

doi:10.3772/j.issn.1000-0135.2016.004.003

基于话语标记理论的在线商品评分修正方法¹⁾

王伟军^{1,2,3} 宋艳秋^{2,3} 张婷婷^{2,3} 黄英辉^{2,3} 陈博洋^{2,3}

(1. 上海外国语大学国际工商管理学院, 上海 200083;

2. 华中师范大学信息管理学院, 武汉 430079;

3. 华中师范大学青少年网络心理与行为教育部重点实验室, 武汉 430079)

摘要 尽管在线商品评论是消费者购买决策的重要依据之一, 然而现实中评论内容与评分间的不一致将对购买决策的制订形成干扰, 折射出解决评论与评分真实性问题的重要性和现实性。当前关于评论与评分偏差的研究主要围绕语法、语义层面进行, 矫正的效果和速度存在较大局限。本文则在语义与语用层面, 通过引入语言学的话语标记理论, 构建商品评论话语标记语库, 设计包含可信度、评论立场、情感态度的评分修正体系, 并采用改进后层次分析法得到各评价指标的权重, 提出评论评分偏差的修正方法; 最后, 通过京东商城的真实评论数据修正结果验证了本方法的可行性和有效性, 从而深化并拓展了对在线商品评论分析的研究。

关键词 在线商品评论 在线商品评分 评分偏差 话语标记

A Correction Method for Score Deviation of Online Product Reviews Based on Discourse Markers Theory

Wang Weijun^{1,2,3}, Song Yangqiu^{2,3}, Zhang Tingting^{2,3}, Huang Yinghui^{2,3} and Chen Boyang^{2,3}

(1. School of International Business & Management, Shanghai International Studies University, Shanghai 200083;

2. School of Information Management, Central China Normal University, Wuhan 430079;

3. Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education, Central China Normal University, Wuhan 430079)

Abstract Online product reviews is an important information source which affects consumer purchasing decisions deeply, but the inconsistency between review content and score could bring troubles to the process of decisions-making. So solving reviews authenticity becomes significant and realistic. The current researches on correction of comments deviation are mainly from the perspective of grammatical or semantic level, thus there is a great limitation on the effect and speed of correction. Therefore, in the semantic and pragmatic level, this paper attempts to introduces one linguistics theory-Discourse Markers -to make a correction by analyzing the content of review. Firstly, this paper builds a library of discourse markers in the field of online product reviews and designs a correction system including credibility, position, emotional attitude. Then, this paper gets the weight of each index through improved analytic hierarchy process and proposes an algorithm to correct deviation of reviews scores. At last, the results of example verification of Jingdong mall show the correction method has better feasibility and validity, which deepens and expands the research of online product reviews.

Keywords online product reviews, online product scores, score deviation, discourse markers

收稿日期: 2015年7月22日

作者简介: 王伟军, 男, 1965年生, 教授, 博士生导师, 主要研究方向: 信息资源管理、电子商务与用户行为、数据分析与个性化服务, E-mail: wangwj919@163.com; 宋艳秋, 女, 1989年生, 硕士研究生, 主要研究方向: 个性化信息服务; 张婷婷, 女, 1990年生, 博士研究生, 主要研究方向: 网络用户行为与个性化服务; 黄英辉, 男, 1993年生, 博士研究生, 主要研究方向: 大数据分析与个性化服务; 陈博洋, 女, 1991年, 硕士研究生, 主要研究方向: 个性化信息服务。

1) 本文系国家自然科学基金项目(71271099、71571084、71471019)和国家留学基金资助项目(201506775025)的研究成果之一。

1 前言

由于电子商务和 Web2.0 迅猛发展,用户生成内容(User Generated Content, UGC)的在线商品评论作为新的网络口碑形式已成为消费者和商家决策的重要信息来源之一。一般来说,在线评分可以给予消费者对于某商品或服务价值的总体感知,在线评论的文本内容则更多地提供一种细节感知。然而,在现实中存在大量的商品评分与评论内容不一致的现象,极大地影响了评论的真实性和有效性,易使消费者陷入一种模糊、矛盾的心理感知状态,进而对消费者购买决策的制定造成干扰。

在线评论的偏差研究中关于评论内容与评论评分不一致的偏差类型并未受到足够重视,国内外学者更多将其作为评论可信度的影响指标进行研究和量化,实现评论信息可信度的排序输出。同时,对于偏差的矫正也主要是采取过滤垃圾评论,提高评论质量和可信度,进而减小评分偏差等间接矫正偏差的方式;另一方面,目前关于在线评论的文本内容分析处理方法主要是利用概率回归模型、向量模型进行产品特征词聚类分析^[1],或者是结合模糊理论,引入本体概念,构建模糊情感本体对评论文本内容进行分析^[2,3],但是这些处理方式大多局限于语法和语义层面,未能对其进行语用层面的讨论和分析。

本文聚焦于评论内容和评论分数不一致的这种偏差类型,尝试将语言学领域的话语标记理论移植到在线商品评论信息分析中,对评论文本内容进行语义和语用分析,进而实现在线商品评分的细粒度修正,试图得到更客观、更真实的评论评分,从而为消费者购买决策提供帮助。

2 相关研究

在线商品评论的研究作为一个热点,相关研究大多集中在技术层面的在线评论的产品特征抽取、聚类 and 情感分析、用户产品属性偏好的识别,以及应用层面的垃圾评论的检测、评论有用性研究和评论质量评估等。其中,关于在线商品评论质量的研究大多又聚焦于评论真实性问题,而归根到底评论的真实性问题则源于评论偏差的存在。

