

基于 CiteSpace 的亚马逊评论外文文献的研究

刘冰琳 张逸敏 孔令建 张菡纯

常州机电职业技术学院 江苏常州 213164

摘要: 选取 2006 至 2024 年间 Web of Science 中关于亚马逊评论领域的 523 篇文献作为样本数据, 利用 CiteSpace 软件通过关键词共现、关键词聚类以及关键词突现等方法, 对发文量、研究热点、研究趋势等进行分析。研究发现: 国外基于亚马逊评论的研究呈现明显上升趋势, 作者间合作较多, 但核心作者群尚未形成; 研究热点主要集中于数据分析, 特别是推荐系统、机器学习和深度学习技术在评论分析中的应用。研究趋势由基础的评论分析转变到利用机器学习和深度学习等技术进行分析。

关键词: 亚马逊; 在线评论; CiteSpace

引言

亚马逊作为全球领先的在线零售平台, 其评论系统为学术界提供了宝贵的数据资源。本文采用 CiteSpace^[1] 知识图谱分析软件, 以 2006 年至 2024 年期间亚马逊用户评论领域文献为研究对象, 通过关键词共现技术、聚类图谱构建以及关键词突现图谱分析等方法, 揭示了亚马逊用户评论研究的核心议题和演进路径, 以期为深化亚马逊用户评论领域的探索提供参考和启示。

1 数据采集

为进行全面探究基于亚马逊用户评论学术研究现状, 以 Web of science 为数据来源, “Amazon online review” 为主题进行检索, 选定文献发表时间为 2006 年 1 月 1 日 -2024 年 11 月 1 日, 共得到 640 篇外文文献。为保证研究的科学性和准确性, 对文献进行筛选, 提出无关文献, 得到 523 篇关于亚马逊与评论的文献, 以此作为研究数据来源。

2 亚马逊评论外文文献基本情况

2.1 外文文献地区分布情况

外文文献发文地区提供了关于文献来源和背景的重要信息, 对于学术研究和文献评估都有着重要的意义。由图 1 可知, 发文量前五的国家分别为美国 (207)、印度 (99)、中国 (44)、韩国 (25)、加拿大 (24)。根据国家共现图谱发现, 国家间的网络较为紧密, 说明国家之间关于亚马逊评论研究的合作较多, 对于亚马逊评论研究的关注度较高。



图 1 发文地区共现图谱

2.2 外文文献年发文量情况

通过对文献数量特征的分析, 可以更加直观的了解特定时期内学者们对相关领域的关注度^[2]。图 2 展示了从 2006-2024 年关于亚马逊评论研究的外文文献发表量。从整体上看, 关于亚马逊评论研究呈上升趋势, 尤其是从 2020 年开始, 年文献发表量突破 50 篇。



图 2 国外基于亚马逊评论的研究发文量

2.3 文献作者网络分析

对亚马逊评论研究领域的发文作者进行分析, 可以了

解该领域的核心作者及其影响力。利用 CiteSpace6.3.R1 软件生成 2006 — 2024 年该领域作者合作网络共现图谱, k 值为 15, 如图 3 为发表量前 60 的作者, 分析可知该领域研究者分散, 但合作的作者较多。为进一步探究该领域核心作者分布情况, 借助普赖斯定律^[3]在同一研究主题中, 一半的论文是由高产作者发布, 核心作者发表的文章应达到该领域总文章的 50%, 作者集合的数量等于总作者数的平方根 ($\sqrt{n}$, 其中, n 代表最高产作者的发文数量)。图中可知亚马逊评论研究邻域发文量最多的作者是 Gu Bin, 共发表 5 篇文章。把 5 带入普赖斯定律的公式中得到 $M=1.675$, 说明发文量 2 篇及以上是核心作者。经过分析发现, 在 523 篇文献中有 48 位作者发文量达到 2 篇及以上, 因此, 核心作者为 48 名, 共发表文献 110 篇占全部文献的 21.03%, 未超过 50%。因此, 亚马逊评论领域的核心作者群尚未形成。



