情报学报 ISSN 1000-0135
第35卷第9期946-954，2016年9月

JOURNAL OF THE CHINA SOCIETY FOR SCIENTIFIC AND TECHNICAL INFORMATION ISSN 1000-0135
Vol. 35 No. 9, 946-954 September 2016

doi:10.3772/j.issn.1000-0135.2016.009.007

线上商品评论有效性分类专业领域知识模型的构建研究[1)]

夏火松[1] 甄化春[1] 张颖烨[2] 杨 培[3]

（1. 武汉纺织大学管理学院，武汉 430073；2. 复旦大学经济学院，上海 200433；
3. 西南财经大学管理学院，成都 611130）

摘要 线上商品评论有效性分类领域知识模型的构建是商品评论有效性分类的一个重要基础性工作，其直接影响分类器的精度与召回率。以往的研究大多集中于情感词典的构建以及领域术语抽取工作，对于一种专门针对线上商品有效性分类的领域知识库的构建研究较少。本文提出了一种基于信息增益技术进行文本有效性分类领域知识模型构建的半监督方法，同时构建了商品领域停用词表。通过对 Amazon、淘宝和京东商城 iPhone 系列手机评论数据利用 Python 语言进行有效性领域知识抽取和分类预测，实验结果发现该方法显著提高了评论有效性预测的精度。

关键词 信息增益 商品评论 有效性分类 领域知识模型

Research on the Domain Knowledge Construction for the Classification of Online Product Reviews Helpfulness

Xia Huosong[1], Zhen Huachun[1], Zhang Yingye[2] and Yang Pei[3]

(1. *School of Management, Wuhan Textile University, Wuhan* 430073;
2. *School of Economics, Fudan University, Shanghai* 200433;
3. *School of Management, Southwest University of Finance and Economics, Chengdu* 611130)

Abstract Domain knowledge construction is anfundamentalwork for the classification of online review effectiveness, which directly affect the precision and recall of the classifier. Previous studies mostly focus on the construction of emotional dictionary and feature extraction work and research on the domain knowledge based online products reviews is rarely encountered. Based on the IG Technology, this paper put forward a semi-supervised domain knowledge construct method and builds the product stop list. Utilizing the Python language to analysis experimental data of iphone6s from Amazon, JD and Taobao, the result shows the referred method significantly improves the effectiveness prediction accuracy.

Keywords information gain, product reviews, classification effectives, domain knowledge model

1 引 言

在线用户评论（Online Review）对于消费者和商家都具有重要意义，其在向潜在消费者传递用户体验信息的同时也向商家传递了消费者的需求特性以及产品的缺陷等关键信息。传统的产品在线说明使消费者对商品本身基本材质、性能和参数等有了一

收稿日期：2015年9月5日

作者简介：**夏火松**，男，1964年生，武汉纺织大学管理学院教授，博士，主要研究方向：知识管理、数据挖掘、物流信息管理和电子商务、DSS，E-mail：bxxhs@sina.com。**甄化春**，男，1989年生，武汉纺织大学管理学院硕士研究生，主要研究方向：数据挖掘、信息管理。**张颖烨**，女，1993年生，复旦大学经济学院硕士研究生，主要研究方向：金融数据分析。**杨培**，男，1990年生，西南财经大学管理学院博士研究生，主要研究方向：数据挖掘、信息管理。

1） 本文受国家自然科学基金项目“大数据情景的 outlier 分析与异类知识管理研究”（71571139）和“24小时知识工厂的知识共享活动模型与服务支持系统研究”（71171153）以及湖北省高等学校人文社会科学重点研究基地－企业决策支持研究中心项目（DSS20150215 &DSS20150108）的资助。

个大致的了解,但是这些信息弥补不了线下购物所特有的用户体验环节。相对于商家的产品说明,在线用户评论反映出更多用户注重的产品细节问题,并为后续潜在消费者提供了的间接的产品体验信息,其可信度比商家的产品说明更高,更有说服力[1,2]。对于电商卖家而言,其可以基于评论内容发现产品或服务缺陷并针对性的改进产品或服务,从而提高产品的用户接纳度和服务满足率。因此,分析消费者的购物评论,从中发现影响消费者满意度的内容要素,并针对性的改进产品和服务对于提升商家的竞争力和经济效益具有重大的商业价值[3]。

然而,并非所有的在线评论都有价值。网络的匿名性以及沟通的成本低廉等特性使得评论的质量参差不齐,垃圾、无效的线上评论的存在给消费者和商家带来的大量的不确定的信息严重削弱了评论内容的效用,降低了评论数据的价值密度,增加了消费者购物的时间和精力成本[4]。大数据时代,如何从海量用户评论中快速挑选出对潜在顾客购买起决定起辅助作用的有效评论信息,是一个亟待解决的问题,而该问题最终可以归结于评论有效性分类问题。

