

融合信息增益和梯度下降算法的在线评论有用程度预测模型



冯进展 蔡淑琴

华中科技大学管理学院 武汉 430074

(fjinzhan@hust.edu.cn)

摘要 由于无法预知产品在线评论的文本内容是否对浏览者有用,大量的无用评论增加了潜在消费者的信息搜索成本,甚至降低了潜在消费者购买产品的可能性。为提高电子商务平台的有用在线评论率,为撰写评论者提供测试功能,建立在线评论有用程度预测模型。根据在线评论的文本特征,所提模型选择在线评论的词语数量、词语的有用值、产品特征数量等3个特征,构建一个预测在线评论有用程度的模型,其中词语的有用值是词语区分在线评论有用程度的信息增益量,然后根据大量在线评论数据利用梯度下降算法解出模型参数。实验结果显示,随着词语数量、词语有用值、产品特征数量的增长,评论有用程度不断提高。实验中把在线评论分为一般、有用、非常有用3个程度,对于一般的在线评论,预测精确率为92.96%;对于“有用”在线评论,预测精确率为94.83%;对于“非常有用”在线评论,预测精确率为67.63%。实验对模型性能进行测试,得到平均精确率为85.05%,召回率为82.81%,F1值为83.72%,该结果验证了所提模型预测在线评论有用程度的可行性。

关键词: 在线评论;有用程度;信息增益;梯度下降法

中图法分类号 TP391

Helpfulness Degree Prediction Model of Online Reviews Fusing Information Gain and Gradient Decline Algorithms

FENG Jin-zhan and CAI Shu-qin

School of Management, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract Because it is impossible to predict whether the text content of online product reviews is helpful for viewers, many reviewers write a large number of unhelpful reviews, which increases the cost of information search for potential consumers, and even reduces the possibility of potential consumers buying products. In order to improve the helpful online reviews rate of e-commerce platform and provide test function for reviewers, a prediction model of online reviews helpfulness is established. According to the text characteristics of online reviews, the model chooses three features of online reviews: the number of words, the helpful value of words, and the number of product features, to construct a model for predicting the helpfulness of online reviews. The helpful value is the information gain of words to distinguish the helpfulness of online reviews. And then, according to a large number of online reviews, by using the gradient descent algorithm, the model parameters are solved. The experimental results show that with the increase of the number of words, helpful value of words and the number of product features, the helpfulness of reviews increases continuously. The online reviews are divided into three levels: general, helpful and very helpful. The general predicted accuracy of online reviews is 92.96%, helpful accuracy is 94.83%, and very helpful accuracy is 67.63%. The average accuracy, recall and F1 of the model are 85.05%, 82.81% and 83.72%, respectively. The results verify the feasibility of the model to predict the helpfulness of online reviews.

Keywords Online reviews, Helpfulness degree, Information gain, Gradient descent algorithm

1 引言

在线评论数量巨大并以几何级增长,潜在消费者需要有用的评论才能做出购买决策^[1],但有用的在线评论被大量无用的评论所掩盖。有用的在线评论不但能够提供大量的产品

特征信息,而且能提高电子商务网站的经济价值,其既是改进产品质量的强力工具,也是电子商务企业最有力的销售工具之一。电子商务网站包含的有用评论越多,就越会吸引潜在消费者访问其网站,增加网站访问量,从而提高访问量转换到销售量的概率^[2]。有用的在线评论能有效减小潜在消费者的

到稿日期:2019-07-03 返修日期:2019-09-27 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(71371081);教育部博士点基金(20130142110044)

This work was supported by the National Natural Science Foundation of China (71371081) and Specialized Research Fund for the Doctoral Program of Higher Education (20130142110044).

通信作者:蔡淑琴(caishuqin@hust.edu.cn)

决策风险,同时还能提高电子商务平台的销售业绩^[3]。对于企业来讲,有用在线评论是一种稀缺资源,占有优势稀缺资源可使企业掌握信息资源分配的置权,从而获得巨大财富。因此,电子商务企业经常开展“有用评论有奖”活动来激励用户撰写有用评论,如拼多多鼓励消费用户基于真实的交易对拼多多商家做出公正、客观、真实的有用评价,遏制无用评论、无效评论,采取有用评价评分制给予优惠激励。但用户在撰写评论时无法得知评论是否对潜在消费者有用,电子商务企业的信息资源管理平台需要提供一种预测评论有用度的功能,帮助用户预测其评论的有用度。因此,预测评论有用程度是企业信息资源管理领域亟需解决的问题。

现有研究主要集中于有用性的影响因素,即“什么因素致使潜在消费者认为评论有用”,但目前鲜有预测在线评论有用程度的研究,相对于在线评论的有用性,其有用程度更为重要,如某产品的一条在线评论只有一个人认为有用,而另一条评论有 100 个人认为有用,从有用性角度看这两条评论都有用,但从有用程度角度看后者更为有用。本文提出一个预测在线评论有用程度的模型,把在线评论分为“一般”“有用”“非常有用”3 个等级。首先选择评论文本内容特征的词语数量、词的有用值之和、产品特征数量作为预测模型的预测因子,其中词的有用值通过信息增益计算每个词语具体的有用值;然后对大量的在线评论非结构数据进行处理,将其转变为数字数据,运用梯度下降法求解模型参数。

