

情报科学

Information Science

ISSN 1007-7634, CN 22-1264/G2

## 《情报科学》网络首发论文

题目: ChatGPT 大语言模型的评论情感分类预测与主题识别研究  
作者: 朱益平, 慕钰, 孙逸宁  
网络首发日期: 2024-12-27  
引用格式: 朱益平, 慕钰, 孙逸宁. ChatGPT 大语言模型的评论情感分类预测与主题识别研究[J/OL]. 情报科学.  
<https://link.cnki.net/urlid/22.1264.G2.20241227.0935.008>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# ChatGPT 大语言模型的评论情感分类预测与主题识别研究

朱益平，慕钰，孙逸宁

(南昌大学 公共政策与管理学院，江西 南昌 330031)

**摘要：**【目的/意义】探究大众对于 ChatGPT 大语言模型的情感和主要关注点，有助于人工智能企业推进 AIGC 技术的开发，可为政府制定相应技术监管策略提供建议。【方法/过程】爬取微博数据并进行预处理后，构建 Bert 模型进行情感分类，并结合 LDA 主题聚类及 ARIMA 时间序列模型，揭示公众对 ChatGPT 大语言模型的关注焦点和态度倾向，预测用户评论情感走向。【结果/结论】大众对以 ChatGPT 为代表的大语言模型的态度因人而异，情感分布较为均衡。用户的主要关注点呈现复杂化的特点，未来以 ChatGPT 为代表的大语言模型在公众的认可度和情感态度方面有着良好预期。【创新/局限】对情感分类结果进行细粒度情感评论 LDA 主题聚类的同时，将粗粒度分类结果与 ARIMA 时间序列模型结合，打破单纯依赖静态情感分析的局限性，多角度、多方面地对大众之于 ChatGPT 大语言模型的态度、关注点和未来预期进行分析，深入探究 ChatGPT 大语言模型的发展趋势和潜在风险。

**关键词：**ChatGPT；大语言模型；机器学习；情感分析；主题聚类

## Sentiment Classification Prediction and Topic Identification of Reviews

### Related to the ChatGPT Large Language Model

ZHU Yiping, MU Yu, SUN Yining

(School of Public Policy and Administration, Nanchang University, Nanchang 330031, China)

**Abstract:** 【Purpose/significance】 This research aims to explore the public's sentiments and key concerns regarding ChatGPT and large language models, which can assist AI companies in advancing AIGC technology development and provide recommendations for governments in formulating corresponding regulatory strategies. 【Method/process】 After web-scraping and preprocessing data from Weibo, a BERT model was constructed

基金项目：本文系国家自然科学基金项目“电力需求侧信息的精准推荐及隐私风险控制研究”(72164026)，江西省高校人文社科项目“突发事件社交媒体用户负面情感早期识别与风险预警机制研究”(编号：TQ22201)的阶段性研究成果。

for sentiment classification. The Latent Dirichlet Allocation (LDA) model was used for topic clustering, and the ARIMA time series model was employed to reveal the public's focal points and sentiment tendencies toward ChatGPT, as well as to predict the emotional trajectory of user comments. **【Result/conclusion】** The public's attitudes toward large language models, represented by ChatGPT, vary among individuals, with a relatively balanced distribution of sentiments. Users' key concerns are characterized by increasing complexity, and there are positive expectations for public recognition and emotional attitudes toward large language models like ChatGPT in the future.

**【Innovation/limitation】**By integrating fine-grained sentiment classification with LDA topic clustering and combining coarse-grained classification results with the ARIMA time series model, this study overcomes the limitations of relying solely on static sentiment analysis. It provides a multi-perspective and comprehensive analysis of the public's attitudes, concerns, and future expectations regarding ChatGPT, offering in-depth insights into the developmental trends and potential risks of large language models like ChatGPT.

**Keywords:** ChatGPT; large language model; machine learning; emotion analysis; topic clustering

## 0 引言

随着人工智能技术的迅速发展和应用,人们的生产和生活方式产生了深刻的变化。这种变化渗透到了多个行业,人工智能技术在多个领域都得到了广泛的应用。在图书馆领域,人工智能技术可以帮助图书馆进行图书自动分类、管理和库存优化,给读者提供智能化服务,清华大学图书馆的“清小图”、深圳图书馆的“小图”等智能机器人可以向读者提供 24 小时不间断的在线咨询服务,南京大学图书馆的“图客”智能机器人能够快速识别需要上架图书的信息及图书所在位置,实现了对图书的智能盘点与架位管理。在金融领域,借助人工智能的强大算力,金融机构可以进行风险管理和欺诈检测,中国平安银行通过大数据和人工智能技术,显著提高了风险管理的质量和效率,有效防范和减少了金融风险<sup>[1]</sup>。据艾瑞咨询统计测算,2021 年,“AI+金融”核心市场规模达到 296 亿元,带动相关产业规模 677 亿元,预计 2026 年,核心市场规模将达到 666 亿元<sup>[2]</sup>。人工智能同样也为高校科研带来了新动能,AI 的介入使得科研更加高效,推动高校科研范式变化,开创了人机共创的科研方法和学习路径<sup>[3]</sup>。斯坦福大学和麻省理工学院等世界知名高校已经开始使用机器学习技术来分析学术论文和专利数据,以便发现科研趋势和热门领域<sup>[4]</sup>。

伴随着各类新技术驱动机器智能创作内容,在众多人工智能技术方向中,人工智能生成内容(Artificial Intelligence Generated Content,AIGC)在数据巨量化、内容创造、跨模态融合、认知交互力等方面具有独特的技术特征,成为一种新的内容创作方式<sup>[5]</sup>。中国信息通信研究院发布的《人工智能生成内容(AIGC)白皮书》将 AIGC 定义为“既是从内容生产者视角进行分类的一类内容,又是一种内容生产方式,还是用于内容自动化生成的一类技术集合。”<sup>[6]</sup>AIGC 通过神经网络建模,可以准确地理解各种形式的内容输入,并生成更加自然准确的输出。在 AIGC 领域中具有代表性的产品便是 OpenAI 于 2022 年 11 月 30 日发布的 ChatGPT-3.5 和一季度后推出的 GPT-4。ChatGPT 有目共睹的强大语言理解和内容生成能力,使得它在人工智能领域具有重要的市场竞争力。

