

基于大模型的电商平台用户评论需求获取与情感分析研究

周 迪

上海理工大学管理学院, 上海

收稿日期: 2025年12月2日; 录用日期: 2025年12月16日; 发布日期: 2025年12月31日

摘要

在互联网时代, 电商平台用户评论数据呈快速增长态势, 如何从海量文本中挖掘用户需求并分析不同需求下隐含的情感, 成为企业提升产品竞争力的关键。本研究提出一种基于大语言模型(LLM)的链式分析框架, 设置三个代理模块系统性地实现用户需求获取与情感分析。三个代理模块分别完成从海量非结构化用户评论中精准提取核心需求陈述、基于提取的需求文本进行细粒度情感倾向分析、需求的多维度分类这三个任务。实验表明, 利用大语言模型实现链式分析, 不仅克服了单模型处理复杂任务的局限性, 还通过代理间的信息流传递形成闭环分析链路。本框架为企业和研究者提供了可扩展、自动化的在线评论的深度分析工具, 对精准把握用户诉求、优化产品设计及制定营销策略具有一定的实践意义。

关键词

在线评论, 用户需求, 大语言模型, 情感分析

Research on Demand Extraction and Sentiment Analysis from User Reviews on E-Commerce Platforms Based on Large Language Models

Di Zhou

Business School, University of Shanghai for Science and Technology, Shanghai

Received: December 2, 2025; accepted: December 16, 2025; published: December 31, 2025

Abstract

In the internet era, user review data on e-commerce platforms has experienced explosive growth. How to mine user demands from massive texts and analyze the implicit user emotions under different demands has become a key factor for enterprises to enhance product competitiveness. This study proposes a multi-agent collaboration framework based on Large Language Models (LLMs), incorporating three agent modules to systematically achieve user demand extraction and sentiment analysis. These three agent modules respectively accomplish three tasks: accurately extracting core demand statements from massive unstructured user reviews, conducting fine-grained sentiment tendency analysis based on the extracted demand texts, and performing multi-dimensional classification of demands. Experiments show that the agent collaboration mechanism implemented by large language models not only overcomes the limitations of single models in handling complex tasks but also forms a closed-loop analysis chain through information flow transmission between agents. This framework provides enterprises and researchers with an extensible and automated tool for in-depth UGC (User-Generated Content) analysis, and holds significant practical implications for accurately grasping user needs, optimizing product design, and formulating marketing strategies.

Keywords

Online Reviews, User Demands, Large Language Models, Sentiment Analysis

Copyright © 2025 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

随着市场竞争的加剧和消费者需求的多样化，许多企业面临的核心挑战在于如何高效、精准地捕捉用户需求并挖掘出隐含的需求与情感倾向。传统需求分析方法通常依赖于问卷调查、焦点小组访谈或专家经验[1]，但这些方法存在局限性：其一，人工主导的流程效率低下且成本高昂；其二，静态的调研手段难以实时追踪用户需求的动态变化；其三，非结构化的用户反馈难以被系统化分析和利用。如何有效利用互联网时代不断增长的非结构化用户在线评论，成为企业提升产品竞争力的关键[2]。然而，现有技术手段对非结构化数据的处理能力不足，导致大量潜在需求信息未被有效提取和利用。所以，现在需要提出一种自动化、低成本的需求提取以及情感倾向分析的方法。

近年来，大语言模型(LLM)在自然语言处理领域的突破为解决上述问题提供了新的可能性。LLM 具备强大的文本理解、生成和推理能力，能够自动化处理非结构化数据，并通过小样本学习(Few-Shot Learning)快速适应特定任务[3]。基于此，本研究聚焦于“基于大模型的电商平台用户评论需求获取与情感分析研究”，旨在构建一套完整的框架，通过设计分阶段的提示模板，结合小样本学习能力，可在不依赖定制模型的情况下实现需求提取与情感分析。这一思路突破了传统单任务模型的局限性，为 LLM 在垂直领域的应用提供了新范式。通过自动化处理用户评论，企业可快速定位高频需求与潜在痛点，将分散的用户反馈转化为清晰的改进方向，通过需求优先级排序与可行性筛选，企业可集中资源开发高价值需求，避免盲目投入[4]。

