

在线商品评论可信性评价方法

李超*, 向静, 向军

(湖北民族学院 信息工程学院, 湖北 恩施 445000)

(* 通信作者电子邮箱 springxun@163.com)

摘要:针对现有商品评论存在数量大、质量参差不齐、可信度差,导致用户难以快速获取有效信息并制定高效的决策,而现有评论可信性评估主要考虑评论来源和投票形式的支持度等问题,提出了一种从评论内在质量角度实现评论可信度评估方法,即通过融合评论者等级、评论支持度和评论观点一致性等实现评论可信性评估。首先基于规则库和方法库完成了评论数据的预处理;然后基于商品特征库、通用词典、情感词典以及方法库,完成了商品特征识别和特征值提取及标准化;最后基于建立的模型完成评论可信度评估。实验结果验证了该方法的可行性,该方法可以应用于其他电商平台实现商品评论可信性自动评估。

关键词:商品评论;评论质量;可信性;一致性;质量评估

中图分类号: TP391.1; TP18 **文献标志码:** A

Assessment method of credibility on online product reviews

LI Chao*, XIANG Jing, XIANG Jun

(Information Engineering College, Hubei University for Nationalities, Enshi Hubei 445000, China)

Abstract: Since there are many troubles such as large quantity, uneven quality and poor credibility in getting helpful information and making effective decision for stakeholders, and the existing research work on credibility assessment mainly considers the sources of reviews and the support of reviews in form of votes, an assessment method on review credibility from perspective of intrinsic quality was proposed. That is, the credibility assessment of reviews was realized by integrating the ratings of reviewers, the support degree of reviews and the consistency in reviews, etc. Firstly, the pre-processing of review data was completed based on rule and method libraries. Then, the feature recognition and the feature value extraction and standardization were completed based on product feature library, generic dictionary, sentiment dictionary and method library. Finally, the credibility assessment of reviews was completed based on the established models. The experimental results verify the feasibility of this method, and it can be applied to assess the credibility of product reviews automatically on other e-commerce platforms.

Key words: product review; review quality; credibility; consistency; quality assessment

0 引言

电商平台如 Amazon.com、Taobao.com 和 JD.com 等为消费者提供了分享其所购买商品质量及服务体验的平台,消费者可以浏览他人发表的评论并对评论进行有用性投票,评论管理系统可以根据评论发表时间、有用性和评论者级别等进行排序与展示^[1-3]。产品制造商、销售商、消费者和其他利益相关人员通过遍历评论获取相关信息,进而实现商品质量评估与预测、客户关系管理、新产品研发等^[4-6]。在电子商务系统中,评论的价值严重依赖于评论质量^[7-8]。

商品评论主要涉及消费者及专家发表的关于商品质量及服务体验的描述,目前多数电商平台都为消费者提供了评价评论“有用性”途径,然而因为误操作、广告宣传、导向性和欺诈行为等导致许多不可靠、不相关评论的存在^[9-10];同时,因为评论者随意性和主观性,评论观点存在不一致或相互矛盾。相关人员要获取有价值信息,往往遍历很多评论,从而判断哪

些评论更加合理和可靠,并进行综合分析决策。在电商平台中,评论存在数量大、质量参差不齐、可信度差、不及时等问题^[11-13],严重阻碍了用户进行有效决策。为了帮助用户高效地获取信息,“有用性”投票和评论者等级成为了评价评论有用性和可靠性的主要指标,然而,“有用性”投票往往只能反映两个极端情况,高“支持率”和高等级评论者发表的评论对商品质量及消费体验描述内容侧重点不同,同时可能存在不符合事实的情况甚至商业欺诈。

评论质量管理是评论排序、观点与情感识别、评论融合、摘要提取、垃圾评论检测、产品质量评估与预测等工作的基础。可信性是评论质量的重要方面,评论的可信性与评论来源(如发布者类别与级别、发布平台等)和评价对象相关,一般而言:1)具有较高等级的评论者发表的评论,可信度往往比较高,但部分评论也可能存在导向性;2)具有更多支持度的评论,评论观点往往与多数用户一致,可信度往往也比较高,但部分评论也可能被操纵而获得较多支持;3)各评论者

