

基于在线评论获取产品优化辅助决策信息的算法研究^①



李 想

(宁波大学 机械工程与力学学院, 宁波 315211)
通讯作者: 李 想, E-mail: 912582020@qq.com

摘 要: 在大数据时代, 如何通过数据分析抓住顾客需求, 增加产品优化的科学性, 对企业有着至关重要的战略意义. 本文将在线评论数据应用于企业产品的辅助优化中, 提出了产品优化信息的获取技术与方法, 从评论中获取产品优化所需要的优化信息. 首先计算在线评论中的顾客关注度和满意度等指标, 构建客户意见的权重算法模型; 然后, 提取出产品特征和顾客意见的词对, 并根据权重算法模型计算出顾客意见的权重; 接着, 通过关联矩阵找到对应的产品优化信息; 最后并通过实例分析验证的方法的可行性.

关键词: 产品优化; 文本挖掘; 权重矩阵; 信息获取; 情感分类

引用格式: 李想. 基于在线评论获取产品优化辅助决策信息的算法研究. 计算机系统应用, 2019, 28(9): 180-184. <http://www.c-s-a.org.cn/1003-3254/7071.html>

Online Comments Based Algorithm Research for Obtaining Product Optimization Assistant Decision Information

LI Xiang

(School of Mechanical Engineering and Mechanics, Ningbo University, Ningbo 315211, China)

Abstract: In the era of big data, how to grasp customer needs through data analysis and increase the scientific nature of product optimization is of strategic importance to enterprises. This study applies online comment data to the assisted optimization of enterprise products, proposes techniques and methods for obtaining product optimization information, and realizes the acquisition of product optimization information. Firstly, we calculate the indicators such as customer attention and satisfaction in online reviews, and construct a weighting algorithm model for customer opinions. Next, the word pairs of product characteristics and customer opinions are extracted, and the weight of customer opinions is calculated according to the weight algorithm model. Then, the corresponding product optimization information is found through the correlation matrix. Finally, the feasibility of the method is verified by an example.

Key words: product optimization; text mining; weight matrix; information acquisition; sentiment classification

在线客户评论是消费者在网络消费平台上发表的, 对某种产品或服务的使用体验和质量评价, 通过在线评论, 消费者可以更好地获取产品质量信息、评估商家信誉^[1]. 一方面, 对在线评论的研究有助于帮助电商平台挖掘顾客消费行为与消费特征^[2,3], 并根据消费者

的差异性为不同类别的消费者提供专属服务; 另一方面, 以顾客需求为主导的消费模式以深入人心, 在线评论包含大量的顾客意见^[4], 对产品的优化改进有重要的参考意义.

目前, 对于在线评论的研究已经取得很多的成果.

^① 收稿时间: 2019-03-03; 修改时间: 2019-03-24; 采用时间: 2019-04-04; csa 在线出版时间: 2019-09-05

苗蕊等从归因理论的视角,对真实评分和平均评分的差异性所产生的有用性影响进行解读,为评论数据的研究提供新的理论支撑^[5];Alaei AR等从数据集的使用和关键评价指标的表现两个方面对旅游业应用的不同情绪分析方法进行了评价,并对情感分析在旅游业中未来的研究趋势进行了总结^[6];修国义等引入信源和信宿对在线评论信息传递效率进行测量,辅助读者对重要的在线信息进行筛选^[7];Cheng XS等通过实证研究了网上评论内容对潜在顾客信任认知的影响,并对信任感知的相关关系进行了研究,发现评论内容正向影响消费者对酒店的信任^[8];于超等提出一种基于在线评论的服务要素的优化配置方法,提取出满足顾客需求程度最大的服务要素^[9]。

不难看出,对在线评论的研究多集中在对消费者的影响上^[10],在企业的产品优化决策方面的应用较少。

在线评论中包含了大量的顾客对产品的意见信息,研究如何从非结构化的评论数据中提取出重要的意见信息,对产品的优化有着重要的指导意义.文章把产品评论数据应用于产品的更新优化中,对产品优化辅助信息获取的技术和方法进行研究,第1节介绍了基于在线评论的产品优化决策信息的获取的分析流程,第2节计算了顾客意见的权值计算方法,第3节介绍了产品优化信息的提取流程,第4节通过实例验证了方法的可行性。

