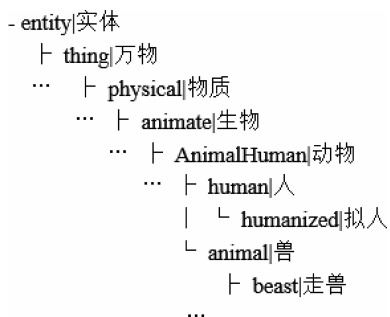


图 6 HowNet 义原层次结构^[16]

本文也通过计算两个义原在如图 6 所示的层次体系中的路径距离来作为义原间的相似度, 计算如式(3)所示^[16]。

$$\text{Sim}(p_1, p_2) = \frac{\alpha}{d + \alpha} \quad (3)$$

其中, p_1 和 p_2 表示两个义原(primitive), d 是 p_1 和 p_2 在义原层次体系中的路径长度, 是一个正整数。 α 是一个可调节的参数, 本文中 α 的取值为 2.0。

由于每个词都可以通过几个概念来描述, 而所有的概念都最终归结于用义原来表示, 所以在式(3)的基础上, 可以得到两个词语相似度计算的方法, 如式(4)所示^[16]。

$$\text{Sim}(W_1, W_2) = \max_{i=1 \dots n, j=1 \dots m} \text{Sim}(S_{1i}, S_{2j}) \quad (4)$$

其中, W_1 和 W_2 表示两个词语, W_1 有 n 个概念: $S_{11}, S_{12}, \dots, S_{1n}$; W_2 有 m 个概念: $S_{21}, S_{22}, S_{23}, \dots, S_{2m}$ 。这里获取 W_1 和 W_2 的相似度为各个概念的相似度之最大值。通过这个方式就可以把两个词语之间的相似度问题归结到了两个概念之间的相似度问题, 再通过式(3)将两个概念之间的相似度问题归结到了两个义原的相似度问题, 逐层分解进行计算。

确定词语之间的相似度计算方法后, 本文基于 HowNet 词典分别构建了一个褒义基准词库 $\text{Set1} = \{\text{commendatory}_1, \text{commendatory}_2, \dots, \text{commendatory}_n\}$ 以及一个贬义基准词库 $\text{Set2} = \{\text{derogatory}_1, \text{derogatory}_2, \dots, \text{derogatory}_m\}$, 并基于上一步提出的相似度计算方法计算形容词的得分。对于某个评价词 word, 本文提出其得分计算如式(5)所示。

$$\text{SO_HowNet}(\text{word}) = \sum_{i=1}^n \text{Sim}(\text{word}, \text{commendatory}_i) - \sum_{j=1}^m \text{Sim}(\text{word}, \text{derogatory}_j) \quad (5)$$

其中, $\text{SO_HowNet}(\text{word})$ 代表 word 的得分值, $\text{Sim}(\text{word}, \text{commendatory}_i)$ 代表 word 与第 i 个褒义基准词的相似度, $\text{Sim}(\text{word}, \text{derogatory}_j)$ 代表 word 与第 j 个贬义基准词的相似度。由该公式可知, 如果评价词 word 与所有褒义基准词相似且和所有贬义基准词不相似, 式(5)中的第一个子式值为 1, 第二个子式值为 0, 反之第一个子式值为 0, 第二个子式值为 1, 因此形容词的得分范围在 $[-1, 1]$ 之间。

3.3 否定副词与程度副词

否定副词和程度副词作为修饰词, 在评论文本的态度倾向中也起着关键作用。否定副词的处理则通过寻找评价词中含有“neg”标注的词语来判断该评价词中是否含有否定副词, 再对其评分; 对于程度副词, 由于抽取评价词时使用的判定标准是 HowNet 的中文程度级别词语词典, 比较泛化, 为了实现程度副词更加合理的分类, 本文还使用董丽丽等^[17]构建的程度副词词典来对已有程度副词进行二次过滤。