收稿日期:2018-06-08;修回日期:2018-09-05;录用日期:2018-09-20。 基金项目:国家自然科学基金资助项目(61040006);湖北省自然科学基金资助项目(2018CFB451);湖北省教育厅项目(B2014245);湖北民族学院博士基金资助项目(MY2017B019)。

作者简介:李超(1979—),男,湖北恩施人,副教授,博士,主要研究方向:数据挖掘、数据质量管理; 向静(1978—),女,湖北恩施人,讲师,硕士,主要研究方向:知识工程; 向军(1978—),男,湖北恩施人,副教授,博士,主要研究方向:移动数据质量管理。

对于相同商品的质量及服务消费体验存在不一致的观点,如果观点存在严重的不一致问题,那么某些评论的可信度较低,但存在不一致性是合理的。

1 相关工作

评论质量被定义为评论满足用户使用评论要求的程度^[14],Zhang 等^[15]认为评论质量应该从相关性、真实性、可靠性、主观性、可理解性和充分性等方面进行度量。评论的数量、可读性和文本内容的正确性对产品销售量有重要的影响^[16],评论信息不一致可以体现新产品独特性反而会增强消费者购买意愿^[17],但评论不一致在不同产品领域影响产品销售方面有不同结论;Lee 等^[18]认为评论质量主要在于可用性,并认为评论有用性主要依赖于评论者的专业性、写作风格和评论及时性等因素,并提出了相应的评价方法;Singh 等^[19]分析了评论语言表达及语义特征等因素对评论有用性的影响;Back 等^[20]分析了评论者等级、评论者名称、评论字符数、否定词比例等因素对评论质量的影响;针对发表越早的评论往往会获得更多关注及投票,而最新发表的评论获得的投票往往较少^[21]等问题,Miao 等^[22]引入了评论新鲜度和相关性实现评论的排序进而满足消费者信息需求。

评论的可信性是评论质量的重要方面,可信度是从定量的角度对可信性进行度量。Racherla 等^[23]从评论者和评论特征信息角度研究了评论者对评论可信度的感知体验。目前,对于 Web 数据可信性评估主要通过 Web 数据来源以及来源相关对象的可信性方面直接或间接地完成评估,但来自相同数据源的不同类别对象其数据值准确和数据可信性方面也存在差异^[24]。目前对于 Web 数据可信性的研究很多^[25-27],评论数据属于一类特殊的 Web 数据,现有用于 Web 数据可信性评估的方法对于评论可信性评估具有很好的借鉴意义。针对评论可信性,现有研究主要考虑评论来源和投票形式的支持度,没有从评论内在质量角度进行评估,因此,本文提出了一种融合评论者等级、评论支持度和评论观点一致性的评论可信性评估方法,从评论内在质量角度实现评论可信度的评估,实验结果验证了该方法的可行性。该方法可以应用于其他电商平台实现商品评论可信性自动评估。

2 在线商品评论的可信性评价模型

2.1 评论可信性评价体系结构

评论可信度评估,首先基于相关规则和方法实现评论数据的预处理,然后基于词典和商品特征库识别商品特征及取值,最后基于评估模型及方法库实现评论的可信性评估,体系结构如图 1 所示。

