

基于文本挖掘的电商产品用户评论分析

——以小米14为例

尚煜力

贵州大学管理学院, 贵州 贵阳

收稿日期: 2024年8月25日; 录用日期: 2024年9月18日; 发布日期: 2024年11月28日

摘要

用户发表的商品评论繁杂多样, 蕴含着用户多方面的体验信息, 具有宝贵的价值。本文基于百度AI大模型的情感分析接口和LDA主题模型, 对京东平台小米14型号手机的商品评论进行了分析, 探讨了手机评论中消费者的满意度及情感分布, 了解消费者对手机产品的核心需求, 为小米手机的产品优化、市场定位及营销策略提供了有力的数据支持。

关键词

用户评论, 情感分析, LDA模型

Analysis of User Comments on E-Commerce Products Based on Text Mining

—Taking Xiaomi 14 as an Example

Yuli Shang

School of Management, Guizhou University, Guiyang Guizhou

Received: Aug. 25th, 2024; accepted: Sep. 18th, 2024; published: Nov. 28th, 2024

Abstract

The product reviews published by users are complicated and diverse, which contain various experience information of users and have valuable value. Based on the emotional analysis interface of Baidu AI big model and LDA theme model, this paper analyzes the product reviews of Xiaomi 14 mobile phone on JD.COM platform, discusses the satisfaction and emotional distribution of consumers in mobile phone reviews, and understands the core demand of consumers for mobile phone products, which provides strong data support for product optimization, market positioning and

文章引用: 尚煜力. 基于文本挖掘的电商产品用户评论分析[J]. 电子商务评论, 2024, 13(4): 5698-5707.

DOI: 10.12677/ecl.2024.1341807

marketing strategy of Xiaomi mobile phone.

Keywords

Online Comments, Emotional Analysis, LDA Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,随着互联网的飞速发展和移动支付的普及,我国网络购物市场迎来了前所未有的繁荣。根据《中国互联网络发展状况统计报告》显示,截至2023年12月,在我国,使用网络购物的用户数量就达到了9.15亿人,几乎覆盖了全国各个角落的消费者,使用率也屡创新高,彰显出网络购物在现代生活中的重要地位[1]。在这一背景下,网购商品评论作为消费者表达购物体验 and 反馈的重要渠道,其特点和蕴含的价值日益凸显[2]。

网购商品评论以其真实性、即时性和互动性为特点,成为连接消费者与商家、产品之间的重要桥梁[3]。通过对评论内容的深入挖掘和细致分析,商家可以及时发现产品的优点和不足,进而采取针对性措施进行改进。同时,网购评论也是商家与消费者沟通的重要窗口,通过积极回应评论中的问题和建议,商家可以建立更加紧密和信任的客户关系[4]。

近年来,越来越多的学者开始关注电商产品评论的分析与挖掘,旨在通过文本分析技术提取出有价值的信息,帮助企业更好地理解客户需求和市场动态。其中,潜在狄利克雷分配(Latent Dirichlet Allocation, LDA)模型作为一种无监督的主题模型,因其能够自动发现文本中的隐藏主题而受到了广泛关注。在电商产品评论分析中,LDA模型能够提取出用户评论中的关键主题,如产品性能、外观设计、售后服务等,这些主题直接反映了用户的关注点和需求[5]。此外,通过对不同主题下的评论进行情感分析,还可以进一步了解用户对产品的情感倾向,从而为企业改进产品和服务提供有针对性的建议[6]。

本文通过八爪鱼模版采集京东平台小米手机官方直营旗舰店的商品在线文本评论数据,利用文本挖掘方法对其进行分析,通过情感分析和词云图,可以较为直观地查看情感分析结果。之后通过LDA模型,了解用户的关注主题,提取评论的关键信息,探索小米手机用户在购买时的情感倾向和关注主题,提出提高用户满意度的建议。