2 相关研究工作

2.1 在线评论有效性相关研究

自 Chatterjee[5] 首次提出在线评论有用性(Helpfulness of Online Reviews)概念以来,在线评论有用性问题便得到了企业界和学术界的关注和探索。根据 Ghose 以及 Mudambi 等的相关研究,一般将评论有用性理解为评论内容对于信息使用者是否有效以及有效程度,即某条评论是否包含用户所需的信息以及包含信息量的多少[6]。

从已有的研究看,当前评论有效性的研究主要围绕何为有效性评论以及如何快速进行有效性评论筛选两个问题。对于评论有效性界定的问题,已有研究成果大多集中于从实证角度定性研究评论有效性影响因素,其中郝媛媛等[7]通过实证验证了产品特征、评论极性以及评论长度等对于评论有效性具有显著正面的影响,殷国鹏[8]通过对已有的研究进行归纳总结发现众多的研究均证明了这一结论。在有效评论筛选方面,基于用户评分机制进行有效评论选择是 Amazon 等电商网站采用的主流方法,该方法参考用户对评论的累积评分统计进行评论有效性排名,其能够帮助用户精准、快速的选出有效评论,但其耗时长的问题被众多学者所诟病[9,10]。从文本有效性分类算法上改进分类器性能进而提高评论筛选效率和精度是当前的重点研究方向,但是同样存在分类精度与分类效率不能够同时达到最优的问题。吴含前等[11]提出了一种单一主题下基于逻辑回归的垃圾评论监测模型,并取得了较好的有效性预测精度但是需要大量的人工进行评论有效性标注。Zhang 等[9]提出了一种拓展的 GARC 算法对评论有效性进行分类,避免采用专家进行有效性标注的低效率和应用普适性问题,但是分类预测精度有待进一步的提高。使用评论特征词库结合进行分类器进行评论分类是评论情感分类和垃圾评论识别中另一种常用的方法[12],但是鲜有学者通过构建评论有效性领域知识进行评论有效性特征提取。因此,通过构建评论有效性领域知识库来进行评论有效性分类具有重要的研究价值。

2.2 领域词典的构建相关研究

领域知识库是根据研究的需要而构建的机器学习语料库,根据用途的不同,其可以分为领域特征词库、领域情感词库、领域停用词库以及领域分类词库等。对于前面三种领域知识的构建已有相当多的研究成果,而从评论有效性分类角度研究其评论有效性领域知识库构建相对较少[12,13]。评论有效性分类领域知识库的构建不同于通常意义上的产品特征领域词典或评论情感词典的构建[7],其综合考虑了评论语句中产品特征领域词和评论用户情感极性词语[8,14]。

目前比较常用的分类词典构建方法有基于统计的特征词语提取方法和基于语言学的方法以及混合式方法[15]。基于统计的特征词语提取方法通常包括基于词频统计的方法、基于逆文档频率的方法和互信息的方法等[16]。Jinadl 等[12]基于词频统计的思想通过对三种类型的垃圾评论进行词频统计,构建了领域垃圾词典,并采用 Logistic 回归模型对英文垃圾评论进行识别,取得了较好的效果。Popescu 等[13]通过抽取评论中频繁出现的名词和名词短语作为候选特征词,并通过 Web PMI 来评估候选词,利用贝叶斯分类提取产品特征,从一定程度上提高了特征词典准确度但是耗时较长。基于语言学的方法在特征词典的构建方面应用的也较为广泛,该方法可以有效解决商品评论中不同词语相同语意的问题,通过计算语意相似度计算来达到降低训练模型

维度,从而提高分类有效性的目的。该方法的缺点是模式覆盖面有限,存在领域与语言适应性问题,且其召回率受到限制[17]。考虑到基于统计方法和基于语言学方法进行特征提取各自的优缺点,在实际应用中一般将两种方式结合。Dailleli[18]利用语言学方法获取候选特征词集,然后通过互信息、LogLib统计方法获得术语。章成志[15]提出基于一体化策略的术语抽取方法,并通过实验证明了利用多层术语度进行特征词库抽取的有效性。

综上所述,前人对于特征领域知识库的构建已取得了较多的理论成果,但是从评论有效性分类视角研究分类领域词典的构建的问题相对较少,缺少一种文本有效性标识的领域特征词库。另外,基于统计的特征提取方法其特征提取精度还有待进一步提高,基于语言学方法的特征提取方法存在样本普适性方面的缺陷。

3 基于信息增益的评论有效性分类领域知识模型的构建方法

本文的工作主要集中于三个方面:第一,通过产品说明、通用领域词以及情感极性词构建种子领域词库并结合部分评论集建立初级有效性分类领域词库;第二,利用信息增益进行特征提取,通过控制过滤阈值来调节特征集合,并通过与初级领域词库对比,增加领域词典的特征数量,从而达到丰富领域词库的目的;第三,实验测试逐次构建的分类知识库对测试评论进行有效性标注并同基于有效性统计的评论数据进行对比,通过分析二者拟合度来测量模型效果。图1是本文提出的基于信息增益的线上评论有效性分类领域知识模型构建的基本框架。