近年来,中国政府在人工智能领域的投入和支持力度也在不断增加,并取得了显著进展。2021 年 4 月,华为发布业界首个 2000 亿参数以中文为核心的预训

练生成语言模型——盘古系列 AI 大数据模型；在 ChatGPT 风靡全球之时，各大厂商的自研大数据模型如雨后春笋般涌现：百度于 2023 年 3 月 16 日发布“文心一言”大语言模型，强调该模型在文学创作、商业文案写作、多模态生成等方面具有广泛的应用价值<sup>[7]</sup>；同年 4 月 7 日，阿里云宣布自研大模型“通义千问”开始邀请用户测试体验，紧随其后的还有知乎、腾讯、网易、科大讯飞等等。2023 年还未过半，国产自研大语言模型便已展现出了强劲的发展势头。

这些自然语言处理产品为用户带来了解放生产力的便利，但同时也引发了人们对于 GPT 类技术可能带来的问题的思考。在推进新一代人工智能发展的过程中，如何才能确保人工智能技术的安全、可靠与可控。互联网上关于人工智能生成内容技术所带来风险的争论愈演愈烈，一部分人持积极态度，认为 AIGC 在塑造人机交互新范式、提升信息处理效率等方面具有较大的发展前景<sup>[8]</sup>；而另一部分则持消极态度，担心 AIGC 技术会引发隐私泄露、数据造假、就业困难等问题<sup>[9]</sup>。

评估 ChatGPT 大语言模型在大众中的舆论表现和影响，深入分析公众的情感态度和主要关注点，对人工智能企业和从业者更好地推进 AIGC 技术的开发和应用具有重要的理论与实践意义，同时可为政府在人工智能生成内容技术监管和风险应对方面提供一定的思路和建议。本研究基于微博大数据，运用深度学习模型进行情感极性分类，同时结合 LDA 主题聚类模型和时间序列模型，更加深入地研究大众对于 ChatGPT 大语言模型的主要关注点，预测舆论走向。本文主要试图解决以下三个问题：①大众对以 ChatGPT 为代表的大语言模型的情感类别分布如何？②大众对于以 ChatGPT 为代表的大语言模型的主要关注主题有哪些？③在未来，公众对以 ChatGPT 为代表的大语言模型的情感倾向会有什么样的演变？

## 1 文献综述

本文从网络热点事件的情感分析与人工智能生成内容（AIGC）两个方面对研究现状进行梳理。

### （1）基于网络热点事件的情感分析

从最早的基于词典和规则，到后来的基于机器学习和深度学习，情感分析方

法不断得到创新和完善。这些技术的发展使得情感分析在社交媒体数据挖掘、产品评论分析、网络舆情监测等领域得到了广泛应用,并且也为网络热点事件情感分析提供了更加高效可靠的技术支持。吴成斌<sup>[10]</sup>提出一种将情感词典与 Agglomerative、K-Means 以及 DBSCAN 聚类相结合的方法,应用于微博热点事件分析,实验结果表明其准确率较高,有助于相关部门掌握舆情趋势并制定对策。针对传统自然语言处理在机器学习中依赖于手动标记的特征、耗时且容易出现维度爆炸等难以解决的问题,黄萍<sup>[11]</sup>等人提出使用基于 CNN 的深度学习技术,通过收集校园热点话题进行预处理以及运用 Word2vec 模型生成词向量后,运用卷积神经网络提取其中的特征并进行情感倾向分类。

除了基于机器学习和深度学习的分析方法,目前还有一些其他的研究方法也被广泛应用于网络热点事件分析中。例如,利用社交网络分析和数据挖掘技术来研究网络热点事件的传播过程、影响因素等。王毅<sup>[12]</sup>以成都 A 校食堂问题事件的微博转发数据为基础,通过构建分层线性模型进行实证分析,寻找社交网络分析中相关联的指标,研究这些指标与网络热点事件影响力的相关性。赵爱<sup>[13]</sup>等人使用社交网络分析方法和 GIS 可视化技术,对微博中信息传播的路径进行了深入探索,研究发现,网络空间的空间结构具有地理特征,中心节点城市能够主导着周边地区甚至全国范围内的舆论方向。

同时,也有一些学者将网络热点事件分析与其他相关领域结合起来,开展跨学科研究。丁晓蔚<sup>[14]</sup>强调了对网民情绪和心理的引导在舆论引导中的重要性,将社会心理学、统计学、信息科学等多个领域的知识融合在一起。此外,也有学者<sup>[15]</sup>提出基于 LDA 模型识别用户所关注的主题,同时结合了 LSTM 模型和 PITSs 来分析并判定这些主题的情感倾向,以及有学者从社会学视角出发<sup>[16]</sup>,结合情报领域定量方法,采用了意见领袖影响力评价、LDA 主题模型和 SnowNLP 情感分析方法全面阐释和解构网络热点事件的社会本质和演进规律。

## (2) 人工智能生成内容 (AIGC) 领域的相关研究

学术界聚焦 AIGC 技术的应用也进行了一系列研究。李白杨<sup>[5]</sup>等人立足于互联网的演化发展,梳理了 AIGC 发展的基础条件,分别探讨了 AIGC 的技术特征、技术要素和发展阶段,认为 AIGC 与技术算法的融合应用,为信息组织、数据管