2. 研究方法和研究设计

2.1. 百度 API 情感分析

情感分析,是自然语言处理和计算语言学的一个研究领域,其专注于从文本数据中提取、分析并理解人们的情感、观点、情绪、态度等主观信息。情感分析广泛应用于社交媒体监控、品牌声誉管理、客户反馈分析、舆情分析、心理健康评估等多个领域。百度API(Application Programming Interface,应用程序编程接口)情感分析是基于自然语言处理(NLP)技术和机器学习算法,特别是深度学习模型,对文本内容进行多层次、多角度的分析。该接口利用神经网络对文本中的语言特征和语义信息进行深入挖掘,自动识别出文本所表达的情感倾向。相比使用传统的SnowNLP进行情感分析来说,具有技术成熟度与稳定性更高、易用性与集成度更好、定制化与灵活性更强、性能与效率更优以及支持的语言与领域更广等优势。

2.2. 主题模型

LDA 主题模型及其各种衍化版本拥有“文档 - 主题 - 词”三层结构，因为其可以更好的获取文本中潜在语义信息的特点，而被广泛用于文本挖掘领域的研究中。Blei D.M.等人通过应用 LDA 模型，不仅解决了 PLSA 的参数估计问题，还极大地提升了模型的泛化能力和灵活性[7]。杜利使用情感分析分类出正负面情感主题词，然后利用 LDA 主题模型来分析用户评论，了解潜在目标用户的需求、意见、购买原因，以及产品的优缺点[8]。王克勤等结合 Kano 的思想建立 LDA 模型，将时间因素加入到文本挖掘过程中，探究了用户需求主题演化过程[9]。李贺等使用 word2vec 技术，提出了一种基于语义相似性的产品需求识别模型[10]。Han Yi 等通过在 BERT 模型中加入新的卷积网络和命名实体识别(NER)层，从而在在线评论中提取产品属性并从中识别用户需求[11]。

由此可见，主题模型在多个领域得到了广泛应用，包括自然语言处理、生物信息学、商业智能、人文社会科学等。随着大数据和人工智能技术的不断发展，主题模型也在不断优化和改进，以适应更复杂的数据分析需求。

2.3. 研究流程

首先利用八爪鱼软件模版功能抓取用户评论，之后通过 Python 软件进行数据预处理，包括删除短评，去除噪声，文本分词和词性标注，去除停用词。之后通过统计词频来识别高频词，并绘制词云图以直观展示。下一步，使用百度 API 情感分析模型进行消费者情感倾向分析，初步分析用户对小米 14 手机的整体评价，并进行正负面情感词词云图绘制。最后，通过 LDA 主题建模，提取评论的关键信息，了解用户的关注热点，提出提高用户满意度的建议。流程图见如下图 1。

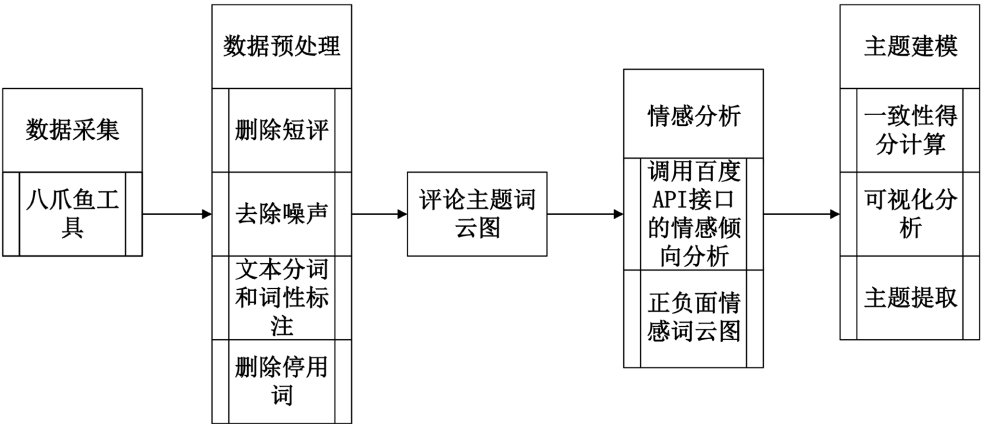


Figure 1. Text mining and analysis flowchart
图 1. 文本挖掘与分析流程图

3. 数据获取预处理

3.1. 数据获取

本文的数据来源于京东电商平台小米手机官方直营旗舰店小米 14 手机的用户评论。通过八爪鱼软件，从商品网站抓取用户评论数据，包括 ID 信息，用户星级、评论内容等信息。

