

挖掘中文网络客户评论的产品特征及情感倾向*

李实¹, 叶强², 李一军², 罗嗣卿¹

(1. 东北林业大学信息与计算机工程学院, 哈尔滨 150040; 2. 哈尔滨工业大学经济与管理学院, 哈尔滨 150001)

摘要: 为探索中文客户评论中的产品特征及相关情感倾向的挖掘, 以帮助生产商和服务商改进产品, 改善服务, 提高竞争力, 提出采用基于 Apriori 算法的非监督型产品特征挖掘算法, 结合监督型情感分析技术, 实现对于评论中产品特征及其情感倾向的综合信息挖掘; 并根据用户的关注权重将产品特征和情感倾向进行排列。采用几种从互联网下载的真实产品评论语料, 对该方法进行了数据实验, 实验结果初步验证了该方法的有效性。

关键词: 用户评论; 产品特征; 数据挖掘; 情感分析

中图分类号: TP311 **文献标志码:** A **文章编号:** 1001-3695(2010)08-3016-04

doi: 10.3969/j.issn.1001-3695.2010.08.054

Mining product features and sentiment orientation from Chinese customer reviews

LI Shi¹, YE Qiang², LI Yi-jun², LUO Si-qing¹

(1. College of Information & Computer Engineering, Northeast Forestry University, Harbin 150040, China; 2. School of Economics & Management, Harbin Institute of Technology, Harbin 150001, China)

Abstract: At mining product features and polarity of customer's opinion from Chinese online customer reviews, in order to help manufactures and service providers improve their products and services, and gain competitive advantages. This paper proposed an approach based on Apriori algorithm which unsupervised feature extraction algorithm and supervised sentiment analysis technology to realize the summery information of customer reviews. And then ranked the work features and opinions based on customer attention strength. Several actual Internet products reviews were tested by an experimental model. Final results demonstrate the validity of the method.

Key words: customer reviews; product features; data mining; sentiment analysis

0 引言

随着电子商务的蓬勃发展,越来越多的客户开始浏览大量的网络评论来了解产品和服务的口碑,帮助作出可靠的决策。同时网络客户评论作为反馈机制也帮助了生产者和销售商提升产品,改进服务,从而获得竞争力。但是网络评论数量的飞速增长,使得信息内容越来越庞杂,造成了客户评论中有用信息难以获取的后果。因此,迫切需要借助一定的技术手段来使这一过程变得更为准确而便捷。以有效获取网络用户评论信息为目标的非结构化数据分析技术——评论挖掘吸引了很多学者的关注。评论挖掘作为非结构化信息挖掘的一个新兴领域,主要涉及情感分析、评论中产品特征挖掘,以及评论中主观内容识别等。在英文评论领域,研究者已经初步取得一些成果,而针对中文网络用户评论的研究仍还处于起步阶段。随着中国电子商务在世界领域内的崛起,亟需对于中文评论中有用信息的自动提取技术。由于文化差异导致语言表达方式不同,使得英文领域的研究成果不能直接应用于中文评论。所以,本文将针对面向电子商务的中文网络客户评论,探索评论内容中产品特征及相应的情感倾向的提取技术,旨在为中文领域内的客户和企业提供更为方便和科学的评论挖掘工具。

单一产品的特征结合正面和负面倾向后,可以直观地得到

客户对于产品整体以及各个属性的满意度;对于不同产品,将其结合情感分析的产品特征相对比,各方面的优劣一目了然。这对于消费者下一步的购买决策,或者企业营销的相对应措施,都能起到非常确切的指导作用,而且可以在行为导向的研究中以此分析评论中的内容与相关商务内容的直接或者间接影响。本研究制订了以下针对于中文评论中产品特征的情感分析挖掘任务:

a) 提取用户关心的产品属性,并根据它们在评论中出现的频率排序;

b) 找到属性所在的评论语句,判断语句的情感倾向是正面还是负面。

1 研究背景

Popescu 等人^[1]将网络评论挖掘问题分为几个子任务:a) 挖掘重要产品特征;b) 挖掘用户对于产品特征的主观观点;c) 判断评论观点的情感导向;d) 根据观点的重要性进行排名。本文的研究内容主要涉及 a)c)d) 这三个子任务。