为了计算 $\text{Sign}(\text{neg})$ 和 $\text{Score}(d)$, 本文给每一个修饰词赋予一个强度值 $\text{Strength}(\text{word})$, 将此强度值赋值给 $\text{Sign}(\text{neg})$ 和 $\text{Score}(d)$ 。只要在相同的基准下计算态度倾向得分, 这些强度的取值将不会改变不同评论文本的倾向差异。本文借鉴文献^[17]对强度值的设计, 对程度副词进行了四个等级的划

分,即“低量”“中量”“高量”“极量”。其中,“中量”是基础等级,首先设其强度值为 1;而后将程度弱于基础等级的模糊量化为其 50%,即“低量”的强度为 0.5;接着将程度高于基础等级的量化为其 50%的增量,即“高量”的强度为 1.5;最后将程度特别高的模糊量化为基础等级二倍的增量,即“极量”的强度为 2。

至此便可根据式(2)计算 $\text{Score}(\text{positive})$ 和 $\text{Score}(\text{negative})$,进而根据式(1)计算出 $\text{Score}(\text{Target})$ 。

4 实验

4.1 实验数据

本文从京东、淘宝等电商网站中爬取 22 465 条评论文本数据作为初始的实验文本集,去掉重复文本、不超过 5 个字的短文本及其他无效的评论文本如“系统默认好评”“此用户未填写评价内容”等,得到有效评论 12 873 条,其中,评分为 3、4、5 颗星的评论共 7 572 条,3 颗星以下的评论共 5 301 条。

4.2 数据预处理

本文使用结巴分词工具对评论文本进行分词,由于文本中出现颜表情的频率较高,本文加入自定义词典调整结巴分词工具的分词策略,使特定的词语被正确分开。分词后去无效字符,包括常用语气词、连词、拟声词、电商评论中常出现的 HTML 标记的无意义代码符号、颜表情、特定节日名词等。

4.3 实验结果

数据预处理之后,我们组织了 5 个本科生组成数据标注小组,分别对数据集的每条评论进行人工标注,最后采用一致性最高的结果作为标准(数据规模如表 7 所示),以验证本文方法的有效性。评价指标采用了准确率(precision)、召回率(recall)以及 F_1 值,三者的计算如式(6)~式(8)所示。

$$P = \frac{|A \cap B|}{|A|} \quad (6)$$

$$R = \frac{|A \cap B|}{|B|} \quad (7)$$

$$F1 = \frac{2PR}{P+R} \quad (8)$$

其中, $|A|$ 表示通过方法抽取到的评价对象或评价对象—评价词总数, $|B|$ 代表人工标注的标准

评价对象或评价对象—评价词总数, $|A \cap B|$ 代表本文方法抽取的与标准集完全匹配的评价对象或评价对象—评价词总数。

表 7 标准数据集规模

类别	个数
评价对象	40
评价词	1 239
评价对象—评价词	2 229

4.3.1 评价对象抽取实验

本次实验共分为三组:第一组是评估 PrefixSpan 算法过滤噪声的有效性;第二组是验证加入动词的有效性;第三组是与其他工作的对比。

(1) 噪声过滤实验

由于直接利用 FP 频繁项集挖掘方法会产生过多的噪声,为此本文引入了 PrefixSpan 算法进行二次过滤。此次实验同时抽取了名词和动词,实验结果如表 8 所示。

表 8 噪声过滤实验结果

词性	$P/\%$	$R/\%$	$F_1/\%$
FP (名词+动词)	10.77	100	19.45
FP+PrefixSpan (名词+动词)	77.78	87.50	82.35