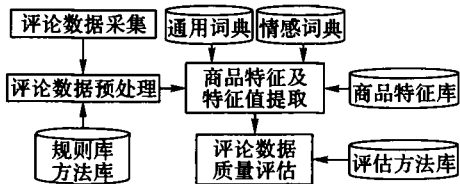


图 1 评论可信性评价体系结构

Fig. 1 Assessment architecture of review credibility

2.2 评论观点一致性评价

相关人员要获取有价值的信息,往往需要遍历很多评论

从而判断评论合理和可靠性,进而进行综合分析以便于决策。商品评论主要涉及消费者及专家发表的关于商品质量及服务体验的描述,因评论者强烈的主观性和随意性以及商业宣传和欺诈等而存在评论观点不一致甚至相互矛盾等现象,因此对评论一致性进行分析和评估变得十分必要。评论一致性评价(目标)对象可以是任两个评论间的一致性、任一评论集(如某个评论者或某评论者团体发表的相关评论集、某时段发表的相关评论集、某商品关联评论集等)特征项取值的一致性,具体分为以下几种情况。

1) 两条评论间观点相似性。

不失一般性,假设任意两条评论 r_1 和 r_2 ,其蕴含的信息可以分别用信息特征项集合 I^1 和 I^2 表示, M_1 和 M_2 分别表示论 r_1 和 r_2 蕴含的特征项个数, $I^1 = i_1^1 \cup i_2^1 \cup \dots \cup i_{m_1}^1 \cup \dots \cup i_{m_1}^1$, $I^2 = i_1^2 \cup i_2^2 \cup \dots \cup i_{m_2}^2 \cup \dots \cup i_{m_2}^2$;假定 $I^{1 \cap 2} = I^1 \cap I^2 = \langle i_1^0, i_2^0, \dots, i_{|I^{1 \cap 2}|}^0 \rangle$,这里 $|I^1|$ 和 $|I^2|$ 分别表示 I^1 和 I^2 的势, $|I^{1 \cap 2}|$ 表示集合 $I^{1 \cap 2}$ 的势; i_k^1 和 i_k^2 分别表示 r_1 和 r_2 的第 k 个信息特征,函数 $Val(i)$ 用来返回信息项 i 的值,特征取值分为数字型和字符型,在进行一致性评估时需要进行归一化处理变换为标准形式。 $Cons(Val(i_k^1), Val(i_k^2))$ 用来表示评论 r_1 和 r_2 在第 k 个信息项上取值的一致性,则评论 r_1 和 r_2 观点一致性表示为 $Con(r_1, r_2)$:

$$Con(r_1, r_2) = \frac{\sum_{i=1}^{|I^{1 \cap 2}|} Cons(Val(i_i^1), Val(i_i^2))}{|I^{1 \cap 2}|} \quad (1)$$

评论 r_1 和 r_2 的一致性取决于评论在各个特征项上取值的相似性。基于 $Con(r_1, r_2)$ 可以用来判定或与给定评论相似的评论集,发现与给定评论关联的评论者持相似观点的评论者集。

2) 任一特征项上观点一致性。

假定 R^p 表示任一商品 p 关联的评论集, $R^p = \{r_1, r_2, \dots, r_K\}$,这里 K 商品 p 相关联的评论数量, R^p 涉及的特征项表示为集合 $I^p = \{i_1^p \cup i_2^p \cup \dots \cup i_n^p \cup \dots \cup i_N^p\}$, $Val(i_{n[j]}^p)$ 表示特征项 i_n^p 在评论集中第 j 条评论相应特征上的取值, $1 \leq j \leq N1$,即评论集中包含特征项 i_n^p 的评论数量为 $N1$ 。当特征值为非数值类型时,则基于规则把特征值统一转换为数值类型。定义商品 p 相关评论集在任一特征项 i_n^p 上观点的一致性表示为取值的一致性 $Con(i_n^p)$:

$$Con(i_n^p) = \delta(i_n^p) = \sqrt{\frac{\sum_{j=1}^{N1} (Val(i_{n[j]}^p) - \overline{Val(i_n^p)})^2}{N1 - 1}} \quad (2)$$

其中: $\overline{Val(i_n^p)}$ 表示特征项 i_n^p 在评论集的均值, $Val(i_{n[j]}^p)$ 表示特征项 i_n^p 在评论集中第 j 条评论相应特征上的取值。 $Con(i_n^p)$ 的取值越大,则相关的评论者在信息项 i_n^p 上分歧越大,观点越不一致。