1.1 产品特征挖掘

网络客户评论中的产品特征挖掘是指通过机器从大量的网络客户产品评论中自动地获取备受关注的产品特征^[2],如在用户评论中出现的“外形”“手感”“做功”。

收稿日期: 2010-01-25; 修回日期: 2010-03-08 基金项目: 国家自然科学基金资助项目(70971033)

作者简介: 李实(1976-),女,黑龙江哈尔滨人,博士,主要研究方向为电子商务、商务智能(shishili@gmail.com);叶强(1972-),男,教授,博士,主要研究方向为商务智能、电子商务;李一军(1957-),教授,博士,主要研究方向为管理信息系统、决策支持系统等;罗嗣卿(1964-),男,副教授,硕士,主要研究方向为管理信息系统。

在英文评论挖掘领域,由于研究人员意识到网络客户评论所带来的巨大影响比较早,研究者已经初步取得一些成果。2004年,Hu等人^[2]首先提出应用关联规则分类方法提取英文评论中的产品特征,利用该非监督型方法对于包括手机、数码相机等产品评论进行挖掘,平均查全率达到80%,平均查准率达到72%,并在此基础上进行了后续的研究,判断了用户对这些特征的情感导向^[2,3]。

也有一些研究人员采用了其他方法实现这一功能,比如Kobayashi等人^[4]采用了一个半自动化的循环方法提取产品特征和用户观点,但是需要大量的人工参与;Popescu等人^[5]利用研发的KonWitAll系统进行贝叶斯分类,从而提取产品特征,提高Hu的准确率,但是查准率却有所下降。

目前对于中文评论中的产品特征挖掘研究主要集中在需要人工参与的监督型方法或者半监督型方法,例如姚天等人^[6]利用本体建立了汽车的产品特征;刘健等人^[7]提出与产品特征挖掘并考虑产品特征的情感倾向这两个功能类似的意见实例抽取(opinion instance extraction)需要建立相关的领域知识。Shi等人^[8]面向中文也作了挖掘产品特征的研究,但是需要人工建立基于产品属性的概念模型。

由于网络评论的信息挖掘是噪声很大的领域,网络上的评论大多由非专业人员写的,文献[9]指出对于评论作者这一传播者的专业性对于网络评论这一口碑传播方式之一是具有很大影响的。而且不论是数量还是风格,网络上的评论变化很快。由于网络客户评论的异步性^[10],既有的知识库结构和内容可能会带来一些延时性误差,需要人工参与的监督型方法对于面向网络评论的挖掘都有所局限,特别是监督型方法对中文评论进行挖掘,需要首先从语言学的角度对这种主观性的文本进行研究,包括语料收集、分析语言规律、研究标注规范和方法等。但在国内的研究中,这方面的基础研究还做得很少^[6],所以非监督型方法更为适合面向电子商务的中文评论产品特征挖掘。

本文在前面研究的基础上^[11],将基于关联规则的非监督型挖掘方法结合情感分析技术来实现更为完整的评论挖掘。

1.2 情感分析

情感分析的目的是判断给定的文本片段(词汇、短语或文档)中所体现的说话者的情感倾向,通常为正面、负面和中性^[12]。目前实现情感分析的技术主要包括机器学习方法及语义方法两类。

最初Pang等人^[12]在研究中提出用机器学习的方法进行情感倾向的挖掘工作,他们以互联网上的电影评论文本作为语料,采用了不同的特征选择方法,应用朴素贝叶斯、最大熵、SVM对电影评论进行分类,准确率达到87.5%。之后的一些学者在此研究的基础上扩展和延伸,并与其他一些技术进行集成,取得了很好的研究结果。对于利用机器学习的方法进行中文的情感分析,由于机器学习方法的通用性,面向英文的很多方法都可以借鉴。近两年国内研究人员在此方面也取得一些成果,唐慧丰等人^[13,14]还特别针对各种情感分类技术包括面向中文文本的方法进行了总结和比较。