2 研究综述

在线评论文本内容特征的有用性影响因素包括评论长度、情感。词语多的评论能提供更多的产品特征信息,比短评论拥有更深刻的产品体验^[4]。一般认为过短的评论为“灌水”行为,而长的评论拥有更多的产品细节信息,增强了评论的有用性^[2]。长评论包含很多细节,而许多细节给出了产品特征情况,对理解产品特征起到了积极作用^[5]。依据前景理论,消费者对风险感知非常敏感,负面信息相对于正面信息更能吸引消费者关注,因此负面信息具有说服力^[6]。但负面信息评论不是越多就越使人信服,当负面信息量占总信息量的比例超过最佳唤醒水平时,再增加负面信息就会降低说服效果。除评论文本特征外,评论者、评论时间也是评论有用性的主要影响因素^[7]。有研究表明,在线评论的各特征项的“贡献度”中,评论词语数、产品特征数量、评论发表时间、评论者排名是影响评论有用性的关键特征^[8],级别高的评论者其评论的有用性较高^[9]。另外,评论所在店铺的特征都会对评论有用性产生影响,如店铺收到的订单和得到的评论越多,其评价就越有用^[10]。发布的评论时间离现在越远,感知有用性就越高,因为人们习惯认为越久的评论的有效性越强,同时,经过长时间使用产品后撰写的评论的有效性越强^[11]。

对于在线评论的有用性预测研究主要聚焦于预测因子的选择和模型构建。Lee 等把在线评论的产品数据、评论特征、评论文本特征作为预测模型的预测因子,通过反向传播方法利用多层感知器神经网络(Multi-Layer Perceptron, MLP)模型预测评论的有用性,其认为评论文本特征中词数量、每句的词数量、词的平均字母数量对评论有用性起着关键作用^[12]。Singh 等对于在线评论的文本内容运用自然语言处理技术,

把星级、评论极性、形容词数量、副词数量、名词数量、动词数量、易读词数量、困难读词数量、词平均长度、错词数量、长字母词、单一字母词作为预测模型的预测因子,通过 Gradient Boosting 算法求最小均方差,结果显示评论极性、评论星级、形容词数量对有用性起着关键作用^[13]。Park 在 LIWC2015 词典的帮助下,得到在线评论的比较词数量、分析词数量、专业词数量、信任词数量、认知词数量、感知词数量、积极情感词数量、消极情感词,以及与评论星级、评论总词数量、句子平均词数一起作为模型预测的预测因子,通过线性回归、支持向量机回归、随机森林回归、M5P 决策树模型分别预测评论的有用性,其中随机森林回归模型的预测效果最好,其中词的总数量、分析词数量起到了关键性作用^[14]。Krishnamoorthy 通过 LCM 语言模型以及人工标注方式,将语言特征类型、元数据特征、可读性、主观性作为模型的预测因子,通过随机森林算法预测在线评论的有用性,以验证所提特征的有效性^[15]。Jiang 等运用复杂网络分析评论的有用性,将评论看作内容相互连接的网络拓扑状,建立基于拓扑势的评价模型 TPBM,通过划分评论网络社区,采用主观评价与网络拓扑相结合的方法度量评论的有用性^[16]。

由上述内容可知,目前在线评论的有用性研究取得了很大成就,但仍存在以下问题:1)大部分研究只是预测在线评论是否具有有用性,鲜有研究评论有用的程度;2)有的研究通过抽取评论文本中不同类型的词语,如形容词、副词、名词、动词、易读词、困难读词、比较词、分析词、专业词、信任词、认知词、感知词等,来试图分析不同类型词语对评论有用性的影响程度,但这些研究只是简单地统计数据,并没有计算出这些词语的有用值,本文通过词语信息增益算法来计算词语的有用值;3)预测模型假设评论的文本内容特征变量与有用性呈线性关系,但实际词语的表达并不是简单两个词的堆积,它们之间可能存在非线性关系。

3 在线评论有用性预测模型的构建

本文研究的目的是鼓励评论者撰写有用程度更高的评论,因此模型预测因子忽略了评论者特征、时间、接受者特征等因素,也不考虑带目的性营销的垃圾评论,只关注撰写在线评论。选择文本内容特征中词语数量、词语有用值、产品特征数量作为预测因子,通过词的有用值来区分不同类型词语的作用。在线评价有用程度预测模型的流程为数据处理、预测因子的测量、模型构建、模型参数求解、模型性能测试,如图 1 所示。