从实验结果来看,在 FP 算法的基础上,通过 PrefixSpan 算法进行噪声过滤,能够显著提升 F_1 值,这表明该过滤方法有显著效果。

(2) 扩充动词实验

本次抽取的评价对象是在频繁项集挖掘结果的基础上,结合评价对象—评价词对的抽取结果,进一步修正整合得到的评价对象,性能如表 9 所示。

表 9 扩充动词实验结果

词性	$P/\%$	$R/\%$	$F_1/\%$
名词	78.79	86.67	82.54
名词+动词	81.13	89.58	85.15

从实验结果来看,通过本文方法扩充的动词,能够在一定程度上弥补仅仅利用名词作为评价对象的不足。此外,对比表 8 的(名词+动词)抽取结果,表 9 中同样词性的抽取结果在准确率和召回率上都有所提升。这表明本文设计的句法规则的抽取结果能够有效地修正和扩充初始的评价对象词典。原因

是结合句法规则的结果可以发现短语类型的评价对象(如名词短语“快递速度”、动词短语“开机反应”)。然而,如果仅仅根据句法规则来抽取评价对象,很难准确定位评价对象的核心词,但其能够对核心词的修饰成分抽取提供依据。上述结果表明融合频繁项集挖掘和句法规则能够获取语义上更加完整的评价对象,从而避免了指向“似是而非”的评价对象,为更精准地细粒度观点分析提供良好的基础。

(3) 与其他工作的对比

我们将本文方法得到的实验结果与其他无监督方法进行对比,结果如表 10 所示。其中,BTM 方法来自于文献[10],Ours 对应表 9(名词+动词)的实验结果(即本文获得的最佳性能)。BTM 方法很好地解决了短文本特征稀疏问题,在无监督获取短文本主题词时取得领域内优越的性能,因此本文选择其作为对比方法。BTM 参数取 Top 100 个词,主题数 $k=50$,这是我们复现代码时经多次测试取得最好结果的一组参数。该方法的准确率计算如式(9)所示。

$$P = (\sum_{i=1}^k N_i) / (40 * k) \quad (9)$$

其中, k 为主题数, N_i 表示第 i 个主题内捕获到的准确评价对象总数,40 为人工标注的评价对象总数(表 7)。召回率为所有主题内捕获到的准确评价对象总数(去重)除以人工标注的评价对象总数 40。

表 10 对比实验结果

方法	$P/\%$	$R/\%$	$F_1/\%$
BTM ^[10]	78.15	75	76.54
Ours	81.13	89.58	85.15

表 10 的对比实验结果表明,BTM 的准确率和召回率都低于本文模型,原因是 BTM 抽取词的原理是基于文本中词共现的概率,它将同时在一个评论文本中出现的高频词对纳入到同一主题中,导致该方法捕获到的主题词包含较多的噪声。

4.3.2 评价对象—评价词对抽取实验

本次实验主要评估规则的有效性、评价对象—评价词对抽取的准确率、规则和数据集的适用性。

(1) 本文方法实验结果

本文设计的规则覆盖率如表 11 所示。由于规则 3 与规则 4 是在规则 1 与规则 2 抽取的基础上为了进一步扩展评价对象和评价词而建立的,只是辅助规则,并不是新的规则,故不考虑其在原生评论上的覆盖率。

由表 11 可知,本文基于 SBV 句法关系设计的规

则 1 在本文语料上的覆盖率达到 75.91%,这表明本文提取的大部分评价对象和评价词的关系符合 SBV 句法关系;本文设计的规则 2 在规则 1 的基础上完善了动词评价对象和评价词对的抽取,构成了比较完备的基础词对抽取规则,提升了覆盖率,达到了 81.56%;但是还有将近 20% 的评论没有覆盖到,这是由于部分评论用语大多是简短、口语化的句子,这给句法分析带来一定挑战,因此难以被规则所覆盖。

表 11 本文规则在本文语料上的覆盖率

规则	覆盖率/%
规则 1	75.91
规则 2	5.65
规则 1+规则 2	81.56

表 12 给出了评价对象“屏幕”对应评价词搭配频数排序(此处忽略了程度副词)。

表 12 评价对象“屏幕”的评价词搭配排序

评价对象—评价词对	频数
<屏幕,清晰>	22
<屏幕,大>	17
<屏幕,好>	8
<屏幕,模糊>	8
<屏幕,差>	7

表 13 给出了通过本文方法获取到的部分评价对象—评价词对的准确率。

表 13 部分评价对象—评价词对抽取结果(本文)