1 研究框架

基于在线评论的产品优化决策信息的获取的分析流程分为几个步骤:评论数据采集,对数据的预处理,对评论数据进行挖掘,数据的应用和展示,如图1所示。

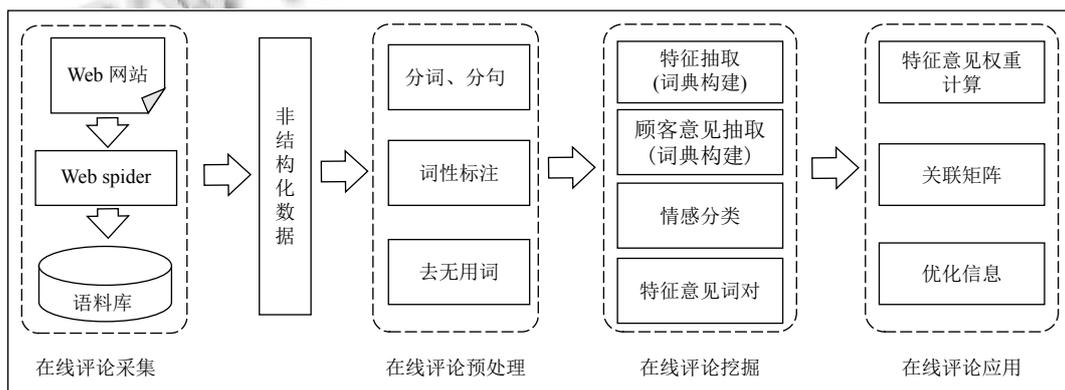


图1 基于在线评论的产品优化决策信息获取的挖掘流程

文章采用 scrapy 爬虫框架从电商网站爬取在线评论数据,并保存到语料库中;然后对语料库中的数据进行清洗和预处理,如去重、分词、分句等;在线评论挖掘是对预处理后的数据进行产品特征抽取和顾客意见抽取,评论数据的情感分类计算,产品特征意见词对的抽取;评论的应用主要是通过特征意见权重计算对特征意见词对进行重要性排序,构建从评论到产品优化的关联矩阵,获取优化信息。

2 在线评论顾客意见的权重计算

2.1 产品特征的关注度分析

词频统计指统计某词在文件中出现的次数,主要方法有 IDF、TF-IDF 等^[11],对评论数据进行词频统计

一定程度上可以反映顾客对产品各维度的关注情况.词频统计的基本思想是根据词频找到重要性较高的词,而通过加权的方式过滤掉重要性低的词,由于是对评论数据进行统计,一个词出现在多条评论中时,其重要性并不需要被降低,所以文章采用基于词频的方式直接进行统计,并通过词频占比的方式计算词的关注度,如式(1)所示。

$$F(t_i) = \frac{f(t_i)}{\sum_{t_m \in T(t)} f(t_m)} \quad (1)$$

2.2 顾客的满意度分析

情感分类是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程^[12-15].文章通过对评论数据的情感分析,探究顾客对产品各特征(外观、功能

等)的看法,确定产品在客户心中的满意程度,根据顾客的满意情况求出产品各特征在优化过程中的情感权重,即满意度和重要性呈负相关,基本步骤如下:

(1) 预处理 通过编写代码自动实现评论数据的去重、清洗、分词、去停用词的工作。

(2) 特征提取 由于机器学习只能对数值或类别数据进行训练,所以需要转化成向量的形式,文章通过Doc2Vec词向量模型^[16]对处理好的数据进行特征向量提取。

(3) 情感分类 通过逻辑回归分类器对提取的特征向量进行训练,生成情感分类模型,实现对评论数据的情感分类。

文章选取已标记的消极评论和积极评论各5000条作为训练集,通过对训练集预处理、特征提取、分类器训练3个步骤,构建分类模型,随机抽取消极评论和积极评论各1000条作为测试集,分类准确率88.6%,召回率87.7%,通过训练好的模型对数据进行情感分类,然后按照式(2)进行顾客的满意度在产品优化中的权重计算。

$$Q(t_i) = \begin{cases} 1 - q_i & q_i \geq q_k \\ \lambda(1 - q_i) & q_i < q_k \end{cases} \quad (2)$$