3.2. 数据预处理

在数据分析之前，需要对原始数据进行预处理，以提高分析的准确性和效率。爬取之后共收集到评

论 5481 条, 经过删除重复评论, 保留 4529 条评论数据, 之后对保留的评论数据进行文本预处理, 主要步骤包括:

1. **删除短评:** 为了提高数据质量, 减少运算负担, 符合分析需求, 将过短的评论数据进行删除, 具体的保留文字下限的标准可按特定语料确定, 在本文中, 通过观察文本数据的长度, 设定文字下限为 6。

2. **去除噪声:** 删除文本中的无关字符、数字、符号等噪声。将爬取后的评论数据复制到 word 中, 利用 word 替换工具来删除无用字符, 标点符号、英文单词和数字, 表情等非中文字符。包括: 14, *, 快递等等。

3. **文本分词和词性标注:** 中文文本没有像英文那样的空格分隔符, 因此需要通过分词将连续的字符序列切分成有意义的词单元。此外, 分词能够将文本分解为更小的单元, 使得 LDA 模型在建模时能够更准确地捕捉文本中的主题信息。如果不对文本进行分词处理, LDA 模型可能会因为无法识别词边界而导致性能下降。本文中使用中文分词工具 Jieba 进行分词处理。

词性标注可以为文本中的词语提供词性信息(如名词、动词、形容词等), 这些信息有助于 LDA 模型更好地理解文本的词法结构, 帮助模型更准确地识别文本中的主题。例如, 名词和动词往往与主题紧密相关, 而介词、连词等则可能对主题识别贡献较小。

4. **删除停用词:** 去除文本中的常见但不包含有用信息的词汇(如“的”、“了”等)。这些词汇通常被称为停用词, 它们在文本中大量出现, 但对文本的含义贡献不大。

3.3. 评论主题词云图

经过上述的文本数据预处理后, 通过词云图, 对处理生成的评论数据进行直观展示, 如图 2 所示。从词云中我们可以看出, 消费者提及最为频繁的五个词汇为“小米”“外形”“外观”以及“系统”和“手感”。这反映出, 对于小米 14 手机, 大多数消费者持有非常积极的评价, 特别是在产品的设计美学(外形与外观)、操作流畅性(系统)以及使用体验(手感)上给予了高度认可。小米作为智能手机市场中的佼佼者, 其综合实力和品牌影响力得到了广大用户的广泛认同。



Figure 2. Comment on the keyword cloud map

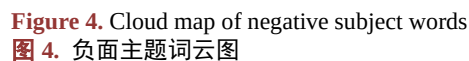
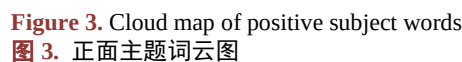
图 2. 评论主题词云图

4. 情感分析及 LDA 模型构建

4.1. 情感分析

为进一步探讨消费者对小米手机的满意程度, 本文借助百度 API 情感分析接口进行情感分析。

通过情感分析, 共得到积极文本数量 3471 条, 占比 76.6%, 消极文本数量 1058 条, 占比 23.4%, 并得出正面情感词和负面情感词。利用微词云进行词云图绘制, 分别如图 3 和图 4 所示:



同时，负面评价也不容忽视，主要集中在性能表现(卡顿、发烫、延迟)、设计缺陷(硌手、屏闪)、续航能力不足以及客服服务等方面。这些问题直接影响了用户的使用体验和满意度，是小米手机需要重点关注和改进的领域。