目前,国内外学者在并未考虑评论真实性存在偏差的前提假设下,发现在线评论对消费者的购买意愿、购买决策等具有显著影响^[4,5]。其中,研究较

多的评论可信度、评论感知有用性、感知风险、感知信任等因素在消费者的购买意愿中起到中介作用^[6-9]。感知有用的可信评论,信息质量较高,可以减少交易过程中的不确定性,降低感知风险,进而作用于顾客的购买意愿,影响购买决策。然而,评论真实性偏差的存在,使消费者陷入矛盾、犹豫的购买决策困境。

本文通过梳理国内外在线商品评论偏差研究的相关文献,发现在线商品评论信息存在评论有用性偏差和评论得分偏差两种类型。其中,评论有用性偏差包括失衡投票偏差、优胜者偏差以及早期评价偏差^[10],其存在导致评论的有用性投票结果并不客观。评论得分偏差按照时间因素、消费者个体因素以及评价的利益相关者的角度进行分类,包含顺序偏差^[11-13]、自我选择偏差^[14]、操纵评论偏差^[15],这些偏差类型的存在直接影响了评论得分的真实性。同时,国内外学者承认评论内容和评分不一致的偏差类型的真实存在,但研究主要侧重于此种类型偏差对评论可信度的影响^[16,17]。

偏差的产生并非单因素作用,更多的则是由心理因素、经济因素和制度因素共同影响而产生的。从心理层面来说,不仅不同的发表动机和从众心理均带来评论不同程度的主观偏差^[18-21];从经济层面来说,在线商品评论的评分的高低通过情感因素的中介作用,间接影响商品销量,卖家因此产生操纵评论得分的动机,人为地推高自身评分或者拉低竞争对手评分^[22]。从制度层面来说,主要涉及到网站评价机制,当前的网站评价机制存在信用度相关因素单一、评价级别太少、评价双方关系不平等、评价率不透明等方面的问题^[23,24],导致评论得分偏差的产生,影响评论的真实性。

针对偏差的矫正方式,国内外学者提出了不同的解决方案,主要涉及的是宏观层面的改进措施、建议和相关算法改进。李雨洁根据在线商品评论的分布,提出商品评分均值作为商品的无偏估计量的条件,以此来纠正偏差,并且从电子商务网站、消费者、卖家三个方面分别提出避免偏差产生的对策^[25]。此外,靳健、陈涛等通过对评论质量等级进行判断,辨识高质量评论,从而尽可能的避免偏差产生^[26,27]。

综上所述,国内外学者不仅开展了在线商品评论与评分偏差类型及其产生原因的研究,而且从语法及语义层面提出一些偏差矫正方法和建议。然而,对于现实中广泛存在评论内容与评分不一致的

这种偏差类型,国内外学者忽视了其存在的广泛性和影响的恶劣性,相关研究较少。相较于评论内容来说,评分反映的是消费者对商品或服务价值的总体感知。当总体感知与细节感知不相符甚至相悖时,不仅影响了评论者真实意图的传达,而且干扰了消费者购买决策的制定。鉴于此,本文引入话语标记理论,从语义和语用层次对此种类型的偏差进行有效修正,以使在线商品评论分数与评论内容的真实意思更匹配。

3 话语标记

作为言语系统中一类重要的语言表达形式,话语标记(Discourse Markers, DM)常常独立于句法之外,具有语义提示作用和语用制约作用,主要用于表明对话单位间各类关系的语言单位^[28]。话语标记不仅是话语表达上起联结作用的形式标记,也是话语理解时起引导作用的形式标记。前者为话语标记的篇章功能,后者为话语标记的人际功能,将篇章功能和人际功能紧密联结起来的是话语标记的元语用功能,即在元语用意识的指导下对语言进行合语境、合交际目的的选择功能^[29]。

国外语言学界对话语标记和类似现象的研究始于20世纪70年代,并在20世纪80、90年代才日渐兴盛起来。与国外相比,汉语话语标记研究从21世纪初才真正引起汉语学界的重视,起步比较晚,语料来源也大多源自北京大学CCL语料库(网络版),研究内容主要是针对不同语言中单个或某种类别话语标记的意义和功能进行详细描述。

本文通过前期的调查分析发现,由于评论发布者发布评论较为随意自然,具有鲜明的口语色彩,属于自然语言的真实应用,因此,在线商品评论中含有丰富的话语标记,也是话语标记语料的重要来源。

3.1 在线商品评论的话语标记类型

在线商品评论中常见的话语标记主要包含坦言性话语标记、阐发性话语标记、理据性话语标记、断言性话语标记、评价性话语标记。商品评论中不同的话语标记语,具有不同的语用功能,传递出不同的信息。

(1) 坦言型话语标记

坦言性话语标记如“说实话”、“说实在的”、“说白了”、“实不相瞒”、“说心里话”、“平心而论”等,突出了信息真实程度、说话人的坦诚程度以及评论

的公平性和公正性等语用信息。例1中的坦言性话语标记“说实在的”,强调评论者坦诚地表达了自身使用商品后的真实感受,使人信服;例2中的坦言性话语标记“平心而论”,突出评论者是出于一个客观公正的角度,对所购买的商品进行评价。

例1:说实在的,试用装的味道有点儿不一样,看不到什么活酵母的作用。

例2:大衣含有近40%的羊毛,样子还算大方,平心而论,400多块的价格很合适。

(2) 阐发性话语标记

阐发性标记语如“以我之见”、“在我看来”、“以我看”、“就我个人而言”等,主要强调个人建议、意见等相对主观的语用信息。评论者在例3表达了产品的使用效果情况,例4中则表达了产品的一些属性相关的信息。“就个人而言”、“在我看来”作为评论中所包含的阐发性话语标记,则进一步强调使用感受是结合自身实际情况的考虑,比较主观,不一定适用于其他消费者。

例3:就个人而言,这款面霜用着舒服,皮肤更多的改善还需要时间,会持续使用的!

