

基于情感计算的商品评论分析系统

林钦和 刘 钢* 陈荣华

(复旦大学软件学院 上海 200433)

摘 要 针对电子商务中的商品评论信息过载问题,运用情感计算理论,通过挖掘商品评论信息中的商品特征及相应的情感褒贬态度,为消费者提供一个商品特征粒度上的情感分析结果,从而帮助消费者从庞杂的商品评论中快速获取有效信息。系统首先采集指定商品的评论集并挖掘商品特征,然后结合情感语料库和词汇相似度计算,利用依存关系找到特征-极性词对以及程度副词和否定词。基于以上结果,考虑程度副词的强度,以及程度副词和否定词共现语序不同造成的语义差异,提出了商品评论情感倾向程度的计算方式。最后,进行系统实现并验证算法的有效性。实验结果表明,系统具有良好的应用效果。

关键词 电子商务 商品评论 关联规则 句法分析 情感计算

中图分类号 TP3 文献标识码 A DOI:10.3969/j.issn.1000-386x.2014.12.009

A PRODUCT REVIEWS ANALYSIS SYSTEM BASED ON AFFECTIVE COMPUTING

Lin Qinhe Liu Gang* Chen Ronghua

(School of Software, Fudan University, Shanghai 200433, China)

Abstract To solve the product review information overload problem in e-commerce, affective computing theory is applied to provide a sentiment analysis result for consumers in product feature granularity by mining the product features from product reviews information and the corresponding praise and derogatory sentiment attitude so that to help the consumers get useful information rapidly from large amount of product reviews. The system first collects the assigned product reviews set and mines the product features, then performs calculation in conjunction with emotional corpus and word similarity, and finds out the feature-polarity word pair as well as degree adverb and negatives according to dependence relation. Based on the above results, considering the strength of degree adverb and the semantic differences caused by different word order of co-occurring of degree adverb and negatives, we present the calculation means of the sentiment inclination degree of product reviews. Finally, the system is implemented and the effectiveness of the algorithm is verified. Experimental result demonstrates that the system has good applied effect.

Keywords E-commerce Product reviews Associate rule Syntax analysis Affective computing

0 引 言

随着电子商务的迅猛发展,越来越多的消费者选择网上购物。Web2.0 概念的兴起,消费者积极在互联网上发布商品评论,而潜在消费者可以从商品评论中获取信息,辅助购物决策。与此同时,海量且庞杂的商品评论带来了信息过载的问题。如何使潜在消费者快速有效地浏览商品评论,成为新需求。在当前主流电子商务网站上,消费者可以对评论进行有用性投票,并可以根据发表时间等进行排序,从而快速浏览 Top N 的商品评论,但该方法略显粗糙,无法全面有效地利用商品评论信息。

针对消费者商品评论信息过载问题,主要有以下研究工作:(1)商品评论推荐。采用评论者的专业度、写作风格、评论时间^[1]或者结构特征、词汇特征、语义特征^[2]等特征来构建和训练评论有用性预测模型。通过商品评论推荐,消费者可以优先看到高质量的商品评论。(2)商品评论过滤。主要研究如何甄别垃圾商品评论,保留真实商品评论。垃圾商品评论过滤方法主要有围绕用户行为的模式识别、基于评论内容的语义分析和

基于打分行为的聚类等^[3]。通过商品评论过滤,可以让消费者集中精力在真实可靠的商品评论上。(3)商品评论分类。主要有主客观分类、褒贬极性分类等。主观语言是指消费者用来表达自己的态度和感情的语言。识别的主要任务是确定主观词语或语句,如针对主观性形容词、名词、动词的自动识别^[4]。褒贬分类的目标是挖掘评论对商品的总体褒贬态度,主要研究可分为基于传统文本分类技术的方法和基于极性词统计的方法。前者主要使用向量空间模型表示文本,并使用机器学习算法预测类别;后者首先识别词语的褒贬性,采用计数或词频加权求和,获得总体褒贬性^[11]。通过商品评论分类,消费者可以快速查看指定类别的商品评论信息。此外,解决消费者商品评论信息过载的方法还有商品评论的归纳和检索等^[5,16]。

针对消费者商品评论信息过载的问题及研究现状,本文提出了基于情感计算的商品评论分析系统 ReviewHelper,旨在为消费者提供商品特征粒度上的情感分析结果,将庞杂的商品评