2. 相关研究

2.1. 情感分析研究现状

情感分析(SA)，也称为意见挖掘，是利用自然语言处理(NLP)和情感分析技术提取和分析人们对主题、产品和服务等不同实体的意见、情绪、态度、看法等的任务[5]。按照分析对象的不同，情感分析可以分为“粗粒度情感分析”和“细粒度情感分析”。传统的情感分析技术主要有如下三种：基于词典的方法、基于机器学习的方法(ML)以及混合方法[6]。现如今，许多学者利用大语言模型技术进行文本情感分析，并取得了显著的效果。

2.2. 用户需求分析研究现状

用户需求(CR)是产品设计与开发的源头，将顾客对产品的需求信息融入产品研发阶段是生产出符合顾客需求产品的关键[7]。用户需求分析作为产品设计与优化的核心环节，其方法体系经历了从传统定性模型到数据驱动范式的显著演进。早期研究依赖人工访谈与问卷的静态数据采集方式存在成本高、时效性差及主观偏差等问题。随着在线评论数据的增长，学者开始转向动态、实时的需求挖掘技术。然而，传统方法在语义深度与跨场景泛化上仍显不足。近年来，大语言模型的兴起为需求分析注入新动能。

3. 研究方法设计

大语言模型技术的不断发展正推动市场研究领域迈入新的阶段。这种技术发展不仅体现在数据处理效率的大幅度提升，更在于其减少了传统调研方式对结构化数据的依赖。通过语义解析与逻辑推理，大模型能够从社交媒体评论、客服对话记录等碎片化数据中识别潜在需求盲区，并生成具有可借鉴意义的策略与建议[8]。

本研究在对产品进行需求提取、分类与情感极性的评估中采用大语言模型的链式分析框架，将爬取到的非结构化的在线用户评论数据转化为需求的结构化数据，并利用大模型强大的文本处理能力进行分类与情感分析，从而形成一个自动化需求提取的流程框架。该框架将三大核心任务分别封装为独立代理，通过严格定义的数据接口与状态转移机制串联为端到端处理链路，核心思想在于通过职能隔离，平衡模型生成能力。

近年相关研究显示，大模型在文本摘要与分析任务中的表现已趋近人类专业水平，这为本研究引入LLM代理机制提供了重要理论支撑。本框架所采用的LLM代理基于大型语言模型构建，通过定制化提示设计实现特定功能需求，体现出显著的功能灵活性与执行效率。框架结构如下图1所示。

在大模型多代理协作学习中，原始用户评论经数据模型封装后，输入至需求提取代理。该代理通过深度优化的提示模板驱动大语言模型生成初步解析结果，输出包含需求条目、情感标签及评分的结果。在此过程中，大模型从原始文本中提取结构化字段。数学上，此层级协作可形式化为函数链式组合：

$$\text{Result}_{\text{final}} = f_{\text{align}} \left(f_{\text{sentiment}} \left(f_{\text{extract}} (C) \right) \right) \quad (1)$$

其中 C 表示原始评论内容， $f_{\text{sentiment}}$ 、 f_{extract} 、 f_{align} 分别对应三个代理的处理函数。下面将详细介绍大模型链式分析过程的具体设计。

(1) 用户需求的提取

首先，本研究采用网络爬虫技术，选择与X公司智能手环产品的电商平台进行用户在线评论的爬取，将爬取到的非结构化文本保存为CSV格式并进行数据清洗。其次，利用大语言模型的文本分析能力，设计相关提示词，将得到的评论数据进行逐条分析，明确评论中具体的用户需求，并进行概括提取。