3.1 初级领域词库的建立

初级领域词库是基于种子领域词典结合有效评论集合进行综合抽取的反映产品特征以及评论情感极性的一类词语,这类词语从一定程度上反映了评论语句的效用。建立领域词典的第一步是构建领域种子词库,其通常由领域专家给出,也有基于产品术语词典[19]、情感领域词典以及评论要素分析来进行有效性分类种子词语的提取[20]。本研究将以Amazon网站中iPhone5s、iPhone6、iPhone6 Plus商品的产品说明结合评论要素分析来进行产品有效性分类特征初步提取并结合台湾大学发布的极性情感词库进行iPhone评论中情感极性词语的提取。

李杰等[10]利用评价要素分析构建了电子商务服装产品的3层树状结构模型,该模型从产品和服务两个维度对线上商品特征词语进行分类。本文同样从产品和服务两个层次对手机评价要素进行划分,将iPhone系列产品说明作为产品评价要素归纳为质量、外观、大小、价格、参数四个维度,从商品描述、服务态度、配送速度、退换货处理来分析商家服务要素特征(图2)。结合上述要素特征和极性词库,我们从Amazon iPhone系列手机(iPhone5s、iPhone6、iPhone6 Plus)共1678条评论中抽取了199条已被标记为有效的评论作为特征提取样本,最终提取了54个特征词语和15个情感极性词语,建立了容量为69的初始有效性分类特征领域词库(表1)。该词库特征提取精度较高,但是不能完全替代总体有效评论集合特征。

表1 初级评论有效性分类领域词库

特征词典	情感词典
屏幕	坑
外观	可靠
大小	信赖
参数	好用
充电器	信任
配置	失望
容量	泪奔
重量	失望
行货	Perfect
港货	碉堡
包装	流氓
物流	快
……	……
网络	爽

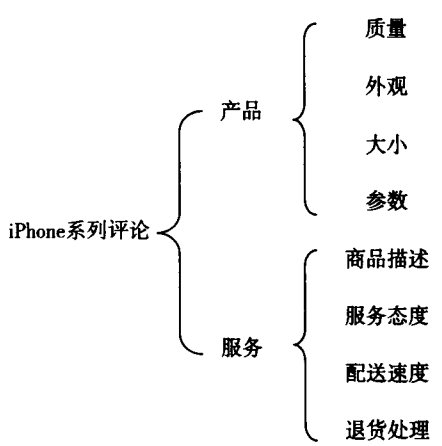

图2 iPhone系列手机的树状特征结构模型

3.2 基于信息增益的特征提取

信息增益是一种基于统计思想的特征提取方法，其通过一定的函数自动计算特征词语对于评论的有效性分类的意义，该种意义通过量化的形式表现出来，通过设定有效性归类阈值（α）来对领域特征词语进行筛选，阈值高低将直接影响特征词语个数，最终决定待处理向量空间模型的维度以及特征词语与初级领域词典之间的匹配度。

信息增益的基本原理是通过计算整个文本在包含与不包含某一特征时信息量的差值，差值越大，代表这个特征对于文本集合越重要。在计算信息增益之前需要计算“熵”，然后计算“条件熵”。对于 N 类问题，“熵”的计算公式如式（1）所示，特征 t 的“条件熵”如公式（2）所示。

$$H(C) = -\sum_{i=1}^{n} P(C_i) \cdot \log_2 P(C_i) \quad (1)$$

$$H(C \mid T) = P(t)H(C \mid t) + P(\bar{t})H(C \mid \bar{t}) \quad (2)$$

其中，$P(C_i)$ 表示 C_i 出现的概率，$P(t)$ 和 $P(\bar{t})$ 分别表示特征 t 在总文本中出现的概率与不出现的概率，而 $H(C \mid t)$ 和 $H(C \mid \bar{t})$ 分别表示在特征 t 出现以及不出现的情况下文本的熵。其计算方法如式（3）、式（4）所示。

$$H(H \mid t) = -\sum_{i=1}^{n} P(C_i \mid t) \log_2 P(C_i \mid t) \quad (3)$$

$$H(H \mid \bar{t}) = -\sum_{i=1}^{n} P(C_i \mid \bar{t}) \log_2 P(C_i \mid \bar{t}) \quad (4)$$

