

虚假评论的识别与过滤：现状与展望



□王 乐¹ 张紫琼² 崔雪莹¹

[1. 西安交通大学 西安 710061; 2. 哈尔滨工业大学 哈尔滨 150001]

[摘要] **【目的/意义】**构建虚假评论识别的理论框架,归纳各个知识模块的局限和争议,并据此提出未来研究方向。对推动虚假评论识别的相关研究有重要理论意义,同时对电商平台有效识别和过滤虚假评论也有一定实践启示。**【设计/方法】**检索发表于国内外知名期刊的470篇中英文文献,并通过主题分析,对这些文献的理论基础、研究方法以及结论进行了归纳,对理论概念之间的逻辑联系进行了详细阐述;最后,对现有文献进行了评述,指出了未来研究的方向。**【结论/发现】**未来虚假评论的治理需要以“防”为主,以“治”为辅。深入了解虚假评论产生的内在动机,降低虚假评论发布的意愿是未来研究的重要方向。

[关键词] 电商平台; 虚假评论; 评论操纵; 文本分析; 半监督学习

[中图分类号] C931

[文献标识码] A

[DOI] 10.14071/j.1008-8105(2021)-1057

Identifying and Filtering Fake Reviews: Current Situation and Prospect

WANG Le¹ ZHANG Zi-qiong² CUI Xue-ying¹

(1. Xi'an Jiaotong University Xi'an 710061 China;

2. Harbin Institute of Technology Harbin 150001 China)

Abstract **[Purpose/Significance]** This paper develops a conceptual framework of fake review identification, and summarizes the limitations, controversies, and future directions of fake review identification. This paper has important theoretical significance to promote the related research of fake review identification, and also has some practical implications for e-commerce platforms to effectively identify and filter fake reviews. **[Design/Methodology]** This paper reviews 470 papers published in well-known Chinese and English journals. With thematic analysis, the theoretical foundations, research methods, and conclusions of these papers are summarized. Then the logical relations between these theories are described in detail and a direction for future research is provided. **[Conclusions/Findings]** The governance of fake reviews should rely on prevention supplemented by punishment. Understanding the intrinsic motivations of fake reviews and reducing the intention to publish fake reviews will be an important direction for future research.

Key words e-commerce platform; fake review; review manipulation; text analysis; semi-supervised learning

随着互联网的普及和物流行业的高速发展,电子商务得到了蓬勃发展。国家统计局数据显示,2020年全国社会消费品零售总额下降3.9%;而线上零售额增长10.9%,达到11.8万亿元^①。网上交易突

破了时间和空间的限制,但买卖双方的信息不对称会阻碍交易的进行。为了应对网上交易的信息不对称问题,各大电商平台建立了评价反馈机制,评价累积形成卖方的“声誉”^[1]。然而,虚假评论的出

[收稿日期] 2021-10-18

[基金项目] 国家自然科学基金重点项目(72131005);国家自然科学基金面上项目(72171189);国家自然科学基金青年项目(71901172)。

[作者简介] 王乐(1988-)男,西安交通大学经济与金融学院副教授、博士生导师;张紫琼(1982-)女,哈尔滨工业大学经济管理学院教授、博士生导师;崔雪莹(1997-)女,西安交通大学经济与金融学院博士研究生。

现严重影响了在线评论对产品质量的诊断力。据中国互联网协会发布的《中国网民权益保护调查报告》^[2]显示,72.7%的消费者遭遇过虚假评论的侵害,虚假评论严重影响了消费者的购买决策。Luca和Zervas^[3]指出,美国著名餐饮点评网站Yelp上16%的评论为虚假评论,Amazon上该比例更是高达42%。

虚假评论增加了消费者的信息搜索成本、商家的销售成本和平台的运营成本。鉴于虚假评论的危害,学术界和业界纷纷研究如何识别和过滤虚假评论。Jindal等^[4]首次提出运用有监督学习的方法对评论进行分类,以此区分虚假评论和真实评论。此后大量研究聚焦于虚假评论的特征并开发了相应的算法。例如,Ott等^[5]创建了第一个大规模、可公开获取的虚假评论研究数据集,并专门聚焦于伪装成真实评论的、迷惑性较高的虚假评论文本的识别。除评论文本特征外,Lim等^[6]提出通过行为特征识别虚假评论者,从而间接识别虚假评论。Mukherjee等^[7]进一步提出虚假评论群组的检测方法。上述研究从不同的方面对虚假评论的识别和过滤进行了系统的研究,但是仍然有一些关键的问题悬而未决。

首先,以往文献主要关注虚假评论的“事后”识别和过滤,对虚假评论的“事先”预防关注不足。学者们竞相开发新的识别算法,各大平台也都开发了相应的虚假评论过滤系统。例如,Amazon开发FakeSpot去识别过滤虚假评论。然而该平台虚假评论的占比由2019年的36%升至2020年的42%^[8]。可见,虚假评论的事后过滤并不能完全杜绝虚假评论。

其次,虽然现有研究从不同角度研究了虚假评论的识别算法,但这些算法对虚假评论的识别精度有待提高。Emerson等^[9]使用不同类型的算法和数据集,对各种虚假评论识别算法的速度和识别率进行了评估。结果发现,各种虚假评论识别算法的正确识别率均介于32%~43%。Yelp的Ghost Algorithm程序每年会过滤掉近11%的虚假评论,然而Luca和Zervas^[3]指出,Yelp上虚假评论的比例超过16%,这意味着约5%的虚假评论未被识别。

一、文献搜集

(一) 文献收集的范围界定

本文的文献主要来源于中国知网和Web of ScienceTM数据库。虚假评论的相关研究,起源于Jindal等^[4]。因此,本文将文献检索时间范围设定为2007年

(1月1日)~2021年(8月31日)。文献的搜索和筛选采用关键词检索和回溯法相结合的检索方法。

中文文献的收集首先以“虚假评论”为关键词在中国知网上进行初步检索,然后在梳理和阅读文献过程中不断扩充关键词库,最终确定的关键词包括:虚假评论、垃圾评论、欺诈虚假评论、异常评论、虚假评论检测、虚假评论识别、虚假评论者检测、虚假评论群组检测。接下来按照选定的关键词逐一搜索,并人工检查文章标题、关键词、摘要等,将检索结果中重复或不符合研究主题的文章剔除。

英文文献的搜集思路类似:首先以关键词review spam为起点在Web of ScienceTM数据库中进行初步检索,然后逐渐扩充关键词库,最终确定的关键词包括:review spam、fake review、opinion spam、deceptive review、review manipulation、spam detection、spam filtering、review spammer、spammer group。接下来,按照关键词库进一步检索。

同时,以Jindal等^[4]的文章为起点,运用回溯法对其被引文献以及被引的被引进行搜索,与根据关键词搜索到的文献进行比对。最终得到中文文献186篇,英文文献284篇(表1)。