理等研究与实践带来了实质性的影响。詹希旎<sup>[17]</sup>等人系统梳理了 AIGC 的概念内涵,探析 AIGC 的交互形态和互动模式的方法,同时提出 AIGC 其本质特征还是内容生态的创新发展。赵浜<sup>[18]</sup>等人利用文献计量分析、信息行为研究等情报领域典型任务对 ChatGPT 和 ChatGLM 两种大型语言模型进行了实证检验,分析评价大语言模型的 9 项能力水平,得出大语言模型在执行多数典型情报任务时普遍具备较强能力的结论。

作为一项新兴技术,AIGC 在提高生产力、减少成本和提升智能化服务的同时,也带来了一些负面影响和风险。王静静<sup>[19]</sup>等人使用逻辑综合方法,探讨了生成式 AI、LLM、GPT 对信息管理与传播的影响,发现生成式 AI 和 GPT 类技术存在着数据偏差问题、透明度问题、隐私问题、恶意使用问题等。陆伟<sup>[20]</sup>等人从信息管理角度出发,认为需要考虑 ChatGPT 等大型语言模型生成过程和内容的可解释性,以及如何将生成的信息形成资源化利用的问题。为了深入研究 AIGC 技术带来的机遇与挑战,Dwivedi<sup>[21]</sup>等人汇集了计算机科学、市场营销、信息系统等多领域专家的 43 篇文章,在承认 ChatGPT 提高生产力的能力的同时,也表示需要考虑到它对隐私安全的威胁,以及滥用和生产错误信息的后果。漆晨航<sup>[22]</sup>针对 AIGC 的虚假信息现实问题,从机制建设、多元治理和科技方法三个层面提出了对策建议。王婷瑜<sup>[23]</sup>等人指出 AIGC 技术颠覆了传统的知识生产和传播方式,在推动知识向多元、平衡发展的同时,也引发了知识膨胀、泡沫化的问题。

在进行理论研究的同时,也有研究者聚焦用户对 ChatGPT 的态度和认知进行了访谈分析。段荟<sup>[24]</sup>等人使用质性研究方法,围绕科研人员对 ChatGPT 的态度与认知展开访谈,访谈结果显示,在信息资源管理科研领域,相关人员对 ChatGPT 所蕴含的技术理念及价值持有高度积极的认知与认同感。张海<sup>[25]</sup>等人以 ChatGPT 用户为研究对象,采用访谈的方式收集数据,构建了 ChatGPT 用户使用意愿影响因素研究模型,并发现主体因素、技术因素、信息因素和社会环境因素是影响用户使用意愿的重要因素。

学界对网络热点事件的情感分析以及 AIGC 技术应用的已有研究为笔者提供了许多有益借鉴,同时给了笔者进一步扩大研究空间的启示。在以“人工智能生成内容”技术为主题的研究中,一方面,由于该技术的代表性产品——Chatgpt-

3.5 自发布时间较短及其应用形式多样，相关研究和分析相对较少；另一方面，对 AIGC 技术的研究多集中于发展过程、实际应用以及风险对抗等方面的理论探讨，围绕该技术的舆论情感研究较为分散和局限。

因此，本文采用 BERT 模型进行情感分类，将其与 LDA 主题聚类模型和 ARIMA 时间序列模型相结合，进行多角度的综合分析，理解和评估 ChatGPT 大语言模型在大众中的舆论表现和影响，旨在为 AIGC 技术的舆论情感研究的拓展提供一定思路，并且为国内 AIGC 技术的可持续发展提供一定参考。

## 2 研究框架与数据处理

### 2.1 研究框架

本文围绕社会热点话题，对关于“ChatGPT 大语言模型”的微博评论进行情感分析研究。首先，爬取有关“ChatGPT 大语言模型”的微博用户评论，进行文本预处理后构造情感分类模型。本文基于 Bert 预训练模型构建情感分类模型，与其他模型进行对比实验，确定最佳情感分类器。

其次，在利用训练完毕的 Bert 模型对数据进行细粒度情感极性标注后，使用 LDA 主题模型分别进行主题聚类，分析大众对于“ChatGPT 大语言模型”的主要关注点。

同时，在粗粒度情感极性标注的基础上从传统的情感分类定性研究延伸至情感分析定量研究，按日期统计不同情感的评论数量，建立情感时间序列，构建时间序列模型进行时间序列分析。

