

产品垃圾评论检测研究综述

孙升芸 田 萱

(北京林业大学信息学院 北京 100083)

摘 要 互联网上的产品垃圾评论混淆视听,误导了潜在消费者。产品垃圾评论检测的目的就是将垃圾评论从评论文本中找到并去除,保留真实的产品评论供用户参考。首先将产品垃圾评论和互联网上其它常见的垃圾信息进行了对比,并把产品垃圾评论的检测和产品评论的质量判断、产品评论的情感分析等相关的工作进行了比较分析。然后从产品垃圾评论检测的数据集、检测方法两个角度对相关工作做了概述和分析。最后,在上述工作的基础上提出了一些产品垃圾评论检测研究中值得进一步关注的问题。

关键词 产品评论,垃圾评论,垃圾评论检测

中图分类号 TP391 **文献标识码** A

Survey on Product Review Spam Detection

SUN Sheng-yun TIAN Xuan

(School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China)

Abstract It is common for e-commerce Web sites to enable their customers to write reviews of products that they have purchased. Unfortunately, reviewers may write some untruthful opinions in order to promote or damage specific products' reputation which called review spam. Product review spam detection makes an attempt to find untruthful opinions and remove them to restore the truthful for readers. The difference between product review spam and other spams was analyzed firstly. The related work of review spam detection was presented including review helpfulness and review sentiment analysis later; and then the data set and methods to detect spam were introduced and summarized. The trend of product review spam detection was concluded finally.

Keywords Product review, Review spam, Product review spam

随着 Web2.0 的蓬勃发展,互联网逐渐发展成为“以用户为中心,用户参与”的开放式架构^[1]。互联网用户由简单的信息接收者,发展成为信息的发布者,甚至是信息的监督者。除此之外,互联网也在很大程度上改变了消费者的消费观以及消费意见反馈的途径。因此,越来越多的消费者青睐于网上购物。同时,大部分电子商务网站允许用户对所购买的产品发表评论,以表达消费者对产品的看法,如批评、赞扬、喜欢、厌恶等。这些评论信息为厂家以及潜在消费者提供了宝贵的信息资源。厂家通过分析这些信息,可以及时了解市场现状以及消费者的反馈意见,潜在消费者也可以以此作为购买产品的重要参考依据。

然而,有些用户出于某些利益会对产品发表一些垃圾评论,刻意的吹捧或者诋毁某些产品,这些垃圾评论在一定程度上影响了评论信息的参考价值,混淆视听,从而会误导潜在消费者。产品垃圾评论检测旨在解决这一问题,将垃圾评论从评论文本中找到并去除,保留真实的产品评论,提高评论情感分析的精度和准确性,为用户提供可靠的参考依据。

本文综合国内外已有的研究成果,首先将产品垃圾评论和互联网上其它常见的垃圾信息进行了对比,并将产品垃圾

评论的检测和产品评论的质量判断、产品评论的情感分析等相关的工作进行了比较分析。然后从产品垃圾评论检测的数据集、检测方法两个角度对相关工作做了概述和分析。最后,在上述工作的基础上提出了一些值得产品垃圾评论检测研究进一步关注的问题。

1 产品垃圾评论与互联网中其它垃圾信息的比较

产品垃圾评论与互联网中其它常见的垃圾信息,如垃圾网页和垃圾邮件有相似之处但又不同。垃圾网页是指通过操纵页面内容或者网页间链接的方式来误导搜索引擎的作弊网页,其目的是为了提高自身搜索排名,从而吸引用户浏览页面^[2]。垃圾网页主要有3类:基于链接的垃圾网页,基于内容的垃圾网页和基于页面隐藏的垃圾网页^[3,4]。垃圾邮件主要是指广告、电子刊物、各种形式的宣传品等宣传性的电子邮件^[5]。

垃圾评论(review spam)也称为 opinion spam、fake review 或者 bogus review,发表此类评论的用户称为 review spammer、opinion spammer 或者 fake reviewer^[6]。产品垃圾评论是由一些用户蓄意发表的不切实际、不真实的、有欺骗性