(二) 文献脉络和趋势

国外对于虚假评论识别的研究起步较早。Jindal等^[4]对虚假评论加以界定,并提出运用机器学习模型对虚假评论进行分类,为该领域的学术研究提供了新的视角。此后相关研究进入快速增长阶段,且至今仍保持着持续增长的态势。近年来,虚假评论相关研究的领域越来越广泛,研究的视角也更加多样。国内对于虚假评论识别的研究起步相对较晚,但发展更为迅速,在2019年达到近几年峰值40篇(图1)。国内在研究虚假评论识别技术的同时,更能紧密结合中文语言表达、中国人行为方式以及电商平台的运作等特点,由此得到更好的研究成果。

中文文献主要集中在学位论文(52%)和计算机领域的学术期刊(21%),经济管理和情报学领域的学术期刊虽然有所涉及,但没有形成群聚效应,总体发文量不高。而英文文献分布相对比较均衡,在计算机(42%)、经济管理(24%)、数学(19%)、电信(11%)领域都有所关注。

二、虚假评论识别的研究框架

近年来虚假评论的文献数量不断增加,研究议题也不断丰富。为系统性总结现有研究的进展,解

表 1 文献主要分类

中文文献			英文文献		
领域	代表性期刊	数量	代表性期刊	数量	
学位论文	\	114	\	\	
计算机领域	《软件学报》	46	《INFORMS Journal on Computing》	121	
	《计算机学报》		《Communications of the ACM》		
	《计算机研究与发展》		《Journal of Artificial Intelligence Research》		
	《计算机科学与探索》		《Journal of Machine Learning Research》		
经济管理类	《管理世界》	10	《American Economic Review》	68	
	《管理科学学报》		《MIS Quarterly》		
	《管理科学》		《Information Systems Research》		
	《中国管理科学》		《Journal of Management Information Systems》		
	《南开管理评论》		《Information & Management》		
情报学	《情报学报》	9	《Information Processing & Management》	9	
	《图书情报工作》		《Information Sciences》		
	《信息资源管理学报》		《Journal of Informetrics》		
	《情报杂志》		《Information Technology for Development》		
电子信息技术	《信息技术与信息化》	4	《Journal of the Association for Information Science and Technology》	30	
	《电子学报》		《Journal of Intelligent Information Systems》		
	《数学的实践与认识》		《Knowledge and Information Systems》		
数学	《经济与管理》	3	《SIAM Journal on Computing》	56	
	《数据分析与知识发现》		《Mathematical Problems in Engineering》		
			《Operational Research》		

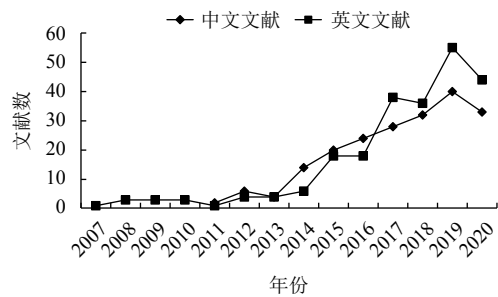


图 1 国内外发文趋势

决开篇提到的争议, 发掘未来的研究方向, 本文构建了一个虚假评论的理论分析框架。在管理与行为科学研究领域中常用的ABC (Antecedents-Behavior-Consequences, 前因-行为-结果) 分析范式的基础上, 本文整合了虚假评论的特征和识别算法, 以帮助我们更为系统地了解虚假评论的前因后

果。这一模型主要包含5个知识模块, 各个知识模块之间的逻辑联系见图2。

(一) 虚假评论的界定和分类

广义上来讲, 虚假评论通常是指与商品本身特征不相符的评论^[10], 包含“虚”——没有任何价值的无关评论和“假”——与事实不符的捏造评论^[11]。Jindal等^[4]依据评论者表达观点的真实性、评论内容与产品的相关性对虚假评论进行界定。一是不真实评论, 即内容相关但观点不真实的评论, 包括为自身利益而发布的正面不实评论和为损害对手声誉而发表的负面不实评论。二是无关评论, 包括仅针对品牌的评论以及不包含任何观点的评论, 如产品的广告宣传和功能介绍或来自对手的相似产品宣传等。由于这类评论很容易被阅读者识别, 所以文本对于消费者决策影响较小, 但是会对系统的评论评

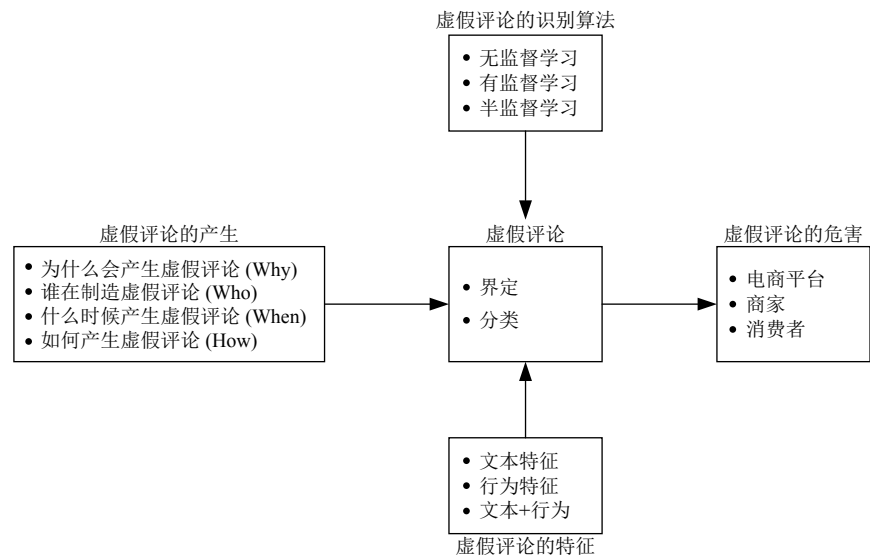


图 2 虚假评论相关研究框架

级产生影响。狭义的虚假评论通常指评论者出于某种目的发表的,与产品事实不符的评论^[12-13]。此类评论通常是商家或者消费者(发泄不满或索要补偿)为影响消费者决策而发表的虚假好评或差评。狭义的虚假评论对消费者的决策影响较大且不易被识别。因此,本文重点关注狭义虚假评论的识别和过滤。