机器学习的方法虽然在目前来讲分类的准确程度比较高,但是它针对每一种产品使用前,训练样本集的建立都需要采用人工方法对大量的评论文章逐一阅读甄别,并进行手工标志,这与利用自动情感分类降低人的阅读负担这一初衷还有着

一定的差距。

因此,近来许多研究者将情感分析研究的重点集中在对训练样本的需求量较低的语义方法上。从情感分类的对象为分类依据,相关研究可以分为以下几个部分:

a)关于词语的情感分类。目前对于词的情感倾向分析的非监督型语义方法主要有以下两种方法:(a)由已有的电子词典或词语知识库扩展生成情感倾向。英文词语情感倾向信息的获取主要是在General Inquirer和WordNet的基础上进行的研究;而中文词语情感倾向信息的获取依据主要有HowNet,但是还处在研究阶段^[15]。(b)基于人工标注语料库的学习方法。

b)对于文档的分类。Turney等人^[16]最早提出将点互信息(pointwise mutual information)与信息检索方法(information retrieval)相结合,借助搜索引擎的后台数据库获得语义倾向信息,得到汽车评论的准确率是84%,电影评论的准确率是66%。其可靠性已经在英文客户情感分类的研究中得到了初步的验证。

由于语言结构的差别,现有的面向英文客户评论情感分析的语义方法无法直接用于中文客户评论。Ye等人^[17]探索了中文环境下对于文档的情感分析理论与方法,在PMI-IR方法的基础上,初步建立了中文语义倾向情感分析方法,获得了接近英文同类研究的分析结果,显示出该方法在中文情感分析上的应用前景。但是利用非监督型的语义方法来判断中文词语的情感分类还没有成熟的方法,只有考虑应用机器学习的方法来判断基于词和短语组成产品特征的情感分析。

2 方法

本文提出了利用基于关联规则的产品特征挖掘方法结合情感分析技术的方法来完整实现中文网络客户评论的产品特征挖掘任务。

2.1 方法步骤

图1为本文所提出的挖掘评论综合信息算法流程。

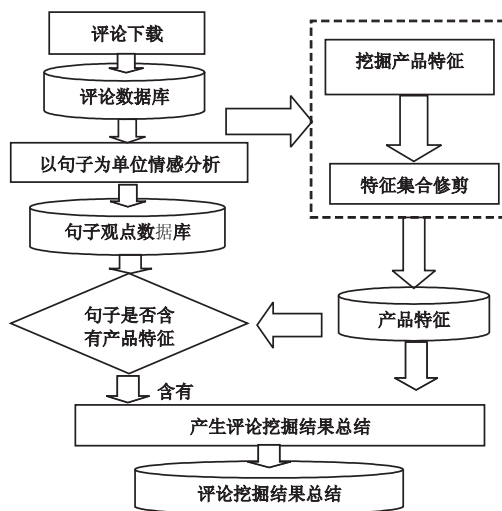


图1 挖掘中文客户评论的信息算法概要流程
具体方法如下:

a)应用中文分析工具对评论语料进行分词和词性标注,包括中文名词短语。

b)根据产品类别训练分类模型,以句子为单位判断句子的情感倾向,这里分类采用LingPipe自然语言分析工具^[18]。利用实验语料进行训练,然后利用情感分类工具将所有评论语

句判断情感倾向,存入用户观点数据库。

c) 利用词性标注后的评论语料创建关联规则事务文件。

d) 基于关联规则的 Apriori 算法,找到频繁项集作为候选产品特征集合 I_0 。

e) 将候选产品特征集 I_0 按照邻近规则剪枝,成为候选特征集 I_1 。

在中文评论中,假设 f 是频繁规则项集,并且 f 包含 n 个名词(或名词短语),假设一个句子 s 包含 f ,而且在 f 中的词(或名词短语)出现在 s 中的顺序为 w_1, w_2, \dots, w_n 。假设 s 中任何两个相连的名词(或名词短语) w_i 和 w_{i+1} 的距离不超过三个词(根据中文分词的结果),则可以说 f 在 s 中是邻近的。