3) 评论集任一评论观点一致性。

相似地,假定从任一评论 r 提取待的信息项表示为集合 $I, I = i_1^r \cup i_2^r \cup \dots \cup i_n^r \cup \dots \cup i_N^r$,评论集 R^p 在信息项 i_n^p 上

取值的均值为 $\overline{Val}(i'_n)$,则评论集 R 中评论 r 的一致性表示为 $Con(r)$:

$$Con(r) = \frac{\sum_{n=1}^N [Val(i'_n) * \overline{Val}(i'_n)]}{\sqrt{\sum_{n=1}^N [Val(i'_n)]^2} * \sqrt{\sum_{n=1}^N [\overline{Val}(i'_n)]^2}} \quad (3)$$

2.3 评论支持度评价

在电子商务平台中,评论往往会引起消费者的关注,如果某条评论获得了较多点击或支持(如点赞、转发等形式),则表示该评论的支持度往往高。假定评论 r 被点击的次数为 $CN(r)$,其中支持和反对的数量分别为 $CNoS(r)$ 和 $CNoO(r)$, $CN(r) = CNoS(r) + CNoO(r)$;同时,假定评论 r_ϕ 获得了最多的点击(“支持”和“反对”两种形式)次数,点击次数为 $CN(r_\phi)$;评论 r_ϕ 获得最多的支持且次数为 $CNoS(r_\phi)$,评论 r 的支持度为 $Sup(r)$:

$$Sup(r) = \frac{CNoS(r)}{CN(r)} * \alpha + \frac{CN(r)}{CN(r_\phi)} * \beta \quad (4)$$

这里 α 和 β 为两个可调因子, $0 \leq \alpha \leq 1, 0 \leq \beta \leq 1$ 且 $\alpha + \beta = 1$ 。相关取值可以由专家赋初值,由反馈系统调节。

2.4 评论可信性评价

评价者等级和评论支持度越高,评论往往越可信,但也在某些评论因为商业操纵行为而获得较高支持度的情况。在获得了相同点击次数情况下,如果点击评论的评论者数量越多,其可信度越高(即在点击次数不变情况下,具有更多评论者的评论或评论集其可信度较高),为此定义评论者覆盖率。假定任一评论集 R (或一条评论),相关评论者集为 $Aut(R)$, $NofAut(Aut(R))$ 表示评论集 R 相关评论者的数量, $NofC(R)$ 表示评论集 R 被点击次数,则评论者覆盖率定义为 $Cov(Aut(R))$:

$$Cov(Aut(R)) = NofAut(Aut(R)) / NofC(R) \quad (5)$$

与评论观点一致性评价相似,评论可信度评价(目标)对象同样可以是任一评论或评论集。通过对相关评论集的可信性进行评估,进而为用户展示在支持度较高、一致性较好的评论。评论者等级经过归一化处理后为 $Rat(Aut(R))$,评论集 R 的可信度表示为 $Cre(R)$:

$$Cre(R) = Con(R) * Sup(R) * Rat(Aut(R)) * Cov(Aut(R)) \quad (6)$$

对于评论集的可信性评估,首先确定参与一致性评估的特征集,然后完成评论集在相关特征集上取值一致性评估。

3 实验结果及分析

3.1 实验数据及环境设置

本文所有实验运行在一个双核 Intel Core i5-4590 CPU 3.3 GHz、内存 8 GB 的 Windows 7 64 位操作系统,实验验证程序采用 Java 语言编写, JRE1.7 环境下运行。通过爬虫从某知名电商平台上爬取了 5 个服装商品相关联的近 10 万条评论并识别出了相应的评论者。