例4:毛呢裙质量过关,当然,以我看,再长点更好。

(3) 理据性话语标记

理据性标记语如“通常情况下”、“理论上来说”、“道理上来说”、“据说”等,主要具有增强说理、劝说的理据性的语用功能。例5中评论者借助于理据性话语标记“普遍来说”,通过强调普遍存在的情况,侧面反映其对于所送礼品的不满,例6中则借“朋友”之口,相对客观委婉地表达了产品的不错功效。

例5:不祈求你们多慷慨,但是普遍来说,送的东西不要太次好吗?

例6:包装很精致,还送了礼品!据朋友说,补水效果还不错。

(4) 断言性话语标记

断言性标记语如“我敢肯定”、“不客气地说”、“我保证”等,主要表达了评论者以言行事的决心,主观性极强。例7、例8中的断言性话语标记“我敢拍胸脯保证”、“我敢肯定”,表达评论者非常肯定自身对于商品所持的看法,主观性极强。

例7:手机触屏太差,完全跟不上节奏与速度。我敢拍胸脯保证,绝对不带恶意中伤。

例8:屏幕边缘触控灵敏,还得多摸索。机子查了,我敢肯定没问题。

(5) 评价性话语标记

评价性标记语如“太棒了”、“值得庆幸的是”、“太可惜了”、“无语的是”等,可表达说话人喜悦、庆幸、惋惜、反讽等不同的情感态度。例 9 中的衣服的质量出乎评论者意料之外,评价性话语标记“令人惊喜的是”,表达了评论者的惊喜之情;例 10 中评论者则借助于评价性话语标记“搞笑的是”,表达了对店家服务态度的嘲讽之意。

例 9:没寻思衣服能好到哪去,令人惊喜的是,收到货后发现可以跟好几百块的衣衣媲美!

例 10:店家迟迟不发货,搞笑的是,我一要求退款,店家却神速发了货!

3.2 基于话语标记理论的在线商品评论体系

鉴于话语标记本身是作用于话语层次的,本文将借助于在线商品评论中蕴含的大量话语标记,构建在线评论的话语标记库,并设计调查问卷,获得不同类型话语标记在可信度、立场以及情感维度的属性值,最后结合本文提出的评价体系,对评论内容进行深层次的挖掘,赋予评论与内容相匹配的评分,对评分偏差进行精细化修正。

具体来说,首先依照在线商品评论的话语标记词库,对评论进行逐条检测,识别评论中所含话语标记;然后,参照问卷调查结果,获取检测出的话语标记在可信度、立场、情感属性值,并进行比较,得到三大维度的属性最大值;最后,利用我们提出的属性最大值加和的 DM 修正算法,结合评论自身评级状态(好评、中评和差评),计算基于话语标记的评论最终评分,如图 1 所示。

3.3 在线商品评论话语标记库

坦言性话语标记、阐发性话语标记、理据性话语标记、断言性话语标记、评价性话语标记内涵丰富,每一类型中话语标记数量巨大,目前并未有学者针对庞大的汉语言话语标记进行系统的总结归纳。针

对这一现状,并结合本文的研究领域和研究目的,构建一个基于在线商品评论的常见话语标记库显得尤为重要。

构建商品评论的话语标记库时,本文主要从两方面开始着手:一方面,收集国内外的有关话语标记的相关论文中涉及的话语标记,并依据坦言性话语标记、阐发性话语标记、理据性话语标记、断言性话语标记、评价性话语标记等五大类别对其进行归类;另一方面,利用网络爬虫程序在京东商城上随机抓取了 5000 条普通商品评论,利用频次统计法,将在线商品评论领域中出现频次较高(频次 ≥ 50 次)的话语标记进行保留。同时,考虑到在线商品评论与传统文学的差异性,剔除文学领域较为常见,而在自然语言中出现频率较低的、非口语化的话语标记,诸如“不瞒你说”、“以我之见”等。

根据不同类型话语标记的主要凸显的不同语义和语用信息,将其划分为可信度-话语标记、立场-话语标记和情感-话语标记,其中可信度-话语标记体现说话人的真实坦诚和公平公正性,主要来源于坦言性话语标记;立场-话语标记主要涉及主观立场和客观立场,分别来源于阐发性话语标记、断言性话语标记和理据性话语标记;情感-话语标记主要包含正向情感、负向情感和中性情感,来源于评价性话语标记,如表 1 所示。

4 基于 DM 的在线评论修正方法

为了获取普通消费者对可信度-话语标记、立场-话语标记(主观立场/客观立场)、情感-话语标记(正向情感/负向情感/中性情感)在可信度、立场和情感三大维度的凸显程度的认知情况,设计不同类型话语标记的感知调查。问卷测试者共计 203 人,回收有效问卷 162 份。其中,有效问卷的测试者年龄均在 20~35 岁,其中男性占 41.6%,女性占 58.4%,且均具有 1 年及以上的网上购物经验。