最后，结合情感分类、主题聚类和情感预测结果，提出具有可操作性的意见和建议，并对未来的研究方向提出展望。

综上，提出技术路线图如图 1 所示。



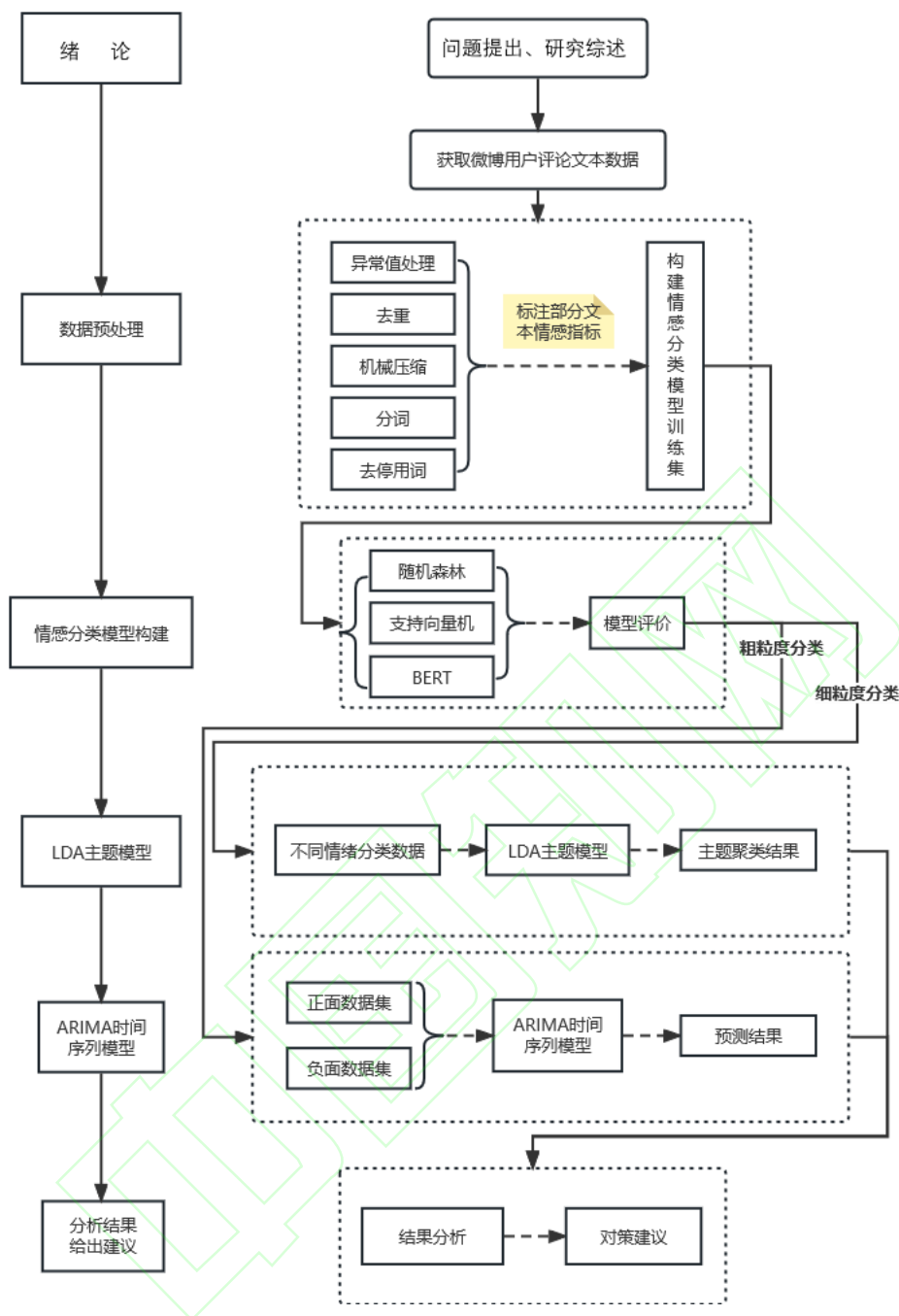


图 1 技术路线图

Figure 1 Technology roadmap

## 2.2 数据获取

微博具有使用简便、内容丰富、更新迅速等特点，符合大数据的基本特征，被广泛认为是大数据时代的典型代表<sup>[26]</sup>。本文聚焦新浪微博这一国内用户量最大、最具代表性的社交平台网站，对微博用户发布的有关 ChatGPT 大语言模型的评论进行数据挖掘，以“chatgpt”“gpt4”作为关键词进行微博内容检索，爬取到 2023 年 1 月 2 日至 2023 年 10 月 29 日共 43 周合计 48 679 条数据，在为确保

评论内容的有效性、提高实验数据的准确性，限定单个评论文本长度为 0-200 个字符，筛选去重后共计 36 677 条，汇总作为实验数据，部分实验数据格式如表 1 所示。

表 1 部分原始数据示例

Table 1 Partial examples of raw data

用户昵称	发布时间	微博内容	IP 国家属 地	IP 省份 属地
财联社 APP	2023/1/28 15:25	无论你是否准备好拥 抱未来，它早已把手伸进 你的前襟了#chatgpt#	中国	上海
研究者 July	2023/1/29 17:53	chatgpt 的诞生无疑又 是划时代的，世界都在进 步，我也不能一直懒惰着	中国	湖南
CemeteryA dministrator	2023/2/13 13:32	科技发展太快似乎不 是件好事	中国	广东
珂奈九十 九	2023/2/14 8:24	这个人工智能……聪 明倒有点儿……吓人，它 是不是已经有自我意识了 啊？？？	中国	江苏
Kanachinn_ TwT	2023/5/29 13:02	我宣布@ChatGPT 是 21 世纪最伟大的发明 我 爱死这个 ai 了	澳大利亚	
Laetitia 喜 悦	2023/7/9 21:44	#chatgpt#真是废话大 师啊	比利时	

2.3 评论情感标注

为了进行情感分析，需要对数据进行人工标注以获得情感倾向信息。在 36677 条实验数据中进行随机抽样，得到 10 000 条人工情感标注目标数据。参考张冬等<sup>[27]</sup>的观点，选取“积极”“中立”“消极”三个情感主类构建情感粗粒度指标。

情感细粒度指标方面，参考大连理工大学的情感词典，初步将积极情感分为“快乐”“惊喜”和“赞扬”，将消极情感分为“悲伤”“愤怒”和“担忧”，中立情感不做进一步的划分，一共七个情感次类。

同时根据专家咨询法，邀请 8 名相关领域专业人员参与情感分类指标的构建工作（其中包括 3 位从事情感分析研究的高校研究者、2 位自然语言处理专家、2 位机器学习工程师、1 位网络媒体评论员）对初步构建的七类情感次类进行讨

论。经过德尔菲法多轮咨询，最终划分出赞扬、惊喜、中立、担忧、愤怒和失望等六个情感次类，同时邀请这 8 名专家进行分工标注。表 2 展示了情感分类指标信息以及部分评论文本的标注示例。