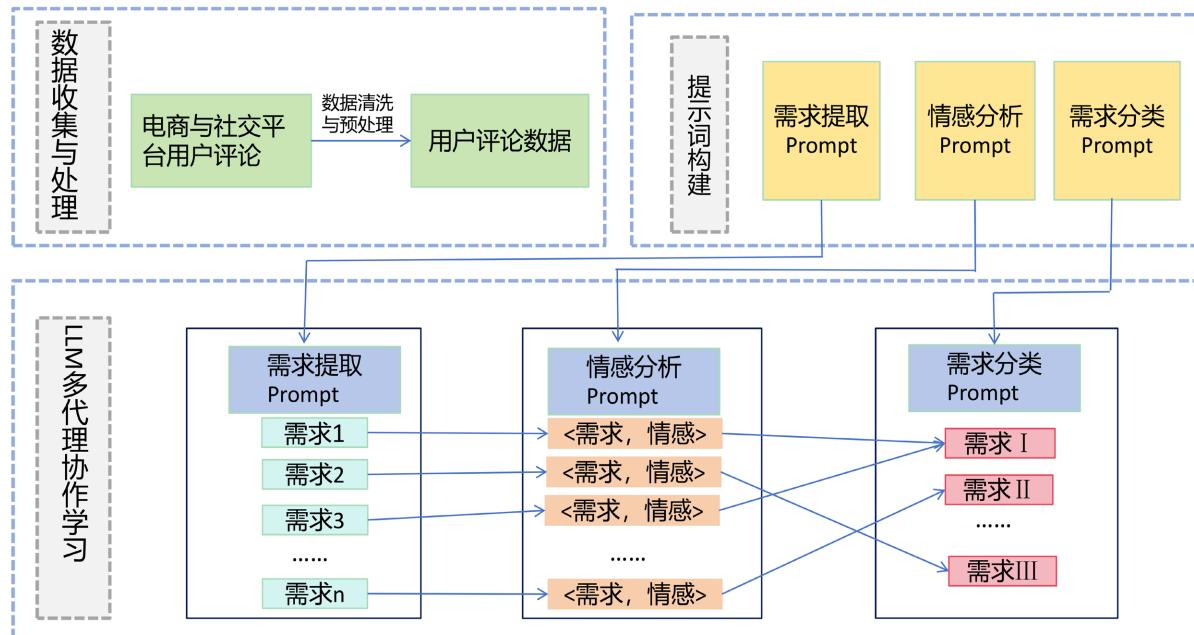


Figure 1. Construction framework diagram of LLM Agent

图 1. LLM Agent 构建框架图

(2) 用户情感极性分析

在提取用户需求的同时, 本研究也采用大模型的文本情感分析能力, 将提取出的不同需求进行情感标注, 并且进行情感评分, 形成 `<需求, 情感极性, 情感评分>` 三元组。

(3) 用户需求分类

需求分类代理接收校准后的需求文本。本研究利用大模型将所提取出的结构化需求文本进行分类, 将所有提取出的需求归类为预定义的功能特性之中。该代理采用“指令 - 类别列表 - 示例”的三段式提示, 将开放域问题转化为多分类任务。

基于大模型的链式分析框架旨在通过任务分解与模块化处理, 实现复杂语义解析流程的高效协同, 其依赖于多个大模型代理, 将利用人工进行的市场调研交给大模型完成, 将调研时间从数周压缩至几小时, 从而帮助企业决策提供多维度的方案[9]。

4. 实例分析

智能手环作为穿戴式智能设备的重要分支, 近年来在功能集成与用户体验方面取得显著突破。用户在评论中会自然表达对智能手环的功能体验、使用场景、外观设计等方面的真实感受, 这种数据为挖掘用户真实需求提供了直接依据。因此本研究选取智能手环为研究对象, 通过对 X 公司智能手环产品在线评论的获取, 利用大语言模型技术将数据进行分析, 从用户评论中准确提取用户需求并分析判断用户需求的情感极性。