图 3 亚马逊评论研究作者合作网络图谱

3 研究热点、趋势分析

3.1 研究热点

3.1.1 关键词共现分析

利用 CiteSpace 软件对亚马逊评论研究进行关键词共现网络分析, 关键词的研究可以为理解研究主题的集中度提供解释, 以便科学直观地了解邻域的研究热点。设置好节点类型关键词 k 值为 16, 进行循进网络分析^[4], 综合整体网络计算得到图 4。图中显示网络的 N (节点数) 为 297, E (节点之间的连线) 为 520, 密度为 0.0118。关键字出现的频率决定了圆圈的大小, 关键词圈层越大, 证明出现频次越高。图中以 “word-of-mouth”、“sentiment analysis”、“impact”、“sales”、“machine learning” 等词为核心向外辐射。其次, 在这图谱中关键词的中介中心性非常重要, 关键词中介中心度表示了它控制的关键词之间信息流的数量以及对整个网络资源的控制程度。关键词的中介中心性越大, 说明这个关

关键词越重要,当中介中心性大于等于 0.1 的节点,说明是一个关键点。图中的中介中心性在 0.1 及以上的共有 19 个关键词,但在这些关键词中它们的频次与中心度并不都呈正比关系。根据表 1 关键词频次和中心度统计表可知,“word of mouth”出现频次最多,但中介中心性较低,相反“information”出现频次低于“word of mouth”,但中介中心性高达 0.34,说明了评论中的信息挖掘对整个亚马逊评论的研究起到了核心支撑作用,“impact”、“product review”等词中心性也较高,是整个网络的基础支撑。因此,在亚马逊研究领域,评论信息的挖掘、影响以及商品评论是该领域的热门话题。

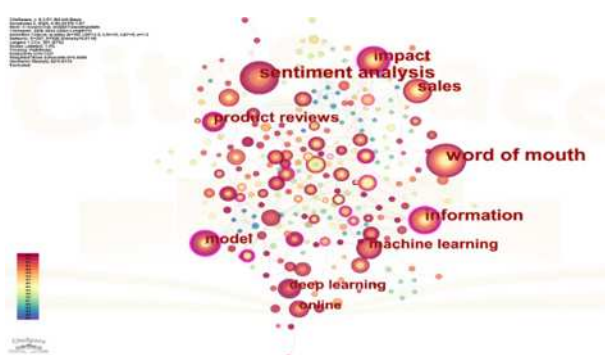


图 4 Amazon online review 关键词共现网络图谱

表 1 亚马逊评论研究高频关键词统计结果

序号	关键词	频次	中介中心性	首发年份
1	Word of mouth	93	0.07	2008
2	Sentiment analysis	85	0.05	2013
3	information	59	0.34	2010
4	impact	53	0.24	2008
5	sales	51	0.05	2010
6	Product reviews	39	0.19	2009

3.1.2 关键词聚类分析

为了对研究热点有更深刻的认识,在绘制高频关键词图谱的基础上,进行其关键词聚类分析,得到#0consumer behavior、#1product reviews、#2recommend systems、#3helpfulness prediction、#4user satisfaction、#5product ranking、#6survey、#7online word of mouth、#8review helpfulness、#9machine learning、#10opinion mining、#11heuristic-systematic model、#12deep learning、#13asymmetric information 共 14 个聚类词语,如图 5 所示。其中,聚类模块值 $Q=0.7513$, 聚类平均轮廓值 $S=0.9069$, 表示此次聚类结构显著且令人信服 ($Q>0.3$ 意味着聚类结构显著, $S>0.5$ 表示聚类合理, $S>0.7$ 表示聚类令人信服) [5]。