根据评论是否通过欺骗消费者,进而对消费者决策造成影响,Ott等^[5]将虚假评论归为两类。一是破坏性评论,其评论本身难以误导消费者,但降低了评论信息质量,甚至影响整体评分和排名;二是欺骗性评论,指为欺骗消费者故意编造的过度夸赞或过分贬低的评论内容,其目的是引导消费者决策。还有研究将其进一步细分,从信息有用性角度将虚假评论划分为四类^[14]。一是推销、诋毁性评论,即为推销自身产品或诋毁其他品牌而发布的不真实评论,其目的是影响消费者观点和行为决策以实现自身利益;二是干扰性评论,包括品牌宣传、产品介绍、广告链接等参考价值很低甚至没有任何参考价值的评论信息,主要用于干扰消费者获取有效信息或者干扰系统的识别;三是无意义评论,其内容未传递任何有用信息,如敷衍性的评论或单纯的情绪宣泄等;四是系统评论,即具有默认好评机制的平台,在消费者一段时间没有给出评论后自动生成好评,该评论对用户来说没有任何参考价值。此外,还有学者根据发布虚假评论的类别将虚假评论者划分为两类^[15]。第一类是随机虚假评论者,其发表的内容没有任何参考价值,虽然可能不是出于恶意的目的,但降低了评论信息质量;第二类是恶意虚假评论者,出于自身目的故意制造与产品事实不符的评论,来诋毁商家声誉、误导消费者决策。

(二) 虚假评论的产生

要识别和过滤虚假评论,首先要了解虚假评论如何产生。以往文献从以下几个方面讨论了虚假评论的起源。

1. 为什么会产生虚假评论(Why)

从宏观上来讲,注入虚假的评分能够提高整体评分水平,从而提升商品的排名,使得商家获得更多的网络流量和点击率^[16-17]。一般来说,消费者在搜索目标产品时,平台会根据产品的评分排名或者销量排名依次展示商品信息,而排在商品首页的商家占据了80%~90%的网络流量^[18]。大量的实证研究证明网络流量和产品销量之间存在正相关关系^[19-21],因而评分和排名对于商家来说至关重要。因此,商家选择采取不正当的竞争方式,通过增加自身的高

分好评或竞争对手的低分差评,来提高产品整体得分和排名,增加产品曝光度,进而提升销量。

从微观上来讲,虚假评论的内容可以通过影响产品口碑、商家声誉,进而对消费者决策产生影响。评论评分是以往消费者对产品总体质量的一个数字化的评价,而评论的文本内容则包含了消费者的主观感受、情感倾向、使用心得、产品优缺点、商家服务等信息。相比于评论评分,评论内容表达的信息量更加丰富,所以大部分消费者在挑选产品时,会参考以往消费者对产品的评价^[22-23]。因此,商家试图利用大量正面评论夸大产品质量,建立正面口碑,或发布负面评论诋毁竞争对手产品,形成负面口碑,从而引导甚至转变消费者的购买意愿。

总的来看,无论是出于提高整体评分和排名以吸引更多流量,还是建立虚假口碑、塑造良好的声誉,发布虚假评论的根本原因都是自身利益。因此,利益驱动是产生虚假评论最主要的动机。

2. 谁在制造虚假评论(Who)

商家是产生虚假评论最主要的来源,商家为牟取更多的经济利益,试图通过发布虚假评分以提升自身竞争力,并企图通过虚假的评论内容误导消费者做出利于商家的决策。大量的研究表明,虚假评论会影响产品绩效^[24-25]。如Petrescu等^[26]对亚马逊网站上某剃须产品的评论进行定量分析,发现在商家进行虚假评论的激励干预之后,产品评论数量和产品销量显著提升。竞争对手发布虚假差评来打击他人的手段更为卑劣,但负面评论能更有效地打击对手,产生的影响也更加恶劣^[27],所以仍是不少商家的惯用手段。Mayzlin等^[28]发现,一定地理范围内具有竞争对手的酒店相对来说更有可能收到虚假差评,且随着一定地理距离之内竞争对手的增加,酒店发布虚假评论的频率也随之增加。虚假评论除了来源于恶性竞争的商家外,消费者也可能出于发泄不满^[29]、索要奖励^[30]、获取小额返现^[31]等目的发布虚假评论。最后,常常被忽略的一方面是来源于平台的默认好评。具有默认好评机制的平台在消费者一段时间没有给出评价后,会自动生成好评,而此类好评可能并未表达消费者真实意愿,但当前学术界对于默认好评是否属于虚假评论尚未有统一的定论。

3. 什么时候产生虚假评论(When)

持续性地增加好评,会导致评论操控成本过高且被识破的风险大大增加。因此,商家会侧重在某些时刻发布虚假评论。一是在商家自身评分下降或竞争对手评分上升时。Luca和Zervas^[3]发现,当产品评分有下降趋势时,商家发布虚假评论的频率增

加。二是商家收到差评时, 此时商家为降低差评的影响, 会选择注入大量好评使负面评价迅速被“淹没”。三是产品处于推广期时^[32], 商家往往在产品推出初期持续地增加好评。一方面好评数量增加可以提高产品评分和排名, 引来更多的流量; 另一方面, 面对新产品时, 消费者试图通过参考评论内容获取更多产品信息, 以降低不确定性。Cui等^[33]研究了在线评论对新兴的电子产品和视频游戏销售的影响, 结果表明在线评论对新产品发布初期的销售有显著影响, 并且这种影响随着时间的推移而减弱。Zhang等^[34]指出, 在产品的成长期增加大量的正面评论可以使产品后续获得更多的正面评论。四是当周围竞争环境变化时, 如面临同类竞争对手增加的横向挤压和替代商家不断出现的纵向挤压时。Liu等^[35]指出, 由于营业范围的部分重合, 周围蛋糕店数量越多, 星巴克操控虚假评论的频率越高。Lee等^[36]研究发现竞争环境对虚假评论发布强度具有调节效应, 当行业竞争水平提高时, 虚假评论发布的频率随之增加。

4. 如何产生虚假评论 (How)

商家可以选择不同的方法发布虚假评论, 主要包括增加自身好评、删除自身差评、给竞争对手注入差评三种方式。首先, 向自身增加虚假好评的方式最为普遍且形式多样, 如商家刷好评、激励好评^[32]、发放免费样本^[37]、进行好评返现^[38]、差评威胁^[31], 甚至要求消费者修改不利评价来制造虚假评论等^[39]。此外, 一些商家还通过删除差评达到自身目的。Zhuang等^[32]研究发现适量的增加好评确实会影响消费者购买意愿, 而达到一定程度后则面临着被消费者怀疑或者识破的风险, 从而产生负效应。删除差评相对较为隐蔽, 不易引起怀疑, 但加剧了商家与消费者之间的信息不对称。最后, 向竞争对手注入差评也是较为常见的方法之一。Mayzlin等^[28]发现, 一定地理范围内竞争对手的数量与收到虚假差评的数量相关, 且虚假的负面评论可能会对商家产生更严重的消极影响^[40]。Lappas等^[17]发现, 向竞争对手注入差评对产品排名的提升比增加自身相同数量的好评对产品排名提升的效率40%。