一般顾客的产品某一特征满意程度越低其在产品优化中的重要性越高,即其重要性和满意度成负相关,式(2)主要用于计算产品各特征的情感权重,表征产品各维度在产品优化中的重要性程度。 t_i 为产品的某一特征, $Q(t_i)$ 表示 t_i 特征的情感权重, q_i 为情感占比, λ 为权重系数, q_k 表示情感阈值,当产品某一维度情感占比低于阈值时,可以根据优化的需要适当提高情感权重,情感权重体现对满意度低于某值的特征的容忍程度。

2.3 在线评论顾客意见的权重计算

顾客对产品某一特征可能会有多个意见,为表征各意见的重要程度,文章把在某一产品特征中的某一顾客意见的重要性用内权值表示,并通过式(3)计算某一产品特征的各意见内权值。由于顾客的表达的多数为非专业术语,同一意见可能适用于多个产品特征,这时需要降低其权重,文章提出了外权值的概念,表示某一意见对各产品维度的重要程度,使用式(4)进行计算。

$$W_{O_j}[In(t_i)] = \frac{o_{ji}}{\sum_{o_{li} \in In(t_i)} o_{li}} \quad (3)$$

$$W_{O_j}[out(t_i)] = \frac{o_{ji}}{\sum_{o_{jn} \in out(t_i)} o_{jn}} \quad (4)$$

式中, t_i 表示产品某一维度, o_j 表示 t_i 的某一顾客意见, o_{ji} 为权重, $In(t_i)$ 为指向 t_i 的意见权重集合, $out(t_i)$ 为 o_j 指向的产品维度的意见权重集合。

产品特征 t_i 的某一顾客意见 o_j 的重要性由 t_i 的关注度、满意度、内权重、外权重共同决定,所以由式(1)–式(4)得出产品特征 t_i 的某一顾客意见 o_j 的总的权值为:

$$W(t_i o_j) = \frac{\lambda \beta o_{ji}^2 \cdot f(t_i) \cdot (1 - q_i)}{\sum_{o_{li} \in In(t_i)} o_{li} \cdot \sum_{o_{jn} \in out(t_i)} o_{jn} \cdot \sum_{t_m \in T(t)} f t_m} \quad (5)$$

通过式(5)即可分别求得各客户意见总体权重,权重越大其特征-意见词对 $t_i o_j$ 的重要程度越高,其中 β 是为了平衡值得大小便于展示,当 $q_i \geq q_k$ 时 λ 的值为1。

3 产品优化决策信息获取

3.1 在线评论中顾客意见提取流程

文章为了能从评论数据中提取出用于产品优化的辅助信息,需要先提取出在线评论中的顾客的具体意见,基本流程如图2所示。

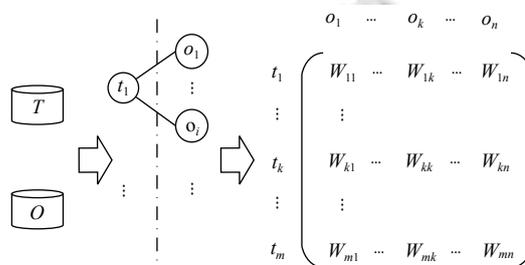


图2 顾客意见提取流程

文章通过词频抽取出来客户关注的产品特征和顾客意见构建特征词库 T 和意见词库 O ,假设每个产品特征都指向所有的客户意见,形成特征-意见词对 $t_i o_j$,再通过权重计算式(5)计算出评论中的特征-意见词对 $t_i o_j$ 的权重矩阵,最后根据权重矩阵提取出重要的顾客意见。

3.2 基于顾客意见的产品优化决策信息获取

评论中的顾客意见一般为顾客对产品某些特征的情感表达或问题现象的描述,表达较为模糊,需要构建客户意见 $t_i o_j$ 与具体优化信息 opt_i 的映射关系,映射关

系及其权重 w 每个企业都有所偏重,一般由专业人士确定,其映射关系如表1所示,其 opt_r 的重要性由 $t_i o_j - opt_r$ 映射关系表中的权重 w 和顾客意见的重要性共同决定.