在情感分类指标的构建过程中，结合了不同专家的知识 and 技能，以确保项目的科学性、客观性和全面性，尽可能避免认知差异带来的偏差。

表 2 情感分类指标信息

Table 2 Information of emotion classification indicators

情感倾向编号	情感主类	情感次类	含义	特征词	例子
0	积极	赞扬	用户对 ChatGPT 的优点和价值的认可	肯定、赞美、赞同、欢乐、兴奋等	chatgpt 写稿子太爽了太爽了昨晚憋了一晚上写了一篇，今天一晚上已经九篇了感谢当初不厌其烦搞账号的我。
1		惊喜	与 ChatGPT 有关的某方面超出了用户的预期	惊喜、惊奇、震撼等	理解透彻，表达清晰，ChatGPT 的精准度令人拍案叫绝！多样的话题处理方式和严谨的逻辑性都让我感到十分惊喜，就像是拥有了一个无所不知且思维细腻的朋友！
2	中立	中立	用户对 ChatGPT 没有太多的情感倾向	中立、客观、平淡、中性等	我每天进行最多的对话是在 chatgpt 上问：xxxxx（某术语）是什么？
3	消极	担忧	用户担心 ChatGPT 的应用会带来一些不良后果或者影响	担忧、忧虑、紧张、不安等	每当和 ChatGPT 聊天，我都会不由自主地想，我的隐私真的安全吗？
4		愤怒	用户对 ChatGPT 的某些表现或者应用方式等感到不满	愤怒、生气、不满、憎恨等	用了 ChatGPT 几次后，发现它真是太受限了，很多问题都回答不上来，细节处理也很差，真是让人愤怒。
5		失望	用户对 ChatGPT 的期望没有得到充分的满足	失望、沮丧、悲伤、绝望等	ChatGPT 的门槛太高了，我等普通用户只能远远观望，太失望了。

人工标注完毕后，通过对标注数据进行随机抽样、重复标注和互相校对等方式对标注结果进行质量控制和验证，进一步确保标注结果的准确性和可靠性。最终标注结果如下：中立情绪数量最多，共有 3067 条；赞扬情绪次之，共有 2743

条；惊喜情绪共有 1031 条；担忧情绪共有 1238 条；愤怒情绪共有 1049 条；失望情绪共有 872 条。同时，将细粒度情感评论按照表 2 的规则整合粗粒度评论，分别得到积极、中立和消极评论 3774、3067 和 3159 条，如图 2 所示。

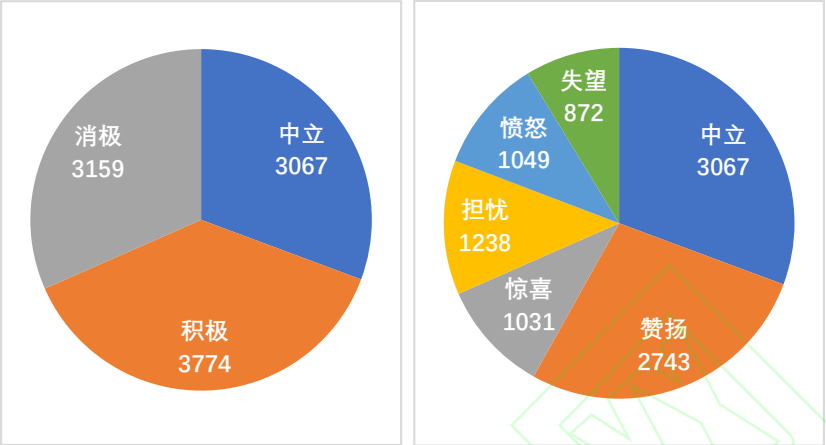


图 2 已标注粗粒度和细粒度评论基本信息

Figure 2 Comments with coarse-grained and fine-grained basic information marked

### 2.4 数据预处理

为过滤无效字符和无用信息，本文调用 Python 中的 re 模块来进行初步清洗，在去除无用信息后，调用 Python 中的 jieba 包进行文本分词操作，同时采用开源停用词表“中文停用词库”“哈工大停用词表”“四川大学停用词表”“百度停用词表”合并处理，并加入自定义字符，构建出共有 2443 个停用词的停用词表，进行停用词的过滤操作。

## 3 基于机器学习的评论情感分类

### 3.1 基于 BERT 预训练模型的情感分类

在本文的分类任务中，首先使用预训练的 BERT（Bidirectional Encoder Representation from Transformers）作为特征提取器，本文选择中文 BERT 模型 BERT-Base-Chinese，并选择 PyTorch 作为深度学习框架，随后在 BERT 之上添加一些全连接层，用于将 BERT 的输出特征进行进一步的处理和调整，在全连接层之后添加一个分类层，将模型的输出映射到不同的情感类别。

当数据量较大，以万为单位时，常见的划分比例一般为训练集、验证集、测试集的分配比例为 6:2:2<sup>[28]</sup>。考虑到模型需要足够的数据学习数据的特征，同时保证验证集和测试集也有足够的数据来准确评估模型的性能，初期以 6:2:2 的比

例进行划分，后来根据实验的结果尝试了不同的比例，观察模型在验证集和测试集上的性能，确保模型既不过拟合也不欠拟合，并且具有良好的泛化能力，最后根据实验的结果选择了 65%、20%、15% 的比例。为尽可能保证各情感评论数量的平衡，在已经人工标注的 10 000 条六种细粒度情感评论里按照此比例进行随机划分。选择 AdamOptimizer 作为优化器，并使用交叉熵损失函数作为目标函数，用于训练和验证模型。交叉熵描述两个概率分布之间的距离，交叉熵越小的情况下距离约小，两个分布越接近，公式如下：

$$\text{Loss} = -(y \log P + (1-y) \log(1-P)) \quad (1)$$

随后，通过调用 evaluate 函数评估模型在验证集上的性能，并保存当前最佳准确率对应的模型参数。

模型构建完毕后，在主程序中调用函数进行训练、验证和测试。由于本次实验限定单个评论文本长度为 0-200 个字符，因此将 Max\_Seq\_Length 设置为 200，默认学习率设置为  $2e^{-5}$ ，EPOCHS 选择 10 轮。为了实现性能和算力的最大化平衡，经过多轮调试，最终赋值 256 作为 Batch\_Size 批处理大小，并绘制出模型在训练集和测试集的准确率曲线图以及相应的损失值曲线图，如图 3 所示。