由于本文所选择的研究对象为智能手环, 其相关评论数据可以从当前的主流电商平台获取, 所以在本研究中选取京东平台作为用户评论数据获取的渠道。

对于数据的收集, 在本研究中利用 Python 编程进行对智能手环的在线评论数据进行爬取, 从而获得原始数据, 如下表 1 所示。获取到每条主评论的核心字段如下: 评论 ID、用户 ID、评论时间、评论文字内容。最后, 将数据的 JSON 格式转化为 CSV 格式进行储存。

Table 1. Sample table of user online reviews**表 1. 用户在线评论示例表**

repid	user_id	date	content
243172096880	180348608	2024/9/26 16:02	震动是我最大的需求，设置闹钟后的震动无论办公室午睡还是早上起床都不会影响到他人。其次就是步数每周会踢 2 场球看下跑了多少，偶尔看下时间。
240753830608	10114642	2024/8/29 9:26	陶瓷真的天天掉，估计是陶瓷重量大氟橡胶更软滑加上腕带扣也是陶瓷的阻尼太小了，经常因为背包包之类的刮到哪里就开扣了。
246795439728	71575613	2024/11/10 22:14	应该可以主动接入各种三方平台，这样才有更多人接受。

4.1. 数据的预处理

由前文可知，本研究利用 Python 代码爬取了京东有关 X 公司智能手环的评论，收集到评论 3478 条。在进行数据预处理时，本研究采用 Python 并结合手工整理去进行数据清洗。首先，大量的重复内容对研究分析是无效的，并且会降低大模型分析用户需求的效率，增加运行成本。所以本文利用 Python 将重复数据删除，删除了评论中文字小于等于五个中文字符的极短评论，并且去除了去除非中文字符和表情符号。其次，在进行数据清洗时也设置了停用词的词典，去除了许多例如“哈哈哈哈”“很好很好”“嘻嘻”等无关词语。最后，进行人工检查，进行筛选，将与研究无关的许多评论删除。最终，将获取的评论数据进行清洗后，得到数据为 1454 条。

接下来，需要对第一轮清洗后的文本进行分词处理。考虑到本研究要处理智能手环相关的在线评论，可以利用 Python 中的 jieba 库来进行分词。本文借助 jieba 库，将构建的智能手环领域自定义词典导入分词工具，完成词典更新后重新执行分词流程。经处理，评论文本被合理切分为若干中文词组。完成分词后，下一步是去停用词。实际操作时，本文导入哈工大的停用词表，并通过人工补充，通过多次测试、调整，把和评论内容无关的词汇逐步过滤掉，这样处理后的数据会更干净，能提升后续分析的效率和准确性。

本研究运用了 TF-IDF 方法对经过数据预处理的在线评论数据进行高频词的统计，并将出现次数多的词语进行整合生成词云图，从图 2 可以更加直观地看出“运动”“屏幕”“睡眠”“心率”“表带”等被用户提及次数较多，是用户所重点关注的方面。

**Figure 2.** Word cloud diagram of online reviews**图 2. 在线评论词云图**

在这里统计了前 30 个高频词, 统计表格如下表 2 所示, 该表中统计了产品的部分功能属性特征和用户情感倾向, 这个表格的关键词可以初步反应用户的需求。

Table 2. Table of high-frequency words
表 2. 高频词表

高频词	频数	高频词	频数	高频词	频数
功能	201	监测	113	简单	61
睡眠	144	表带	106	外形	60
运动	143	喜欢	105	不准	58
质量	132	灵敏度	103	体验	58
操作	121	外观	98	已经	57
屏幕	120	做工	98	性价比	57
心率	120	容易	64	陶瓷	50
手机	115	健康	64	方便	47
好看	115	产品	63	智能	47
续航	113	准确性	62	难易	47

4.2. 用户在线评论主题挖掘

LDA 主题模型作为一种可从文本集合中挖掘潜在主题结构的无监督学习方法[10], 能够实现对 X 公司智能手环产品需求主题的自动化识别。从图 3 可以看出, 主题数为 8 时, 主题一致性虽略低于部分峰值点, 但仍保持在相对可观的水平, 表明主题内部词汇间具有一定关联性。因此, 主题数为 8 时, 既避免了主题数过少导致的文本特征覆盖不足, 也未因主题数过多陷入主题碎片化、过拟合的问题, 实现了较好平衡, 能够较为全面且清晰地挖掘出文本中的潜在主题结构。