此外, 商家常用的发布虚假评论的渠道主要有两种。一是自己制造虚假评论, 如注册新的账号伪装成真实用户发布虚假评论, 或者使用各种激励手段促使消费者发布虚假好评; 二是雇佣专业的造假团队或者职业差评师, 这种方式相对来说成本更高, 其手段更加卑劣, 造成的负面影响也更为恶劣。

(三) 虚假评论的危害

对电商平台来说, 虚假评论的存在降低了评论的信息质量。消费者在浏览到虚假评论后, 会降低对平台和在线评论系统的信任, 进而损害了平台的声誉。而电商平台为了降低虚假评论带来的恶劣影响, 尽可能维持商家间的公平竞争关系、维护消费者权益、保护平台声誉, 势必要采取各种办法识别并过滤虚假评论。如亚马逊和Yelp都已经部署了各自的虚假评论识别系统, 京东日前也公开申请了“虚假评论检测方法”专利, 从而虚假评论大大增加了平台的运营成本。

对商家来说, 无论是自行发布虚假评论还是雇佣专门的造假团队, 都会增加自身的销售成本。同时, 发布虚假评论的商家还会面临被识破的风险, 从而损害商誉, 甚至面临平台和法律的惩罚。而对于被诋毁的商家, 虚假的低分差评会损害产品的平均评分和排名以及商家的声誉, 最终导致销量的下降。

对消费者来说, 虚假评论的存在降低了评论有用性。为规避虚假评论的误导, 消费者需要花费更多的时间和精力辨别评论真实性。

(四) 虚假评论的特征

尽管虚假评论具有较高迷惑性, 不易被识别, 但其评论内容和发布者行为的特征仍然会露出破绽, 抓住这些特征对于辨别虚假评论至关重要。现有研究不断挖掘有效的识别特征并将各种特征组合运用到模型中, 实现对虚假评论的识别。根据特征选取角度的不同, 主要包括评论文本特征和评论者行为特征。具体特征及描述见表2。由此, 可将虚假评论的识别路径大致分为三种。第一种主要依据评论的文本特征, 从内容本身出发识别虚假评论。第二种依据评论者行为特征, 对虚假评论人或者群组进行识别。第三种将文本与行为相结合对虚假评论进行识别。

1. 评论文本特征

尽管虚假评论的发布者试图尽可能地模仿真实的体验, 但是在一些文字细节上仍然会有一些破绽, 如文本相似性和冗余程度高, 单一化的写作风格等^[40-41]。因此, 越来越多的文献利用评论文本特征来识别虚假评论。

早期的研究主要运用文本相似性、重复性等特征识别虚假评论^[4]。但随着虚假评论的隐蔽性越来越高, 学者们开始寻找更多的文本特征以提高虚假评论识别的精度。如在文字层面使用文本长度、词汇的复杂度^[42]、不同词性的数量^[42]、客观性单词比

表 2 基于评论文本和评论者行为特征识别体系

来源	特征	属性	描述
文本	重复度	相似性	评论文本与该商品其他文本的重复程度
	相似度	相似性	评论文本与该商品其他文本的相似程度
	文本长度	复杂性	评论文本中文字个数
	字符冗余性	冗余性	评论文本中字符频率
	数字冗余性	冗余性	评论文本中数字频率
	第一人称比重	词性 词频	评论文本中第一人称词汇占比
	不同词性比重	词性 词频	评论文本中动词、名词、代词等词性占比
	不同符号比重	符号	评论文本中感叹号、表情符号、特殊符号等占比
	文本复杂度	复杂性	评论文本使用词汇的复杂性
	互动性	内容特征	评论获得的反馈数、有用反馈数、反馈百分比
	一致度	一致性	评论初评和追评的一致程度
	描述相符性	一致性	评论文本与产品特征描述的符合程度
	主客观性	内容特征	评论文本中包含的主客观内容比例
	品牌提及率	内容特征	评论文本中品牌相关内容比例
	情感极性	情感分析	评论文本中的情感强度
	情感词依存	情感分析	情感词之间的依存关系
	用户信誉	个人信息	评论者所发表的评论被其他消费者采纳的数量、有用性投票等
	个人信息	个人信息	用户的注册信息以及用户相关信息,包括用户名ID号,是否实名,注册时间,主页链接,自我评价,朋友数和粉丝数等
评论者	历史文本相似度	评论行为	评论者所发表过评论文本的相似度
	评论频繁度	评论行为	单位时间发布评论数
	突发性评论比例	评论行为	短时间内评论数量占评论总数的比重
	评论极端数	评论行为	历史发布的极端评论数量(1分或5分)
	历史评分差异	评论行为	历史发布评分的方差
	针对某一产品	评论行为	评论者对于某一产品重复打分次数、每次打分差异、内容相似度
	针对一类产品	评论行为	评论者在某一时段内对某一类产品打分次数、极端评分比例
	共同评论频率	群组行为	多人群组共同评论同一产品的频率
	共同评论相似性	群组行为	多人群组共同评论产品内容的相似程度
	共同评论同步性	群组行为	多人群组共同评论发布时间的同步程度

例和主观性单词比例、感叹句数量^[43]等特征。Jindal等^[4]的虚假评论识别模型中,在文本特征方面重点考虑了评论和产品特征描述的符合程度、评论中的品牌提及率、评论文本的主客观性以及评论获得的反馈情况等。Zhao等^[42]总结了虚假评论的六个文本特征,包括评论文本长度和文本复杂程度等。

此外,评论文本的情感是识别虚假评论的语义特征中的重要组成部分^[44-45]。Deng和Chen^[46]认为绝对正面或绝对负面的评论极有可能是虚假评论。任亚峰等^[47]指出,可以利用情感极性和第一人称代词词频来区分虚假评论和真实评论:情感极性越强,第一人称词频越少,该评论为虚假评论的概率越大。文本中不同词性数量一定程度上也可反应心理认知过程,从而帮助识别虚假评论^[5]。Gregorio等^[48]构建了负向情感词库,以此对虚假评论文本进行分析,取得了比人工识别更好的效果。情感词间的关系也能够用于识别虚假评论,Evans等^[49]对评论中情感词的依存关系进行了研究,运用量化情感算法进行分析能够有效地识别出虚假评论。因此,引入情感分析在一定程度上能够提升识别的准确率和原有模型的效果^[50]。

2. 评论者行为特征

虚假评论往往模仿真实评论的写作方式,给依靠文本的虚假评论识别算法造成了极大困难。张文

等^[51]认为,评论文本所能提供的信息是有限的,而虚假评论者的行为可以提供较多的信息并且检测更为容易。因此,越来越多的研究者通过分析评论者的行为特征来识别恶意账户,以此锁定虚假评论。已有研究表明,虚假评论者的行为不同于真实用户的行为。例如,邓胜利和汪奋奋^[52]指出,虚假评论者会赞扬或贬低某一特定品牌的产品。Mukherjee等^[53]发现,虚假账户往往会在短时间内写出大量的评论。所以评论发布时间越集中,突发性评论占比越大,账户为虚假评论者的可能性越高。此外,虚假评论者与真实评论者比较而言,多给出极端评价^[54-55],即历史评分在极端评论处的分布较为集中^[4],且常常对特定产品重复评分^[25],或对不同产品的评论内容重复^[56]。