论信息转化为结构化的商品特征褒贬信息,并加以归纳总结,从而解决消费者商品评论信息过载的问题。

本文的主要内容如下:

(1) 基于模块化设计思想,提出了基于情感计算的商品评论分析系统框架。

(2) 介绍基于关联算法挖掘商品特征的算法,通过该算法挖掘商品特征,同时实现领域词典的半自动化更新。

(3) 介绍基于依存关系的极性词、程度副词和否定的挖掘算法。

(4) 提出商品评论的情感倾向程度的计算方式,并着重考虑了否定词的作用,程度副词的强度差距,以及程度副词、否定词共现语序不同造成的语义差异。

(5) 系统实现及效果展示,结合真实的电子商务网站商品评论数据进行实验分析。

1 系统框架

ReviewHelper 商品评论分析系统包含四大功能模块:商品评论采集模块、商品特征挖掘模块、评论倾向分析模块和评论分析结果的展示模块,如图 1 所示。每个模块对应系统处理流程的一个步骤,左侧为该模块的输出结果。

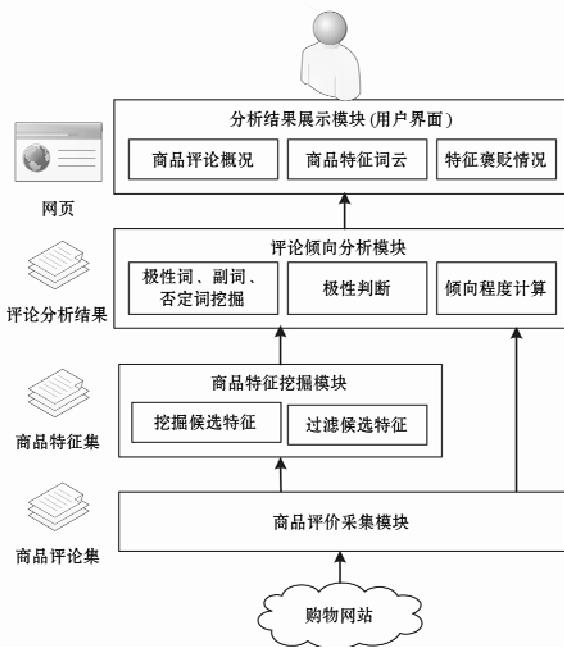


图 1 ReviewHelper 系统框架图

步骤 1 商品评论采集模块根据消费者的购物网址,自动采集和整理商品评论数据集,即第 2.1 节中定义的商品评论集 R 。该模块为后续的特征挖掘和倾向分析提供数据基础。

步骤 2 商品特征挖掘模块基于采集到的商品评论集 R ,利用第 2.2 节的算法挖掘商品特征。商品特征作为消费者的评论对象,是分析消费者对商品具体特征所持褒贬态度的前提。因此,商品特征挖掘的准确性和全面性非常重要。

步骤 3 评论倾向分析模块基于上一个模块得到的商品特征,利用第 2.3 节的算法从商品评论集中获得商品特征-极性词对,及相应的程度副词和否定词,并分析情感极性,逐级计算情感倾向程度。极性词的抽取和倾向判断是商品评论倾向分析的前提和基础,其准确性会影响商品特征及整个商品评论的褒贬

分析。

步骤 4 分析结果展示模块针对用户的浏览需求,组织商品评论分析结果进行页面展示。页面展示包括商品评价概况、商品特征词云和不同商品特征的褒贬评价情况,并允许用户与商品特征之间进行交互。用户可以选择感兴趣的特征,从而查看该特征的情感褒贬情况以及相关的商品评论。

2 关键算法设计

2.1 基本定义

商品特征挖掘和评论倾向分析的基础是商品评论集。 $R = \{r_1, r_2, \dots, r_m\}$ 表示某个商品的评论集合,其中 $m = |R|$ 是商品评论集包含的评论个数。 $F(R) = \{f_1, f_2, \dots, f_n\}$ 表示评论集 R 中的商品特征集合,其中 n 是特征数量。单个特征评价元组 $o = (feature, oWord, dWord, nWord, isNDP)$,其中 $feature$ 为特征词, $oWord$ 为极性词, $dWord$ 为程度副词, $nWord$ 为否定词, $isNDP$ 表示否定词和程度副词的语序关系。每条评论对应的特征评价分析结果集 $ra = \{o_1, o_2, \dots, o_j\}$,其中 j 为该条评论包含的特征评价元组数量。