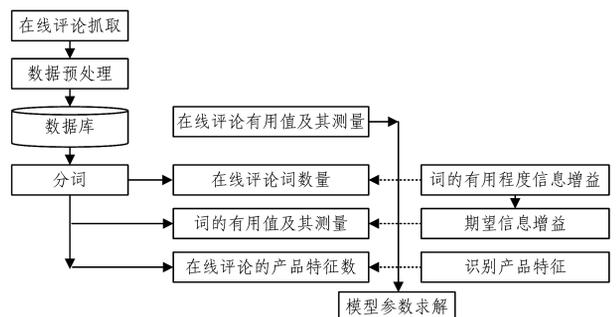


图 1 在线评论有用程度预测模型

Fig. 1 Prediction model of helpfulness of online reviews

3.1 在线评论的有用程度及测量

在线评论有用程度表示评论对潜在消费者的有用程度。如在亚马逊手机评论中,A评论有2人认为有用,而B评论有56人认为有用,这两条评论都有有用性,但有用程度是不同的。本文将在线评论的有用度分为 t 个等级。将某产品的在线评论有用得票数与其最高得票数的比值作为在线评论的有用值 y_i ,如式(1)所示:

$$y_i = score_i = \frac{vote_i}{MAX_1^N(vote_i)} \quad (1)$$

其中, $vote_i$ 表示评论 i 得到的有用票数, $score_i$ 表示评论 i 的有用得分。在线评论的有用值 y_i 根据式(2)进行有用程度分类。

$$hd_i = \begin{cases} 0 \leq score_i < h, & 1 \\ h \leq score_i < l, & 2 \\ \dots & \\ p \leq score_i < q, & t-1 \\ q \leq score_i \leq 1, & t \end{cases} \quad (2)$$

其中, hd_i 表示评论 i 的有用程度, $h, l, p, q \in score$ 。

3.2 评论的词语数量及其有用值测量

在线评论的词语数量是影响有用程度的重要因素,词语数量越多,传达的产品信息就越多,从心理角度讲词语数量多的评论的撰写者被认为对产品认识深刻、态度真诚,而另长的评论使阅读者能够很好地理解评论者表达的意思。齐普夫的最省力原则认为人们在参加各种活动时,都有意无意地选择一条最省力的途径到达目的地,即选择一条符合自己的路径使自己付出的“力”最小。在语言表述思维上,如同两个不同的相反作用力,即“单一化力”和“多样化力”^[17],一方希望尽量简短描述,而另一方希望能够完全理解对方的表达,表达者希望用一个词来表达所有概念最省力,而阅读者希望每个概念都用不同的词表达为好。从可读性角度讲,评论的词语数量是衡量可读性的指标,评论的词语数量多则能更好地刺激潜在消费者对评论进行阅读^[18]。一般认为至少一行文字的评论有阅读性,因此本文运用以10为底的对数函数表示在线评论的词数量的有用值, x_{i1} 表示在线评论词语的数量, a 表示 x_{i1} 的有用增强系数,则在线评论词的有用值测量表达式为:

$$\log_{10}(1 + ax_{i1}) \quad (3)$$

3.3 词的有用信息熵及有用值测量

许多研究试图从评论文本中划分的不同词语中找出其对有用性的影响力,其本质是求解每个词语在评论中的有用值。本文首次通过信息增益来衡量词语对在线评论的有用值。信息论之父香浓提出了自信息的信息,解决了信息的量化度量难题,自信息是不确定性的度量,如果信息消除的不确定性越大,则信息含有的信息量越大。自信息用该事件概率倒数的对数来表示其所带来的信息量。信息熵是各离散自信息的期望值,用于表示信息的价值。词语的有用值可以用消除在线评论有用程度的不确定性来衡量,词语确定在线评论的有用程度越大,其有用值就越大。在线评论的有用程度划分为 c 个等级, $H(s)$ 表示确定评论的有用程度级别的信息量, $p(c_i)$ 表示评论有用程度类别 c_i 的概率:

$$H(s) = - \sum_{c=1}^c p(c_i) \log(p(c_i)) \quad (4)$$

已知某个词 w 在 c_i 类的条件下,词 w 确定评论有用程

度的信息量,即词 w 的信息熵为:

$$H(S|w) = - \sum_{i=1}^c p(c_i|w) \log(p(c_i|w)) \quad (5)$$

词 w 信息增益表示在已知词 w 的情况下,在线评论划分确定到有用程度类别 c_i 的预期熵减少量,即:

$$Ingain(w) = H(S) - H(S|w) \quad (6)$$

则词对在线评论的有用程度确定的信息增益为:

$$\begin{aligned} Ingain(w) &= H(S) - H(S|w) \\ &= - \sum_{i=1}^c p(c_i) \log(p(c_i)) + p(w) \sum_{i=1}^c p(c_i|w) \log \\ &\quad (p(c_i|w) + p(\bar{w}) \sum_{i=1}^c p(c_i|\bar{w}) \log(p(c_i|\bar{w})) \end{aligned} \quad (7)$$