随着电子商务的发展,虚假评论发布者的规模也在不断扩大,甚至出现多人协同形式的虚假评论群组。群组凭借大规模地发布虚假评论,从而引导评论导向,因此更具破坏性^[7]。由此对于虚假评论群组识别的研究应运而生。对于虚假评论群组的识别同样可依据其行为特征。Mukherjee等^[7]首次以群组为目标进行识别,指出虚假评论群组通常具有各成员共同评论同一产品的特征。Xu等^[57]选取了八个群组特征识别虚假评论,涉及内容相似性与偏差、群组规模和数量、时间上的评论差异等方面的

指标。

3. 评论文本与评论者行为相结合

单独使用评论文本特征或评论者行为特征可能难以应对复杂的情况, 为提高识别的准确度, 越来越多的学者将二者相结合, 建立更为全面的识别方法。Mukherjee等^[53]使用Yelp数据集, 验证了只选取评论文本特征情况下与加入行为特征情况下识别准确度的差异, 结果显示, 加入评论者行为特征后准确率得到显著提升。Li等^[43]对选取的文本及行为特征采用朴素贝叶斯方法和联合训练机制对虚假评论进行识别, 使用Epinions网站的评论数据进行检测得到了61.3%的F1值。杨超等^[58]将评论文本特征与评论者行为特征融合, 采取卷积神经网络对所选特征数据进行处理, 通过实验验证了该方法的有效性和对虚假评论较高的识别率。

(五) 虚假评论的识别算法

对于虚假评论识别早期的研究, 由于评论形式较为简单, 因此研究识别的对象主要针对重复性评论和广告性评论。而随着互联网的蓬勃发展和电子商务的兴起, 虚假评论数量激增, 形式也更为多样, 识别的复杂性随之增加。因此, 更为高效、智能的机器学习技术开始广泛应用到虚假评论的识别, 下面从无监督学习、有监督学习和半监督学习三方面对虚假评论识别算法进行归类。

1. 无监督学习

无监督学习只使用未标注真假的数据样本进行学习, 通过学习评论数据之间的相似性, 将数据分成两个组: 虚假评论和非虚假评论, 然后根据数据间的相似性和差异性对数据的分类进行预测。基于图结构的方法是无监督学习中的典型方法, Mukherjee等^[7]率先使用无监督学习算法, 采用频繁项挖掘方法来筛选可疑的候选组, 然后通过建立模型进一步锁定虚假评论群组。宋海霞等^[59]利用F统计量对K均值算法进行改进, 实现评论数据的自适应聚类这一无监督学习, 然后通过寻找异常簇的方法来识别虚假评论, 并验证了该方法的有效性。

2. 有监督学习

监督学习利用已标注真假的评论数据训练分类器, 进而对评论进行分类。Ott等^[5]通过在数据集上统计真实评论和虚假评论的惯用词词频情况, 构建SVM分类器和朴素贝叶斯分类器进行识别。邓莎莎等^[41]在由评论者分别撰写的真实评论和虚假评论语料上, 建立支持向量机、朴素贝叶斯、决策树等有监督学习算法, 检验词性特征在虚假评论识别中的有效性。陈燕方^[60]分别从评论文本特征、评论者

行为特征、商家特征三个维度选取了十个指标, 在此基础上通过实验证明, 采用DDAG-SVM多分类支持向量机能够较好地兼顾评论识别的效率和准确率。

3. 半监督学习

由于人类无法通过先验知识有效识别虚假评论, 手工标注的数据集必定存在一定数量的误例。因此, 简单使用传统的有监督学习算法识别虚假评论并不合理, 但可以通过启发式的规则获取少量真实评论和大量的未标注评论。基于真实评论集和未标注评论集, 可建立一个半监督学习的分类器用于识别虚假评论。半监督学习是利用大量未标注真假的评论和少量有标注真假的评论训练分类器。协同训练算法作为一种典型的半监督学习方式, 可以结合评论文本特征和评论者行为特征训练分类器, 大大提高了虚假评论的识别效率^[51,61]。任亚峰等^[62]先使用朴素贝叶斯有监督学习和人工标注的标签来评价不同特征建模性能, 选出最好的特征组合, 然后设计Co-training和Tri-training两种半监督学习算法充分利用大量未标注文本, 以提高识别性能。随后任亚峰等^[63]还提出一种创新的PU学习框架来识别虚假评论, 使用两种主流多核学习算法SILP和LPSOLVE来训练分类器, 实验证明所提方法可以有效用于虚假评论的识别。

目前虚假评论的识别算法以全监督框架下的分类方法为主, 但全监督学习需要大量标注真假的评论作为训练集, 标注数据集的缺少及人工标注耗时费力等问题为全监督框架下的算法带来极大局限性。一些学者尝试使用无监督学习方法, 该方法解决了标注评论集缺失的问题, 但却存在识别精度普遍偏低的情况。而半监督学习很好地克服了全监督学习与无监督学习的缺点。

三、现有研究局限

(一) 以往文献“重治轻防”

以往文献主要关注虚假评论的“事后”识别和过滤, 未能从根本上抑制虚假评论产生的内在动机。本文系统性梳理了虚假评论的相关研究发现: 93%的中文论文和89%的英文论文主要关注虚假评论的特征、分类和识别(图3)。少量研究从法律法规和平台制度的角度研究了虚假评论的治理(中文2篇、英文4篇), 但仅限于定性理论分析。然而仅靠法律约束和平台监管并不能从根本上有效抑制虚假评论的产生。朱星圳等^[64]研究发现, 加大平台

管控力度短期内能够约束商家行为,但长期来看,该约束效果不明显。因此,深入研究虚假评论产生的内在动机,减少虚假评论的产生是未来研究的重点。

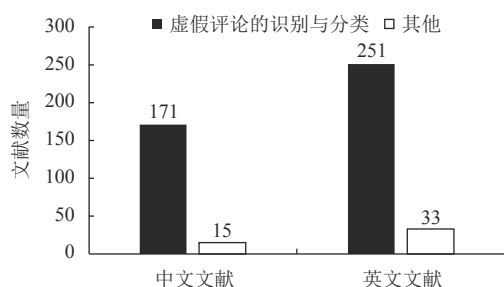


图3 虚假评论相关研究数量

(二) 虚假评论识别算法训练集不充分不平衡,致使算法精度不高

采用监督学习训练分类器的基础是具备已标注真假的数据集,而虚假评论识别的研究中所使用的数据集主要分为两种:一种是真实数据集,另一种是人工构造的数据集。真实数据集的获取方式有如下两种:一是直接使用亚马逊、Yelp等平台的公开数据集,二是研究人员通过网络爬虫技术搜集的数据。两种方式收集到的数据集都需要对数据进行人工标注,而人工标注工作量巨大,且标注真假与真实情况相比存在一定的误差,由此影响识别的精度。也有一部分研究采用人工数据集^[5],即人工制造真实评论和虚假评论以此训练分类器。人工构造的数据集虽然避免了标注数据导致的偏差,但由于人工模拟的评论与现实中的虚假评论情况仍有差异,由此训练出来的分类器实践效果不甚理想^[53]。另外,虚假评论的相关算法面临好评和差评数量不平衡问题。通常情况下,产品的好评数量远大于差评,造成了算法训练样本的严重不平衡。从而训练所得的分类器往往更擅长识别好评的真假,而对差评是真是假的诊断力不强,大大削弱了虚假评论识别的精度。