评价对象	评价词个数	$P/\%$
做工	30	93.3
运行	28	92.9
散热	42	90.5
颜值	26	88.4
屏幕	91	87.9
内存	25	76
硬盘	15	73
鼠标	8	62.5

由表 13 可知,动词性评价对象“做工”“散热”“运行”抽取到的评价词准确率较高,能达到 90% 以上,这表明本文对动词类的评价对象—评价词对抽取方法是有效的。其余比较常见的名词性评价对

象—评价词对(如“颜值”“屏幕”)的抽取准确率也达到 85%以上,但是涉及到一些比较具体的硬件(如“内存”“硬盘”“鼠标”)时抽取准确率有所下降,原因是这部分的评价词简短、灵活多变,在分词、词性标注和依存句法分析的一个或多个环节中遭受破坏,因而无法识别。

(2) 与其他工作的对比

我们还选择了郁圣卫等^[9]提出的有监督方法 BiLSTM-CRF 在本文数据集上进行实验,结果如表 14 所示。该方法需要人工标注大量的数据,由于标注方式的差异,结果可能与原作者所提的方法相比会有所差异。

表 14 部分评价对象—评价词对抽取结果(文献[9])

评价对象	评价词个数	P/%
做工	172	66.93
运行	158	74.88
散热	267	79.23
颜值	170	85.00
屏幕	173	56.54
内存	44	68.75
硬盘	29	65.91
鼠标	69	86.25

对比表 14 和表 13,我们发现文献[9]中的有监督方法,获取到的评价词数目在 8 个评价对象上都比本文增加不少,但是对应每个评价对象所捕捉到的评价词的准确率反而出现明显下降(除“鼠标”这个对象外)。原因是该方法出现了大量的评价对象漏检情况,导致无法准确配对,从而在 8 个对象中有 7 个对象的配对准确率明显低于本文的方法。

(3) 规则和数据集适用性实例分析

为了验证本文提出的规则在其他数据集上的适用性,我们还选择来自 ChineseNLPcorpus 的公开语料——手机评论数据集进行实验,规则的覆盖率如表 15 所示。从结果可以看出,本文的规则在其他语料上也具有一定的适用性。

表 15 本文规则在其他语料上的覆盖率

规则	覆盖率/%
规则 1	75.50
规则 2	19.75
规则 1+规则 2	78.31

另外,我们还选择了文献[18]所提规则在本文数据集上进行测试,结果如表 16 所示。

表 16 其他规则在本文语料上的覆盖率

规则	覆盖率/%
规则 1-6(SBV 关系)	54.80
规则 4、9、10 (SBV、ATT)+VOB 关系)	2.84
规则 7、8(AVD、DE 关系)	6.99
总覆盖率	64.63

由表 16 可知,本文基于 SBV 关系设计的规则 1 覆盖率更高(表 11),规则更完善;文献[18]使用了 VOB 与其他关系组合来间接抽取词对,覆盖率只有 2.84%,而本文基于 CMP 关系设计的规则 2 覆盖率达到 5.65%(表 11);另外,使用 ADV、DE 规则抽取出的词对虽然覆盖率接近 7%,但是通过对抽取结果的进一步分析,发现这两个规则抽取出的词对并不完整,无法进行后续的属性评分。总体来看,本文规则的覆盖率更高,并且本文还补充了其他扩展规则,其适用性更广。

4.3.3 观点得分实例分析

为了评估所提出的观点计算方法的有效性,本文选择两个电脑品牌“联想”和“惠普”进行细粒度口碑对比,由于评价对象数量众多,这里只选择 10 个常见的商品属性进行对比,结果如图 7 所示,图中横坐标表示商品属性,纵坐标表示已购买用户对该属性的观点倾向得分,根据式(5)可知,正分值意味着用户的好评数量比差评数量多,反之则出现负分值。