为了深入探究产品评论文本间隐含的语义联系，并了解消费者的关注主题，下一步使用 LDA 模型，挖掘并提炼出评论中蕴含的核心主题词，以便更全面地理解消费者反馈。

4.2. 构建 LDA 主题模型

将预处理后的评论文本作为 LDA 的训练数据集。之后使用 LDA 主题模型分析，LDA 分析的基本步骤通常包括以下几个方面：

4.2.1. 构建词袋模型

使用 Doc2Bow 将每个文档转换为一个词频向量，忽略词序。每个文档的表示是一个词汇表中词汇的出现次数的列表。

4.2.2. 确定主题数量

主题数量的选择会影响模型的效果和解释性，通常可以通过以下方法之一来确定：

人为判断：根据文档集合的内容和目的，主观地选择一个主题数量。

模型评估：使用诸如困惑度、一致性分数等指标来评估不同主题数量下的模型性能，并选择最佳的一个。

启发式方法：如肘部法则，通过观察模型性能随主题数量变化的曲线来确定。

本文主要采用一致性分数来确定主题数量，一致性分数的原理可以概括为：通过增加主题数量来提高评估的稳健性和准确性。

4.2.3. 训练 LDA 模型

使用 Gensim 或其他 NLP 库中的 LDA 来训练模型，调整参数，输出主题词分布。

在使用一致性分数确定主题数量时，得分情况和主题数量的关系如图 5 所示。

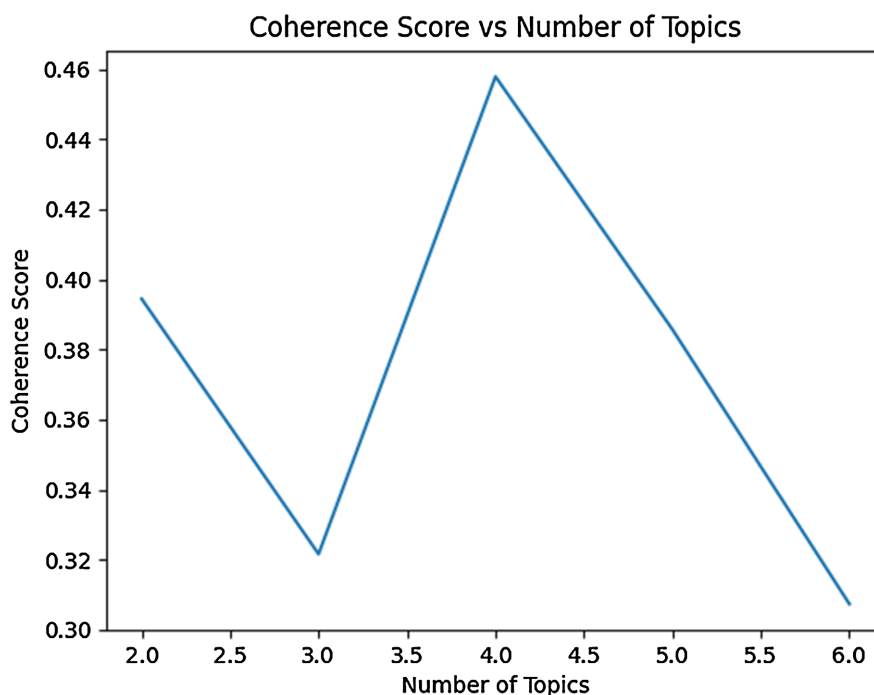


Figure 5. Relationship between consistency score and number of topics

图 5. 一致性得分和主题数量关系图

由图 5 可知，对于评论数据，当主题数为 4 时，一致性得分达到最高。因此，对评论文本，所设主题数量为 4。

4.3. 实验结果与讨论

设置好参数后，运行代码，经过 LDA 主题分析后，可得 LDA 主题分析结果，图 6 为 LDA 可视化结果中主题 1 的结果。

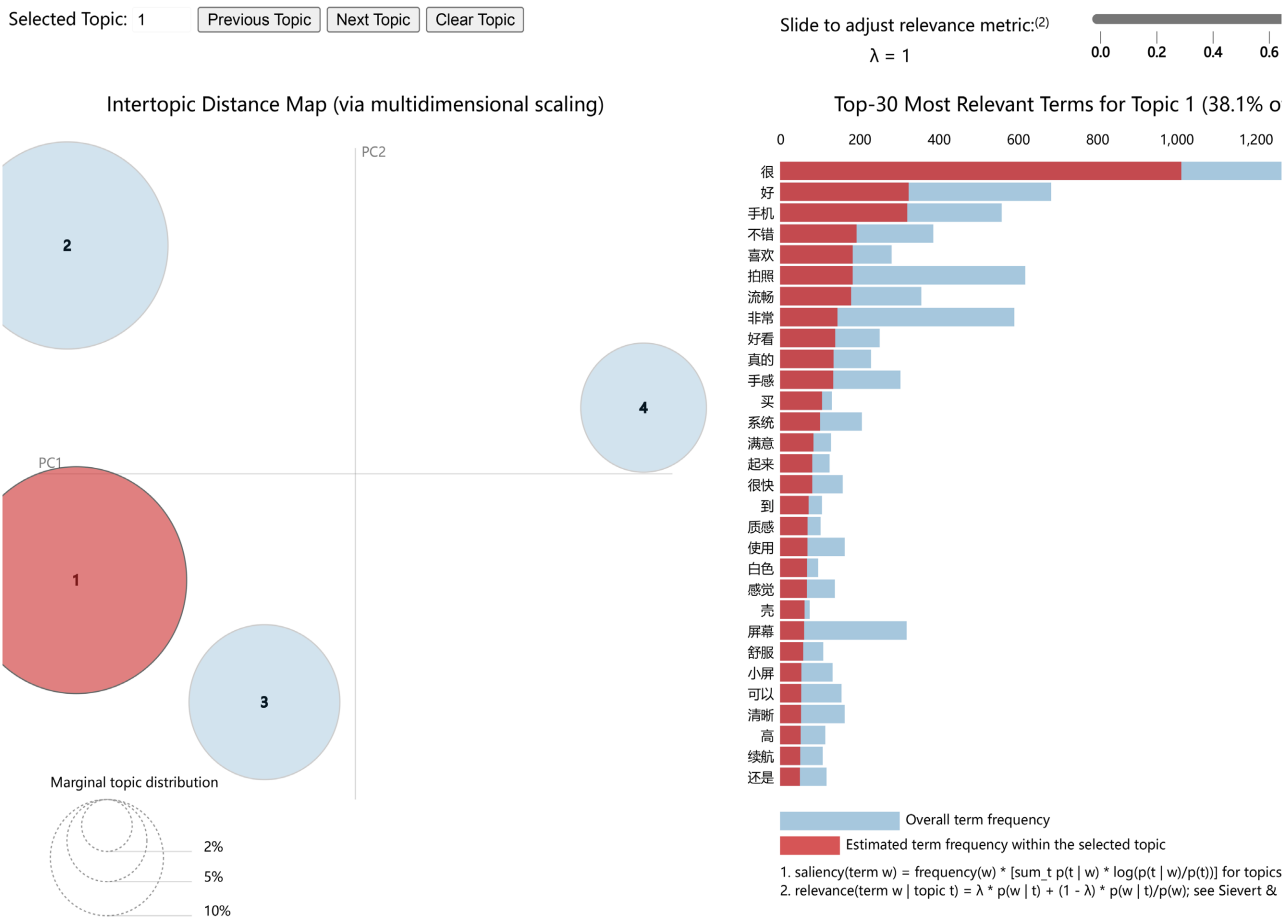


Figure 6. LDA model visualization results (K = 4)
图 6. LDA 模型可视化结果(K = 4)