其中，$P(C_i \mid t)$ 和 $P(C_i \mid \bar{t})$ 分别表示在特征 t 存在的条件下类别 C_i 出现的概率与特征 t 不存在的条件下类别 C_i 出现的概率。有上述公式得到特征 t 的信息增益公式如式（5）所示。

$$IG(t) = H(C) - H(C \mid T) \quad (5)$$

对每个特征都可以用这个方法计算出其信息增益量，对于信息增益量小于“阈值”的特征项去掉该特征，可以根据不同的情况设定不同的阈值。

3.3 评论有效性分类领域模型的构建方法

信息增益不需要建立领域词典，特征提取速度快，能够从大样本数据集合中自动提取本特征集合，但是该方法特征提取精度不高，受停用词典质量影响较大。本研究使用信息增益提取有效性分类领域词，然后同初级领域词典对比将具有评论有效性识别特征但未被初级领域词典覆盖的领域词汇添加到初始领域特征词库，同时将每次将信息增益选取的特征词中不能体现评论有效性分类特征但是信息增益大于 α 的词语加入到停用词库，利用每次新建的领域词库重新对 iPhone 系列手机在线评论进行有效性分类自动标注并将测试集预测结果同基于统计的有效性标签进行对比，如此反复，比较每次新建分类器的分类精度，选择分类效果最好的分类有效性领域词库。α 的取值由实验确定，具体有效性分类特征领域知识词语提取流程如图 3 所示。

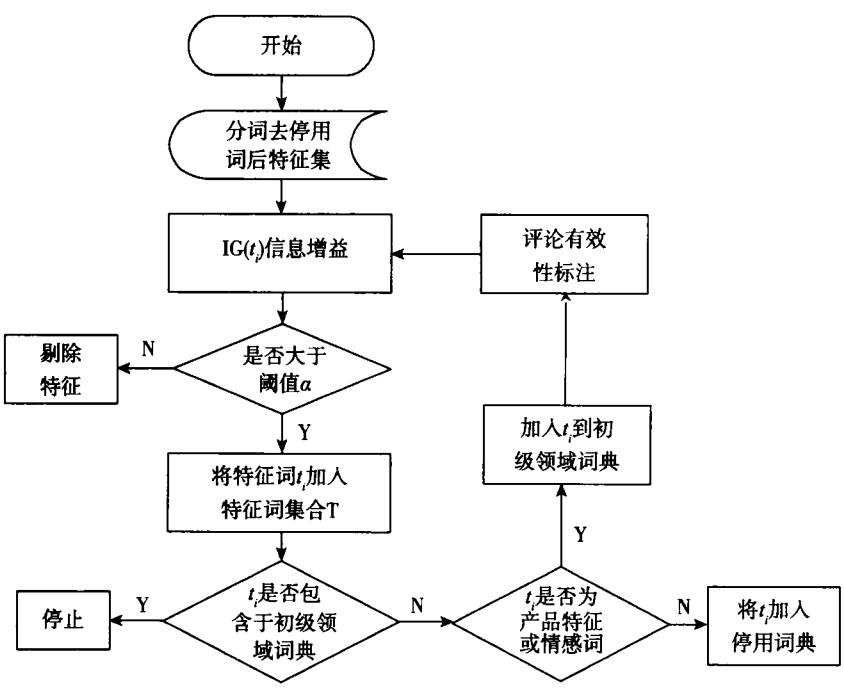

图 3 评论有效分类特征提取流程图

4 实验设计及结果分析

4.1 实验设计

特征领域词语的覆盖度以及对有效评论的特征表示都直接影响模型预测分类的精度,本实验通过评论有效性分类精度来衡量领域知识库的性能。评论有效性识别是文本分类问题的一种,其通常包括文本数据的获取以及预处理、样本类别标注、文本特征模型的表示、分类器的训练及结果的预测四项工作。

(1)数据的获取及预处理

本研究的实验数据是通过网络爬虫软件GooSeeker[20]在Amazon、京东和淘宝网站上分别获取iPhone5s、iPhone6、iPhone6s的用户产品评论集共1723条、1953条和2486条评论,剔除空白、重复评论后分别得到1678条、1288条和2391条用户评论数据。表2是本研究的实验数据统计表。文本的预处理包括分词、去停用词的工作,经预处理将产品评论以词语集合的形式表现出来,同时去掉没有实际含义的功能性词汇(如"由此可见","总而言之"等)以及标点符号和使用频率非常频繁的单汉字。在文本预处理的过程中,我们采用Python编程语言结合"哑巴分词"作为分词工具,使用"四川大学机器智能实验室通用词库"来进行停用词的去除工作。

表2 实验数据统计表

iPhone系列评论网站	原始评论数量	剔除重复后数量	训练集	测试集
亚马逊(Amazon)	1723	1678	839	839
淘宝(Taobao)	2486	2391	1195	1196
京东(JingDong)	1953	1288	644	644