其中, c_i 表示有用程度类别中的第 i 类; $p(c_i)$ 表示在线评论在类别 c_i 中的概率; $p(w)$ 表示包含词 w 的评论占整个评论的概率; $p(\bar{w})$ 表示不含词 w 的评论占所有评论的概率; $p(c_i|w)$ 表示含有词 w 在 c_i 类的评论中占有词 w 评论的概率; $p(c_i|\bar{w})$ 表示不含词 w 在 c_i 类的评论占没有含词 w 所有评论的概率。 $Ingain(w)$ 可以计算出各个词对评论在各个有用程度类别中的增益信息,但是增益信息一般偏好与其出现类别概率较大的类别。如 w 在一般程度类别中出现的概率最大而在其他类别中出现的概率较小,且其增益信息也较大,但不能将其增益信息作为词的有用程度,需要通过其期望值来计算词的有用值 $h(w)$:

$$h(w) = Ingain(w) \sum_{i=1}^c score_i * p(c_i|w) \quad (8)$$

其中, $h(w)$ 表示词在评论中的有用值; $score_i$ 表示类别 c_i 的有用值在其所在的区间取平均值, $p(c_i|w)$ 表示含有词 w 的评论在 c_i 中的概率。一个词传达的信息可能模糊,而几个词会呈现出现清晰的信息,如“小桥流水人家”,如果词“小桥”“人家”“流水”单独出现,画面就比较模糊,但几个词连在一起就是一个清晰的清新幽静的画面,多词语链接在一起使用存在马太效应,因此本文采用指数函数来表示词在评论中的有用值。一个评论是否有用,取决于评论中所有词语的组合,一个评论所有词的有用值之和表示其有用值。 x_{i2} 表示在线评论所有词的有用值之和,即 x_{i2} 是一个评论中所有词的有用值之和, $b^{cx_{i2}}$ 等于 x_{i2} 用 b 为底数、 $\sum_{i=1}^n h(\omega_i)$ 为指数、 c 为系数的指数函数表示在线评论词的有用值:

$$b^{cx_{i2}} = b^{c \sum_{i=1}^n h(\omega_i)} \quad (9)$$

3.4 在线评论产品特征及有用值测量

潜在消费者阅读在线评论的主要目的是获取产品特征的信息,因此评论中介绍的产品特征体验越多,其有用程度就越高。但在在线评论产品特征需要通过信息抽取得到,Wang等采用粗糙集的方法抽取在线评论中固定搭配对的产品特征和情感词^[19];Yu等提出FP-Growth算法挖掘获得种子产品特征词,然后通过增量迭代发现新的产品特征词^[20]。本文建立产品特征为本体,用于识别评论中提到的产品特征的数量。首先将产品官方网站参数作为产品特征词,通过自然语言处理技术处理后统计高频率的名词、动词组成候选词,最后进行人工统计与筛选。对手机的在线评论构建产品特征本体的流程如图2所示。

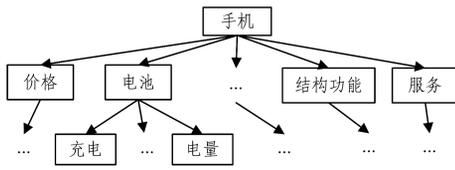


图2 产品特征本体-手机

Fig. 2 Product feature ontology: mobile phone

本文通过产品领域本体识别在线评论的产品特征,计算其描述的产品特征个数 x_{i3} ,以 e 为底数的对数函数来表达产品特征个数对评论的有用值,其中 d 表示产品特征个数的有用值系数。

$$\log(1 + dx_{i3}) \tag{10}$$

3.5 在线评论有用性预测模型

综上所述,本文将在在线评论的词数量、词的有用值、产品特征作为预测因子,在线评论的有用值 y_i 为预测目标,预测因子的系数分别为 $\lambda_1, \lambda_2, \lambda_3$,则在线评论预测模型为:

$$f(x_{i1}, x_{i2}, x_{i3}) = \lambda_1 \log_{10}(1 + ax_{i1}) + \lambda_2 b^{cx_{i2}} + \lambda_3 \log(1 + dx_{i3}) \tag{11}$$

其中, x_{i1}, x_{i2}, x_{i3} 分别表示第 i 条评论的词数量、词的有用值、产品特征。

4 实验与模型检验

4.1 数据收集及处理

实验收集在线评论数据,以手机为实例。编写 Python 爬虫程序抓取京东、亚马孙网站关于荣耀、苹果、三星、魅族等品牌手机的在线评论,选择条件为产品的评论数大于 50 条,抓取内容为评论文本内容、有用得票数,经清理得到 3538 条评论。用 Python 进行数据处理,得到模型的 $y_i, x_{i1}, x_{i2}, x_{i3}$ 。

(1)通过式(1)计算得到各条在线评论的有用值 y_i ;