表1 顾客意见和优化辅助信息的映射关系

	opt_1	opt_2	opt_3	opt_4	...
$t_1 o_1$	w_{11}	w_{12}			
$t_1 o_2$		w_{22}			
$t_2 o_1$		w_{32}	w_{33}		
$t_2 o_2$			w_{43}	w_{44}	
...					

假设任何一个的 $t_i o_j$ 都指向所有 opt ,通过 w 表示其相关性程度,如表1所示.映射关系组成的矩阵中,一般一个 $t_i o_j$ 对应一个或几个 opt_r ,所以映射关系矩阵是一个多数值为零的矩阵.通过权重矩阵提取出重要的 $t_i o_j$,再由映射关系表提取出对应的 opt_r ,实现由顾客模糊的情感表达达到产品优化所需要的决策信息的映射.

4 实例验证

通过 scrapy 框架爬取 vivo X20 手机在天猫网站上的产品评论信息,构建分析模型对这些评论语句进行分词等预处理操作,计算顾客关注的词频占比 $F(t_i)$,词频统计的结果可视化展示如图3所示.



图3 评论数据的词云图

根据词频统计提取出的客户关注度较高的产品特征,更新特征词库,对特征词进行情感分类,并计算客户满意度 $Q(t_i)$,情感阈值设置为0.5,情感权重系数设置为2,情感分类结果的可视化展示如图4所示.

在图4中, x 轴以上的为正面评论数, x 轴以下的为负面的评论数.通过情感分类结果可以直观的了解顾客对产品和服务属性的满意情况.

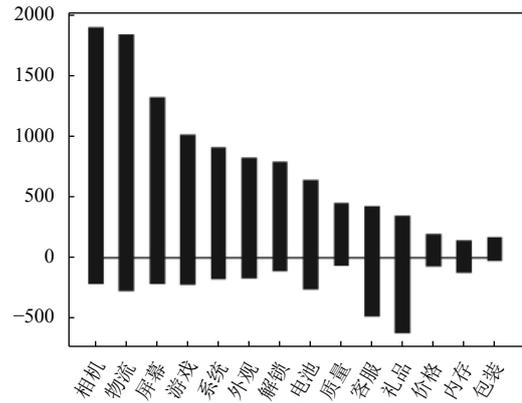


图4 在线评论的情感分类结果

更新顾客意见词库,构建产品特征和顾客意见的映射关系,根据权重计算式(5)把评论中的特征-意见词对 $t_i o_j$ 转化为权重矩阵,对重要的客户意见进行降序排列,提取前5个特征-意见词对,如表2所示.

表2 顾客意见及其权重表

t	o	w
客服	不理人	3.429 349 1
充电	发热	0.475 831 42
内存	不够用	0.387 583 12
电池	不耐用	0.360 130 41
游戏	卡顿	0.193 765 86

根据 $t_i o_j - opt_r$ 映射关系,提取出产品优化所需要的辅助参考信息 opt_r ,文章设置信息映射关系表中的权重都为1,其 opt_r 重要性全由 $t_i o_j$ 的重要性决定, opt_r 的提取如表3所示.