(2)样本类别标注

评论有效性研究是一个二分类问题,在分类器训练以及信息增益特征选择以及分类结果评估中都要用到文本分类标签。通常对于有效的评论标记为1,无效的评论则标记为-1。评论数据的有效性标注存在基于评论有效性特征测量的标注和人工主观性标注两部分,其中第一部分是用来训练分类器并对测试样本进行有效性分类预测,第二部分标注是用来对比分析和评估第一部分测试样本标签的分类标注结果,同时第二部分的标注还可以用来检验评估实验中分类器的分类效果。其中,基于评论有效性特征测量的样本类别标注主要是通过上述构建的领域知识模型进行评论样本类别标注。而人工主观性标注则是在人工反复阅读理解的基础上进行的,而不同消费者对于文本有效性的判定具有趋同性的同时也具有个体差异性,文本有效性评定受到参与人数的影响较大而且需要耗费大量的时间。因此,实验仅以人工标注的评论有效性结果作为基于领域进行评论有效性预测结果的参考。试验中选取了Amazon网站消费者评论有效性投票统计结果作为评论有效性标记依据,而对于淘宝和京东的评论效用的标注则是选取了10位在网上购买过iPhone手机的消费者来对淘宝和京东的在线评论进行有效性标记,并以标记之和的正负分别作为有效和无效标注的结果。对于Amazon网站有效性投票统计标注方法的具体操作如下:

Amazon网站消费评论中提供了针对用户的调查问项——"这条评论对您有用吗?"以及调查统计数据,如"354人中有323人认为以下评论非常有用"。通过对Amazon iPhone系列手机评论词条以及其有用性统计情况进行分析发现,该网站用户评论评价阅读人数较少,在阅读人数大于9人且有用性比例为0.8时可以取得的绝对有效评论文本仅有36条,从该类评论中提取的词条有效性区分度较高,但对于样本总体的代表性不强。我们通过对阅读人数、评价为有用的比例以及过滤的条数研究发现在单条评论阅读人数大于或等于4人且认为有用人数比例大于或等于0.7时,该条评论绝对有效(图4)。同时,在评论阅读人数大于或等于4人且认为有效人数比例小于或等于0.2时,该条评论绝对无效[9]。

对此,结合在线调查数据与文献[7~8]提出的评论有效性影响因素,文章提出了一种文本自动标注方法,具体如下:

1)对于某条评论若阅读过该评论的人数大于或等于4人且认为该条评论有用的人数占到总人数的比例$\alpha \geq 70\%$时,认为该评论有效。

2)对于某条评论若阅读过该评论的人数大于或等于4人且认为该条评论有用的人数占到总人数的比例$20\% < \alpha < 70\%$时,若该条评论中含有领域词则认为该条评论有效,否则无效。

3)对于某条评论若阅读过该评论的人数大于或等于4人且认为该条评论有用的人数占到总人数

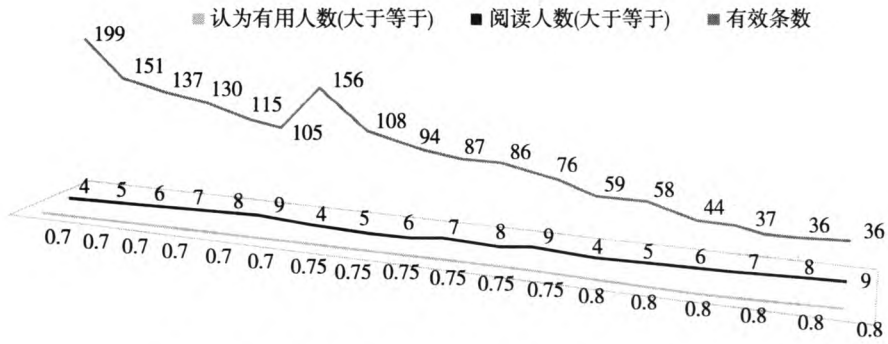

图4 有效评论与评论阅读人数以及有用性比例变化情况

的比例 $\alpha \leqslant 20\%$ 时,认为该条评论无效。

4)对于某条评论若阅读过该评论的人数小于4人,如果其含有领域词则认为该评论有效,否则认为该条评论无效。

(3)文本特征模型的表示

在现有的研究中,文本特征通常以向量空间模型(VSM)的形式表示出来。在线用户每一评论可以映射为一个特征向量 $V(d) = (t1, w1(d); t2, w2(d); \ldots ; tn, wn(d))$,其中 $ti(i = 1,2,\cdots,n)$ 表示在信息增益 α 阈值下该评论中剩余互不雷同的词条项,$wi(d)$ 为 ti 在 d 中的权值,一般定义为 ti 在 d 中出现评论 $tfi(d)$ 的函数,即 $wi(d) = W(tfi(d))$。在信息检索中,常用的词条权重计算方法有布尔函数、平方根函数、对数函数、TF 算法以及逆文档频率算法(TF-IDF)等,这里我们选用 TF-IDF 作为特征词语权重计算方式。TF-IDF 由 Salton 于 1973 年首次提出,其主要思想为:一个词语在特定文档中出现的频率越高,说明它在区分该文档内容属性方面能力越强,即 TF;一个词语在文档集合中出现的范围越广,说明其区分文档内容的属性越低,即 IDF[19]。经典的 TF-IDF 具体表现形式如公式(6)所示:

$$W_{ij} = tf_{ij} \times idf_j = tf_{ij} \times \log(N/n_j) \tag{6}$$

其中,tf_{ij} 指特征项 t_j 在文档 d_i 中出现的次数;idf 指出现特征项 t_j 的文档倒数。N 表示文档数目 n_j 表示出现特征项 t_j 的文档数目。