基于以上定义,下面首先介绍基于关联规则的商品特征挖掘算法的主要步骤,接着重点讨论基于情感计算的商品评论倾向分析算法,其中包括极性词、程度副词和否定词的挖掘,以及情感倾向程度的计算方式。

2.2 基于关联规则的商品特征挖掘

商品特征挖掘的一般方法是利用关联规则挖掘频繁特征^[5,6,10],但没有考虑评论对象的句法结构和领域相关性,产生较多噪音;而基于模板的特征提取方法,需耗费大量人力^[15];也有学者利用点互信息和最大熵等进行特征提取,准确率有所提高,但计算量巨大。

本文采用已有的关联规则算法首先得到候选特征集,接着根据依存关系过滤非短语特征,通过最小独立支持度过滤冗余特征。在此基础上,本文引入自定义领域词典,补充对非频繁特征词的挖掘,从而获取更全面准确的商品特征。同时通过对挖掘出的频繁特征进行反馈实现领域特征词典的半自动化更新,节省了大量人力。

(1) 利用关联规则挖掘候选特征集

首先对评论集 R 的评论 r 逐一进行分词和词性标注。将评论的名词成分作为基本事务,过滤单字名词^[6],创建基本事务集 $D = \{(TID, Noun(r_{TD}))\}$ 。然后,使用 Apriori 算法得到候选商品特征频繁 1-项集 L_1 和频繁 2-项集 L_2

(2) 过滤非短语特征

关联规则的频繁项集挖掘算法不考虑名词在评论中的位置关系,因此频繁 2-项集的两个名词可能无法组成短语。考虑关系为 NP 的名词短语^[9],过滤后得到新的频繁 2-项集的集合 L'_2 。

(3) 过滤高频非领域特征和补充低频特征

该步骤从候选商品特征集 L_1 中过滤掉包含在非领域特征词典中的频繁 1-项集,同时补充包含在领域特征词典中的非频繁 1-项集的集合 L'_1 。

(4) 过滤冗余特征

最小独立支持度是指包含该特征词的频繁 1-项集的支持度减去包含该词的频繁 2-项集的绝对支持度。本文设置最小

独立支持度阈值为 2,若小于该阈值则过滤掉。最终得到满足独立支持度约束的频繁 1-项集的集合 L_1'' 。

(5) 更新领域特征词典和领域非特征词典

该步骤通过人工反馈更新词典。用户可以自由选择是否对获得的高频非登录词列表进行判断:若确认该词为领域特征,则该词属于领域特征词典的权重增加;反之,则该词属于非领域特征词典的权重增加。根据累计的多用户反馈综合计算,决定该候选特征的词典归属。该步骤非必选。种子词典经过循环迭代后,可以大大提升同类商品评论特征的挖掘性能。人工反馈可以由领域专家做,也可以将其整合到用户界面中,通过不断得到用户反馈来更新词典。

经过商品候选特征挖掘、剪枝过滤和非频繁项集的补充后,最终得到商品特征集 $F(R) = L_2' \cup L_1' \cup L_1''$ 。

2.3 基于情感计算的商品评论倾向分析

情感计算研究试图让计算机能够感知、识别和理解人的情感。本文从极性词出发,让系统自动识别商品评论的情感倾向。极性词是指句子中带有感情色彩的词语,用来表达评论发表者的态度。

极性词的倾向判断主要分为基于词典和基于语料库的方法。前者主要利用词典来判断待定词和基准极性词的词汇相似程度,如在给定种子极性词后,利用 WordNet 查找同义词和近义词扩大情感词词典^[16];后者则主要利用待定词和基准词的共现模式或修饰模式,如利用点对互信息^[17]、句法依存关系和启发规则等。在特征和极性词的关系挖掘上,主要方法有句法结构(评论对象和极性词的句法依存关系)、距离窗口(选取商品特征一定距离内的极性词)、基于模板和启发规则等^[8,9,10]。