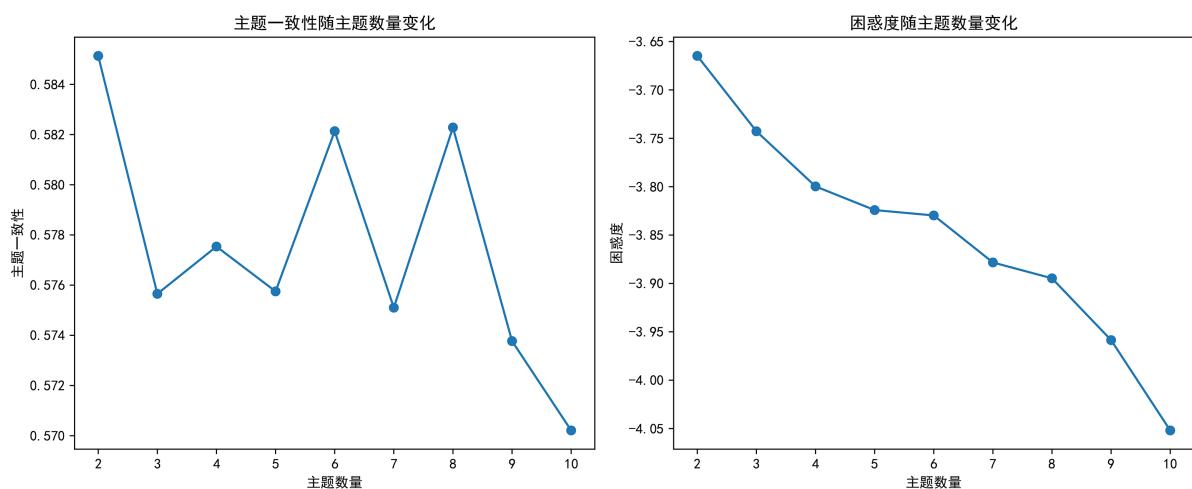


Figure 3. Diagram of topic coherence and perplexity
图 3. 主题一致性与困惑度图

为确保主题代表词的有效性与精准性, 进一步通过人工筛选的方式去除无意义词汇及重复项并将高频且表示产品特征的名词归类到下表 3 所示的八种不同属性类别中, 构建出“产品属性 - 特征词”双层结构。

Table 3. Topic representative words of user reviews
表 3. 用户评论主题代表词

客户需求属性	关键词
健康监测 C1	睡眠、不准、检测、睡眠监测、时间
运动辅助 C2	心率、运动、功能、检测、游泳
外观材质与个性化 C3	好看、颜色、喜欢、外观、不错
电池与续航 C4	充电、电池、电量、有点、续航
操作性能 C5	nfc、蓝牙、屏幕、操作、灵敏度
通知交互 C6	震动、振动、自定义、马达、闹钟
硬件与配置 C7	表带、手腕、问题、上带、陶瓷
兼容性 C8	支持、公交、连接、功能、下载

4.3. 基于大模型的需求提取及情感分析

本小节将详细阐述应用“基于大语言模型的链式分析框架”对清洗过后的在线评论数据进行需求提取、情感分析与需求分类。本文采用链式分析框架，设置三个代理模块分别完成三个不同的任务。

本研究中，基于大语言模型的链式分析在稳定的计算环境中运行。核心计算任务主要部署在本地高性能服务器上进行。该服务器配备了满足计算需求的 CPU 和足够容量的内存，同时确保高速互联网连接，以满足频繁调用远程 API 的低延迟要求。