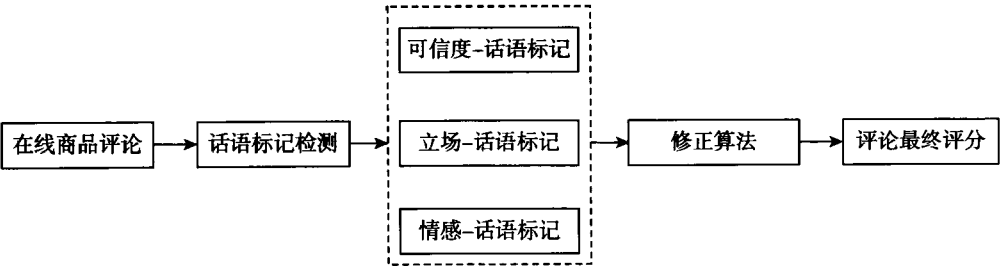


图 1 基于话语标记的在线商品评论修正体系

表1 常见在线商品评论的话语标记库举例

话语标记	属性	实例		类别
可信度类话语标记	真实坦诚性	实话实说/说实话;说实在的/实在 xx;说心里话;真心 xx/真的 xx/真是 xx		坦言性话语标记
	公平公正性	说句良心话;凭良心说;平心而论		
立场类话语标记	主观立场	个人觉得/感觉/认为;就个人而言;我敢说/我肯定;肯定是/简直是/绝对是/必定是/确实是/必须是/完全是		阐发性话语标记、断言性话语标记
	客观立场	据 XX 说/听 XX 说/XX 都说;普遍来说/一般来说;应该是		理据性话语标记
情感类话语标记	正向情感类	强	让人惊喜/又惊又喜/惊奇/惊讶的是;大、超爱/赞;太、很、非常、特别棒/满意/喜欢;棒棒哒/超值	评价性话语标记
		中	挺好的/还不错/还可以/凑合吧	
		弱	好在;庆幸的是/幸运的是/幸好	
	负向情感类	强	搞笑的是/好笑的是/讽刺的是 无语了/无言/无力吐槽	
		较强	忍无可忍的是;太差/不爽/郁闷	
		中	失望的是	
		弱	还行吧/就这样吧/还可以吧;就是 XX/只是 XX;可惜的是/遗憾的是	
	中性情感类	竟然/居然;出乎意料/出人意料/出乎意外/意想不到/没想到的是		

4.1 话语标记理论修正体系

4.1.1 DM 的可信度、立场和情感维度的指标权重

本文选取了涵盖了坦言性话语标记、阐发性话语标记、理据性话语标记、断言性话语标记、评价性话语标记等五大话语标记类型的 28 种话语标记,进行了问卷调查。其中问卷中选取的 28 个话语标记主要包含 7 种可信度类话语标记(“实话实说/说实话”、“说实在的”、“说心里话”、“真心觉得”、“说句良心话”、“凭良心说”、“平心而论”),6 种立场类话语标记(“个人觉得/感觉”、“就个人而言”、“我敢说”、“我肯定”、“据 XX 说/听 XX 说”、“普遍来说”),6 种正向情感类话语标记(“让人惊喜的是”、“太/很/非常/特别棒”、“太/很/非常/特别满意”、“大赞”、“好在”、“庆幸的是/幸运的是”),7 种负向情感类话语标记(“失望”、“可惜的是/遗憾的是”、“还行吧”、“忍无可忍的是”、“太差”、“搞笑的是/讽刺的是”、“太无语了”)和 2 种中性情感类话语标记(“出乎意料”、“没想到”)。

问卷中 28 个话语标记均采用双例句形式,调查用户在阅读含有话语标记的评论后对评其所体现出的可信度、主观/客观立场和情感倾向的感知情况。如“实话实说”:

例句 I “衣服做工 XX,超 XX。缺点也有,帽子里线头 XX。”

例句 II “衣服做工 XX,超 XX。实话实说,缺点也有,帽子里线头 XX。”

问卷填写者,结合自身网购经验,观察相较于例句 I,添加了话语标记“实话实说”的例句 II,分别在真实坦诚性、公平公正性、主观性、客观性、情感倾向(正向、负向)等方面,是否有凸显作用,用户若感知到“实话实说”对上述相关属性具有强调凸显作用,则在相应属性栏目进行勾选。

问卷回收后,将统计数据汇总,按照表 1 中话语标记实例的类型和程度轻重进行归类合并处理,然后对归类后的话语标记对可信度、主观性、客观性、正向情感和负向情感等各方面的突显程度进行均值求取,结果如表 2 所示。其中值得注意的是,中性情感话语标记自身情感倾向不明显,其凸显的正向情感和负向情感需要根据实际语境中具体判

表 2 在线商品评论的话语标记的属性值

话语 标记	实例列举	可信度	主观性	客观性	正向 情感	负向 情感
可信度类 话语标记	实话实说/说实话;说心里话;说实在的/实在 xx;说句良心话;真心 xx/真的 xx/真是 xx;凭良心说;平心而论	0.684	0.135	0.181	0	0
立场类 话语标记	个人觉得/感觉/认为;就个人而言	0.509	0.406	0.085	0	0
	我敢说/我肯定;肯定是/简直是/绝对是/必定是/确实是/必须是/完全是	0.531	0.420	0.049	0	0
	据 XX 说/听 XX 说/XX 都说;应该是;普遍来说/一般来说	0.437	0.074	0.489	0	0
正向情感 类标记	让人惊喜/又惊又喜/惊奇/惊讶的是;大、超爱/赞;太、很、非常、特别棒/满意/喜欢;棒棒哒/超值	0.175	0.235	0.021	0.569	0
	挺好的/还不错/还可以/凑合吧	0.192	0.208	0.048	0.48	0.072
	好在;庆幸的是/幸运的是/幸好	0.214	0.07	0.095	0.443	0.178
负向情感 类标记	搞笑的是/好笑的是/讽刺的是 无语了/无言/无力吐槽	0.149	0.11	0.066	0	0.675
	忍无可忍的是;太差/不爽/郁闷	0.146	0.157	0.029	0	0.668
	失望的是	0.19	0.191	0	0	0.619
	就是 XX/只是 XX;可惜的是/遗憾的是	0.239	0.148	0.084	0.026	0.503
中性情 感类标记	竟然/居然;出乎意料/出人意料/出乎意外/意想不到/没想到的是	0.214	0.208	0.022	0.556/0	0/0.556