3.2 实验分析

基于规则库和方法库完成了评论数据的预处理,然后基于商品特征库、通用词典、情感词典以及方法库,完成了商品特征识别和特征值提取及标准化(商品特征及特征值的提取

方法本文不作介绍,可见前期工作^[28-29])。通过统计分析与评论挖掘,得到了不同等级评论者发布评论数量情况、评论特征集、评论特征值一致性、评论者级别与一致性相关性以及任一评论或评论集可信度评估结果。

1) 评论者等级及发表评论分析。

通过对爬取评论进行分析,统计得到了不同等级评论者发表评论所占比例,如图 2 所示。

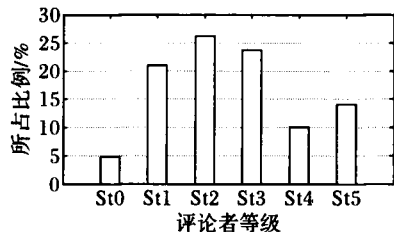


图 2 不同等级评论者发表评论所占比例

Fig. 2 Proportion of reviews published by reviewers belonged to each star rating

图 2 中 St0、St1、St2、St3、St4 和 St5 表示评论者等级,不同等级评论者针对相关商品发表评论的平均次数,如图 3 所示。

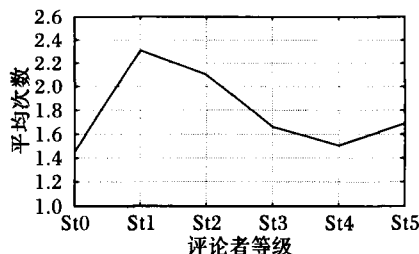


图 3 不同等级评论者发表评论平均次数

Fig. 3 Average times of reviews published by reviewers belonged to each star rating

从图 2 可以看出,多数评论者处于等级 1、2 和 3;从图 3 可以发现,等级为 1、2 和 3 的评论者发表评论的平均次数要多于其他级别。换句话说,处于中等级别的评论者更乐于分享他们的购买观念和消费体验质量。

2) 评论特征值的一致性。

基于用 Apriori 算法、商品特征库和 SCEPD(中国科学院发布的“简化中文情感词典”),从评论集半自动提取了频繁信息项并选择了 15 个重要的特征项,它们分别是 quality、baby、price、size、logistics、color、seller、service、style、packaging、fabric、workmanship、color、dress、odor,分别用 ft1 到 ft15 表示。通过计算得到上述评论集在这 15 个信息项上的一致性结果如图 4 所示。

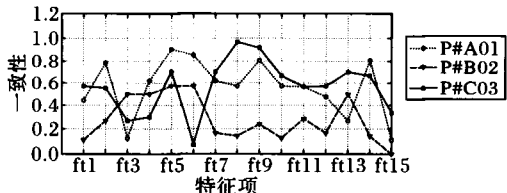


图 4 三个商品评论集中特征值一致性

Fig. 4 Consistency in value on three feature sets

从图 4 可以看出:a) 商品 P#A01 的评论集中关于“logistics”和“style”的观点差别很大,而在“odor”和“price”比较一致;b) 商品 P#B02 的评论集对于“logistics”和“color”也有显著区别,但相对于 P#A01 来说,分歧要小;c) 商品 P#C03

相关的评论集中,在“service”和“style”上观点不同,在“price”和“color”趋于一致。同时统计得到了关于这三个商品一致性取值的平均值,分别为 0.569、0.286 和 0.570,这意味着评论者对于商品 P#B02 的观点相对于其他两种商品更加一致。

3) 评论者级别在不同特征项取值的一致性。

通过分析不同等级评论者发表 3 种商品相关评论集在上述 15 个特征项上取值一致性,得到了在每个特征项上的一致性取值结果,分别如表 1 所示。表 1 中间各个单元都表示具有相同等级的评论者发表的评论在相关特征项上观点的一致性取值。