本文受中央高校基本科研业务费专项资金(YX2011-30),北京林业大学新进教师科研启动基金(BLX2w8019)资助。

孙升芸(1988—),女,本科生,主要研究方向为文本意见挖掘;田 萱(1976—),女,博士,讲师,CCF 会员,主要研究方向为智能信息检索、知识工程,E-mail:tianxuan@bjfu.edu.cn(通信作者)。

质的评论,其目的是为了 提高或者毁坏某一产品或某一类产品的声誉,从而误导潜在消费者,使其做出风险性的购买决策,或者干扰评论意见挖掘和情感分析系统的分析结果,降低精确度^[7]。

- Jindal N 等人将产品垃圾评论分为 3 类^[8]:
- a) 欺骗性的评论 (untruthful opinions), 指蓄意提高或毁坏产品声誉的不真实的评论, 过度吹捧产品的评论称为 hyper spam, 过度贬低产品的评论称为 defaming spam;
 - b) 不相关的评论 (reviews on brands only), 指评论的对象仅仅是品牌、生产商、销售商等与产品本身无关的评论;
 - c) 非评论信息 (non-reviews), 如广告、读者的问题和回答等。

垃圾网页和垃圾邮件通过人工方法可以识别, 但是一个垃圾评论发表者可以刻意地掩饰、乔装自己所发表的垃圾评论, 使其与正常的评论看似没有差别, 导致产品垃圾评论即使采用人工方法也难以检测, 所以, 现有的检测垃圾网页和垃圾邮件的方法^[3,9-16] 大都不能用于检测产品垃圾评论。

2 评论质量的判断

针对网上的评论信息, 评论质量的判断也是一个热点研究领域。高质量的产品评论是对产品详细、全面地评价, 具有说服力, 是潜在消费者做购买决策的主要参考。低质量的评论则相反, 不但不能让潜在消费者全面地了解产品, 而且浪费读者的时间和精力^[17]。Liu J 等人根据评论是否全面、详细以及令人信服的程度将评论分为 4 类: 优秀的评论 (best review)、良好的评论 (good review)、中等评论 (fair review) 和差评 (bad review), 其中前 3 种属于高质量评论 (high-quality review), 第 4 种属于是低质量评论 (low-quality review)。评论质量高低的判断是一个分类问题。例如, 文献^[18]根据评论的信息量 (informativeness)、可读性 (readability) 和主观性 (subjectiveness) 这 3 个方面的特征, 使用支持向量机 (SVM, Support Vector Machines) 法建立机器学习模型来区分高质量评论和低质量评论。

检测低质量的评论与评估评论的有用程度类似^[19,20]。大部分电子商务网站允许读者对所读过的评论进行评价, 如 Amazon.com 中, “3,204 of 3,272 people found the following review helpful” 表示有 3,272 位读者对这条评论做了评价, 其中 3,204 位用户认为这条评论有用。Kim S-M 等人根据评论所得的有用的投票比例来评估评论的质量^[19]。Ghose A 等人用两种排名机制来预测评论的有用程度: 以消费者为导向的排名机制和以商家为导向的排名机制^[17]。

评论质量的判断与垃圾评论检测有类似之处但又不同。前者是为了区分不同有用程度的评论, 在意见挖掘系统中, 质量越高的评论所占的比重应该越高; 将高质量的评论呈现给读者, 能更好地帮助读者全面了解产品, 并且减少阅读低质量评论的时间。然而, 低质量的评论并不一定是垃圾评论, 高质量的评论中也可能包含垃圾评论, 因为垃圾评论发表者可以刻意掩饰垃圾评论, 使其看似与高质量的评论相同。