本节在商品特征集合 F 的基础上,重新遍历商品评论集 R ,综合词典和句法分析两种方法,采用 HowNet 情感词语库和依存关系来挖掘极性词及与商品特征的关系,并结合词汇相似度计算^[18]和同义词词林识别未收录词的情感极性。在此基础上,考虑程度副词的强度差距、以及程度副词和否定词共现语序引起的语义差异,最后逐级计算情感倾向程度。词汇相似度计算和同义词词林的引入,提高了极性词识别的查全率。而程度副词和否定词的考虑,则提高了情感倾向判断的准确性。

2.3.1 基于依存关系挖掘极性词

本节首先对商品评论集的每条评论进行句法分析,得到句法分析树,并基于以下依存关系进行极性词挖掘:从 dobj 结构挖掘极性动词,从 nsubj 和 rcmode 中挖掘极性形容词。

(1) 例句:“我喜欢这本书的纸张质量”

依存关系为:

```
...
assm(这本书 -3, 的 -4)
nn(质量 -6, 纸张 -5)
dobj(喜欢 -2, 质量 -6)
```

特征词:质量

正面情感词语:喜欢

结论:动宾结构,在 Stanford parser 中的类型为 dobj(直接宾语)。

(2) 例句:“书的包装不错”

依存关系为:

```
assmod(包装 -3, 书 -1)
assm(书 -1, 的 -2)
nsubj(不错 -4, 包装 -3)
```

特征词:包装

正面评论词语:不错

结论:主谓结构,在 Stanford parser 中的类型为 nsubj(名词性主语)。

(3) 例句:“比较平淡的一本书籍”

依存关系为:

```
advmod(平淡 -2, 比较 -1)
rmod(书籍 -6, 平淡 -2)
...
```

特征词:书籍

正面评价词语:平淡

结论:的字结构,在 Stanford parser 中的类型为 rcmode(关系从句修饰)。

基于以上结论,遍历商品评论集 R ,对每条评论 r 使用以下步骤进行极性词挖掘。

输入:商品评论 r

输出:特征评价分析结果集 ra (缺少程度副词和否定词)

步骤 1 利用第 2.2 节的商品评论基本事务集,判断该评论是否包含商品特征,即 $Noun(r) \cap F(R) \neq \emptyset$ 。若 r 包含商品特征,进行后续步骤;反之,对单条评论的极性词挖掘过程结束。

步骤 2 得到商品评论 r 的句法分析树。

步骤 3 取得句法分析树的第一个依存关系。

步骤 4 判断是否为 dobj 或 nsubj 依存关系结构,且包含了商品特征 $feature$ 。若是,则得到候选极性词 $oWord$;反之,跳到步骤 6。

步骤 5 如果依存关系是 dobj,查找扩展情感词语库;如果依存关系是 nsubj,查找扩展评价词语库。若查找结果为真,即扩展词库存在该候选极性词 $oWord$,则创建新的特征评价元组 $o = (feature, oWord, "", "", false)$,并更新该评论的分析结果集 $ra = ra \cup \{o\}$ 。此时,程度副词和否定词暂时设置为空。若查找结果是扩展词库不存在该词,则计算该词与种子情感词语库中每个词的相似性,当最大值超过一定阈值时,则将该词加入扩展词库,仍创建新的特征评价元组。本文设置相似度阈值为 0.8。

步骤 6 若仍有依存关系结构,取得下一个依存关系,跳到步骤 4;反之,则对单条评论的极性词挖掘过程结束。

本文使用了 HowNet 的情感语料库。其中,种子情感词语库由“正面情感词语(中文).txt”和“负面情感词语(中文).txt”建立而成,种子评价词语库由“正面评价词语(中文).txt”和“负面评价词语(中文).txt”建立而成。依此构建扩展情感词库的初始极性词索引表,索引为词语,值为极性,负面为 -1,正面为 +1。在此基础上,使用哈工大的同义词林,将同义词也加入到索引表中,从而提高算法的召回率。

2.3.2 挖掘极性词的程度副词和否定词

2.3.1 节挖掘商品特征对应的极性词,本节考虑修饰极性词的程度副词和否定词。在挖掘极性词的同时,抽取句法分析树中的副词词和否定词:程度副词采用 advmod(副词修饰)结构进行抽取,而否定词则使用 neg(否定词)结构。

以“书的质量可以,是正品,不过里面的内容简单化了,对内容不是很满意”为例,部分依存关系如下:

```

.....
neg(满意 -21, 不 -18),
cop(满意 -21, 是 -19),
advmod(满意 -21, 很 -20),
.....