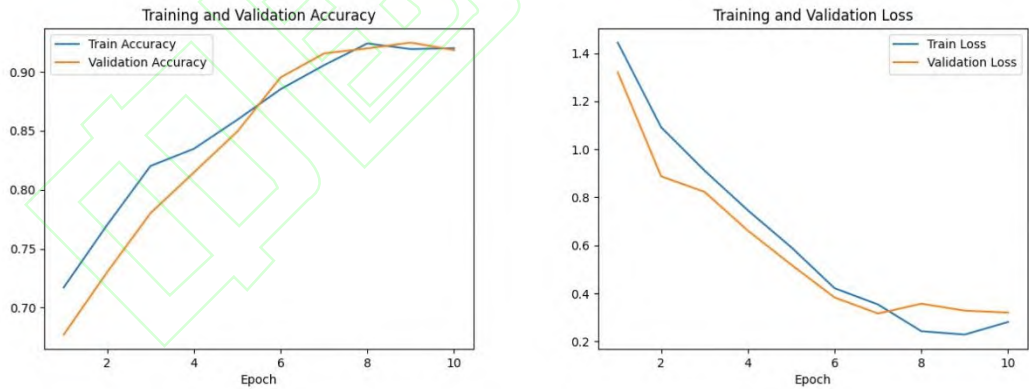


图 3 准确率曲线及损失值曲线变化情况

Figure 3 Accuracy curve and loss curve

由图 3 可知，BERT 模型在测试集上的准确率从第 1 个训练周期开始就有所提升，并且在前 9 个训练周期中不断增加，最终达到了 92.48%；同时，训练集准确率一直紧密围绕在测试集准确率曲线上，没有出现过拟合现象，这说明模型



具有较好的泛化能力。

另外，在测试集损失函数值方面，它在前 7 个训练周期中不断下降，最终在第 7 个周期达到了最小值 0.3165，并趋于稳定。这表明模型在训练过程中逐渐学习到了有效的特征表示，并取得了较好的分类效果。

综上所述，BERT 模型在经过 9 个训练周期的训练后，表现出较好的性能和泛化能力，可以用于实际应用场景中。

3.2 其他情感分类模型

本部分应用支持向量机和随机森林算法进行文本情感分类，利用词袋模型将评论文本进行向量化处理，将文本数据转换为 TF-IDF 矩阵，获取特征词列表。同时，引入 FastText 词嵌入方法进行比较分析，下载使用中文预训练模型 cc.zh.300.bin。在 Python 中使用 Pandas 库加载已标注的训练集和测试集，使用 fasttext.load\_model 进行词向量模型调用，再定义函数获取单词对应词向量，将句子中所有词语的词向量进行加权平均，最终得到每个句子的向量表示，如表 3 所示。

表 3 部分句向量示例

Table 3 Examples of partial sentence vectors

文本内容	句向量表示
小试 ChatGPT，惊艳 …… 审核员 邮件 内容 复制 回复 …… 应声虫 级 算 得 体	[ 1.15752611e-02 5.20742945e-02 3.40225071e-01 -3.23876515e-02……]
终于 chatGPT 震撼 感觉 再也 理由 说 时间 整理 资料 写 memo 写 文件 改 合同 写 稿子 饭 嘴里	[-1.11070052e-02 1.50568234e-02 4.19823229e-01 -7.93736354e-02……]
ChatGPT 或者说 暴露 撒谎 胡说八道 不负责任 本来 NLP 模型 自然语言 natrual language processing	[ 2.96160523e-02 4.47431803e-02 3.18308145e-01 -7.35705420e-02……]
事实证明 我要 失业 chatgpt	[-7.95356408e-02 9.30919051e-02 5.01374483e-01 -2.02526897e-03……]

3.2.1 基于支持向量机的情感分类

在 Python 中导入 sklearn 库的 SVC 类和 GridSearch 模块，以实现网格搜索和交叉验证，用于在超参数空间中进行参数搜索和模型选择，找到最佳的模型超参数组合<sup>[29]</sup>。支持向量机模型的超参数及其对应候选取值如表 4 所示。其中，C

表示惩罚项系数，较小的 C 值会导致较大的间隔超平面，可以容忍更多的错误分类，防止过拟合，而较大的 C 值则会追求更高的训练准确率，但可能会导致过拟合；kernel 表示核函数类型；gamma 表示 rbf 和 poly 核函数的参数<sup>[30]</sup>。

表 4 支持向量机超参及候选值

Table 4 Support vector machine hyperparameters and candidate values

参数	候选值
C	0.1,1,10,100
kernel	linear,rbf,poly
gamma	0.01,0.1,1,auto

实例化 SVM 模型并用模型输出概率，随后便调用 GridSearchCV 进行网格搜索，交叉验证设置为 5 折，加载经 Word2Vec、GloVe 和预处理好的 FastText 词向量及其对应的标签对模型进行拟合。使用 grid\_search 找出的最佳超参数组合为：C=0.1，kernel=rbf，gamma=1。使用最佳模型对测试集进行预测，并生成分类性能报告。

3.2.2 基于随机森林模型的情感分类

从 sklearn.ensemble 模块中导入随机森林分类器，并同样使用网格搜索和交叉验证找到最佳的模型超参数组合。支持向量机模型的超参数及其对应候选取值如表 5 所示。其中，n\_estimators 表示决策树数量，增加数量可以提高模型的稳定性，但过多可能导致过拟合；max\_depth 表示决策树最大深度，控制深度有助于防止过拟合、节省计算资源；min\_samples\_split 表示节点分裂所需最小样本数；min\_samples\_leaf 表示叶节点所需最小样本数，增加此值可以防止模型过度细分，防止过拟合情况出现<sup>[31]</sup>。

表 5 支持向量机超参及候选值

Table 5 Support vector machine hyperparameters and candidate values

参数	候选值
n_estimators	100, 300, 500, 800, 1000
max_depth	5, 10, 15, 20
min_samples_split	2, 4, 6, 8, 10
min_samples_leaf	1, 2, 3, 4, 5