断,可借助台湾黄金兰学者的 LIWC 词典对中性情感类话语标记的上下文进行情感极性识别,辅助判定该话语标记的情感倾向。比如“出乎意料”作为中性情感类话语标记,既可表示超出预想的满意,也可表示不如预想的不满,利用极性词典自动识别匹配出该话语标记的下文出现“XX 出乎意料的舒服”中的“舒服”一词,根据“舒服”一词的极性,判定该评论中话语标记“出乎意料”在此处凸显了正向情感,并借助表 2 可知正向情感的凸显程度为 0.556,负向情感的突显程度为 0。

4.1.2 基于 DM 的评分修正算法

每个话语标记本身内涵丰富,本文将其定义为五维向量 $\text{word}_m = (a_m, b_m, c_m, d_m, e_m)$,五个维度分别是可信度、主观立场、客观立场、正向情感、负向情感,并用英文字母 a, b, c, d, e 表示。

假设单条评论中含有 n 个话语标记,每个话语标记均用五维向量表示,从 word_1 到 word_n 。不同向量的五个维度的值,各自代表了不同话语标记在可信度、主观性、客观性、正向情感、负向情感的突显程

度,具体数值来源于表 2 在线商品评论的话语标记的属性均值。

$$\begin{aligned}\text{word}_1 &= (a_1, b_1, c_1, d_1, e_1) \\ \text{word}_2 &= (a_2, b_2, c_2, d_2, e_2) \cdots \\ \text{word}_n &= (a_n, b_n, c_n, d_n, e_n)\end{aligned}$$

判断比较不同话语标记(word_1 到 word_n)的同一属性的不同均值,分别获取可信度属性、立场属性、正向情感属性、负向情感属性的最大值 q_1, q_2, q_3, q_4 ,计算方式如下所示。由于评论中说话者立场相对独立,主要倾向于为主观或客观,因此本文将立场属性最大值 q_2 取自于主观立场属性最大值和客观立场属性最大值之间的较大值。

$$\begin{aligned}q_1 &= (a_1, a_2, \cdots, a_n) \\ q_3 &= (d_1, d_2, \cdots, d_n) \\ q_2 &= (b_1, b_2, \cdots, b_n, c_1, c_2, \cdots, c_n) \\ q_4 &= (e_1, e_2, \cdots, e_n)\end{aligned}$$

由于可信度、立场与情感态度是正向影响的关系,因此情感态度($q_3 + q_4$)的符号决定了评论话语标记的权重值的符号,即如果评论说话人立场坚定,内容真实可信,那么对评论所蕴含的情感态度具有

加强作用,呈现出一种评论倾向的“马太效应”。同时,基于之前的问卷数据统计的结果,发现可信度、立场和情感态度在我们提出的评分修正体系中存在一定的比例关系,即可信度:立场:正向情感:负向情感=0.291:0.279:0.175:0.256。此外,由于人们对

于不同评分等级(好评、中评、差评)的打分界限相对模糊,结合实例考察,我们设定好评的评分区间为[3, 5],中评的0评分区间为[2, 4],差评的评分区间为[1, 3]。基于以上所述,我们设计出单条评论多项话语标记的权值总和 W_{DM} 计算公式为:

$$\begin{aligned} a &= \text{sgn}(q_3 + q_4) \\ \beta &= |1 - (1 - q_1) \times (1 - 0.291)| \\ \gamma &= |1 - (1 - q_2) \times (1 - 0.279)| \\ \delta &= \begin{cases} -|1 - (1 - q_4) \times (1 - 0.259)| & q_3 = 0, q_4 \neq 0; \\ |1 - (1 - q_3) \times (1 - 0.175)| & q_3 \neq 0, q_4 = 0; \\ |1 - (1 - q_3) \times (1 - 0.175)| - |1 - (1 - q_4) \times (1 - 0.259)| & q_3 \neq 0, q_4 \neq 0; \end{cases} \\ W_{DM} &= a \cdot (\beta + \gamma + \delta) \end{aligned}$$

然后在计算评论最终得分时,同时兼顾考虑情感极性和评论原始的评级状态(好评、中评、差评),设计以下的评论最终评分的计算方法,即情感最终极性为正向时,评论最终得分由评论的所处评价状态区间的最小临界值与话语标记的权重加和求得(此时,权值为正数);若情感最终极性为负向时,评论最终得分由评论所处评价状态区间的最大临界值与话语标记的权重加和求得(此时,权值为负数)。

Praise: if $(q_3 + q_4) > 0$, $Score = 3 + W_{DM}$; if $(q_3 + q_4) < 0$, $Score = 5 + W_{DM}$;

Neutral: if $(q_3 + q_4) > 0$, $Score = 2 + W_{DM}$; if $(q_3 + q_4) < 0$, $Score = 4 + W_{DM}$;