(4)分类器的训练及预测

文本分类常用的分类器有支持向量机(SVM)、贝叶斯分类(Naïve Bayes)、最大熵以及 n 元语言模型等,刘志明等[21]通过实验对比证明采用 TF-IDF 权重计算方法结合信息增益进行特征提取并通过 SVM 进行分类可以得到较好的分类效果。本实验拟采用台湾大学林智仁教授等开发的 LibSVM 软件包[22]在 Matlab R2009b 平台下进行模型的训练和分类预测。为防止模型欠拟合或过度拟合,试验中均使用 50% 训练样本和 50% 的测试样本。

4.2 结果分析

对于分类结果的评测,采用信息检索领域普遍使用的精度(Precision)、召回率(Recall)、准确率(Accuracy)和 $F1$ 值[23],具体如下:

$$Precision = \frac{a}{a+b} \tag{7}$$

$$Recall = \frac{a}{a+c} \tag{8}$$

$$Accuracy = \frac{a+d}{a+b+c+d} \tag{9}$$

$$F_{\beta=1} = \frac{(1+\beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall} = \frac{2a}{2a+b+c} \tag{10}$$

其中,a、b、c、d 分别对应的是有效评论且被识别为有效评论的个数,是无效评论但被识别为有效的评论数,是有效评论但被识别为无效评论的评论个数,是无效评论且被识别为无效的个数。

(1)基于 IG 领域知识的评论效用分类结果分析

通过上述实验过程,我们得到表3中的 Amazon iPhone 系列评论集合在初级领域词典和信息增益改进后的领域词典下评论有效性的预测结果。从表中的数据可以看出,在信息增益阈值 α 为 0.0055 至 0.0035 之间时,分类器对于样本标签的分类预测准确率得到显著提升,其中对于有效性样本的识别精度 p 总体呈上升趋势,分类精度及 $F1$ 值随着特征数量的增加而得到显著提升;在相同的阈值 α 下,基于信息增益改进的领域词典标注的评论文本有效性预测准确率比初级领域词典标记的文本有效性分类有效性预测的精度要高,其说明在相同阈值下通过信息增益改进的领域词典可以显著改变文本有效性预测精度。

表3 基于领域知识库的评论有效性分类预测结果

	阈值 α	精度(P)	召回率(R)	$F1$ 值	准确率(A)	领域词个数
初级领域词	0.0055	0.913	0.642	0.714	0.766	69
	0.0050	0.888	0.673	0.766	0.768	75
	0.0045	0.904	0.704	0.791	0.786	83
	0.0040	0.916	0.730	0.813	0.805	89
	0.0035	0.916	0.744	0.821	0.801	94
基于信息增益领域词典	0.0055	0.921	0.656	0.766	0.774	75
	0.0050	0.901	0.702	0.789	0.784	83
	0.0045	0.912	0.714	0.801	0.795	89
	0.0040	0.926	0.742	0.824	0.805	94
	0.0035	0.930	0.765	0.839	0.814	98

表4反映的是在初级领域词库和信息增益改进领域词库标注下基于有效性分类词典自动标记预测标签分别同基于Amazon官网统计的潜在消费者对于iPhone手机评论有效性自动标注的拟合度。实验数据表明,在阈值 α 为0.0055~0.0035时,随着 α 的减小初级分类标注词典和改进的分类标注词典对评论有效性预测同基于统计有效性标注的拟合度逐渐增加,同时 $F1$ 值显著提高;在相同的阈值 α 下基于信息增益的领域词典比初级领域词典有效性标记预测的拟合度相对较高。

表4 预测指标同基于Amzon官网统计标签匹配情况

	阈值 α	精度(P)	召回率(R)	$F1$ 值	拟合度
初级领域词	0.0055	0.909	0.624	0.744	0.749
	0.0050	0883	0.652	0.749	0.748
	0.0045	0.898	0.683	0.776	0.766
	0.0040	0.909	0.708	0.796	0.785
	0.0035	0.909	0.725	0.806	0.782
基于信息增益领域词典	0.0055	0.915	0.636	0.750	0.755
	0.0050	0.896	0.681	0.774	0.764
	0.0045	0.905	0.692	0.784	0.783
	0.0040	0.919	0.723	0.809	0.787
	0.0035	0.924	0.747	0.826	0.797

表3、表4的数据共同说明通过信息增益改进的领域词典对于分类器预测准确率以及有效性预测值同实际结果的拟合度均有显著的促进作用。

(2)基于不同网购平台的评论效用分类结果分析

对于淘宝和京东(JD)评论效用的评估,研究根据基于评论有效性特征测量的样本类别标签利用LibSVM分类器对测试数据进行预测并同人工标注的测试数据标签对比,从而评估分类器性能同时观察不同购物平台评论效用差异性及其影响因素。表5和表6分别是淘宝实验数据和京东实验数据的训练和测试情况。