通过 `rf_model = RandomForestClassifier()` 实例化随机森林模型，调用 `GridSearchCV` 进行网格搜索，交叉验证设置为 5 折，同样加载经 `Word2Vec`、`GloVe` 和预处理好的 `FastText` 词向量及其对应的标签对模型进行拟合。使用 `grid_search` 找出的最佳超参数组合为：`n_estimators=500`，`max_depth=10`，`min_samples_split=4`，`min_samples_leaf=2`。使用最佳模型对测试集进行预测，并生成分类性能报告。

### 3.3 模型评价

#### 3.3.1 混淆矩阵

对于多分类问题，混淆矩阵<sup>[32]</sup>是一种常见的评估模型性能的方法。通过分析混淆矩阵，可以计算出各种评估指标，如准确率、召回率、精确率和 F1 值等，从而更全面地评估模型在每个类别上的性能表现<sup>[33]</sup>，结果如表 6 所示。

表 6 混淆矩阵

Table 6 Confusion matrix

真实值	预测值				
	Class 1	Class 2	Class 3	.....	Class N
Class 1	TP11	TP12	TP13	.....	TP1N
Class 2	TP21	TP22	TP23	.....	TP2N
Class 3	TP31	TP32	TP33	.....	TP3N
.....	.....	.....	.....	.....	.....
Class N	TPN1	TPN2	TPN3	.....	TPNN

(1) 准确率 (Accuracy)。准确率衡量了模型预测正确的样本在所有样本中所占的比例。

$$Acc = \frac{TP_{11} + TP_{22} + TP_{33} + \dots + TP_{NN}}{\text{总样本数}} \quad (2)$$

(2) 召回率 (Recall，也称为灵敏度、查全率)。召回率衡量了模型能够正确预测某一类别样本的能力。对于每个类别  $i$ ：

$$R = \frac{TP_{ii}}{TP_{i1} + TP_{i2} + TP_{i3} + \dots + TP_{iN}} \quad (3)$$

(3) 精确率 (Precision)。精确率衡量了模型在预测为某一类别时的准确性。

对于每个类别  $i$ :

$$P = \frac{TP_{ii}}{TP_{ii} + TP_{2i} + TP_{3i} + \dots + TP_{Ni}} \quad (4)$$

(4) F1 值 (F1-score)。F1 值综合考虑了精确率和召回率，是一个综合评估指标。

$$F1 = \frac{2PR}{P+R} \quad (5)$$

经过综合计算评估，汇总不同模型各项评价指标如表 7 所示。其中，BERT 模型表现最好，其次是 SVM+FastText 模型，而 SVM+GloVe 模型表现最差。整体来看，在该文本分类任务中，使用 FastText 词嵌入向量的模型较使用 TF-IDF 特征提取器和 GloVe 词向量的模型表现更好；使用支持向量机分类器的模型相对于使用随机森林模型表现也相对更好。

表 7 不同模型各项评价指标

Table 7 Evaluation indicators of different models

模型	准确率	精确率	召回率	F1-Score
BERT	0.9048	0.9197	0.9012	0.9104
SVM+TFIDF	0.7519	0.7432	0.7568	0.7499
SVM+GloVe	0.6911	0.7023	0.6749	0.6883
SVM+FastText	0.8315	0.8231	0.8543	0.8384
RF+TFIDF	0.7023	0.7341	0.7267	0.7303
RF+GloVe	0.7121	0.6958	0.7314	0.7132
RF+FastText	0.7968	0.8199	0.8256	0.8227

3.3.2 ROC 曲线和 AUC 值

在机器学习中可视化分类模型性能的常用工具有 ROC (Receiver Operating Characteristic) 曲线<sup>[34]</sup>和 AUC (Area Under the Curve)。

除了绘制 ROC 曲线并计算 AUC 值外，本次实验还使用了宏平均<sup>[35]</sup> (macro-average) 来计算 ROC。宏平均是一种计算多个类别的性能指标的方法，它将每个类别的性能指标单独计算，然后取平均值。考虑到本次实验的训练集、测试集中，不同情感倾向的评论数量差距较大，而宏平均提供了对整个多类别分类模型

性能的总体评估，因此在处理类别不平衡的问题时有着优良表现。

首先，将所有类别的 FPR 进行汇总并去除重复值；然后对于每个类别，使用 `interp()`函数在这些点上插入它们的 TPR 值。；最后，求平均值并计算 AUC 绘制函数曲线。结果如图 4 所示

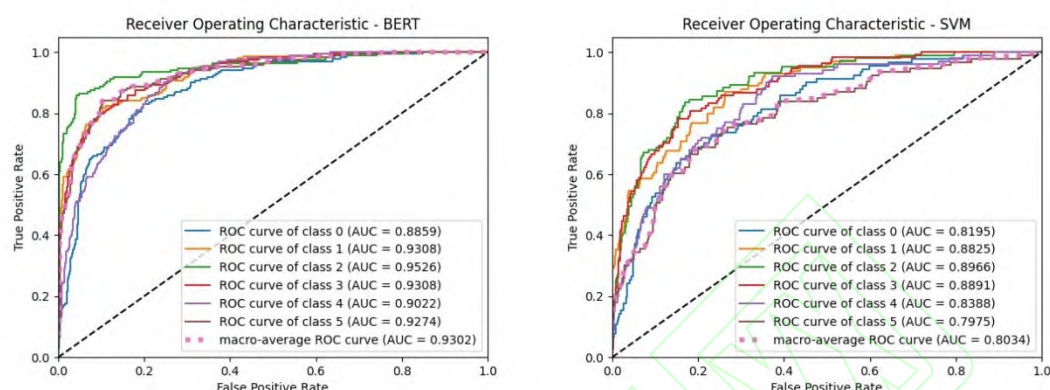


图 4 Bert 模型和 SVM 模型的 ROC、AUC

Figure 4 Bert Model and SVM Model's ROC and AUC

从图 4 来看，使用 FastText 作为词嵌入方法的 SVM 模型，其 ROC 曲线在某些类别上偏离理想位置，存在类别不平衡的问题；BERT 模型在所有类别上的平均性能较好，这表明 BERT 在处理多类别分类问题时具有较高的稳定性。从类别敏感性角度来说，BERT 模型整体 TPR 较高，表明能够有效地识别这些类别的正例，保障分类的准确性；尽管 SVM 在某些类别上的 TPR 较高，但整体来看，其在某些类别上的 FPR 也较高，这意味着模型在识别这些类别的正例时存在一定的误判风险。