从图 7 可以看出,惠普的内存、屏幕、硬盘、运行等属性的得分均比联想的得分高,但从鼠标、做工、散热等属性得分可以看出,联想笔记本更胜一筹。从总体上来看,惠普笔记本的硬件设备受到更多的好评。我们还挑选了一些属性进一步调研,发现:关于“内存”,联想用户大多存在加内存的情况,因此评论中对于内存的评价比惠普稍差;“颜值”方面,联想的小新系列外观得到许多人的青睐,而惠普外观的设计也层出不穷,两者差别不大;对于“价格”来说,惠普的价格比较亲民,性价比高,因此得到更多好评。

另外,本文在各个评价对象上计算的观点得分,基本和人工在对应评价对象上标注的好评、差评数量趋势相符,说明该计算方法能够在一定程度上体现用户对商品属性的偏好程度。

综上可知,相比正负极性的粗粒度商品评论情感分类,本文对商品属性的观点评分,能够综合性地获

取已购买用户群体对商品属性的偏好程度,尤其方便商家更加清晰地对比自身产品和竞争对手家的差异。

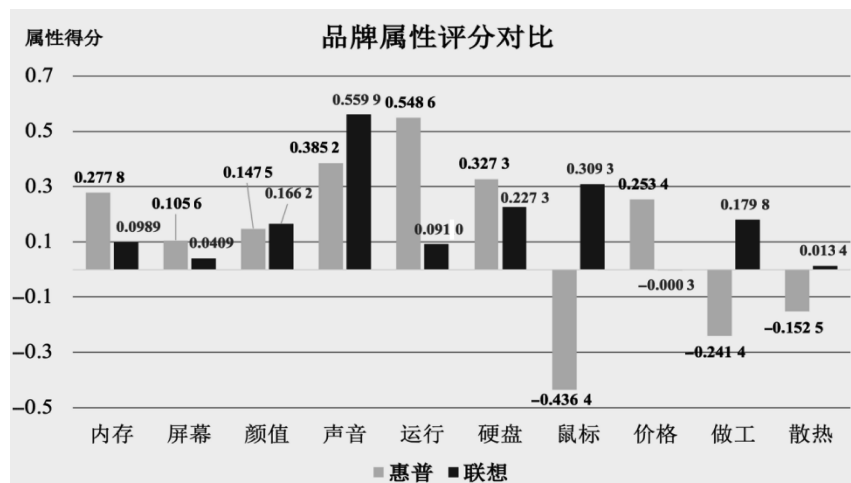


图7 品牌属性评分对比

5 结论

本文提出了一种基于句法规则和 HowNet 的商品评论细粒度观点分析方法。首先,基于频繁项集挖掘方法构建初始的评价对象词典,便于后续评价对象—评价词对抽取时重用。其次,基于词性标注和句法规则设计了评价对象—评价词对的抽取规则,根据抽取结果还能够对初始的评价对象进行进一步的扩充和修正。最后借助 HowNet 词典计算了每个评价对象的总体观点得分。本文的方法无须人工标注数据,充分利用评论语料资源,除了一般性的名词评价对象,还扩充了动词性词语作为评价对象。我们还发现,通过融合频繁项集挖掘和句法规则的方法有助于获取到语义更完整的评价对象—评价词对知识库。另外,本文的观点评分能够为用户商品购买或者商家营销管理提供一定的决策依据。

本文提出的规则主要基于电商评论进行设计,尽管不同种类的商品评论在用语词汇上有着变化,但受中文用语习惯的限定,其句式基本相同,遵循一定的句法规则,本文虽仅选取了两个语料集作为案例,但是对于其他种类的商品评论也是适用的,可以拓展到其他数据集上。然而,本文设计的规则主要基于中文语境进行设计,因此在跨语言分析时可能会因语法的不同而使本文的规则不适应。