(三) 以往虚假评论的识别算法忽略了评论与环境的匹配性

以往虚假评论识别算法重点关注评论文本特征、评论者行为特征或二者的交互对虚假评论识别的影响。这些研究忽略了评论和评论环境的匹配性对虚假评论识别的影响。评论和评论环境的匹配性有三个方面:一是评论评分与产品平均评分的匹配性;二是评论情感与其他评论情感的匹配性;三是评论评分与产品生命周期的匹配性。Ansari等^[65]指出,虚假评论往往与产品的平均评分差异较大,因

此可以采用评论评分和产品平均评分的差异来识别虚假评论。Zhao等^[42]指出,虚假评论的情感往往较为极端并且与真实评论差异较大。因此,可以采用评论情感和其他评论情感的差异来检测虚假评论。此外,研究表明产品推出的早期发布虚假评论对产品销量的影响更大,商家常常选择在产品推出初期发布虚假评论^[33]。产品推出初期的好评更有可能是商家自身产生的虚假好评,而差评更有可能是竞争对手产生的恶意差评。因此,评论和产品生命周期的匹配性也是虚假评论鉴别的重要变量之一。

(四) 新用户缺失历史数据,传统的行为特征难以有效识别

已有研究表明,评论者的行为特征是识别虚假评论者的重要依据,加入评论者行为特征能够有效提升虚假评论识别的准确率^[6]。即使恶意评价者的行为特征对虚假评论的识别具有良好效果,但观察恶意评价者的行为特征需要一段时间的积累。如评论频繁度、突发性评论比例、历史评分差异等等,都需要根据评论者行为的历史数据获得。然而,虚假评论发布者为了降低被识破的风险,会定期更换新账号伪装成真实用户发布虚假评论。这些新用户无历史数据可观测,导致现有研究所选取的行为特征不适用于检测新注册的用户。

四、未来研究展望

(一) 虚假评论的治理需要以“防”为主,以“治”为辅

以往文献主要聚焦于虚假评论识别算法的精度和速度。虽然虚假评价识别系统可以帮助电商平台快速识别过滤虚假评论,但却无法抑制虚假评论产生的内在动机。每天仍有大量的虚假评论产生。此外,虚假评论的发布者也会根据虚假评论识别算法发布更为隐蔽的虚假评论,从而形成恶性循环。因此,虚假评论的治理需要以“防”为主,以“治”为辅。深入了解虚假评论产生的内在动机,降低虚假评论发布者的意愿是未来研究的重要方向。另外,虚假评论发布者通过大量的虚假评论来提高产品的排名,带来更多的网络流量。因此,电商平台可以升级现有的产品排名算法,以提升产品排名系统对虚假评论攻击的稳健性。最后,预防虚假评论的产生,还需要法律法规、平台政策、商家自律的共同努力。对虚假评论的治理需惩前毖后并重,威慑理论认为,惩罚的威慑力取决于感知惩罚确定性和感知惩罚严重性^[66],所以防治虚假评论还需不断

完善相关法律、加大惩罚力度。平台除制定相关制度外, 还可以开发有效的反虚假评论机制, 提高虚假评论制造的难度和成本。

(二) 采用半监督学习解决训练标注样本不足的问题

训练集样本标注不足问题是阻碍虚假评论识别领域快速发展的重要原因之一。现有研究大多使用全监督学习^[5,41,60], 不仅需要大规模已标注的数据集, 而且通常存在标记误差问题。为彻底避开样本标注问题, 还有研究采用无监督学习^[7,60], 自动处理样本寻找隐含的规律, 从而实现样本的分类, 但却难以取得较好的识别精度。而半监督学习介于二者之间, 可以在少量标注真假的样本基础上, 充分利用大量未标注样本提高识别性能, 很好地兼顾了标注样本不足和提高识别精度的问题。根据不同的情况和数据特征, 还可以选择最合适的半监督学习方法。自训练(Self-training)可视为标准的半监督学习模型, 它可以从未标记的数据中根据一定的置信水平标注数据, 从而自动扩大标记集, 且具有易于和其他分类算法相结合的优点^[67]。协同训练(Co-training)是自训练的延伸, 它具有两视图识别的优势, 研究证明某些情况下, 这种两视图的方式有助于提升分类器的性能^[43]。有学者进一步提出多视图的方法(Multi-view Learning), 可用于多种资源、线索的结合。生成式方法(Generative Methods)通过统计学方法(最大似然估计)处理缺失信息, 该方法在标注数据极少的情况下, 具有相对较好的识别性能。此外, 半监督学习还可以与全监督学习结合使用。任亚峰^[62]的识别算法首先采用全监督学习筛选最合适的特征组合, 然后配合半监督学习扩充标注样本, 从而二者优势互补, 以提高算法的有效性。因此, 采用半监督学习是解决训练样本标注不足问题的重要途径, 值得未来研究深入探索。

(三) 虚假评论识别的多领域结合

即使现有算法已经能够在一定程度上实现对虚假评论的有效识别, 但与此同时商家也在根据算法做出相应的策略改变来试图躲避识别, 并不断寻找更隐蔽的方式。因此, 对于虚假评论识别的方法不应局限于计算机领域。一方面虚假评论的识别需要不断了解虚假评论的新特征, 提高算法精度, 升级现有识别技术, 以应对虚假评论形式和策略的变化。另一方面, 虚假评论的识别需要多领域的交叉融合。

首先, 考虑与心理学相融合。一般来说, 虚假

评论者和正常评论者相比, 评论的心理过程和外在表现势必存在差异。有学者从心理学角度出发, 发现虚假评论的情感倾向更加极端^[46]。此外, 不同的用词、表情符号的使用、评论的自信度等特征都能反应评论者的心理、情感和动机^[5,47,68]。还有研究应用LIWC文本分析工具^[62]和IBM Watson语气分析工具^[68], 利用心理语言学对评论文本深入挖掘, 从而探索评论中隐含的评论者的情感过程、认知过程、社会过程等。因此, 虚假评论的识别与心理学的结合有助于升级识别算法, 提升识别的准确率。