```

基于上述两种依存关系结构,结合 2.3.1 节中极性词挖掘的挖掘步骤,可同时获得极性词相应的程度副词和否定词。

输入:商品评论 r ,商品评论分析元组 ra

输出:完整的商品评论分析元组 ra

在程度副词的挖掘中,使用 HowNet 提供的“程度级别词语(中文).txt”建立程度词库。

本节补充了上节特征评论分析结果集 ra 中每个特征评论元组 o 缺失的后三项:程度副词、否定词,以及程度副词和否定词的共现语序关系,从而得到完整的评论分析结果集 ra 。

2.3.3 情感倾向程度分析

本节基于商品评论特征分析结果,进行情感倾向程度分析和计算。首先考虑程度副词和否定词的极性参数设定,接着考虑程度副词否定词共现语序不同时造成的语义差异,最终定义情感倾向程度的计算方法。

考虑程度副词的强度差距,文献[11]将程度副词分成两类,并设置 4 个强度等级。本文基于 HowNet 的程度词语库,为 6 种类别的程度副词设置不同的极性参数,如表 1 所示。

表 1 程度副词分类及极性参数表

类别	极性参数	词语(部分)	数量
1.最	1.6	非常、极其	69 个
2.很	1.4	格外、分外	42 个
3.较	1.2	更加、较为	37 个
4.稍	0.8	略、稍微、挺	29 个
5.欠	0.6	半点、相对	12 个
6.超	1.6	超、过、过于	30 个

对于否定词的处理,单纯地将相应极性词的极性简单取反是不合理的。姚天昉等人^[8]对否定前缀的情感极性处理方法是先取反后除以 2。但按照该算法,“不很满意”和“很不满意”的情感倾向程度是一样的。而事实上,两者是不一样的,该算法并没有考虑否定词和程度副词共现时语序不同造成的语义差异。因此,本文首先设置否定词的极性参数为 -0.5,然后考虑否定词和程度副词共现语序。两者的共现语序可分为两种:NDP 和 DNP,其中 N 表示否定词, D 表示程度副词, P 表示极性词。

对于 DNP 形式,即程度副词在否定词之前,是对否定词所否定程度的确定。例如:

- (1) 手机屏幕不好;
- (2) 手机屏幕很不好;
- (3) 手机屏幕极其不好。

以上三个例句的否定程度是逐渐递增的。

对于 NDP 形式,即否定词在程度副词之前,是对某个程度的否定,作用是把原来的程度降低。例如:

- (1) 手机屏幕不非常好;
- (2) 手机屏幕有点好。

以上两个例句的量级大体相同,在语义上能够相互

推衍^[19]。

基于以上规则,设计否定词与程度副词共现时的情感极性程度计算:

$$\text{deg}(DNP) = \text{deg}(d) \times \text{deg}(n) \times \text{deg}(p) \quad (1)$$

$$\text{deg}(NDP) = 0.8 \times \text{deg}(d) \times \text{deg}(p) \quad (2)$$

其中, d 为程度副词, n 为否定词, p 为极性词。

以上,考虑了程度副词的强度差距,以及程度副词、否定词的共现语序。下面定义商品评论的情感倾向程度计算方式。

定义 1 特征评价元组的情感倾向程度

$\text{degree}(o) = \text{pol}(o\text{Word}) \times \text{deg}(d\text{Word}) \times \text{deg}(n\text{Word})$,其中 $\text{pol}(o\text{Word})$ 为极性词的原极性(+1 或 -1), $\text{deg}(d\text{Word})$ 为程度副词的极性参数, $\text{deg}(n\text{Word})$ 为否定词的极性参数。当程度副词或否定词为空时,极性参数相应设置为 1。当程度副词和否定词共现时,则根据式(1)和式(2)进行计算。

定义 2 单评论句的情感倾向程度

$$\text{degree}(r) = \text{degree}(ra) = \frac{\sum_{i=1}^n \text{degree}(o_i)}{n}, o_i \text{ 为评论 } r \text{ 对应的}$$

的特征评价分析结果集 ra 的第 i 个特征评价元组, n 为 ra 中的特征评价元组总数。

定义 3 商品评论的情感倾向程度

$$\text{degree}(p) = \frac{\sum_{j=1}^m \text{degree}(r_j)}{m}, r_j \text{ 为商品 } p \text{ 的评论集中的第 } j \text{ 条}$$