后续的研究工作主要从以下几个方面考虑: ①进一步提取隐式评价对象,完善评价对象词典; ②根据评论语料特点,对规则进行扩展与完善,提高规

则的覆盖率; ③考虑观点得分与评论星级之间的关联关系,进一步增强评价对象观点得分的置信度; ④探索更好的对未登录评价词计算得分的方法。

6 致谢

本文研究工作还有本科生陈嘉棋、江郭远、梁嘉晴、叶海铷的参与,在此对他们表示感谢。另外,感谢文献[9]作者提供源代码让本文的对比实验更加充分,感谢审稿人提供的宝贵意见。

参考文献

- [1] 彭云,万红新,钟林辉.一种语义弱监督 LDA 的商品评论细粒度情感分析算法[J]. 小型微型计算机系统, 2018,39(5): 116-123.
- [2] 陈放,王颖,梁爽,等.面向商品评论的二元情感认知模型[J]. 中文信息学报,2018,32(8): 135-142.
- [3] 廖祥文,陈兴俊,魏晶晶,等.基于多层关系图模型的中文评价对象与评价词抽取方法[J]. 自动化学报,2017, 43(3): 462-471.
- [4] 江腾蛟,万常选,刘德喜,等.基于语义分析的评价对象-情感词对抽取[J]. 计算机学报. 2017,40(03): 617-633.
- [5] Wei Wei J A G. Sentiment learning on product reviews via sentiment ontology tree[C]//Proceedings of ACL, 2010: 404-413.
- [6] Jiang L, Yu M, Zhou M, et al. Target-dependent Twitter sentiment classification[C]//Proceedings of ACL, 2011: 151-160.

- [7] Bauman K, Liu B, Tuzhilin A. Aspect based recommendations: Recommending items with the most valuable aspects based on user reviews [C]//Proceedings of SIGKDD, 2017: 717-725.
- [8] Ma Y, Peng H, Cambria E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM [C]//Proceedings of AAAI, 2018: 5876-5883.
- [9] 郁圣卫, 卢奇, 陈文亮. 基于领域情感词典特征表示的细粒度意见挖掘[J]. 中文信息学报, 2019, 33(02): 112-121.
- [10] Yan X, Guo J, Lan Y, et al. A biterm topic model for short texts [C]//Proceedings of WWW, 2013: 1445-1455.
- [11] 李勇敢, 周学广, 孙艳, 等. 中文微博情感分析研究与实现[J]. 软件学报, 2017, 28(12): 3183-3205.
- [12] Xu L, Liu K, Lai S, et al. Mining opinion words and opinion targets in a two-stage framework [C]//Proceedings of ACL, 2013: 1764-1773.
- [13] Moghaddam S, Ester M. Opinion digger: An unsupervised opinion miner from unstructured product reviews [C]//Proceedings of CIKM, 2010: 1825-1828.
- [14] Liu B. Web data mining: Exploring hyperlinks, contents, and usage data [M]. Springer, 2007.
- [15] Hu M, Liu B. Mining and summarizing customer reviews [C]//Proceedings of SIGKDD, 2004: 168-177.
- [16] 刘群, 李素建. 基于《知网》的词汇语义相似度计算 [J]. 中文计算语言学, 2002, 7(2): 59-76.
- [17] 董丽丽, 赵繁荣, 张翔. 基于领域本体、情感词典的商品评论倾向性分析 [J]. 计算机应用与软件, 2014, 31(12): 104-108.
- [18] 丁晟春, 侯琳琳, 王颖. 基于电商数据的产品知识图谱构建研究 [J]. 数据分析与知识发现, 2019, 3(03): 45-56.



韦婷婷(1986—), 博士, 讲师, 主要研究领域为自然语言处理。

E-mail: weitingting@scau.edu.cn



胡勇军(1978—), 通信作者, 博士, 讲师, 主要研究领域为自然语言处理。

E-mail: hyjsdu96@126.com



陈伟生(1997—), 本科, 主要研究领域为自然语言处理。

E-mail: cwsccmail@163.com