其次, 与行为学的融合同样重要。研究发现, 正常评论者发表评论的时间、对象一般是随机的, 而虚假评论者为达到特定目的, 往往发布评论的时间、数量、评价对象存在一定规律^[69]。如虚假评论者通常会在特定时间段内高频率地发布虚假评论, 且大多聚集于某个品牌或某类产品^[70], 从而快速对产品评分或网络口碑产生影响。另外, 虚假评论者常在产品发布早期较为活跃, 由于早期的评论信息影响范围更大、影响持续时间更长, 所以成为虚假评论的高发期。因此, 将虚假评论识别与行为学相融合, 有助于初步缩小识别与过滤范围, 提高识别效率, 还能够通过寻找有标识的行为特征, 帮助提高识别的准确率。

(四) 探索有效特征解决新用户历史数据缺失问题

现有研究所选取的行为特征需要根据历史行为观察、计算得到可用指标, 但是对于新注册账号的虚假评论者的有效识别是现有研究的缺失。未来研究需充分考虑新用户的识别问题, 寻找适用于新用户的行为特征弥补上述空白。如账号注册至首次评论的时间间隔、注册后短期内的评论次数等。一般来说, 新用户由于操作不熟悉或尚未建立平台信任感等, 短时间内处于观望状态, 且消费者通常从产品下单购买到对产品产生使用感受具有一定的时间间隔。因此, 新注册的用户较少会立即且频繁地发表评论。而虚假评论者注册新用户, 为达到使用效用最大化, 会频繁使用此账户发布虚假评论。因此, 首先可以参考新用户从注册至首次评论发布的时间间隔, 若此间隔时间很短, 则是虚假评论者的可能性较大。其次, 还可以考虑新账户短时间内发布的评论数, 若新注册账号短时间内频繁操作则更有虚假评论者的嫌疑。最后, 还可以着重关注新用户评论的极端程度及其与大众观点的偏差, 寻找可疑用户并根据后续行为排查, 尽可能降低新用户识别不足带来的负面影响。

五、总结

虚假评论是电商平台声誉机制的必然产物,也是电子商务健康发展的顽疾。随着虚假评论数量的快速增长,虚假评论的识别和过滤成了学术界和业界关注的热点。虚假评论的识别算法广泛应用到各大电商平台,电商平台的制度和相关法律法规也在不断完善。然而,评论操控不降反升,成为商家恶性竞争的惯用手段。本文首先对虚假评论的相关概念进行界定,然后对虚假评论的研究现状进行归纳分析,总结了现有研究的进展和争议,指出了未来研究方向,并对现有研究的争议提出了具体解决方案。本文的研究结论对推动虚假评论识别的相关研究有重要理论意义,对电商平台的健康发展有一定实践启示。

注释

- ① <https://data.stats.gov.cn/easyquery.htm?cn=A01>

参考文献

- [1] 刘伶俐. 评论操纵对声誉的影响[D]. 大连: 东北财经大学, 2019.
- [2] 2015年中国网民权益保护调查分析[J]. 互联网天地, 2015(9): 77-86.
- [3] LUCA M, ZERVAS G. Fake it till you make it: reputation, competition, and Yelp review fraud[J]. *Management Science*, 2016, 62(12): 3412-3427.
- [4] JINDAL N, LIU B. Analyzing and detecting review spam [C]. Omaha: IEEE International Conference on Data Mining. IEEE, 2007: 547-552.
- [5] OTT M, CHOI Y, CARDIE C, et al. Finding deceptive opinion spam by any stretch of the imagination[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland: ACL, 2011: 309-319.
- [6] LIM E P, NGUYEN V A, JINDAL N, et al. Detecting product review spammers using rating behaviors[C]. Toronto: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. ACM, 2010: 939-948.
- [7] MUKHERJEE A, LIU B, Glance N. Spotting fake reviewer groups in consumer reviews[C]. Lyon: The 21st International Conference on World Wide Web, WWW, 2012: 191-200.
- [8] ISABELLE. Amazon fake reviews reach holiday season levels during pandemic, Bloomberg News [EB/OL]. [2020-10-19]. <https://www.bloomberg.com/news/articles/2020-10-19/amazon-fake-reviews-reach-holiday-season-levels-during-pandemic>.
- [9] EMERSON F C, RENATO M S, TIAGO A. Towards automatic filtering of fake reviews[J]. *Neurocomputing*, 2018, 309(October): 106-116.
- [10] 武雅利, 徐勇, 焦梦蕾, 等. 基于在线交易的虚假评论与网络水军研究综述[J]. 现代计算机, 2019, 21(1): 37-40.
- [11] 孟美任, 丁晟春. 虚假商品评论信息发布者行为动机分析[J]. 情报科学, 2013, 31(10): 100-104.
- [12] 刘璇. 虚假评论对消费者购买意愿的影响研究[D]. 南京: 南京大学, 2015.
- [13] HUSSAIN N, MIRZA H T, RASOOL G, et al. Spam review detection techniques: a systematic literature review[J]. *Applied Sciences*, 2019, 9(5): 987.
- [14] 李霄, 丁晟春. 垃圾商品评论信息的识别研究[J]. 现代图书情报技术, 2013, 29(1): 63-68.
- [15] 董宇蔚. 电子商务中的评论挖掘及应用研究[D]. 成都: 电子科技大学, 2014.
- [16] FENG N, SU Z, LI D, et al. Effects of review spam in a firm-initiated virtual brand community: evidence from smart phone customers[J]. *Information & Management*, 2018, 55(8): 1061-1070.
- [17] LAPPAS T, SABNIS G, VALKANAS G. The impact of fake reviews on online visibility: a vulnerability assessment of the hotel industry[J]. *Information Systems Research*, 2016, 27(4): 940-961.
- [18] KAYE L. 95 percent of web traffic goes to sites on Page 1 of Google SERPs [EB/OL]. (2013-6-21). <https://www.brafton.com/news/95-percent-of-web-traffic-goes-to-sites-on-page-1-of-google-serps-study>.
- [19] DOLEGA L, ROWE F, BRANAGAN E. Going digital? The impact of social media marketing on retail website traffic, orders and sales[J]. *Journal of Retailing and Consumer Services*, 2021, 60(5): 102501.
- [20] SUN H, CHEN J, FAN M. Effect of live chat on traffic-to-sales conversion: evidence from an online marketplace[J]. *Production and Operations Management*, 2020, 42(1): 1-18.
- [21] LUO X, ZHANG J, DUAN W. Social media and firm equity value[J]. *Information Systems Research*, 2013, 24(1): 146-163.
- [22] DAI H, CHAN C, MOGILNER C. People rely less on consumer reviews for experiential than material purchases[J]. *Journal of Consumer Research*, 2020, 46(6): 1052-1075.
- [23] LIU X, LEE D, SRINIVASAN K. Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning[J]. *Journal of Marketing Research*, 2019, 56(6): 918-943.
- [24] FLOSTRAND A, PITT L, KIETZMANN J. Fake news and brand management: a Delphi study of impact, vulnerability and mitigation[J]. *Journal of Product and Brand Management*, 2020, 29(2): 246-254.
- [25] ANANTHAKRISHNAN U M, LI B, SMITH M D. A tangled web: should online review portals display fraudulent