表2 模型的 $y_i, x_{i1}, x_{i2}, x_{i3}$

Table 2 $y_i, x_{i1}, x_{i2}, x_{i3}$ of model

评论	x_{i1}	x_{i2}	x_{i3}	y_i
这个/ 价格/ 挺/ 合适/, / 媳妇/ 很/ 喜欢/ !	7	0.0722683	1	0.005
电池/ 不算/ 耐用/, / 拍照/ 的/ 效果/ 一般/ 没/ 预计/ 的/ 好/, / 运行/ 处理速度/ 快/, / 千元/ 机里/ 来说/ 不错/, / 我/ 挺/ 喜欢/. / 物流/ 发货/ 快/, / 也/ 没有/ 损坏/, / 赞/ 一个/。	29	0.8509548	5	0.04
手机/ 速度/ 很快/, / 就是/ 耗电/ 有点/ 快/, / 不够/ 我/ 那台/ x20p/ 耗电/ 慢点/, / 但/ 没事/, / 指纹/ 解锁/ 真的/ 很/ 好/, / 应该/ 不/ 玩游戏/ 能撑/ 一天/ 吧/. / 喜欢/ 的/ 朋友/ 快快/ 入手/ 哦/。	32	1.6100575	6	0.11
不错/, / 非常/ 不错/ ! / 我/ 居然/ 抢到/ 了/ ! / 屏幕/ 是/ 魅族/ 手机/ 有史以来/ 最好/ 的/ 2k/ 屏幕/ ! / 国产手机/ 中/ 最好/ 的/ 屏幕/ 没有/ 之一/ ! / 手感/ 贴个度/ 可以/ 单手操作/, / 我手/ 比较/ 大/. / 不像/ 魅族/ pro6plus/ 那样/ 单手操作/ 非常/ 吃力/ ! / 握持/ 感/ 非常/ 棒/ ! / 信号/ 也许/ 是/ 魅族/ 手机/ 中/ 信号/ 最好/ 的/ 一部/ 吧/ ! /	52	2.2519576	11	0.19
在/ VIVOX21/ 和/ OPPOR15/ 之间/ 选择/, / 感觉/ R15/ 主打/ 拍照/, / X21/ 屏幕/ 指纹/ 功能/ 是/ 创新/, / X21/ 在/ 性能/, / 功能/ 包括/ 音乐/, / 拍照/ 各/ 方面/ 综合/ 比较/ 好/, / 所以/ 最终/ 我/ 还是/ 选择/ 了/ X21/ 屏幕/ 指纹/ 版/. / VIVO/ 目前/ 除了/ 是/ NBA/ 的/ 赞助商/, / 还是/ 俄罗斯/ 世界杯/ 的/ 赞助商/, / 现在/ 还有/ 库里/ 给/ X21/ 代言/, / 所以/ 必须/ 选它/ 支持/ 国货/ ! / 用到/ 现在/ 都/ 还/ 不错/, / HIFI/ 音质/ 很/ 好/, / 玩游戏/ 也/ 不卡/, / 拍照/ 和/ 其他/ 功能/ 都/ 挺/ 好/ ! / 流畅/ 挺/ 不错/ 的/ 一款/ 手机/ ! /	85	4.2193415	13	0.44
机器/ 很/ 精致/, / 金属/ 框/ 和/ 旁边/ 的/ 实体/ 键能/ 看/ 出来/ 华为/ 的/ 工艺/ 上/ 进步/ 了/ 很多/, / 赞/ 一个/ ! / 5.9/ 寸/ 也/ 没有/ 想象/ 的/ 那么/ 大/, / 占屏/ 比/ 很/ 高/. / 这一版/ 的/ U/ 做/ 的/ 很/ 用心/, / 虚拟/ 键/ 大部分/ 程序/ 都/ 可以/ 做到/ 沉浸/, / 屏幕/ 相比/ 之前/ 舒服/ 多/ 了/, / 个别/ 程序/ 需/ 进一步/ 升级/ 完善/. / 随意/ 按/ 了/ 下/ 快门/, / 放大/ 后/ 床单/ 上/ 布纹/ 清晰/ 的/ 纹理/ 惊到/ 我/ 了/ ! / 莱卡/ 绝非/ 浪得虚名/ ! / 外放/ 横屏/ 立体声/ 音质/ 虽然/ 比/ 其他/ 用过/ 的/ 华为/ 手机/ 好些/, / 我/ 觉得/ 华为/ 还是/ 应该/ 在/ 音质/ 上/ 再/ 多/ 下点/ 功夫/. / 机器/ 直接/ 贴/ 了/ 保护膜/, / 送/ 了/ 背壳/, / 多送/ 了/ 一个/ 数据/ 转接头/, / 耳机/ 一应俱全/, / “美酒”/ 真的/ 很/ 用心/. / 亚马逊/ 也/ 是/ 一如既往/ 的/ 贴心/, / 送/ 了/ 64G/ 的/ 三星/ U1/ 高速/ 卡/ 和/ 数据线/. / 总之/ 对/ 手机/ 和/ 亚马逊/ 都/ 很/ 满意/ ! /	145	14.6038	23	0.83

(2) y_i 根据式(2)得到各评论的有用度 hd_i, c_i 分为一般、有用、非常有用 3 个级别;

(3)通过哈尔滨工业大学自然语言技术平台(LTP)对每条评论文本进行分词处理,对照百度停用词表去掉评论的停用词,统计每条评论的不同的词数量 x_{i1} ;