评论, m 为该商品评论集的评论总数。

定义 4 商品特征的情感倾向程度

$$\text{degree}(\text{feature}) = \frac{\sum_{j=1}^k \text{degree}(o_j)}{k}, o_j \text{ 为第 } j \text{ 个包含 feature 的}$$

特征评价元组, k 为包含该 feature 的评价元组总数。

基于以上定义,遍历评论分析结果集 ra 的集合,逐步获得特征评价元组、评论句、商品以及特征的情感倾向程度。情感倾向程度的正负分别表示褒义和贬义。绝对值越大,强度越大。

3 实验分析

3.1 系统展示

系统采用 B/S 架构,用 Java 语言实现。商品评论采集使用开源数据抽取工具 WebHarvest,预置评论信息抽取规则,实现信息采集自动化,汉语分词和词性标注采用了中科院计算所的汉语分词系统 ICTCLAS^[13],情感语料库使用 HowNet“情感分析用词语集(Beta 版)”,词汇相似度计算使用基于知网的 WordSimilarity 开源工具包,同义词词库使用“HIT-ITLab 同义词词林(扩展版)”,句法分析则选用了 Stanford Parser v2.0.4^[14]。

系统界面效果如图 2 所示。消费者可以将商品详情页链接复制到右上方输入框。采集模块根据预设抽取规则进行评论采集,然后进行特征挖掘和情感倾向程度分析,最终组织分析结果。界面分三大部分:第一部分为商品评论概况,包含商品评论的总体情感倾向得分以及评论摘要;第二部分将商品特征分成两个正负特征词云;第三部分展示当前选中的商品特征相关的极性词以及与包含该商品特征的所有评论。



图 2 ReviewHelper 系统界面

3.2 实验数据

实验语料从亚马逊网站上选择 4 种商品评论数据,分别为图书类和电脑数码类,总计 4788 条评论,再从每种商品取出 300 条作为测试预料。实验首先需要测试语料进行商品特征和评论倾向的人工标注。以“摸着石头过河”为例,商品特征人工标注结果如表 2 所示。

表 2 商品特征的人工标注结果(《摸着石头过河》)

商品名称	逐句人工标注出的商品特征集合	数量
摸着石头过河	封面,包装,内容,排版,纸张,送货,态度,书质,语言,文字,见解,话语,速度,质量,物流,故事,性价比,快递,意见,配送,字数,书本,插图,用词,行距,纸质,发货,叙述,质感,运输,手感,印刷,设计,可读性,页边距,用字间距,容量	37

3.3 商品特征挖掘实验结果及分析

为评估系统性能,采用查准率、查全率和 $F-measure$ 作为评价指标,计算方式如下所示:

$$\text{查准率 } precision = \frac{A}{A+B}$$

$$\text{查全率 } recall = \frac{A}{A+C}$$

$$F-measure = \frac{2 \times precision \times recall}{precision + recall}$$

其中 A 为系统挖掘出的真实商品特征数, B 为系统挖掘出的错误商品特征数, C 为系统未挖掘到的真实商品特征数。

由于挖掘频繁项集的 min_sup 会影响到性能,实验时令 min_sup 取评论总数 1% 到 2% 的不同值,测试 $F-measure$ 所能达到的最大值。以“三星 S7562 手机”为例,各项指标随 min_sup 变化如表 3 所示,在 $min_sup = 4$ 时 $F-measure$ 达到最大值。系统在实现时取评论总数的 1% 作为默认值。

对 4 种不同商品的挖掘性能如表 4 所示,平均查准率为 57.1%,平均查全率为 65.5%。对比其他研究者的实验效果(如文献[6]为 63% 和 74%,文献[9]则为 56% 和 42%),说明了系统具有一定的有效性。

查全率的比较结果显示,在进行第一种同类产品的挖掘后,查全率略有提高,可能源于领域词典的更新。此外,由表 4 可以

看出后两种商品的总体性能较前两种好,主要原因是商品分属不同类型。前两种商品是图书类别,属于享乐型商品,而后两种商品是手机数码类,属于实用型商品。从评论特点上看,享乐型商品的评论比实用型商品更为复杂和冗长,包含了各种叙述和修饰手法,且名词众多,导致查准率较低,而华丽的辞藻和复杂句式则直接影响到查全率。