为了更加深入研究聚类结果群集中包含的内容，经过进一步的分析，将该研究领域热点归为以下 3 个方面。

基于亚马逊评论对消费者行为、用户满意度的影响的研究（#0、#1、#3、#4、#7、#8）：如 Gurney,L 等（2019）以 218 名亚马逊顾客评论为研究对象，探讨了信息诊断对消费者购买意愿的影响。Dan Ke 等（2021）对亚马逊中 241000 条评论进行分析，得出消费者过去的购物体验和不满意水平显著影响回购意愿。Refaie N（2021）明确了消费者参与维度（认知、情感和行为）在在线价值破坏之前、期间和之后的作用。

基于推荐系统与机器学习亚马逊评论分析研究（#2、#9、#10、#11、#12）：如：Jeong J（2021）以 1.41 亿条亚马逊评论的数据集为分析对象，通过机器学习分析消费者对可编程恒温器的数字足迹（DF），以识别和预测未观察到的消费者偏好。TK Aslanyan 等（2021）将潜在因子模型与潜在狄利克雷分配相结合，通过对不同规模的 Amazon 数据集上进行测试，发现在处理中等和大型数据集上，文本评论与评分结合起来可以提高推荐系统的预测准确性，有助于优化用户体验，提高销售转化率和客户满意度。Zarzour H 等（2022）使用深度学习来预测可解释推荐系统中的正面评论，便于用户可以获得对推荐项目的直观解释，提高正面解释的透明度。

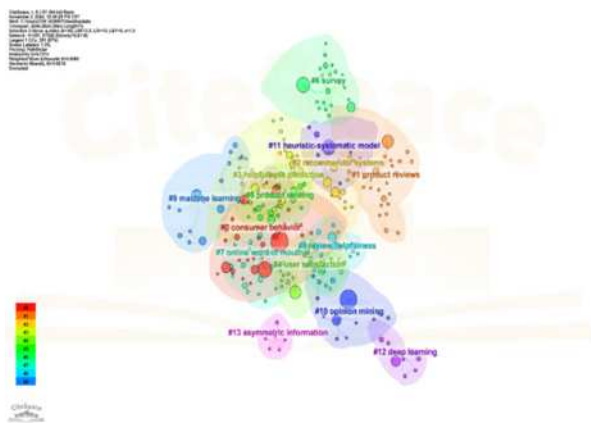


图 5 Amazon online review 研究关键词聚类图谱

亚马逊评论中的信息经济学进行市场分析的研究（#5、#6、#13）：如 Grinberg I（2012）对亚马逊上 10 个类别的 20 种产品的评论分析发现，在电子产品类别中，评论数量显著影响其排名，在线口碑指标通过提高产品知名度和说服消费者来影响购买决策，有助于企业提升产品销量，更好地

应对市场变化趋势；Sudhakar Vijayakumar,G 等（2021）研究分析了亚马逊和 Flipkart 上的口罩和消毒剂的在线评论，证明了在线零售商评论的特征分布存在显著差异，且这种信息不对称为在线消费者的消费行为转换创造了条件。

3.2 研究演进趋势分析

3.2.1 关键词突现分析

关键词突现可以反映出某个研究领域在某个时间段的研究热点和发展趋势。利用 CiteSpace 软件绘制出关键词突现图谱，由于突现的关键词较少，将伽马值调整为 0.7，形成的突现图谱，如图 6 所示，共出现 19 个突现词，一共可以大致分为 6 个阶段。第一阶段（2006–2010），研究主要关注“common identity”和“product reviews”，说明是基于评论对身份认同的初步研究，研究处于初始探索阶段；第二阶段（2010–2013），随着互联网、大数据的发展，结合互联网海量信息，对评论的研究成为研究热点，这一阶段见证了电子商务和在线市场的快速增长，以及对信息管理的重视；第三阶段（2013–2017），研究开始以人为本，主要关注“behavior”和“dynamics”，分析消费者行为以及通过评论进行市场预测；第四阶段（2018–2019）， “online review”和“impact”成为研究热点，研究者主要关注评论对品牌和消费者决策的影响。第五阶段（2020–2022），随着 ChatGPT 人工智能技术的横空出世，“machine learning”和“experience”成为研究热点，研究者借助先进的技术对评论进行多维度分析，更加注重用户的消费体验；第六阶段（2022–2024），利用“natural language processing”技术提取评论中的相关信息，准确高效的理解消费者行为和优化产品，是对文本分析和理解技术的深入探索。