(4)根据式(4)计算各评论的有用程度,分类需要的信息量 $H(S)$;

(5)将所有评论出现的词构建成为词典,根据式(6)、式(7)计算每个词 w 的信息增益,再根据式(8)计算每个词的有用值,如表 1 所列;

(6)根据建立的手机产品特征本体,计算每条评论包含的产品特征词的数量;

(7)得到模型的 $y_i, x_{i1}, x_{i2}, x_{i3}$,如表 2 所列。

表1 部分词的有用值

Table 1 Helpful values of some words

词	有用值	词	有用值
功能键	0.642365	划痕	0.146329
照片	0.401478	帧	0.140450
手机电池	0.240887	真机	0.129807
数据	0.179983	握感	0.121352
续航力	0.240887	说实话	0.117465
无卡顿	0.162420	传感器	0.105414
音效	0.061029	光圈	0.060395
处理器	0.056308	外包装	0.054450
设置	0.045042	屏	0.044966
接口	0.102868	指纹识别	0.109412
显示屏	0.102264	操作系统	0.102868
分屏	0.102264	强烈推荐	0.102264
数据线	0.094693	电池容量	0.093284
铃声	0.084217	摄像头	0.081598
缓存	0.077673	屏保	0.071376
晒图	0.060395	卡机	0.061029
裸机	0.0516781	做工	0.045953
内置	0.0442935	识别系统	0.044198

4.2 模型参数求解

本文建立了一个损失函数 $L(y, f(X, \theta))$,采用预测值与

真实值的差的平方作为损失函数值。 θ 表示模型参数, m 表示样本数量,损失函数为 $R(\theta)$:

$$R(\theta) = \sum_{i=1}^m (y_i - f(x_i, \theta))^2 \quad (12)$$

将已知在线评论的处理数据作为样本经验,通过样本经验逐渐逼近模型的参数,即目标变成求 $R(\theta)$ 的最小值,并求出模型中的所有参数,其中参数都大于 0,当误差接近最小时参数 θ 就是预测模型的解:

$$\theta = \operatorname{argmin}(R(\theta)) \quad (13)$$

梯度下降算法是解决大规模数据的有效学习方法之一,具有更高的准确性。由于样本独立同分布,目标函数的梯度需遍历一次所有的样本数据,即计算所有的样本来求解梯度。 ρ 表示学习率,批量梯度下降的迭代式为:

$$\beta = \lambda_1 \log_{10}(1 + ax_{i1}) + \lambda_2 b^{cx_{i2}} + \lambda_3 \log(1 + dx_{i3}) \quad (14)$$

$$\lambda_1 = \lambda_1 - \rho \sum_{i=1}^m \beta * \log(1 + x_{i1}) \quad (15)$$

$$a = a - \rho \sum_{i=1}^m \beta * \lambda_1 * \frac{1}{(1 + ax_{i2}) \log(10)} \quad (16)$$

$$\lambda_2 = \lambda_2 - \rho \sum_{i=1}^m \beta * b^{cx_{i2}} \quad (17)$$

$$b = b - \rho \sum_{i=1}^m \beta * x_{i2} * \lambda_2 * c * b^{(cx_{i2}-1)} \quad (18)$$

$$c = c - \rho \sum_{i=1}^m \beta * \lambda_2 * b^{cx_{i2}} * \log(cx_{i2}) * x_{i2} \quad (19)$$

$$\lambda_3 = \lambda_3 - \rho \sum_{i=1}^m \beta * \log(1 + dx_{i3}) \quad (20)$$

$$d = d - \rho * \sum_{i=1}^m \beta * \frac{x_{i3}}{1 + dx_{i3}} \quad (21)$$

对 3538 条评论数据进行模型参数求解,迭代模型的求解结果如图 3 所示,迭代到 1311 次后,损失函数的值从 203 降至 51,模型得到收敛,此时的参数如表 3 所列。

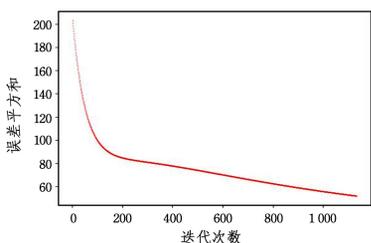


图 3 迭代次数与误差平方和

Fig. 3 Iteration times and sum of squares of errors

表 3 模型参数

Table 3 Model parameters

λ_1	λ_2	λ_3	a	b	c	d
0.167356	0.002513	0.048997	0.086271	1.00351	0.0124394	0.043715

在线评论有用程度预测模型为:

$$y = 0.167356 \log(1 + 0.086271x_{i1}) + 0.002513 * (1.00351)^{0.0124394x_{i2}} + 0.048997 \log(1 + 0.043715x_{i3})$$

4.3 模型检验

预测模型的效果评估采用常见的评价标准:准确率、召回率、 F_1 值。 TP 表示将正类预测为正类数; FP 表示将其他类预测为正类数误报; FN 表示将正类预测为其他类数。