在这一步骤中,遍历二维和三维频繁项 f ,如果 f 出现在评论数据库的 m 个句子中,而且至少在两个句子中是邻近的,那么可以称 f 是一个邻近的特征名词短语,加入到候选项集合 I_1 中。

f) 将候选产品特征集 I_1 按照独立支持度规则进行修正,形成候选特征集 I_2 。

在中文评论中名词或者基本名词短语 f_r 的独立支持度(p -support)是包含 f_r 的,而且句子中不包含 f_r 的父集作为频繁特征项的句子数量。本研究中独立支持度采用 3。

g) 建立常见中文频繁项名词却非产品特征的集合,将 I_2 过滤形成特征集合 I_3 ; 从 I_3 中去掉单字名词的候选项,包括单字名词的 n 项频繁项($n \leq 3$),形成最后的产品特征集合 I_0 。

h) 在用户观点数据库中查找所挖掘到的产品特征所在的句子,提取出句子的情感倾向。

i) 依据每一个产品特征出现的频率按照降序排列,并总结里面包含的正面句子数量和负面句子数量。

表 1 为本文所提出来的属性及其情感倾向挖掘方法与其他两个文献所提出的方法对比。相对面向中文评论的 Shi 等人的研究,本文方法不用建立产品属性的概念模型,对于产品特征的挖掘更为通用,鲁棒性强。

表 1 本文研究方法与其他两个研究方法对比

项 目	本文方法	Hu 等人方法 ^[2]	Shi 等人方法 ^[8]
语言	中文	英文	中文
观点词类型	用 n-gram 特征表示,不用考虑观点词类型	形容词	副词,形容词,动词
观点类型	正面,负面	正面,负面	正面,负面,中性
属性词集合	使用关联规则挖掘方法,以及人工过滤技术	使用关联规则挖掘方法,以及人工过滤技术	人工建立概念模型
粒度	句子	句子	短语
选择	处理所有句子,然后提取含产品属性的句子	包含产品属性的句子	包含观点词的句子
步骤	首先确定产品属性;同时判断句子的情感倾向;在句子中如果包含产品属性,将其情感倾向赋予产品属性;产品属性和句子的情感倾向都有保存	首先确定产品属性;然后判断属性的情感倾向;只有包含属性的词语才判断情感倾向	找到情感强烈的词语和相关产品属性;找到普遍的情感倾向;找到的属性和观点建立关系;没有属性,可以使用普遍情感倾向代替

2.2 评估方法分析

对于结合了情感分析的评论产品特征信息挖掘性能评价,Hu 等人的研究中同时采用了三个指标:a) 特征提取的效果;

b) 主观句提取的效果;c) 句子情感倾向的判断准确率。

这种英文挖掘方法中判断产品属性的情感倾向是利用属性所在句子的形容词,然后根据英文词网进行的。然而对于中文文本,针对产品属性或者句子的细粒度情感分析前面已经讨论过,没有成熟的非监督型方法,所以本文采用对于句子的监督型分类方法。本文研究对于系统的性能评价参考了 Hu 的评价方式,但是采用了更适合监督型情感分析方法的两个指标:

a) 细粒度情感分析的性能评估。在以文本分类为基础的情感分析列联表的基础上,建立以句子为单位的性能评估列联表(表 2);评价指标的计算根据式(1)~(5),其中参数 p 和 n 分别代表函数值针对正面情感句子和负面情感句子。

$$\text{查准率 (precision) } (p) = A / (A + B) \quad (1)$$

$$\text{查全率 (recall) } (p) = A / (A + C) \quad (2)$$

$$\text{查准率 (precision) } (n) = D / (C + D) \quad (3)$$

$$\text{查全率 (recall) } (n) = D / (B + D) \quad (4)$$

$$\text{准确率 (accuracy) } = (A + D) / (A + B + C + D) \quad (5)$$

b) 对于产品属性判断的性能。对于产品属性的挖掘相当于对于属性词语进行二维分类,即真实属性还是非真实属性,所以也采用二维列联表(表 3),指标的计算形式与式(1)(2)一致。