Bad: if $(q_3 + q_4) > 0$, $Score = 1 + W_{DM}$; if $(q_3 + q_4) < 0$, $Score = 3 + W_{DM}$;

5 实例例证

前文提出了基于话语标记的在线商品评分修正算法,可针对评论内容与评分不相符的这类偏差进行修正,降低评论内容和评分之间差距,实现在线商品评论的评分修正。在此我们再以京东网站上的商品评论作为研究对象,分别从可行性和有效性对该算法进行验证。

5.1 可行性验证

为了对 DM 修正算法的可行性进行验证,本文随机选取京东商城“韩都衣舍”的一款女装,2015年2月22日下载该商品下的全部评论共111条(好评88条、中评13条、差评10条)。

依照前面所构建的基于在线商品评论的话语标记库,对111条评论逐条进行话语标记检测,发现含有话语标记的评论共计50条,覆盖率达到

45.05%。同时,50条评论中单条评论所含话语标记在1至3个之间。以评论编号为4的真实评论为例,计算机自动检测发现该条评论含有话语标记“就是”,利用 DM 修正算法,获得该条评论的可信度指标、立场指标和情感指标分别为0.460、0.386、-0.434,结合该条评论原始评级状态(好评)和情感极性(负向),获得修正后最终评分为3.72分。

结合111条评论原始评级状态,利用上述的评分修正算法,对评论的可信度、立场和情感倾向进行分析计算,获得与评论内容匹配的最终评分,显示了该算法具备可行性。其中,原始评级状态为好评且含有话语标记的42条评论修正后的最终评分区间为[3.28, 4.62],原始评级状态为中评且含有话语标记的4条评论修正后的最终评分区间为[2.55, 3.47],原始评级状态为差评且含有话语标记的3条评论修正后的最终评分区间为[1.36, 1.66],详细指标见附表1。

5.2 有效性验证

为验证 DM 修正算法的有效性,本文再次从京东商城的女装、美容护肤、书籍、手机、家具五大类目下分别随机选取一款商品,并收集上述五款商品525条在线评论,获取每条评论的网站原始评分、DM 修正评分和人工评分三组数据。其中,人工评分数据来源于笔者在华中师范大学信管学院、心理学院和社会学院进行的志愿者招募实验。通过招募,40个参与者随机分为5组,分别对上述的五款商品的评论进行在线阅读,并根据评论真实内容所反映的满意程度进行人工判断打分。计算每条评论的人工打分均值,该数据可用来代表评论本身的真实评分。

首先计算5款不同商品中 DM 修正算法的适用

范围,然后针对已收集的上述 5 款商品评论的网站原始评分、DM 修正评分和人工评分这三组数据,利用 SPSS 的独立样本 t 检验,验证该算法修正结果的有效性。

5.2.1 基于 DM 修正算法的适用范围

依照前面所构建的基于在线商品评论的话语标记库,对 5 款商品 525 条评论逐条进行话语标记的自动识别和检测,发现含有话语标记的评论共计 306 条,总体覆盖率达到 58.29%。针对每款商品来说,含有话语标记可通过 DM 算法实现商品评分的精细化修正的评论数量占该款商品评论总量的比例,反映了该算法的适用范围。通过表 3 的简单的统计分析表明,该算法在 5 款商品评论总体的适用范围为 48.24% ~ 64.81%。

表 3 基于 DM 修正算法的适用范围

类别	评论数量	含话语标记 评论数量	DM 修正算法的 适用范围
女装	113	71	62.83%
美容护肤	96	53	55.21%
电子数码	85	41	48.24%
书籍	123	71	57.72%
家具	108	70	64.81%

5.2.2 基于 DM 算法修正效果的 T 检验

在进行 t 检验的之前,为保证测验结果的一致性、稳定性及可靠性,需对人工评分数据先进行信度测验,具体结果如表 4 所示。

表 4 人工评分数据的信度测验

序号	类别	Cronbach's Alpha
1	女装	0.953
2	美护	0.940
3	数码	0.870
4	书籍	0.836
5	家具	0.951

五款商品的问卷 Cronbach's Alpha 系数 (Cronbach's Alpha 系数是检验信度的指标,通常该指在 0.8 以上,被认为可信度较高)均在 0.8 以上,这表明人工评分数据信度比较高。然后对收集的数据进行独立样本 t 检验,分别用 t_1 表征原始评分, t_2 表征人工评分, t_3 表征修正评分, t 检验结果汇总如表 5 所示。

以“女装”为例,在 95% 的置信度下,原始评分 t_1 和人工评分 t_2 的显著性水平 $P = 0.036 < 0.05$,存在显著性差异;人工评分 t_2 和修正评分 t_3 的显著

表 5 基于 DM 算法的评分修正效果 t 检验

类别	比较项	均值方程的 t 检验						
		t	df	Sig. (双侧)	均值 差值	标准 误差值	差分的 95% 置信区间	
							下限	上限
女装	$t_1 \sim t_2$	2.12	138.672	0.036	0.472	0.223	0.032	0.912
	$t_2 \sim t_3$	-1.184	117.542	0.239	-0.23	0.194	-0.615	0.155
美护	$t_1 \sim t_2$	2.083	104	0.04	0.419	0.201	0.02	0.818
	$t_2 \sim t_3$	0.777	104	0.439	0.135	0.173	-0.209	0.478
数码	$t_1 \sim t_2$	2.032	80	0.045	0.59951	0.29501	0.01242	1.18661
	$t_2 \sim t_3$	-0.574	80	0.568	-0.14146	0.24658	-0.63218	0.34925
书籍	$t_1 \sim t_2$	2.063	115.388	0.041	0.25268	0.12246	0.01011	0.49524
	$t_2 \sim t_3$	0.803	140	0.423	0.07239	0.09016	-0.10586	0.25065
家具	$t_1 \sim t_2$	2.284	126.089	0.024	0.28971	0.12685	0.03869	0.54074
	$t_2 \sim t_3$	1.553	138	0.123	0.14257	0.09181	-0.03896	0.3241