LDA 主题分析结果如表 1 所示。

Table 1. LDA model topic analysis results
表 1. LDA 模型分析结果

主题 1	主题 2	主题 3	主题 4
很	效果	紧急	性能
好	外观	好	使用
手机	速度	手感	电池
不错	运行	珍惜	流畅

续表

喜欢	外形	系统	运行
舒服	屏幕	不错	续航
流畅	音效	使用	设计
手感	清晰	真的	功能
好看	快	颜值	效果
很	效果	紧急	性能

根据 LDA 模型运行结果，我们可以尝试为每个主题命名并进行分析，分析结果如下：

主题 1：整体好评与手感体验

这个主题包含了大量正面评价的词汇，如(“很”、“好”、“不错”、“喜欢”)，以及关于手机手感(“手感”、“舒服”)和拍照效果(“拍照”、“好看”)的词汇。反映了用户对小米 14 手机的整体好评，特别是对手感和拍照效果的满意。这些评价通常来自于用户在日常使用中的直观感受，是产品吸引力和用户满意度的重要体现。

主题 2：外观与性能表现

这个主题涵盖了与手机外观(“拍照”、“外观”、“外形”、“屏幕”)和性能(“速度”、“运行”)相关的词汇。展示了小米 14 手机在外观设计和性能表现方面的特点。用户可能特别关注手机的视觉效果(如屏幕显示和外观设计)以及其在运行速度和性能稳定性方面的表现。这些方面共同构成了用户对手机整体品质的评价基础。

主题 3：真实感受与系统体验

这个主题包含了表达真实感受的词汇(“紧急”、“很”、“真的”)，以及对手机系统和使用体验的正面评价(“手机”、“手感”、“系统”、“使用”、“喜欢”)。强调了用户在使用小米 14 手机过程中的真实感受。用户可能特别关注手机系统的稳定性和易用性，以及使用过程中的舒适度和便捷性。这些方面直接影响了用户的整体满意度和忠诚度。

主题 4：性能亮点与续航设计

这个主题包含了与手机性能(“紧急”、“性能”、“使用”、“流畅”)、拍照(“拍照”)、续航(“续航”)和设计(“设计”)相关的词汇，并出现了“突出”、“让”、“能力”等表达强调和影响的词汇。突出了小米 14 手机在性能、拍照、续航和设计方面的亮点。用户可能特别关注手机在这些方面的独特优势 and 创新能力，以及这些优势如何提升他们的使用体验。此外，“紧急”一词可能暗示了手机在应对紧急情况时的表现能力，也是用户关注的一个重要方面。

基于以上对情感和主题的分析，这款手机的优点总结如下：外观好，产品性能优秀，整体体验突出。

5. 结论与建议

5.1. 结论

本文展示了如何分析电商产品评论。通过 Python 和 LDA 模型，对小米 14 手机的用户评论进行了分析，首先进行文本预处理、分词、去停用词等操作，之后利用百度 API 情感分析工具进行了情感分析，识别出了该手机客户评价倾向。最后，通过 LDA 模型对评论内容进行分析，揭示出该手机客户的关注点。结果表明，用户对小米 14 的性能，拍照，续航，设计等方面给予了较高评价。这些发现为商家提供了有价值的反馈和建议，有助于商家抓住吸引顾客的要点，提升用户满意度和忠诚度。

5.2. 建议

对此，基于 LDA 主题分析的结果以及小米自身的品牌定位、目标客户群体和市场竞争环境，以下是一些针对性的营销策略建议：

5.2.1. 强化品牌形象与差异化优势

品牌故事与情感共鸣：利用小米“探索科技，为发烧而生”的品牌理念，结合主题 1 中的正面评价和手感体验，打造情感化营销内容，如用户故事、真实使用场景等，增强用户与品牌之间的情感联系。

差异化优势突出：针对主题 4 中提到的性能亮点与续航设计，通过技术解读、专业评测、对比视频等方式，突出小米 14 在性能、拍照、续航等方面的独特优势，形成差异化竞争优势。