3 文本情感分析

评论质量的判断和垃圾评论的检测都是服务于评论的文本情感分析。文本情感分析也称为意见挖掘, 是对带有情感

色彩的主观性文本进行分析、处理、归纳和推理的过程^[1]。简言之, 产品评论的挖掘就是抽取有价值的情感信息, 如评价对象、评论发表者、意见词、情感词等, 甄别评论内容的情感趋势, 即确定评论内容的情感极性是正向还是负向, 对情感趋向进行统计归纳。

文本情感分析技术首先自动获取大量相关的评论信息, 进而挖掘出主要的产品属性 (如像素)、意见词 (如高) 和情感词 (如喜欢), 最终通过统计、归纳、推理提供给用户该产品各个属性的评价意见, 方便用户全面了解产品并做出最终的购买决策^[21-25]。如第 2 节中所述, 评论质量高低的判断是为了区分评论在意见挖掘系统中所占的比重, 提高系统的精确性。

当前的文本情感分析工作主要是应用自然语言处理的方法提取正面和负面的评论意见, 而对评论本身的可信度的研究比较少。由于任何人都可以发表评论, 评论自由度高, 因此导致产生了很多垃圾评论, 在一定程度上会左右文本情感分析的结果, 导致潜在消费者做出风险性购买决策。

在文本情感分析过程中, 产品垃圾评论检测可以很好地解决这一问题, 将垃圾评论从评论文本中找到并去除, 保留真实的产品评论, 提高文本情感分析的精度和准确性, 为用户提供可靠的参考依据。垃圾评论的识别可以看作是文本情感分析中的去噪过程。

4 产品垃圾评论检测

4.1 数据集

产品垃圾评论的研究是随着互联网上电子商务网站的兴起而兴起的。纵观国内外的相关研究 (如表 1 所列), 大多数研究采用亚马逊电子商务网站 (www.amazon.com) 的相关数据。主要原因在于亚马逊的数据量大而且产品种类丰富。在这些数据集中, 由于图书和工业制造品的评论条目多, 便于观察和统计垃圾评论的特点, 因此占的比重比较大, 其中有 3 项采用了图书评论集, 4 项采用了工业制造品 (如数码相机) 评论集。

表 1 数据集对比

时间	作者	数据来源	产品类别	评论条目
2006	Hu N 等 ^[32]	Amazon.com	图书	237,086
			DVD	289,754
			Video	58,323
2007	Jindal N and Liu B ^[7]	Amazon.com		5800,000
2007	Liu J 等 ^[18]	Amazon.com	数码相机	23,141
			图书	2493,087
			音乐	1327,456
2008	Jindal N and Liu B ^[8]	Amazon.com	DVD	633,678
			工业制造品	228,422
2009	Danescu-niculescuz-mizil C 等 ^[33]	Amazon.com	图书	4000,000
2009	Baccianella S. 等 ^[31]	TripAdvisor	旅馆	15,000
2010	Jindal N 等 ^[27]	Amazon.com	工业制造品	228,422
2010	Jindal N 等 ^[28]	Amazon.com	工业制造品	228,422
2010	Gilbert E and Karahalios K ^[34]	Amazon.com		1000,000
2010	Guangyu Wu 等 ^[29]	TripAdvisor	旅馆	30,000

Amazon.com 中每条评论由以下 8 个属性组成^[8]:

〈Product ID〉〈Reviewer ID〉〈Rating〉〈Date〉〈Review Title〉〈Review Body〉〈Number of Helpful Feedbacks〉〈Number of Feedbacks〉

另外,大部分电子商务网站还允许用户为产品打分(rating),例如,Amazon.com中,打分范围为1~5,1代表最低分,5代表最高分。

除了评论内容本身外,其他的7个属性及组合也常常被用来检测垃圾评论。

国内知名的电子商务网站的评论模式与Amazon.com类似。当当网中每条评论由以下10个属性组成:

〈产品名称〉〈评论标题〉〈评论发表者〉〈发表时间〉〈打分〉〈心情指数〉〈阅读场所〉〈评论文本〉〈认为该评论有用的读者数〉〈认为该评论没用的读者数〉

京东商城中每条评论由以下10个属性组成:

〈产品名称〉〈评论发表者〉〈购买时间〉〈评论标题〉〈打分〉〈优点〉〈不足〉〈使用心得〉〈认为该评论有用的读者数〉〈认为该评论没用的读者数〉

上述电子商务网站中往往还允许读者对所读过的评论进行评论。这些评论信息也成为检测产品垃圾评论的重要研究点。表1中Jindal N等人利用包括认为该评论有用的读者数、评价过该评论的所有读者数以及前者在后者中所占比例在内的24个特征来表述一条评论^[8]。

4.2 检测方法

4.2.1 检测垃圾评论发表者

如果一个用户是垃圾评论发表者,那么他发表的评论有很大的可能性是垃圾评论。垃圾评论发表者的识别是以用户为中心,以用户的行为为驱动的垃圾评论检测方法。检测垃圾评论发表者相对于直接检测垃圾评论而言更为简单,因为一条评论仅仅涉及到一个用户和一个产品,信息量比较少。Lim E-P等人根据用户的打分(rating)行为定义了4种模型^[26]。

a)以某一个产品为目标发表垃圾评论(Targeting Product, TP),针对某一产品发表垃圾评论,一个用户对同一个产品重复打分的次数越多,并且每次打分差距较大,评论内容相似度越高,该用户是垃圾评论发表者的可能性就越大;

b)以某一类产品为目标发表垃圾评论(Targeting Group, TG),针对某一类产品(如同一个品牌的数码相机,同一个出版社出版的图书)在一个时间窗内连续发表垃圾评论,从过高的打分和过低的打分两方面进行建模;

c)总体打分偏差(General rating Deviation, GD),即一个用户对某一产品的打分与所有用户打分的平均值之间的偏差,偏差越大,越有可能是垃圾评论发表者;

d)早期打分偏差(Early rating deviation, ED),与GD类似,与打分的时间顺序有关。

Jindal N等人通过挖掘用户的行为,发现反常的评论模式来分析用户是垃圾评论发表者的可能性^[27,28]。如,一位用户对同一个产品重复的发表正向的评论,则其行为可疑性较大,有很大可能是垃圾评论发表者。

从识别垃圾评论发表者的角度出发识别垃圾评论会丢失部分非垃圾评论并遗漏部分垃圾评论。一方面,垃圾评论发表者发表的评论并不一定都是垃圾评论,在删除垃圾评论发

表者发表的垃圾评论的同时也会丢失一部分非垃圾评论。另一方面,一部分用户虽然不是垃圾评论发表者,但是也有可能发表垃圾评论。

4.2.2 检测垃圾评论

Jindal N等人从识别重复的评论(duplicates)出发,将重复的评论作为训练集,建立机器学习模型对评论进行分类,分为垃圾评论和非垃圾评论两类^[7]。他们采用shingle算法识别重复的评论,相似度得分大于0.9的评论被认为是重复性的评论;采用logistic回归建立检测学习模型。由于垃圾评论用人工检测的方法难以识别,因此训练集比较难以标记。

在上述基础上,Jindal N等人将评论垃圾分为3类,即欺骗性的评论、不相关的评论和非评论信息。考虑到评论文本本身的特征有限,Jindal N等人手工收集了评论文本、评论发表者、产品3个方面的24个特征^[8]。对于第二类 and 第三类垃圾评论,采用手工标记训练集的方法标记出470条垃圾评论,应用logistic回归建立机器学习模型来识别第二类和第三类垃圾评论。对于第一类垃圾评论采用文献^[7]中的方法,即通过识别重复性的评论,将重复的评论作为正向的训练集建立机器学习模型来识别第一类垃圾评论。这种用重复的评论作为训练集的缺点在于会遗漏一部分非重复性的垃圾评论。