注: t_1 = 原始评分, t_2 = 人工评分, t_3 = 修正评分

性水平 $p = 0.239 > 0.05$, 无显著差异。综合表 5 中五大类别商品的 t 检验结果, 均显示未使用 DM 修正算法时原始评分与人工评分之间存在显著差异, 使用 DM 算法后的最终修正评分与人工评分无显著差异, 说明了有无使用 DM 算法对评论评分的影响存在显著不同, 表明了经过 DM 修正算法后的商品评分更接近于评论的真实评分, 也进一步验证了将 DM 修正算法应用于评论评分偏差纠正的有效性。

6 不足与展望

本文尝试引入语言学领域的话语标记理论, 并借助于其所蕴含的语义和语用信息, 为解决电子商务领域在线商品评论内容与评论评分不一致的评论失衡偏差的现实难题做出了积极探索。借助于话语标记理论, 我们首次构建了基于在线商品评论的话语标记库, 并对基于 DM 的修正算法进行了计算机代码实现, 可自动识别检测出评论中的话语标记, 并对真实评论内容反映出的可信度、评论立场和情感倾向进行了量化测量, 进而对在线商品评论评分完成了精细化的修正, 使之与评论真实内容更加匹配。同时, 前文的实证分析也极好地证明了该方法的可行性和有效性。

然而, 在实践的过程中, 我们也遇到一些难题, 存在以下不足: 其一, 话语标记词库规模较小。由于单就从相关论文和 5000 条商品评论入手去构建话语标记库, 渠道少, 词库收录的话语标记数量有限; 其二, 问卷调查的样本类型单一。在问卷调查过程中, 选取的样本人群局限于高校人群, 导致数据的代表性和科学性有待验证; 其三, 在基于 DM 的修正算法中, 可信度、立场和情感这三大维度之间有可能存在交叉影响的关系, 而该修正算法中并未将此考虑其中。

当前存在的问题为我们指明了下一步前进的方向, 我们不仅要增加更多的话语标记收集渠道, 扩建现有的词库, 同时增加调查样本, 改进修正算法, 使之更加科学适用。

参 考 文 献

- [1] 孙春华, 刘业政. 基于产品特征词关系识别的评论倾向性合成方法[J]. 情报学报, 2013, 32(8): 844-853.
- [2] 那日萨, 李媛. 基于在线评论的消费者模糊情感计算与推理[J]. 情报学报, 2011, 30(4): 417-423.
- [3] 史伟, 王洪伟, 何绍义. 基于知网的模糊情感本体构建

- 研究[J]. 情报学报, 2012, 31(6): 595-602.
- [4] Hankin L. The effects of user reviews on online purchasing behavior across multiple product categories[J]. Master's final project report, UC Berkeley School of Information, <http://www.ischool.berkeley.edu/files/lhankinreport.pdf>, 2007.
- [5] Senecal S, Nantel J. The influence of online product recommendations on consumers' online choices[J]. Journal of Retailing, 2004, 80(2): 159-169.
- [6] Chau P Y K, Hu P J H, Lee B L P, et al. Examining customers' trust in online vendors and their dropout decision: an empirical study[J]. Electronic Commerce Research and Applications, 2007, 6(2): 171-182.
- [7] 李念武, 岳蓉. 网络口碑可信度及其对购买行为之影响的实证研究[J]. 图书情报工作, 2009, 53(22): 133-137.
- [8] 毕继东. 网络口碑对消费者购买意愿影响实证研究[J]. 情报杂志, 2009, 28(11): 46-51.
- [9] 张宁. 在线评论对经济型酒店顾客购买决策的研究[D]. 大连: 东北财经大学, 2011.
- [10] Liu J, Cao Y, Lin C Y, et al. Low-Quality Product Review Detection in Opinion Summarization[C]// Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, Association for Computational Linguistics. 2007: 334-342.
- [11] Rabin M, Schrag J L. First impressions matter: A model of confirmatory bias[J]. The Quarterly Journal of Economics, 1999, 114(1): 37-82.
- [12] Kapoor G, Piramuthu S. Sequential Bias in Online Product Reviews[J]. Journal of Organizational Computing and Electronic Commerce, 2009, 19(2): 85-95.
- [13] Sikora R T, Chauhan K. 2012. Estimating sequential bias in online reviews: a Kalman filtering approach[J]. Knowledge-Based Systems, 2012, 27: 314-321.
- [14] Li X X, Hitt L M. 2008. Self-Selection and Information Role of Online Product Reviews[J]. Information Systems Research, 2008, 19(4): 456-474.
- [15] 赵衍. 网络虚假评论研究述评[J]. 上海管理科学, 2014, 36(4): 85-88.
- [16] Hyunmi B, Joong H A, Youngseok. Helpfulness of Online Consumer Review Readers' Objectives and Review Cues[J]. International Journal of Electronic Commerce, 2012, 17(2): 99-126.
- [17] 王倩倩. 一种在线商品评论信息可信度的排序方法[J]. 情报杂志, 2015, 34(3): 181-185.
- [18] 张丽. 在线评论的客户参与动机与评论有效性研究[D]. 天津: 南开大学, 2011.