表 3 商品特征挖掘实验结果

最小支持度	特征数量	查准率	查全率	F-measure
2	130	0.392	0.646	0.488
3	98	0.490	0.608	0.542
4	83	0.566	0.608	0.586
5	76	0.579	0.557	0.568
6	68	0.618	0.481	0.541

表 4 商品特征挖掘实验结果

商品名称	人工标注商品特征数量	查准率	查全率	F-Measure
摸着石头过河	37	0.458	0.595	0.518
少有人走的路	35	0.444	0.628	0.518
三星 S7562 手机	79	0.770	0.608	0.679
iPhone4 (8G) 白色	52	0.612	0.788	0.689
平均值	50.75	0.571	0.655	0.601

3.4 评论倾向判断实验结果及分析

对于评论倾向判断实验,只将包含商品特征的句子设定为评价句,倾向判断相当于做分类工作,因此重新定义查准率和查全率,并加上正确率作为评价指标,计算方式如下所示:

$$\text{查准率 } precision = \frac{TP}{TP + NP}$$

$$\text{查全率 } recall = \frac{TP}{TP + FN}$$

$$\text{正确率 } accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

首先,分别对四种商品评论进行了倾向判断实验。然后,将本文方法和基于词频加权统计^[11]的方法(Base-TFIDF)以及不考虑否定词程度副词共现语序的基本特征-极性词^[10]的方法(Base-Feature)进行比较,并对实验结果进行分析说明。

表 5 的实验结果表明,本文方法的平均正确率为 81%,正极性判断平均查准率为 94%,平均查全率为 85%。通过对比其他研究者的实验效果(如文献[9]中正确率在 75% 到 84% 之间,文献[10]约为 90%),表明了系统的有效性。查全率不够高,原因是倾向分析基于特征挖掘和句法分析结果,有累加错误效应。另外,算法无法准确判断复杂句式,如句子“从中能学到以前不被重视的细节及态度”中,系统识别出(细节,重视,不),判定为消极性。同时,同一词语在不同语境下褒贬不同,也会影响分析结果,如评论“字数不多,看了之后有想冒险的感觉”中,“多”为正面词,“不”为否定词,判定为消极性。针对该点,可后续考虑增加领域或商品

特征相关的情感词汇表。

表 5 商品评论倾向判断实验结果

商品名称	正确率	查准率 (正极)	查全率 (正极)	F-Measure (正极)
摸着石头过河	0.858	0.983	0.886	0.932
少有人走的路	0.788	0.926	0.829	0.874
三星 S7562 手机	0.769	0.895	0.842	0.868
iPhone4 (8G) 白色	0.815	0.955	0.841	0.894
平均值	0.808	0.940	0.850	0.892

表 6 的实验结果表明,本文提出的方法相对基于词频加权统计的方法略有提升,主要原因是基于词频加权统计的方法只是单纯从极性词的极性出发,而本文提出的方法选取描述商品特征的极性词,同时考虑了程度副词和否定词,因此效果较佳。

表 6 同类方法用于商品评论倾向判断的实验结果

褒贬判断方法	正确率	查准率 (正极)	查全率 (正极)	F-measure (正极)
Base-TFIDF	0.742	0.863	0.817	0.839
Base-Feature	0.783	0.931	0.841	0.884
本文方法	0.808	0.940	0.850	0.892

在此基础上,考虑程度副词和否定词的共现语序,进一步提升了算法性能。但由于 NDP 出现的频次低,因此提升程度不明显。

4 结 语

本文针对消费者商品评论信息过载问题,提出了基于情感计算的商品评论分析系统,使用了商品特征挖掘和情感倾向分析等关键算法,其中领域词典的半自动化更新、程度副词和否定词的共现语序考虑,具有一定的创新性。在应用上,系统能够自动根据指定商品购物地址采集商品评论信息,并将庞杂的商品评论信息转为为对商品特征的褒贬评价信息,解决评论信息过载的问题,从而有效地辅助消费者决策,具有较强的应用价值。另外,系统可以将情感倾向分析结果汇总,作为情感特征丰富商品的多属性描述,后续可继续应用到商品推荐或商家绩效管理系统中。

目前系统仍有许多不足,如特征挖掘查准率较低,噪音较多,可考虑结合其他方法进行交叉验证。此外,在挖掘的过程中发现有相同或相似的商品特征,可以考虑做聚类分析,有待进一步研究。而如何进行多数据源的评论采集,更精确地计算情感倾向程度,更好地展示评论分析结果以符合消费者思维习惯等等,都将成为本文下一步的工作。

参 考 文 献

[1] Liu Y, Huang X J, An A J, et al. Modeling and predicting the helpfulness of online reviews[C]//Proceeding of the 8th IEEE International Conference on Data Mining, Washington, DC, USA; IEEE Computer Society, 2008; 443 - 452.