Top 19 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2006 - 2024
common identity	2008	2.25	2008	2013	
product reviews	2009	2.35	2009	2010	
information	2010	2.58	2010	2011	
internet	2011	4.9	2011	2017	
dynamics	2010	4.37	2012	2017	
behavior	2013	2.32	2013	2017	
online review	2014	2.82	2014	2015	
impact	2008	2.39	2015	2019	
naive bayes	2017	2.53	2017	2019	
consumers	2008	2.41	2017	2020	
performance	2018	2.85	2018	2019	
decision making	2018	2.6	2018	2019	
customer reviews	2019	2.96	2019	2024	
experience	2019	2.45	2019	2022	
machine learning	2019	5.44	2020	2022	
perceptions	2020	3.04	2020	2021	
determinants	2020	2.99	2020	2022	
opinion mining	2009	2.19	2020	2022	
natural language processing	2020	4.24	2022	2024	

图 6 Amazon online review 研究关键词突现

4 结论

通过 CiteSpace 软件对 Web of science 数据库中 523 篇亚马逊评论文献进行分析,从总体上探究该研究邻域的基本特征、研究热点、研究前沿,以期为后续研究奠定基础。

第一,在文献发表方面,美国的研究成果尤为显著,发表量已超过 200 篇。不同国家在亚马逊评论领域的研究合作日益频繁,但尚未形成一个稳定的核心作者群体。第二,研究热点多元化,主要集中于消费者行为、推荐系统、用户满意度、机器学习、观点挖掘和深度学习等,研究更侧重数据分析,着重于推荐系统、机器学习和深度学习技术在评论分析中的应用,以及非对称信息在市场分析中的作用。第三,数据分析方法更新迭代频繁,如使用机器学习、深度学习模型、自然语言处理和启发式系统性模型对亚马逊评论进行分析;第四,研究重点应逐步从单一的评论内容分析拓展到用户行为的预测、市场趋势的分析以及品牌国际化战略的探讨。

参考文献:

- [1] 李杰.陈超美 CiteSpace: 科技文本挖掘及可视化 [M] 北京:首都经济贸易大学出版社 2016:1-89.
- [2] 姜琳,盛帅铎.基于 CiteSpace 国内智慧物流演进的可视化分析 [J]. 物流科技,2024,47(21):16-18.

[3] D·普赖斯,张季娅.洛特卡定律与普赖斯定律 [J]. 科学与科学技术管理,1984,(09):17-22.

[4] Olmeda-Gómez C, Perianes-Rodríguez A. Altmetrics as a research specialty (Dimensions, 2005-2018)[J]. Profesional de la información, 2019, 28(6).

[5] 唐赫,唐艳.乡村振兴背景下我国农村电商物流的研究热点与趋势——基于 CiteSpace 可视化分析 [J]. 物流工程与管理,2023,45(09):107-110+116.

作者简介:

刘冰琳(2003—),女,汉族,江苏常州人,大专,研究方向为跨境电商

张逸敏:常州机电职业技术学院跨境电子商务专业学生。

孔令建:副教授,博士,常州机电职业技术学院跨境电子商务专业专任教师。

张菀纯:讲师,硕士,常州机电职业技术学院专业大数据与会计专业专任教师。

基金项目:

常州机电职业技术学院学生创新创业培育计划项目(小驿站服务大民生—快递驿站物流服务质量的调研)资助。