表 1 3 种商品 6 个等级评论者发表评论特征值一致性

Tab. 1 Feature value consistency in six ratings on reviews of three commodities

商品	等级	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15
P#A01-N	St5	0.60	0.39	0.43	0.40	0.57	0.86	0.62	0.43	0.76	0.57	0.45	0.64	0.66	0.00	0.20
	St4	0.33	0.59	0.33	0.62	0.81	0.55	0.67	0.62	0.62	0.67	0.50	0.48	0.27	0.43	0.11
	St3	0.48	0.45	0.42	0.57	0.57	0.64	0.42	0.57	0.81	0.42	0.27	0.50	0.43	0.66	0.27
	St2	0.51	0.50	0.33	0.22	0.50	0.43	0.55	0.22	0.57	0.52	0.33	0.42	0.57	0.55	0.00
	St1	0.62	0.27	0.33	0.62	0.66	0.28	0.50	0.62	0.65	0.57	0.11	0.52	0.42	0.00	0.11
	St0	0.41	0.32	0.11	0.55	0.41	0.36	0.32	0.57	0.41	0.26	0.27	0.00	0.27	0.36	0.14
P#B02-N	St5	0.33	0.25	0.42	0.50	0.43	0.42	0.27	0.24	0.25	0.28	0.39	0.86	0.67	0.00	0.20
	St4	0.22	0.33	0.39	0.33	0.29	0.55	0.32	0.18	0.34	0.33	0.29	0.48	0.27	0.43	0.11
	St3	0.27	0.29	0.48	0.48	0.42	0.50	0.33	0.20	0.35	0.15	0.20	0.50	0.43	0.66	0.27
	St2	0.17	0.22	0.42	0.43	0.33	0.37	0.18	0.14	0.24	0.20	0.17	0.42	0.57	0.55	0.00
	St1	0.30	0.17	0.25	0.37	0.47	0.57	0.17	0.30	0.20	0.17	0.30	0.52	0.42	0.00	0.11
	St0	0.11	0.27	0.33	0.40	0.50	0.40	0.15	0.14	0.24	0.13	0.29	0.17	0.50	0.14	0.00
P#C03-N	St5	0.67	0.40	0.62	0.43	0.41	0.57	0.45	0.46	0.66	0.29	0.36	0.39	0.11	0.50	0.37
	St4	0.81	0.50	0.67	0.62	0.50	0.67	0.50	0.48	0.27	0.43	0.33	0.59	0.33	0.62	0.33
	St3	0.55	0.48	0.42	0.57	0.67	0.42	0.27	0.50	0.43	0.66	0.48	0.45	0.42	0.57	0.48
	St2	0.50	0.43	0.55	0.22	0.33	0.52	0.33	0.42	0.57	0.55	0.51	0.50	0.33	0.22	0.51
	St1	0.42	0.28	0.50	0.62	0.39	0.57	0.11	0.52	0.42	0.00	0.62	0.27	0.33	0.62	0.62
	St0	0.57	0.30	0.27	0.30	0.50	0.02	0.42	0.47	0.25	0.33	0.24	0.50	0.70	0.67	0.33

从表 1 的结果可以发现,相同评论者对于相同对象发表的评论关于商品的不同特征其一致性存在差异;相同评论者对于不同的评论对象发表的评论在一致性方面也存在差异;具有较高等级的评论者发表的评论往往整体上具有较高的一致性,但具有最高等级的评论者发表的评论总体一致性并不是最高的。

4) 评论可信度评估。

评论可信度评价(目标)对象可以是任一评论或评论集(如任意选择的评论子集、评论者相关的子集、某时段相关的子集、某商品关联评论集等)。这里模型(3)在不考虑评论发布者等级及相关评论支持度差异的情况下(即 $Sup(r)$ 和 $Rat(Author(r))$ 都取常量 a) 和不区分商品特征项重要情况下,得到了不同等级评论者发表评论的可信度($a = 1$ 时),如图 5 所示。