[19] Sundaram D S, Mitra K, Webster C. Word-of-mouth communications: A motivational analysis[J]. Advances in consumer research, 1998, 25(1): 527-531.

[20] 李兆飞. 在线消费者产品评论发表动机的研究[D]. 哈尔滨: 哈尔滨工业大学, 2011.

[21] 李敏乐. 在线评论对消费者从众行为的影响机制研究——神经科学视角[D]. 杭州: 浙江大学, 2010.

[22] Hu N, Koh N S, Reddy S K. Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales [J]. Decision Support Systems, 2014, 57: 42-53.

[23] 罗英, 李琼. C2C 信用评价体系的存在问题及对策[J]. 现代管理科学, 2013(8): 83-86.

[24] 谈晓勇, 任永梅. C2C 电子商务网站信用评价中的主要问题及其对策研究[J]. 全国商情, 2008(19): 39-40.

[25] 李雨洁, 廖成林, 李忆, 等. 在线商品评论可信吗——在线商品评论的偏差分析及矫正策略[J]. 营销科学学报, 2013, 9(2): 111-125.

[26] 靳健, 季平. 用于在线产品评论质量分析的 Co-training 算法[J]. 上海大学学报(自然科学版), 2014, 20(3): 289-295.

[27] 陈涛, 谢丽莎. 在线评论文本信息质量等级的测量探析——基于模糊综合评价法[J]. 科技创业月刊, 2012(7): 50-52.

[28] 刘丽艳. 汉语话语标记研究[M]. 北京: 北京语言大学出版社, 2011.

[29] 周明强. 坦言性话语标记语用功能探析[J]. 当代修辞学, 2013(5): 57-64.

附表 1

韩都衣舍女装的实例分析

编号	原始评级	话语标记	可信度	立场	情感	最终评分
4	好评	就是	0.460	0.386	-0.434	3.72
9	好评	挺不错的;就个人而言;就是	0.652	0.572	-0.059	3.72
10	好评	挺不错的;只是	0.460	0.429	-0.059	4.05
12	好评	挺不错的;可惜的是;真的	0.776	0.429	-0.059	3.74
17	好评	真心;挺不错	0.776	0.429	0.261	4.47
19	好评	真心;挺不错;很喜欢	0.776	0.448	0.335	4.56
21	好评	还不错	0.427	0.429	0.261	4.12
24	好评	还不错;没有想到的是(负面)	0.443	0.429	-0.099	4.03
25	好评	还不错	0.427	0.429	0.261	4.12
27	好评	朋友都说;就是;个人建议	0.652	0.632	-0.434	3.28
29	好评	还可以	0.427	0.429	0.261	4.12
30	好评	不错;比想象中好(正面)	0.443	0.429	0.324	4.20
31	好评	不错;遗憾的是	0.460	0.429	-0.059	4.05
32	好评	真心;棒棒哒	0.776	0.448	0.388	4.61
35	好评	让人惊喜;平心而论	0.776	0.448	0.388	4.61
37	好评	挺好的	0.427	0.429	0.261	4.12
38	好评	挺喜欢的	0.415	0.448	0.388	4.25
39	好评	挺好的;不错的	0.427	0.429	0.261	4.12
40	好评	庆幸的是;很满意	0.443	0.448	0.256	4.15
42	好评	嫂子说;凭良心说	0.776	0.632	-0.081	3.51

续表

编号	原始评级	话语标记	可信度	立场	情感	最终评分
44	好评	还可以	0.427	0.429	0.261	4.12
46	好评	出乎意料(负面)	0.443	0.429	-0.495	3.63
50	好评	朋友都说	0.601	0.632	0	4.23
54	好评	很满意	0.415	0.448	0.388	4.25
57	好评	不错	0.427	0.429	0.261	4.12
58	好评	妹妹说;很喜欢	0.601	0.632	0.388	4.62
59	好评	就是	0.460	0.386	-0.434	3.72
60	好评	非常喜欢;只是	0.460	0.448	0.014	3.92
63	好评	惊喜	0.415	0.448	0.388	4.25
66	好评	很棒	0.415	0.448	0.388	4.25
70	好评	还可以;就是	0.460	0.429	-0.059	4.05
71	好评	还可以	0.427	0.429	0.261	4.12
72	好评	还不错	0.427	0.429	0.261	4.12
73	好评	她说;很喜欢	0.601	0.632	0.388	4.62
77	好评	凑合吧	0.427	0.429	0.261	4.12
79	好评	朋友都说	0.601	0.632	0	4.23
82	好评	同事们都说	0.601	0.632	0	4.23
84	好评	还可以	0.427	0.429	0.261	4.12
85	好评	不错	0.427	0.429	0.261	4.12
86	好评	她说;可惜的是	0.601	0.632	-0.434	3.33
87	好评	很喜欢;真的	0.776	0.448	0.388	4.61
88	好评	不错	0.427	0.429	0.261	4.12
92	中评	实话实说;还可以	0.776	0.429	0.261	3.47
93	中评	太郁闷了;居然	0.443	0.429	-0.578	2.55
98	中评	还可以;就是	0.460	0.429	-0.059	3.05
99	中评	没有想象中;失望	0.443	0.429	-0.542	2.59
103	差评	搞笑的是	0.397	0.358	-0.583	1.66
104	差评	太无语了	0.397	0.358	-0.583	1.66
111	差评	没想到;完全是;太失望了	0.667	0.429	-0.542	1.36

(责任编辑 王海燕)