5.2.2. 精准定位目标客户群体

年轻科技爱好者：针对小米品牌一直以来吸引的年轻科技爱好者，通过社交媒体、短视频平台等渠道，采用潮流、创新的营销方式，如 KOL 合作、挑战赛、话题互动等，提升品牌曝光度和用户参与度。

追求品质生活的用户：结合主题 2 中的外观与性能表现，针对追求品质生活的用户群体，强调小米 14 在设计美学、性能稳定性等方面的优势，通过高端体验店、产品发布会等方式，提升品牌形象和用户体验。

5.2.3. 多元化营销渠道与活动

线上线下融合：结合线上电商平台和线下实体店的优势，开展线上线下联动的营销活动，如新品发布会直播、限时抢购、线下体验会等，提升用户购买体验和品牌忠诚度。

跨界合作与联名款：与其他知名品牌或热门 IP 进行跨界合作，推出联名款或限量版产品，吸引更多潜在用户的关注，同时提升品牌影响力和市场渗透率。

5.2.4. 用户体验与售后服务优化

用户反馈机制：建立完善的用户反馈机制，及时收集并分析用户对小米 14 的使用反馈，不断优化产品功能和用户体验。

售后服务升级：提供高效、便捷的售后服务，如快速维修、延保服务、在线客服等，增强用户购买信心和品牌信任度。

5.2.5. 社会责任与可持续发展

环保理念宣传：在营销活动中融入环保理念，如使用环保材料、减少包装浪费等，展示小米作为负责任企业的形象。

公益项目参与：积极参与社会公益活动，如教育支持、灾害救援等，通过实际行动传递正能量，提升品牌形象和社会影响力。

综上所述，针对小米 14 手机的营销策略应围绕强化品牌形象、精准定位目标客户群体、多元化营销渠道与活动、优化用户体验与售后服务以及承担社会责任等方面展开，以全面提升市场竞争力和用户满意度。

参考文献

- [1] 中国互联网络信息中心. 第 53 次中国互联网络发展状况统计报告[R]. 北京: 中国互联网络信息中心, 2023.
- [2] 朱丽叶, 袁登华, 张静宜. 在线用户评论质量与评论者等级对消费者购买意愿的影响——产品卷入度的调节作用[J]. 管理评论, 2017, 29(2): 87-96.
- [3] Hong, D., Chiu, D.K.W., Shen, V.Y., Cheung, S.C. and Kafeza, E. (2007) Ubiquitous Enterprise Service Adaptations

Based on Contextual User Behavior. *Information Systems Frontiers*, **9**, 343-358.

<https://doi.org/10.1007/s10796-007-9039-2>

- [4] 周欢, 秦天琦. 基于在线评论情感分析与 LDA 的物流服务质量影响因素研究[J]. 重庆工商大学学报(社会科学版), 2021, 38(6): 27-38.
- [5] 张东鑫, 张敏. 图情领域 LDA 主题模型应用研究进展述评[J]. 图书情报知识, 2022, 39(6): 143-157.
- [6] 郭晓姝, 吴孟珊. 基于文本分析的在线评论对产品性能提升作用探析[J]. 中国管理信息化, 2021, 24(7): 193-196.
- [7] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.
- [8] 杜利. 基于 LDA 模型的电商用户评价分析[J]. 科技创业月刊, 2023, 36(2): 176-179.
- [9] 王克勤, 高智姣, 乔亚楠, 李靖, 同淑荣. 在线评论中的用户需求识别及其演化趋势挖掘[J]. 机械科学与技术, 2023, 42(7): 1070-1080.
- [10] 李贺, 谷莹, 刘嘉宇. 数据驱动下基于语义相似性的产品需求识别研究[J]. 情报理论与实践, 2022, 45(5): 99-106.
- [11] Han, Y. and Moghaddam, M. (2020) Eliciting Attribute-Level User Needs from Online Reviews with Deep Language Models and Information Extraction. *Journal of Mechanical Design*, **143**, Article ID: 061403.
<https://doi.org/10.1115/1.4048819>