表5 淘宝评论数据效用分析结果

阈值	维度	无效评论	精度(P)	拟合度
0.0015	208	276	87.63%	87.77%
0.0020	136	276	87.63%	87.77%
0.0025	112	276	87.62%	87.70%
0.0026	105	276	87.61%	87.70%
0.0027	91	276	87.59%	87.68%
0.0028	84	276	87.58%	87.65%
0.0029	82	276	87.54%	87.55%
0.0030	76	276	87.54%	87.55%

注:训练集1195,测试集合1196,评论长度范围为1~269字,平均长度为41.86字

表5中数据表明,当 α 处于0.0015~0.0030时,随着阈值的增加信息有效性分类领域词语从208个减少到76个,即向量空间模型维度降低。由于淘宝测试样本数据集合较大,少量高频特征词语对于仅含未登录特征词语的评论并不能有效的进行标注,故在一定范围内增加领域词语个数分类器的

分类效果和分类标签的匹配度将逐渐得到提高。表6的实验结果同样说明了上述问题，并且该评论集随着特征领域词语数量的增加，分类器的预测精度变化和预测标签拟合度提高速度相对更快。

表6 京东评论数据效用分析结果

阈值	维度	无效评论	精度(P)	拟合度
0.0020	235	455	83.93%	84.63%
0.0025	167	455	83.23%	83.23%
0.0030	141	455	83.46%	84.32%
0.0035	120	455	82.22%	84.20%
0.0040	91	455	82.14%	83.22%
0.0050	76	455	80.67%	83.85%

注：训练集644，测试集合644，评论长度范围为0~256字，平均长度为20.88字

综合对比亚马逊、淘宝和京东评论数据实验结果，我们发现以下问题：第一，亚马逊iPhone6s评论数据有效性分类精度要高于淘宝和京东评论数据。从预测标签拟合度来说，淘宝实验数据标签的拟合度要优于京东和亚马逊。导致上述结果的原因在于淘宝和京东评论初级领域特征词库的选择是以亚马逊最终领域词典为依据的，故亚马逊在线评论机器分类效果要优于上述二者。淘宝机器分类预测效果优于京东的原因在于淘宝评论量虽大，但是其评论特征聚合度大，而京东评论特征词语较为离散导致领域词典中未登录词语量较大导致很多仅含有未登录特征词语的评论被标记为无效评论。对比表5和表6可以看到淘宝总评论数为2391条，无效评论比例为11.54%，而京东总评论数1288条，无效评论比例则达到35.33%。对于京东和淘宝评论数据预测标签的拟合度高于亚马逊实验数据预测标签的拟合度的问题部分原因在于三者采用的评论人工标注方式不同。第二，评论长度对评论分类预测精度有重要的影响。单条评论越长其涵盖的领域特征词语的概率要大于较短的评论数据，总体评论集合长度越长则特征词语聚合度越高，分类预测精度也就越好。对比实验数据可以看到，亚马逊的评论长度范围在8~622字符，评论长度为57个字符，远大于淘宝的41.86个字符和京东的20.88个字符，故三者的分类预测精度呈下降趋势。第三，对于同一商品，不同评论网站的特征词语存在一定的差异性。将亚马逊最优的评论有效性分类特征词语应用于京东和淘宝评论数据的实验结果要低于亚马逊实验效果。

以上实验数据的分析结果表明了基于本研究提出的线上商品评论有效性分类领域知识模型在提高评论分类准确率和分类器精度上的显著促进作用。同时，基于研究爬取的亚马逊、淘宝和京东购物网站iPhone系列手机评论数据分析结果可以将三者平台评论可信度和有效性排序为亚马逊>淘宝>京东。

5 结 论

本文设计、发展和评价了一种基于信息增益的评论有效性分类领域知识库构建方法。具体地，通过从实验样本中抽取少量用户评价为绝对有效的评论并结合iPhone系列手机树状结构模型以及情感领域词库构建了评论有效性分类初级领域词库；利用初级领域词库进行评论有效性标记并结合信息增益技术通过设置信息阈值α来调节领域词典个数，并逐步丰富初级领域词库；第三，运用Amazoniphone系列手机评论数据来对本研究提出的基于信息增益建立的有效性分类领域词典进行验证，从分类准确率(A)、拟合度、精度(P)、召回率(R)和$F1$值几个指标同初始分类有效性词典进行对比，证明了基于信息增益构建的有效性分类词典比前者具有显著的优势；最后，通过对比亚马逊、京东和淘宝线上iPhone系列手机评论有效性分类实验结果印证了文章构建的评论有效性特征模型的效用，同时也验证了评论长度对评论效用的影响，即较长的评论其有效性特征词语密度高，分类预测精度越好。