4.2.3 检测不公平的打分

评论的打分(rating)属性也常常被用来检测垃圾评论^[26,28-31]。Dellarocas C通过聚类的方法将打分分为不公平的过高的打分(unfairly high ratings)和不公平的过低的打分(unfairly low ratings)^[30]。信誉低的评论发表者所发表的评论被认为是公平的。Singletons是指评论发表者发表的唯一的一条评论,Wu Guangyu等人利用正向的singletons在一个产品的所有评论中所占的比例和这些singletons时间聚集程度来分析评论发表者的可疑行为^[29]。

结束语 本文对产品垃圾评论检测的研究进行了综述,着重分析比较了前沿的垃圾评论检测方法。产品评论垃圾检测是近两年新兴的一个研究方向,相应的研究工作比较少,而且不够成熟,尚有许多值得深入探索的问题。在本文的最后,提出一些值得进一步挖掘的研究点和问题。

关于产品垃圾评论检测,需提高垃圾评论的识别精度。可以从以下两方面做进一步研究:

a)由于不同类别的产品,其属性不同,用户的评论对象不同,评论风格各异,给研究造成了一定困难,因此可以将评论情感分析和语义分析应用到产品垃圾评论的检测中。

b)由于发表评论成本低廉,一些广告推销员为了吸引读者,会在评论文本中添加热门搜索词或将热门搜索词作为其昵称,以达到推广产品的目的,因此可以将热门搜索词作为一个出发点来检测垃圾评论。

参考文献

- [1] 赵妍妍,秦兵,刘挺. 文本情感分析[J]. 软件学报,2010(8)
- [2] 魏小娟,李翠平,陈红. Co-Training——内容和链接的Web Spam检测方法[J]. 计算机科学与探索,2010(10)
- [3] Gyöngyi Z, Garcia-Molina H. Web Spam Taxonomy[R]. 2005
- [4] 蒋涛,张彬. Web Spam 技术研究综述[J]. 情报探索,2007(7)
- [5] 中国互联网协会. 中国互联网协会反垃圾邮件规范 [OL]. <http://www.isc.org.cn/20020417/ca134119.htm>, 2003