表 2 句子情感分析性能评价的列联表

评论句子	实际正面评论句子	实际负面评论句子
方法预测的正面评论句子	A	B
方法预测的负面评论句子	C	D

表 3 产品属性评估方法性能的列联表

产品属性数	真正产品属性数	非真正产品属性数
本文方法挖掘出的产品属性数	A	B
本文方法没有挖掘出的产品属性数	C	D

3 实验验证及评价

3.1 语料数据

本研究针对一种典型数码产品评论进行实验研究:手机,这类评论属性突出,在电子商务领域应用比较广泛。

手机数据采用了 Moto1200,从 it168(www.it168.com)网站下载 100 篇评论,对于产品特征挖掘实验的参照属性采用人工标注的方法:对这些评论中所提到的该商品属性进行识别和标注,根据最小—最大覆盖原则建立一个最小的属性集合,使这个集合可以覆盖所有这 100 个评论中提到的该商品的属性。对于情感分析的语料随机选取 100 个句子作为训练集合,50 个句子作为测试集合。

3.2 实验结果

1) 手机评论的产品特征挖掘结果 依据关注程度排名前 10 位的产品属性挖掘结果如表 4 所示。

表 4 手机评论所挖掘出来的属性(按关注程度排序)

排名	属性名称	正面句子数	负面句子数	句子总数	正面句子比率/%
1	摩托罗拉	15	4	19	79
2	诺基亚	3	9	12	25
3	摄像头	8	3	11	73
4	智能手机	6	4	10	60
5	收音机	4	4	8	50
6	蓝牙耳机	3	5	8	38
7	扬声器	3	5	8	38
8	文件夹	8	0	8	100
9	联系人	5	2	7	71
10	数据线	3	3	6	50

表 5 即为手机评论的产品特征挖掘结果列联表。通过表中的数据可以计算出挖掘性能达到查全率为 81.8%,平均查准率为 62.8%。

表 5 手机 MOTO1200 评论挖掘性能的列联表

产品属性数	真正产品属性数	非真正产品属性数
本文方法挖掘出的产品属性数	27	16
本文方法没有挖掘出的产品属性数	2	D

2)手机评论的情感分析实验结果 由于实验采用基于 SVM 监督型算法,培训集合的大小直接关系到分析准确性。训练集合为 100 个句子,测试集合为 50 个句子时的性能,实验结果列联表如表 6 所示。结果正面评论句子的查全率为 70%,查准率为 83%;负面评论句子的查全率为 86%,查准率为 74%;总体准确率为 78%。

表 6 手机 MOTO1200 评论情感分析结果

评论句子	实际正面评论句子	实际负面评论句子
方法预测的正面评论句子	35	7
方法预测的负面评论句子	15	43

3)本研究挖掘结果与其他研究结果的比较 表 7 是针对手机这一类产品评论,本文所提出的方法与其他两个研究的产品特征实验结果比较。为了综合评价结果,利用了 F-score 值:

$$F\text{-score} = (2 \times \text{recall} \times \text{precision}) / (\text{recall} + \text{precision}) \quad (6)$$

表 7 针对手机类产品评论相关研究产品特征挖掘结果比较 /%

性能指标	本文实验结果	Hu 等人 ^[2] 实验结果	Shi 等人 ^[8] 实验结果
查全率	81.8	76.1	68.35
查准率	62.8	71.8	70.72
F-score	71.05	73.89	69.51

Hu 等人的研究对象集合是英文评论,通过比较结果可以看到本研究的产品特征性能已经接近针对英文的同类研究,具有一定的有效性。对于 Shi 等人的中文评论挖掘结果,列出的是他们面向句子的实验结果,同样是针对手机这一产品挖掘结果可以看出,本研究在产品特征挖掘部分查全率优于他们的结果,综合挖掘性能比较好。

对于手机类评论句子的情感分析的预测结果,本实验结果为 78%,比较接近 Hu 等人的综合准确率 76.4%。Shi 等人的手机评论情感分析查全率和查准率分别为 59.71% 和 61.79%。本实验中情感分析的实验结果分别考虑了面向正面句子和负面句子的两种情况,无论是根据哪种情况的指标显示本文结果都比较好。