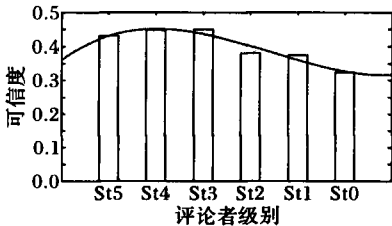


图 5 不同等级评论者发表评论的可信度

Fig. 5 Review credibility published by each rating reviewers

通过图 5 可以看出,处于最高等级发表的评论其一致性并不是最高的,但总体来说,具有较高等级评论者发表的评论集其评论观点具有更高的一致性。在进行可信度评估时,如果融合了评论支持度和评论者等级得到结果与该结果总体趋

于一致。

4 结语

针对评论可信性,现有研究主要考虑评论的来源和投票形式的支持度,没有从评论内在质量角度进行评估,因此,本文提出了一种在考虑评论者等级、评论支持度情况下,融合评论观点一致性的评论可信性评估方法,通过实验统计及分析给出了评论者所属等级、不同等级评论者发表评论的一致性以及特定评论集可信度评估结果。实验结果表明,高等级的评论者发表的评论其支持度和一致性往往较高,但来自相同数据源的不同类别对象其评论可信性存在差异。

还有许多工作有待完善:1) 本文没有考虑评论者专业性等方面不同对发布评论可信性的影响;2) 评论数据集的来源局限于特定领域,相关方法在其他领域的适应性方面还有待改进;3) 实验中涉及的评论集在时间上间隔时间较短,后期工作可以分析较长时间范围内相关评论者发表评论在一致性方面的变化,比如验证具有较长评论发表历史的评论者,其发表的评论是否具有更好的一致性。

参考文献 (References)

[1] GHOSE A, IPEIROTIS P G. Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(10): 1498 - 1512.

[2] HUANG A H, CHEN K, YEN D C, et al. A study of factors that contribute to online review helpfulness [J]. Computers in Human Behavior, 2015, 48: 17 - 27.