- [6] Liu B. Detecting Fake Reviews and Fake Reviewers: Opinion Spam Detection [OL]. <http://www.cs.uic.edu/~liub/FBS/fake-reviews.html>
- [7] Jindal N, Liu B. Review spam detection [C]//Proceedings of the 16th international conference on World Wide Web. Banff, Alberta, Canada; ACM, 2007; 1189-1190
- [8] Jindal N, Liu B. Opinion spam and analysis [C]//Proceedings of the international conference on Web search and web data mining. Palo Alto, California, USA; ACM, 2008; 219-230
- [9] Abernethy J, Chapelle O, Cactillo C. Graph regularization methods for Web spam detection [J]. *Mach Learn*, 81(2); 207-225
- [10] Baeza-Yates R, Carlos C, L'ope V. Pagerank Increase under Different Collusion Topologies [J]
- [11] Paul-Alexandru C, Jörg D, Wolfgang. MailRank: Using Ranking for Spam Detection [C]//Proceedings of the 14th ACM international conference on Information and knowledge management. 2005; 373-380
- [12] Ntoulas A, Najork M, Manasse M, et al. Detecting spam web pages through content analysis [C]//Proceedings of the 15th international conference on World Wide Web. Edinburgh, Scotland; ACM, 2006; 83-92
- [13] Wu B, Goel V, Davision B D. Topical TrustRank: using topicality to combat web spam [C]//Proceedings of the 15th international conference on World Wide Web. Edinburgh, Scotland; ACM, 2006; 63-72
- [14] Fettei I, Sadeh N, Tomasic A. Learning to detect phishing emails [C]//Proceedings of the 16th international conference on World Wide Web. Banff, Alberta, Canada; ACM, 2007; 649-656
- [15] Becchetti L, Castillo C, Donato D, et al. Link analysis for Web spam detection [J]. *ACM Trans Web*, 2008, 2(1); 1-42
- [16] Cortez P, Correia A, Sousa P, et al. Spam email filtering using network-level properties [C]//Proceedings of the 10th industrial conference on Advances in data mining: applications and theoretical aspects. Berlin, Germany; Springer-Verlag, 2010; 476-489
- [17] Ghose A, Ipeirotis P G. Designing novel review ranking systems: predicting the usefulness and impact of reviews [C]//Proceedings of the ninth international conference on electronic commerce. Minneapolis, MN, USA; ACM, 2007; 303-310
- [18] Liu J, Cao Y, Lin C-Y, et al. Low-Quality Product Review Detection in Opinion Summarization [C]//Proceedings of the Joint Conference on Empirical Methods in Natural Language and Computational Natural Language Learning. Prague, 2007; 334-342
- [19] Kim S-M, Pantel P, Chklovskit T, et al. Automatically assessing review helpfulness [C]//Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Sydney, Australia; Association for Computational Linguistics, 2006; 423-430
- [20] Zhang Z, Varadarajan B. Utility scoring of product reviews [C]//Proceedings of the 15th ACM international conference on information and knowledge management. Arlington, Virginia, USA; ACM, 2006; 51-57
- [21] Pang B, Lee L. Opinion Mining and Sentiment Analysis [J]. *Found Trends Inf Retr*, 2008, 2(1/2); 1-135
- [22] Stoyanov V, Cardie C. Topic identification for fine-grained opinion analysis [C]//Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Manchester, United Kingdom; Association for Computational Linguistics, 2008; 817-824
- [23] Titov I, McDonald R. Modeling online reviews with multi-grain topic models [C]//Proceeding of the 17th international conference on World Wide Web. Beijing, China; ACM, 2008; 111-120
- [24] Rao D, Ravichandran D. Semi-supervised polarity lexicon induction [C]//Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Athens, Greece; Association for Computational Linguistics, 2009; 675-682
- [25] Su F, Markert K. Subjectivity recognition on word senses via semi-supervised mincuts [C]//Proceedings of Human Language Technologies; The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Boulder, Colorado; Association for Computational Linguistics, 2009; 1-9
- [26] Lim E-P, Nguyen V-A, Jindal N, et al. Detecting product review spammers using rating behaviors [C]//Proceedings of the 19th ACM international conference on Information and knowledge management. Toronto, ON, Canada; ACM, 2010; 930-948
- [27] Jindal N, Liu B, Lim E-P. Finding atypical review patterns for detecting opinion spammers [R]. 2010
- [28] Jindal N, Liu B, Lim E-P. Finding unusual review patterns using unexpected rules [C]//Proceedings of the 19th ACM international conference on Information and knowledge management. Toronto, ON, Canada; ACM, 2010; 1549-1552
- [29] WU G, Greene D, Smyth B, et al. Distortion as a Validation Criterion in the Identification of Suspicious Reviews [C]//1st Workshop on Social Media Analytics. Washington, DC, USA, 2010
- [30] Dellarocas C. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior [C]//Proceedings of the 2nd ACM conference on Electronic commerce. Minneapolis, Minnesota, United States; ACM, 2000; 150-157
- [31] Baccianella S, Esuli A, Sebastiani F. Multi-facet Rating of Product Reviews [C]//Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval. Toulouse, France; Springer-Verlag, 2009; 461-472
- [32] Hu N, Pavlou P A, Zhang J. Can Online Word-of-Mouth Communication Reveal True Product Quality? [J]. *Experimental Insights, Econometric Results, and Analytical Modeling*, 2006
- [33] Danescu-Niculescu-Mizil C, Kossinets G, Kleinberg J, et al. How opinions are received by online communities; a case study on amazon.com helpfulness votes [C]//Proceedings of the 18th international conference on World wide Web. Madrid, Spain; ACM, 2009; 141-150
- [34] Gilbert E, Karahalios K. Understanding deja reviewers [C]//Proceedings of the 2010 ACM conference on computer supported cooperative work. Savannah, Georgia, USA; ACM, 2010; 225-228