(1) 精确率(Precision)

$$P = \frac{TP}{TP + FP} \quad (22)$$

(2) 召回率(recall)

$$R = \frac{TP}{TP + FN} \quad (23)$$

(3) 综合评价指标(F1-Measure)

$$F_1 = \frac{2PR}{P + R} \quad (24)$$

模型结果按照式(2)转为有用程度级别,在线评论有用程度分为一般[0,0.2)、有用[0.2,0.4)、非常有用(0.4以上)3个程度等级,通过模型预测的有用程度,用 P,R,F1 进行衡量。模型分别选择词语数量、产品特征数量、词语数量+词语有用值、产品特征数量+词语有用值、词语数量+词语有用值+产品特征数量作为特征,进行特征对比以检验有效性,实验结果如表 4 所列。

表 4 模型性能的测试结果

Table 4 Results of model performance test

模型特征	有用程度	P	R	F1
词语数量	一般	0.9989	0.7227	0.8389
	有用	0.6801	0.2128	0.3241
	非常有用	0.7692	0.2127	0.3332
	平均值	0.8160	0.3827	0.4987
产品特征数量	一般	0.7490	0.9910	0.8561
	有用	0.7739	0.3187	0.4514
	非常有用	0.7567	0.1985	0.3144
	平均值	0.7598	0.5027	0.5406
产品特征数量+词语有用值	一般	0.8124	0.9881	0.8916
	有用	0.4611	0.7410	0.5684
	非常有用	0.3714	0.7428	0.4975
	平均值	0.5483	0.8239	0.6525
词语数量+词语有用值	一般	0.9879	0.9376	0.9620
	有用	0.9767	0.8200	0.8915
	非常有用	0.2269	0.7805	0.3515
	平均值	0.7305	0.8460	0.7350
TF-IDF(随机森林)	一般	0.9417	0.9649	0.9532
	有用	0.8786	0.8566	0.8674
	非常有用	0.7037	0.5428	0.6218
	平均值	0.8413	0.7881	0.8111
词语数量+词语有用值+产品特征数量	一般	0.9296	0.9698	0.9492
	有用	0.9483	0.8132	0.8755
	非常有用	0.6763	0.7013	0.6872
	平均值	0.8505	0.8281	0.8372

当选择词语数量为特征时,其准确率为 81.6%,能够有效区分在线评论的有用程度,但其 F_1 值为 0.4987 较低;当选择产品特征作为模型特征时,其准确率为 75.98%,能够在一定程度上区分在线评论的有用程度,其 F_1 值为 0.5406,也较低。当将产品特征数量和词语有用值两个特征放入模型后, F_1 值上升到 0.6525;当将产品特征数量和词语有用值两个特征放入模型后, F_1 值上升到 0.7350,说明特征加入词语有用值提升了模型的综合评价指标。当词语数量、词语有用值、产品特征数量作为特征加入模型时,精确率为 85.06%、召回率为 82.81%、 F_1 值为 83.72%,模型能够对在线评论的有用程度进行有效预测,说明模型选择的 3 个特征是有效的。从本文模型选择的 3 个特征的结果可以看出,模型识别在线评论的“有用”的精确率最高,其能有效区分“一般”评论和“非常有用”评论。但对“非常有用”评论的精确率较低,主要原因是部分很有用的评论由于发布时间与阅读者浏览时间发生冲突,被淹没在大量“一般”评论中,没有得到投票,但模型预测是非常有用的,因此导致精确率很低,“一般”和“有用”的精确率、召回率、 F_1 都很高,说明模型很好地识别出了“一般”和“有用”。为进一步检验模型的有效性,选择 800 个关于手机评论的关键词作为在线评论的空间向量特征,采用关键词的 TF-IDF 作为向量空间特征值。由于数据的维度为 800,采用

支持向量机时分类效果较差,而随机森林算法在处理高维数据时分类表现很优异,具有人工干预较少等优点^[21],因此采用随机森林算法对在线评论的有用程度进行分类,数据分为训练集与测试集。本文提出的模型比随机森林模型的分​​类准确率、召回率、F1 都高,分别提高了0.91%,4%,2.61%。

结束语 本文从提高电子商务企业的在线评论信息资源管理水平的目的出发,希望评论者用心撰写有用评论,针对评论者撰写时无法知道其评论被浏览者认为的有用程度的问题,提出构建基于文本内容的在线评论自动预测有用度模型,以解决这一信息管理问题。在已有文献研究的基础上,选择在线评论文本的词数量、文本中的词语有用值之和、产品特征的数量作为模型预测因子,其中词语有用值通过信息增益来计算;评论的有用程度分为一般、有用、非常有用3个有用程度等级。模型能够自动识别出撰写评论的有用程度,测试模型性能得到平均精确率为85.06%、召回率为82.81%、F1值为83.72%,该模型能够对在线评论的有用程度进行有效预测。由模型结果得到以下结论:1)评论的词语数量越多、词有用值越大、提到的产品特征越多,在线评论的有用程度就越高;2)在线评论的词语数量增加会导致评论有用程度等级上升,在撰写评论时需要认真选择表达的词,即选择有用值较大的词描述自己的产品体验;3)在线评论提到的产品特征越多,越能够提高评论的有用程度等级,因此在撰写在线评论时需要把自己的有深刻体验的产品特征表达出来。