[3] 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.
- [4] 刘云, 孙宇清, 李明珠. 面向社会化媒体用户评论行为的属性推断[J]. *计算机学报*, 2017, 40(12): 2762–2736. (LIU Y, SUN Y Q, LI M Z. User attributes inference based on reviews on social media [J]. *Chinese Journal of Computers*, 2017, 40(12): 2762–2736.)
- [5] KWARK Y, CHEN J, RAGHUNATHAN S. Online product reviews: implications for retailers and competing manufacturers [J]. *Information Systems Research*, 2014, 25(1): 93–110.
- [6] YAN Z, XING M, ZHANG D, et al. EXPRS: an extended PageRank method for product feature extraction from online consumer reviews [J]. *Information & Management*, 2015, 52(7): 850–858.
- [7] LI M, HUANG L, TAN C H, et al. Helpfulness of online product reviews as seen by consumers: source and content features [J]. *International Journal of Electronic Commerce*, 2013, 17(4): 101–136.
- [8] 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.
- [9] LIU L, ZHAO X, WANG H, et al. Research on identification method of anonymous fake reviews in e-commerce [J]. *TELKOMNIKA Telecommunication Computing Electronics and Control*, 2016, 14(4): 1510–1520.
- [10] SIKORA R T, CHAUHAN K. Estimating sequential bias in online reviews: a Kalman filtering approach [J]. *Knowledge-Based Systems*, 2012, 27: 314–321.
- [11] HU N, BOSE L, GAO Y, et al. Manipulation in digital word-of-mouth: a reality check for book reviews [J]. *Decision Support Systems*, 2011, 50(3): 627–635.
- [12] HU N, LIU L, SAMBAMURTHY V. Fraud detection in online consumer reviews [J]. *Decision Support Systems*, 2011, 50(3): 614–626.
- [13] 李璐璐, 秦兵, 刘挺. 虚假评论检测研究综述[J]. *计算机学报*, 2018, 41(4): 946–968. (LI L Y, QIN B, LIU T. Survey on fake review detection research [J]. *Chinese Journal of Computers*, 2018, 41(4): 946–968.)
- [14] KWON O, LEE N, SHIN B. Data quality management, data usage experience and acquisition intention of big data analytics [J]. *International Journal of Information Management*, 2014, 34(3): 387–394.
- [15] ZHANG K Z, CHEUNG C M, LEE M K. Examining the moderating effect of inconsistent reviews and its gender differences on consumers' online shopping decision [J]. *International Journal of Information Management*, 2014, 34(2): 89–98.
- [16] DEVI J I. Estimating the helpfulness and economic impact of product reviews [J]. *International Journal of Innovative Research and Development*, 2012, 1(5): 232–236.
- [17] HE S X, BOND S D. Why is the crowd divided? attribution for dispersion in online word of mouth [J]. *Journal of Consumer Research*, 2015, 41(6): 1509–1527.
- [18] LEE S, CHOE H J Y. Predicting the helpfulness of online reviews using multilayer perceptron neural networks [J]. *Expert Systems with Applications*, 2014, 41(6): 3041–3046.
- [19] 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.
- [20] BAEK H, AHN J, CHOI Y. Helpfulness of online consumer reviews: readers' objectives and review cues [J]. *International Journal of Electronic Commerce*, 2012, 17(2): 99–126.
- [21] LI H, CHEN Z, MUKHERJEE A, et al. Analyzing and detecting opinion spam on a large-scale dataset via temporal and spatial patterns [C]// ICWSM 2015: Proceedings of the Ninth International AAAI Conference on Web and Social Media. Menlo Park, CA: AAAI Press, 2015: 634–637.
- [22] MIAO Q, LI Q, DAI R. AMAZING: a sentiment mining and retrieval system [J]. *Expert Systems with Applications*, 2009, 36(3): 7192–7198.
- [23] RACHERLA P, FRISKE W. Perceived ‘helpfulness’ of online consumer reviews: an exploratory investigation across three services categories [J]. *Electronic Commerce Research and Applications*, 2012, 11(6): 548–559.
- [24] 马如霞, 孟小峰. 基于数据源分类可信性的真值发现方法研究[J]. *计算机研究与发展*, 2015, 52(9): 1931–1940. (MA R X, MENG X F. Truth discovery based credibility of data categories on data sources [J]. *Journal of Computer Research and Development*, 2015, 52(9): 1931–1940.)
- [25] LU T C, YU T, CHEN S H. Information manipulation and Web credibility [C]// Proceedings of the 2017 International Symposium on Distributed Computing and Artificial Intelligence. Berlin: Springer, 2017: 86–95.
- [26] ABDOLRAHMANI A, KUBER R, EASLEY W. Web search credibility assessment for individuals who are blind [C]// Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility. New York: ACM, 2012: 369–370.
- [27] FLANAGIN A J, METZGER M J, PURE R, et al. Mitigating risk in ecommerce transactions: perceptions of information credibility and the role of user-generated ratings in product quality and purchase intention [J]. *Electronic Commerce Research*, 2014, 14(1): 1–23.
- [28] LI C, ZHOU H, ZHOU X F. Design and implement of hydrological data quality assessment system based on business rules [C]// Proceedings of the 2012 IET International Conference on Information Science and Control Engineering. Stevenage, UK: IET, 2012: 925–929.
- [29] LI C, ZHOU H, ZHOU X F. Data quality assessment in hydrological information systems [J]. *Journal of Hydroinformatics*, 2015, 17(4): 640–661.
- This work is partially supported by the National Natural Science Foundation of China (61040006), the Natural Science Foundation of Hubei Province (2018CFB451), the Education Department of Hubei Province (B2014245), the Doctor Foundation of Hubei University for Nationalities (MY2017B019).
- LI Chao**, born in 1979, Ph. D., associate professor. His research interests include data mining, data quality management.
- XIANG Jing**, born in 1978, M. S., lecturer. Her research interests include knowledge engineering.
- XIANG Jun**, born in 1978, Ph. D., associate professor. His research interests include mobile data quality management.