本研究的理论贡献在于提出了一种基于半监督的文本有效性分类领域词库的构建方法，该方法从一定程度上解决了基于监督的高分类精度、低效率和基于统计的高效率低精度问题以及样本整体代表性不足的问题。从管理实践的角度看，利用该有效性分类词典的构建方法，电商企业可以对阅读人数较少的评论以及最新的评论快速进行有效性分类排序，从而为消费者提供更具参考价值的潜在商品信息及用户体验信息。同时，利用本研究中三个平台的iPhone系列手机评论数据的对比分析结果，消费者可以根据自己购物意向择优选择购物网站平台，而上述三个电商平台及其入驻的卖家则可参照分析结果改变竞争策略。

本研究也存在一些局限与不足，这也是后续将要继续研究的内容。第一，研究样本相对有限，后续研究可以使用大样本数据文章提出的有效性分类领

域知识模型进行验证。第二,后续研究可以结合评论长度、评论者特点来研究评论有效性,从而进一步提高评论有效性预测精度。

参 考 文 献

[1] Park D H, Lee J. overload and its effect on consumer behavioral intention depending on consumer involvement [J]. Electronic Commerce Research & Applications, 2008, 7(4):386-398.

[2] Bickart B, Schindler R M. Internet forums as influential sources of consumer information [J]. Journal of Interactive Marketing, 2001, 15(3):31-40.

[3] Wang Hongwei, Yin Pei, Zheng Lijuan, et al. Sentiment classification of online reviews: using sentence-based language model [J]. Journal of Experimental & Theoretical Artificial Intelligence, 2014, 26(1): 13-31.

[4] Chen H, Chiang R H L, Storey V C. Business Intelligence and Analytics: From Big Data to Big Impact[J]. MIS Quarterly, 2012, 36(4):1165-1188.

[5] Patrali Chatterjee. Online Reviews: Do Consumers Use Them? [J]. Advances in Consumer Research, 2001, 28(1):129-133.

[6] Ghose A, Ipeirotis P G. Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics [J]. IEEE Transactions on Knowledge & Data Engineering, 2011, 23(10): 1498-1512.

[7] 郝媛媛, 叶强, 李一军. 基于影评数据的在线评论有用性影响因素研究[J]. 管理科学学报, 2010, 13(8): 78-88.

[8] 殷国鹏. 消费者认为怎样的在线评论更有用? ——社会性因素的影响效应[J]. 管理世界, 2012, (12): 115-124.

[9] Zhang Zunqiang, Ma Yue, Chen Guoqing, et al. Extending associative classifier to detect helpful online reviews with uncertain classes[C]// IFSA-EUSFLAT, Spain, 2015: 1134-1139.

[10] 李杰, 张向前, 陈维军, 等. C2C电子商务服装产品客户评论要素及其对满意度的影响[J]. 管理学报, 2014, 11(2): 261-266.

[11] 吴含前, 朱云杰, 谢珏. 基于逻辑回归的中文在线评论有效性监测模型[J]. 东南大学学报(自然科学版), 2015, 45(3): 433-437.

[12] Jindal N, Liu B. Opinion spam and analysis[C]// Proceedings of the first ACM international conference on Web search and data mining, 2008: 219-229.

[13] Popescu A M, Etzioni O. Extracting product features and opinions from review[C]// Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing, Stroudsburg, USA: Association for Computational Linguistics, 2005: 339-346.

[14] Ngo-Ye T L, Sinha A P. The influence of reviewer engagement characteristics on online review helpfulness: A text regression model[J]. Decision Support Systems, 2014, 61: 47-58.

[15] 章成志. 基于多层术语度的一体化术语抽取研究[J]. 情报学报, 2011, 28(3): 275-285.

[16] 李丽双, 党延忠, 张婧, 等. 基于条件随机场的汽车领域术语抽取[J]. 大连理工大学学报, 2013, 53(2): 267-272.

[17] Kit C, Liu X Y. Measuring mono-word termhood by rank difference via corpus comparison[J]. Terminology, 2008, 14(2): 204-229.

[18] Daille B. Study and Implementation of Combined Techniques for Automatic Extraction of Terminology// Klavans J L, Resnik P. The Balancing Act: Combining Symbolic and Statistical Approaches to Language[M]. Cambridge, MA: MIT Press, 1996: 49-66.

[19] 何燕, 惠志方, 段慧明, 等. 基于专业术语词典的自动领域本体构造[J]. 情报学报, 2007, 26(1): 65-70.

[20] 集搜客 GooSeeker 网页抓取软件[EB/OL]. [2015-05-19]. http://www.gooseeker.com

[21] 刘志明, 刘鲁. 基于机器学习的中文微博情感分类实证研究[J]. 计算机工程与应用, 2012, 48(1): 1-4.

[22] Chih-ChungChang, Chih-JenLin. LIBSVM: a library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 1-27.

[23] 史伟, 王洪伟, 何绍义. 基于微博的产品评论挖掘: 情感分析的方法[J]. 情报学报, 2014, 33(12): 1311-1321.

(责任编辑 车 尧)