[2] Kim S M, Pantel P, Chklovski, et al. Automatically assessing review

helpfulness[C]//Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney; Association for Computational Linguistics, 2006; 423 - 430.

- [3] Jindal Nitin, Liu Bing. Analyzing and detecting review spam[C]//Seventh IEEE International Conference on Data Mining, 2007; 547 - 552.
- [4] 张紫琼,叶强,李一军. 互联网商品评价情感分析研究综述[J]. 管理科学学报, 2010, 13(6): 84 - 96.
- [5] Hu Mingqing, Lin Bing. Mining Opinion Features in Customer Reviews [C]//Proceedings of AAAI-2004, 2004; 755 - 760.
- [6] 李实,叶强,李一军,等. 中文网络客户评论的产品特征挖掘方法研究[J]. 管理科学学报, 2009, 12(2): 142 - 152.
- [7] 姚天昉,娄德成. 汉语语句主题语义倾向分析方法的研究[J]. 中文信息学报, 2007, 21(5): 73 - 79.
- [8] 娄德成,姚天昉. 汉语句子语义极性分析和观点抽取方法的研究[J]. 计算机应用, 2006, 26(11): 2622 - 2625.
- [9] 刘鸿宇,赵妍妍,秦兵,等. 评价对象抽取及其倾向性分析[J]. 中文信息学报, 2010, 24(1): 84 - 88, 122.
- [10] 倪茂树,林鸿飞. 基于关联规则和极性分析的商品评论挖掘[C]//第三届全国信息检索与内容安全学术会议, 2007; 635 - 642.
- [11] 徐琳宏,林鸿飞,杨志豪. 基于语义理解的文本倾向性识别机制[J]. 中文信息学报, 2007, 21(1): 96 - 100.
- [12] Dong Z, Dong Q. HowNet [EB/OL]. 2013. http://www.keenage.com/html/c_index.html.
- [13] ICTCLAS 汉语分词系统 [EB/OL]. 2013. <http://ictclas.cn/index.html>.
- [14] The Stanford Parser: A statistical parser [EB/OL]. 2013. <http://nlp.stanford.edu/software/lex-parser.shtml>.
- [15] Kobayashi N, Inui K, Matsumoto Y, et al. Collecting Evaluative Expressions for Opinion Extraction [C]//IJCNLP 2004, LNAI 3248, 2005: 596 - 605.
- [16] Hu Mingqing, Lin Bing. Mining and summarizing customer reviews[C]//Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, Seattle, WA, USA, 2004: 168 - 177.
- [17] Turney P D. Thumbs up or thumbs down: semantic orientation applied to unsupervised classification of reviews[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, Philadelphia, Pennsylvania, 2002: 417 - 424.
- [18] 刘群,李素建. 基于《知网》的词汇语义相似度计算[C]//第三届中文词汇语义学研讨会论文集, 2002.
- [19] 尹洪波. 否定词与副词共现的句法语义研究[D]. 北京: 中国社会科学院研究生院, 2008.

(上接第 28 页)

- [6] 戴胡斌. 上海大众波罗轿车维修手册[M]. 北京: 机械工业出版社, 2010; 32 - 38.
- [7] Wu Jianjun, Hu Juan. Application of CAN-bus Network in Intelligent parking[J]. Physics Procedia, 2009; 615 - 628.
- [8] 李华嵩. 基于 OBD-II 的车载故障监控诊断仪硬件平台及网络接口单元的设计[D]. 成都: 电子科技大学, 2007.
- [9] Cavalieri S. Meeting real-time constraints in CAN[J]. IEEE Transactions on industrial informatics, 2005, 1(2): 124 - 135.
- [10] 李维臻,王猛. 基于 LIN 总线的车灯控制设计[J]. 计算机应用与软件, 2012, 29(5): 298 - 300.