系统开发采用 Python 3.10 作为主要编程语言，大语言模型能力通过调用深度求索(DeepSeek)提供的 DeepSeek-Chat 模型 API 获得。本研究中明确指定使用的模型版本为 DeepSeek-V3-0324。在调用 DeepSeek 模型 API 时，关键的参数设置直接决定了模型生成响应的确定性和一致性。其中，温度(Temperature)参数被设定为 0。最大输出令牌数(Max Tokens)控制模型输出的最大长度。文本根据各代理任务预期的输出长度结合实际 API 限制进行了合理设定，设置为 512，确保了绝大多数完整响应均能被返回。

本研究通过构建自动化处理流水线实现链式分析。系统以 JSON 格式批量读入经预处理的用户评论数据集后，进行三阶段链式处理流程。每个代理模块通过标准化 API 接口与大语言模型服务连接通过，严格遵循预设的指令模板完成各项任务，并通过结构化数据管道实现信息传递。

以需求提取阶段作为流程起点，需求提取的代理模块接收原始评论文本后进行解析。在实验中，Prompt 模板包含三重指令层：首先明确定义代理角色为“智能手环产品需求挖掘专家”；其次要求识别提取智能手环产品的主要需求点，仅提取与产品功能改进或缺陷相关的需求，忽略其他与产品功能无关的表达；最后强制规定 JSON 输出格式。

在需求提取完成后，情感分析代理去实现需求细粒度情感分析。该代理模块接收前代理输出的需求点列表后，采用并行调度策略同时发起多个 API 请求。每个请求包含三项关键信息：原始需求文本、所属评论 ID 及关联上下文片段。在明确代理为“在线评论情感分析专家”后，其提示词设计引入情感锚定技术：“请基于用户对[具体需求点]的表述强度判断情感极性。输出采用五分类法(非常积极/积极/非常消极/消极/中性)并附带置信度评分，如{“sentiment”: “消极”, “confidence”: 0.87}。”

需求分类代理衔接前期聚类的结果，将前文经聚类得到的八类分类体系融入需求分类的提示词设计之中。首先定义代理为“智能手环产品需求分类专家”，其次将需求提取代理提取出的需求点分类至各个类别点之中。最终输出采用分级标签格式如：{“category”: “续航能力”, “subtype”: “电池容量不足”}。

4.4. 稳健性检验

为了验证大语言模型对智能手环产品需求提取以及情感分析的准确性，本文采用 F1 值作为评估模型

性能的关键指标。首先，将大语言模型应用于产品功能需求分类任务，其次，将其应用于情感分析任务中。将 LLM Agent 完成的对 X 公司的智能手环产品功能分类与情感分类与人工标注的情感进行对比。研究得到的大模型对需求分类的评估结果如下表 4 所示：

Table 4. Evaluation results of demand classification

表 4. 需求分类评估结果

	precision	recall	F1-score
健康监测	0.885	0.892	0.889
兼容性	1.000	0.619	0.765
外观材质与个性化	0.832	0.886	0.858
操作性能	0.889	0.990	0.936
电池与续航	0.963	0.994	0.978
硬件与配置	0.879	0.753	0.812
运动辅助	1.000	0.933	0.966
通知交互	0.866	0.895	0.829
其他	0.724	0.700	0.712
accuracy	0.839	0.839	0.839
macro avg	0.838	0.785	0.895
weighted avg	0.842	0.839	0.838

从需求分类评估报告来看，大模型在整体上表现出了较高的准确性。其中，健康监测类别的精确率较高，这表明大模型在识别这类需求时，既能准确地将该类需求从众多文本中挑选出来，又能尽可能不遗漏实际属于该类的需求，整体性能出色。当然，部分需求类别也暴露出一些问题。兼容性类别的召回率远低于其他大部分类别，虽然其精确率达到了 1.0，但较低的召回率意味着有相当一部分兼容性需求被模型遗漏，可能是该类需求的表述方式较为多样，模型未能充分学习到所有可能的特征。