总之,从实验结果来看不管是情感分析部分还是产品特征挖掘部分,方法的有效性都可以得到验证。

4 结束语

本文的研究将中文网络客户评论中的产品特征挖掘方法结合情感分析技术,提出了可以从大量的网络客户评论中总结基于产品特征的信息,该信息包括对于产品特征的评论句子数量,正面意见的数量,负面意见的数量,正面意见所占比例,并按照产品特征所受到的关注程度排名。使用这种方法进行的评论挖掘实验表现出良好的性能,为生厂商和制造商以及客户提供了一种实用的决策工具。

参考文献:

[1] POPESCU A M, YATES A, ETZIONI Q. Class extraction from the

World Wide Web[C]//Proc of AAAI-04 Workshop on Adaptive Text Extraction and Mining. San Jose, CA: American Association for Artificial Intelligence, 2004: 1-6.

[2] HU Ming-qing, LIU Bing. Mining and summarizing customer reviews [C]//Proc of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2004:168-177.

[3] LIU Bing, HU Ming-qing, CHENG Jun-sheng. Opinion observer: analyzing and comparing opinions on the Web[C]//Proc of the 14th International Conference on World Wide Web. New York: ACM Press, 2005: 342-351.

[4] KOBAYASHI N, INUI K, MATSUMOTO Y, et al. Collecting evaluative expressions for opinion extraction [C]//Proc of the 1st International Joint Conference on Natural Language Processing. Berlin: Springer,2005: 596-605.

[5] POPESCU A M, ETZIONI Q. Extracting product features and opinions from reviews[C]//Proc of HLT-EMNLP. Morristown, NJ: Association for Computational Linguistics,2005: 339-346.

[6] 姚天,程希文,徐飞玉,等. 文本意见挖掘综述[J]. 中文信息学报,2008,22(3):71-80.

[7] LIU Jian, WU Geng-feng, YAO Jian-xin. Opinion searching in multi-product reviews[C]//Proc of the 6th IEEE International Conference on Computer and Information Technology. Washington DC: IEEE Computer Society,2006:25-30.

[8] SHI Bin, CHANG Kui-yu. Mining Chinese reviews[C]//Proc of the 6th IEEE International Conference on Data Mining. Washington DC: IEEE Computer Society,2006:585-589.

[9] 徐伟青,黄孝俊. 口碑传播的影响力要素及其对营销创新的启示[J]. 外国经济与管理,2004,26(6):26-30.

[10] LITVIN S W, GOLDSMITH R E, PAN Bing. Electronic word-of-mouth in hospitality and tourism management [J]. *Tourism Management*,2008,29(3):458-468.

[11] 李实,叶强,李一军,等. 挖掘中文网络客户评论中的产品特征方法研究[J]. 管理科学学报,2009,12(2):142-152.

[12] PANG B O, LEE L, VAITHYANATHAN S. Thumbs up? sentiment classification using machine learning techniques [C]//Proc of Conference on Empirical Methods in Natural Language Processing. Morristown, NJ: Association for Computational Linguistics,2002:79-86.

[13] 唐慧丰,谭松波,程学旗. 基于监督学习的中文情感分类技术比较研究[J]. 中文信息学报,2007,21(6):55-94.

[14] TANG Hui-feng, TAN Song-bo, CHENG Xue-qi. A survey on sentiment detection of reviews[J]. *Expert Systems with Applications*, 2009,36(7):10760-10773.

[15] 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.

[16] TURNEY P D, LITTMAN M L. Measuring praise and criticism: inference of semantic orientation from association [J]. *ACM Trans on Information Systems*,2003,21(4):315-346

[17] YE Qiang, SHI Wen, LI Yi-jun. Sentiment classification for movie reviews in Chinese by proved semantic oriented approach [C]//Proc of the 39th Annual Hawaii International Conference on System Sciences,2006. Washington DC: IEEE Computer Society,2006: 1-5.

[18] BALDWIN B, CARPENTER B. LingPipe [EB/OL]. (2007) [2007-06-20]. <http://www.alias-i.com/lingpipe/>.