本文研究的不足之处是,本文只从文本角度研究评论的有用程度,但没有考虑有用程度影响的因子,以及评论上传图片、追评、评论者身份等条件。另外,没有对模型参数是否有取值范围展开研究。针对在线评论的有用程度目前还没有具体方法,因此本文没有展开模型的对比,只是按照随机森林算法进行对比。

词语有用值对在线评论的有用程度等级有很大的影响,但本文没有聚类分析不同大小有用值的词的具体特征,后续研究将对此进行完善。

参 考 文 献

- [1] MIN H J, PARK J C. Identifying helpful reviews based on customers mentions about experiences[J]. Expert Systems with Applications, 2012, 39(15): 11830-11838.
- [2] SHAN Y. How credible are online product reviews? The effects of self-generated and system-generated cues on source credibility evaluation[J]. Computers in Human Behavior, 2016, 55: 633-641.
- [3] PENG L, ZHOU Q H, QIU J T. Research on the Model of Helpfulness Factors of Online Customer Reviews[J]. Computer Science, 2011, 38(8): 205-207.
- [4] PAN Y, ZHANG J Q. Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews[J]. Journal of Retailing, 2011, 87(4): 598-612.
- [5] FILIERI R. What makes an online consumer review trustworthy? [J]. Annals of Tourism Research, 2016, 58: 46-64.
- [6] HOMER P M. Message Framing and the Interrelationships among Ad-Based Feelings, Affect, and Cognition[J]. Journal of Advertising, 1992, 21(1): 19-33.
- [7] WU T Y, LIN C A. Predicting the effects of eWOM and online brand messaging: Source trust, bandwagon effect and innovation adoption factors[J]. Telematics & Informatics, 2017, 34(2): 470-480.
- [8] WANG H W, MENG Y. Helpful Features Identification of Online Reviews Quality on GBDT Feature Contribution[J]. Journal of Chinese Information Processing, 2017, 31(3): 109-117.
- [9] LI C, XIANG J, XIANG J. Assessment method of credibility on online product reviews[J]. Journal of Computer Applications, 2019, 39(1): 187-191.
- [10] HU X G, CHEN F X, ZHANG Y H. Research on impact factors of online reviews' helpfulness based on product reviews data[J]. Application Research of Computers, 2016, 33(12): 3559-3561.
- [11] SINGH J P, IRANI S, RANA N P, et al. Predicting the "helpfulness" of online consumer reviews[J]. Journal of Business Research, 2017, 70(1): 346-355.
- [12] LEE S, CHOEH J Y. Predicting the helpfulness of online reviews using multilayer perceptron neural networks[J]. Expert Systems with Applications, 2014, 41(6): 3041-3046.
- [13] SINGH J P, IRANI S, RANA N P, et al. Predicting the "helpfulness" of online consumer reviews[J]. Journal of Business Research, 2017, 70: 346-355.
- [14] PARK Y J. Predicting the Helpfulness of Online Customer Reviews across Different Product Types[J]. Sustainability, 2018, 10(6): 1735.
- [15] KRISHNAMOORTHY S. Linguistic features for review helpfulness prediction[J]. Expert Systems with Applications, 2015, 42(7): 3751-3759.
- [16] JIANG W, ZHANG L, DAI Y, et al. Analyzing Helpfulness of Online Reviews for User Requirements Elicitation[J]. Chinese Journal of Computers, 2013, 36(1): 119-131.
- [17] QIU J P. Information Metrology (5) Lecture 5: The Law of Frequency Distribution of Document Information Words-Zipf's Law [J]. Information Studies: Theory & Application, 2000(5): 77-81.
- [18] ZHANG Y H, LI Z W, ZHAO J C. How the Information Quality Affects the Online Review Usefulness? -An Empirical Analysis Based on Taobao Review Data[J]. Chinese Journal of Management, 2017, 14(1): 77-85.
- [19] WANG Z H, JIANG W. Online Reviews Sentiment Analysis Model Based on Rough Sets[J]. Computer Engineering, 2012, 38(16): 1-4.
- [20] YU M Z, NARISA Z. Feature extraction method based on mutual self-expanding mode[J]. Application Research of Computers, 2017, 34(4): 977-980.
- [21] XU Q, ZHANG X, YU S H, et al. Multi-feature-based classification method using random forest and superpixels for polarimetric SAR images[J]. Journal of Remote Sensing, 2019, 23(4): 685-694.



FENG Jin-zhan, born in 1981, postgraduate. His main research interests include business intelligence, management information and network complaint handling.



CAI Shu-qin, born in 1955, Ph.D. professor, Ph.D supervisor. Her main research interests include business intelligence and management information system.