- reviews[J]. *Information Systems Research*, 2020, 31(3): 950-971.
- [26] PETRESCU M, O'LEARY K, GOLDRING D, et al. Incentivized reviews: promising the moon for a few stars[J]. *Journal of Retailing and Consumer Services*, 2018, 41(2): 288-295.
- [27] 郑春东, 韩晴, 王寒. 网络水军言论如何左右你的购买意愿[J]. *南开管理评论*, 2015, 18(1): 89-97.
- [28] MAYZLIN D, DOVER Y, CHEVALIER J. Promotional reviews: an empirical investigation of online review manipulation[J]. *American Economic Review*, 2014, 104(8): 2421-2455.
- [29] ANDERSON E T, SIMESTER D I. Reviews without a purchase: low ratings, loyal customers, and deception[J]. *Journal of Marketing Research*, 2014, 51(3): 249-269.
- [30] THAKUR R, HALE D, SUMMEY J H. What motivates consumers to partake in cyber shilling[J]. *Journal of Marketing Theory and Practice*, 2018, 26(1-2): 181-195.
- [31] 韩菁, 蔡寻, 滕新玉. 价值与风险感知对好评返现行为影响的演化分析[J]. *中国管理科学*, 2019, 27(9): 205-216.
- [32] ZHUANG M Z, CUI G, PENG L. Manufactured opinions: the effect of manipulating online product reviews[J]. *Journal of Business Research*, 2018, 87(1): 24-35.
- [33] CUI G, LUI H K, GUO X N. The effect of online consumer reviews on new product sales[J]. *International Journal of Electronic Commerce*, 2012, 17(1): 39-57.
- [34] ZHANG L, WANG S F, LIN Z Z, et al. Online ballot stuffing: influence of self-boosting manipulation on rating dynamics in online rating systems[J]. *Telematics and Informatics*, 2019, 38(2): 1-12.
- [35] LIU A X, STEENKAMP J, ZHANG J. Agglomeration as a driver of the volume of electronic word of mouth in the restaurant industry[J]. *Journal of Marketing Research*, 2018, 55(4): 507-523.
- [36] LEE S Y, QIU L, WHINSTON A. Sentiment manipulation in online platforms: an analysis of movie tweets[J]. *Production and Operations Management*, 2018, 27(3): 393-416.
- [37] LIN Z J, ZHANG Y, TAN Y. An empirical study of free product sampling and rating bias[J]. *Information Systems Research*, 2019, 30(1): 260-275.
- [38] 徐兵, 张阳. 基于好评返现和差评偏好的体验型产品两阶段定价决策研究[J]. *管理工程学报*, 2020, 34(2): 21-29.
- [39] 王宁, 宋嘉莹, 杨学成. C2C电商平台中在线评论偏离真实性的诱因及应对策略[J]. *软科学*, 2017, 31(4): 100-103.
- [40] HU N, BOSE I, KOH N S, et al. Manipulation of online reviews: an analysis of ratings, readability, and sentiments[J]. *Decision Support Systems*, 2012, 52(3): 674-684.
- [41] 邓莎莎, 张朋柱, 张晓燕, 等. 基于欺骗语言线索的虚假评论识别[J]. *系统管理学报*, 2014, 23(2): 263-270.
- [42] ZHAO J, HONG W. Detection of fake reviews based on emotional orientation and logistic regression[J]. *CAAI Transactions on Intelligent Systems*, 2016, 11(3): 36-42.
- [43] LI F T, HUANG M L, YANG Y, et al. Learning to identify review spam[C]. Vienna: The 22nd International Joint Conference on Artificial Intelligence. AAAI, 2011: 2488-2493.
- [44] 李实, 叶强, 李一军, 等. 挖掘中文网络客户评论的产品特征及情感倾向[J]. *计算机应用研究*, 2010, 27(8): 3016-3019.
- [45] 王祖辉, 姜维, 李一军. 在线评论情感分析中固定搭配特征提取方法研究[J]. *管理工程学报*, 2014, 28(4): 180-186.
- [46] DENG X L, CHEN R Y. Sentiment analysis based online restaurants fake reviews hype detection[C]// Lecture Notes in Computer Science. Changsha: 16th APWeb Workshops, 2014: 1-10.
- [47] 任亚峰, 尹兰, 姬东鸿. 基于语言结构和情感极性的虚假评论识别[J]. *计算机科学与探索*, 2014, 35(3): 313-320.
- [48] GREGORIO F, FOX A K, YOON H J. Pseudo-reviews: conceptualization and consumer effects of a new online phenomenon[J]. *Computers in Human Behavior*, 2021, 114(1): 1-10.
- [49] EVANS A M, STAVROVA O, ROSENBUSCH H. Expressions of doubt and trust in online user reviews[J]. *Computers in Human Behavior*, 2021, 114(1): 106556.
- [50] DONG L Y, JI S J, ZHANG C J, et al. An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews[J]. *Expert Systems with Applications*, 2018, 114(1): 210-223.
- [51] 张文, 王强, 步超骐, 等. 基于Co-training 协同训练的在线虚假评论识别研究[J]. *系统工程理论与实践*, 2020, 40(10): 2669-2683.
- [52] 邓胜利, 汪奋奋. 互联网治理视角下网络虚假评论信息识别的研究进展[J]. *信息资源管理学报*, 2019, 9(3): 73-81.
- [53] MUKHERJEE A, VENKATARAMAN V, LIU B, et al. What yelp fake review filter might be doing[C]. Washington: Proceedings of the International AAAI Conference on Web and Social Media. AAAI, 2013: 409-418.
- [54] SAVAGE D, ZHANG X, YU X, et al. Detection of opinion spam based on anomalous rating deviation[J]. *Expert Systems with Applications*, 2015, 42(22): 8650-8657.
- [55] LIU Y, PANG B. A unified framework for detecting author spamcity by modeling review deviation[J]. *Expert Systems with Applications*, 2018, 112(2): 148-155.
- [56] 邱云飞, 王建坤, 邵良杉, 等. 基于用户行为的产品垃圾评论者检测研究[J]. *计算机工程*, 2012, 38(11): 7-14.
- [57] XU C, ZHANG J, CHANG K Y, et al. Uncovering collusive spammers in chinese review websites[C]. San Francisco: The 22nd ACM International Conference on Information and Knowledge Management. ACM, 2013: 979-988.