表3 顾客意见和优化辅助信息的相关矩阵

	提升服务质量	降低能耗	增强散热	增加内存容量	增加电池容量	提高游戏性能
客服不理人	3.4293					
充电发热		0.4758	0.4758			
内存不够用				0.3875		
电池不耐用		0.3601			0.3601	
游戏卡顿		0.1937	0.1937	0.1937		0.1937

通过分析可知,这款产品在服务质量、电池、内存等方面需要优化提升,具体优化方案如上表所示,优化信息中的列权重之和表征优化信息的重要性,在实际应用中,每个企业的优化方法有所侧重,这时专家可以适当的调整映射关系矩阵中的权重系数,提高本方法的适用性.

最终根据相关矩阵中权重的列和,对产品优化信息进行降序排列,本章节仅取 top4 进行展示,如表4所示.

表4 产品优化方案

vivo X20	优化方案	重要度
1	提升服务质量	3.43
2	降低能耗	1.03
3	增强散热	0.67
4	增加内存	0.58

5 结语

文章对产品优化辅助信息获取的技术和方法进行了研究,通过构建的客户意见的权重算法模型,实现了产品优化辅助信息提取,为企业中产品的优化提供参考.通过实例分析提取出产品优化所需要的顾客反馈信息,在节省的大量的人力成本的情况下为产品的优化提供数据支持,提高了产品优化的科学性.

参考文献

- 邵景波, 胡名叶, 许万有. 在线评论情感属性的动态变化——基于 iPhone 4 的实证研究. 预测, 2016, 35(5): 9–15.
- 朱丽叶, 袁登华, 张静宜. 在线用户评论质量与评论者等级对消费者购买意愿的影响——产品卷入度的调节作用. 管理评论, 2017, 29(2): 87–96.
- 王阳, 王伟军, 刘智宇. 在线负面评论信息对潜在消费者购买意愿影响研究. 情报科学, 2018, 36(10): 156–163.
- 相薏薏, 郭顺利, 张向先. 面向用户信息需求的移动商务在线评论效用评价研究. 情报科学, 2018, 36(2): 132–138, 158.
- 苗蕊, 徐健. 评分不一致性对在线评论有用性的影响--归因理论的视角. 中国管理科学, 2018, 26(5): 178–186.
- Alaei AR, Becken S, Stantic B. Sentiment analysis in tourism: Capitalizing on big data. Journal of Travel Research, 2019, 58(2): 175–191. [doi: 10.1177/0047287517747753]
- 修国义, 王俭, 过仕明. 引入信息传递效率的在线评论效用评价. 情报科学, 2019, 37(1): 43–50.
- Cheng XS, Fu SX, Sun JS, *et al.* An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust. Tourism Management, 2019, 71: 366–377. [doi: 10.1016/j.tourman.2018.10.020]
- 于超, 樊治平. 基于顾客在线评论的服务要素优化配置方法. 计算机集成制造系统, 2019, 25(3): 714–725.
- Elwalda A, Lü K, Ali M. Perceived derived attributes of online customer reviews. Computers in Human Behavior, 2016, 56: 306–319. [doi: 10.1016/j.chb.2015.11.051]
- 秦海菲, 杜军平. 酒店在线评论数据的特征挖掘. 智能系统学报, 2018, 13(6): 1006–1014.
- Verhagen T, Nauta A, Feldberg F. Negative online word-of-mouth: Behavioral indicator or emotional release? Computers in Human Behavior, 2013, 29(4): 1430–1440. [doi: 10.1016/j.chb.2013.01.043]
- 蔡淑琴, 秦志勇, 李翠萍, 等. 面向负面在线评论的情感强度对有用性的影响研究. 管理评论, 2017, 29(2): 79–86.
- Geetha M, Singha P, Sinha S. Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis. Tourism Management, 2017, 61: 43–54. [doi: 10.1016/j.tourman.2016.12.022]
- 郑丽娟, 王洪伟. 基于情感本体的在线评论情感极性及强度分析: 以手机为例. 管理工程学报, 2017, 31(2): 47–54.
- 逯万辉, 谭宗颖. 学术成果主题新颖性测度方法研究--基于 Doc2Vec 和 HMM 算法. 数据分析与知识发现, 2018, 2(3): 22–29. [doi: 10.11925/infotech.2096-3467.2017.1012]