研究得到的大模型对情感极性判断的评估结果如下表 5 所示：

Table 5. Evaluation results of sentiment polarity judgment

表 5. 情感极性判断的评估结果

	precision	recall	F1-score
中性	0.978	0.801	0.881
消极	0.916	0.915	0.916
积极	0.906	0.904	0.896
accuracy	0.942	0.942	0.942
macro avg	0.933	0.873	0.898
weighted avg	0.941	0.941	0.940

情感分类评估报告显示，大模型在情感分类任务上的表现更为优异。在各个具体的情感类别中，消极情感的分类性能最佳，这表明模型在判断消极情感时非常可靠，能够精准地识别出用户文本中的负面情绪。积极情感的分类效果也很好，虽然精确率相对消极情感略低，但召回率较高，说明模型很少会遗漏积极情感的文本。中性情感的分类相对来说存在一定挑战，召回率相对较低可能是因为中性情感的界定本身比较微妙，部分带有轻微情感倾向的文本容易被误判为积极或消极情感，或者模型对一些较为含

蓄的中性表达识别能力不足。

5. 结语

本研究围绕用户生成内容的需求挖掘与情感分析核心问题,构建了基于大语言模型的链式分析框架,通过需求提取、情感分析、需求分类三大模块的协同运作,实现了非结构化评论数据的自动化、精细化处理。以智能手环产品为实证对象的研究表明,该框架有效突破了传统人工调研效率低、结构化分析难的局限,将调研周期从数周压缩至数小时。

该研究成果兼具理论与实践价值:理论上,为大语言模型在垂直领域的多任务协同应用提供了可借鉴的范式;实践中,为企业提供了低成本、高效率的需求分析工具,助力精准捕捉用户痛点与潜在需求,优化产品设计与营销策略。

需指出的是,研究仍存在一定局限,未来可通过优化提示词设计、扩充领域语料等方式完善模型性能,进一步拓展该框架在其他消费电子产品或服务领域的应用场景,提升其跨领域泛化能力,为在线评论数据分析提供更全面的技术支撑。

参考文献

- [1] 涂海丽,唐晓波,谢力. 基于在线评论的用户需求挖掘模型研究[J]. 情报学报, 2015, 34(10): 1088-1097.
- [2] Ghasemaghaei, M., Hassanein, K. and Turel, O. (2017) Increasing Firm Agility through the Use of Data Analytics: The Role of Fit. *Decision Support Systems*, **101**, 95-105. <https://doi.org/10.1016/j.dss.2017.06.004>
- [3] Zheng, Z., He, Z., Khattab, O., Rampal, N., Zaharia, M.A., Borgs, C., *et al.* (2024) Image and Data Mining in Reticular Chemistry Powered by GPT-4VU. *Digital Discovery*, **3**, 491-501. <https://doi.org/10.1039/d3dd00239j>
- [4] 刘意,谢康,邓弘林. 数据驱动的产品研发转型:组织惯例适应性变革视角的案例研究[J]. 管理世界, 2020, 36(3): 164-183.
- [5] 孟园,王洪伟,王伟. 网络口碑对产品销量的影响:基于细粒度的情感分析方法[J]. 管理评论, 2017(1): 144-154.
- [6] Jakob, N. and Gurevych, I. (2010) Extracting Opinion Targets in a Single and Cross-Domain Setting with Conditional Random Fields. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Cambridge, 9-11 October 2010, 1035-1045.
- [7] Wu, T., Liu, X., Qin, J. and Herrera, F. (2021) An Interval Type-2 Fuzzy Kano-Prospect-Topsis Based QFD Model: Application to Chinese E-Commerce Service Design. *Applied Soft Computing*, **111**, Article ID: 107665. <https://doi.org/10.1016/j.asoc.2021.107665>
- [8] Wu, J.H., Zhao, Y. and Zong, C.Q. (2023) Analysis of ChatGPT's Capabilities and Future Prospects. *Bulletin of National Natural Science Foundation of China*, **37**, 735-742.
- [9] Brown, T.B., Mann, B., Ryder, N., *et al.* (2020) Language Models Are Few-Shot Learners. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ACM, 1877-1901.
- [10] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.