综上所述，BERT 模型在情感六分类任务上表现出色，较使用 FastText 作为词嵌入方法的 SVM 分类模型来说具有更高的准确性和稳定性。

### 3.4 基于 BERT 模型的情感分类预测

经过模型评价，相较于传统的机器学习模型，BERT 预训练模型在各项指标方面都有着更优良的表现，因此，本实验选择 BERT 预训练模型作为评论文本情感预测模型。使用 `torch.load` 加载已训练好的 `pth` 模型，调用 `model.eval()` 将模型设置为评估模式，然后遍历数据加载器中的数据进行预测。对每一个批次的数据，模型会将其发送到指定的设备进行计算，并返回预测结果和类别概率值，同时将



这些结果存储在列表中。

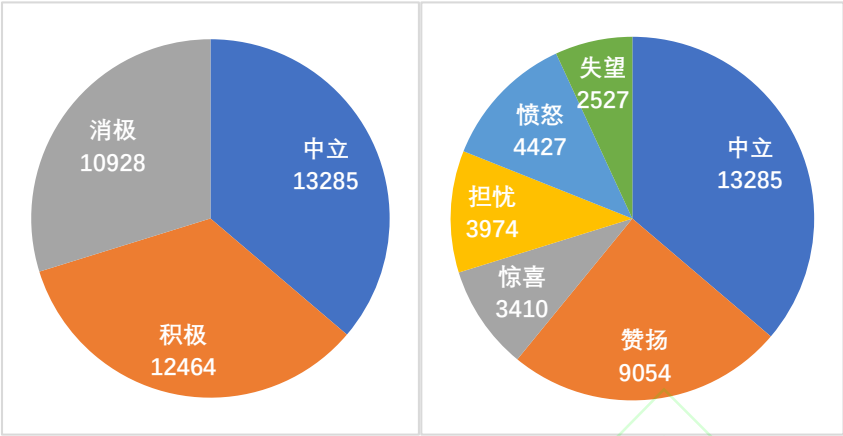


图 5 整体评论集各情感类别评论数量

Figure 5 Number of comments for each emotion category in the overall review set

各情感类别的评论数量结果如图 5 所示。整体评论集共有 36 677 条评论，积极、中立和消极情感评论占比相对平均，分别约为 34.0%、36.2%和 29.8%。在所有评论中，赞扬类别有 9054 条评论，约占总评论数的 24.7%；惊喜类别有 3410 条评论，约占总评论数的 9.3%；中立类别有 13 285 条评论，比例最高，约占总评论数的 36.2%；担忧类别有 3974 条评论，约占总评论数的 10.8%；愤怒类别有 4427 条评论，约占总评论数的 12.1%；失望类别有 2527 条评论，比例最低，约占总评论数的 6.9%。

综上所述可以看出，ChatGPT 作为一个人工智能技术产品受到了广泛关注，不同用户基于自身的经验、期望和价值观对 ChatGPT 做出的反响呈现出复杂且多元化的特点。这种态度的形成，究其原因是人们对新兴技术的期待与担忧并存，同时也反映了人们在接受新技术时的理性思考。理解这种多元化的情感态度，不仅可以作为改进技术的重要参考，对制定政策和指导未来的发展方向也具有不可忽视的价值。

## 4 基于 LDA 的评论主题聚类

### 4.1 主题聚类结果

本部分利用 Python 拟合 LDA 模型，在拟合模型之前，通过计算不同主题数下的困惑度(Perplexity)、一致性(Coherence)来确定最优主题数，本文使用 Python 自编程序进行两个指标的计算，以情感类别为“中立”的共 12 285 条评论集为例，

运用 LDA 主题模型抽取的不同主题数量的困惑度变化趋势、一致性变化趋势如图 6 所示。

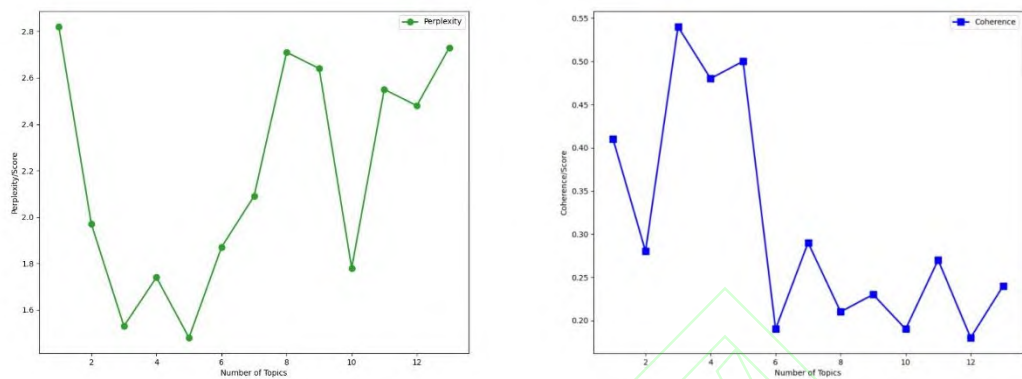


图 6 困惑度和一致性的分析结果

Figure 6 Perplexity results and coherence results

模型评价指标中，困惑度是衡量主题模型对新文档的拟合能力的指标，较低的困惑度意味着模型能更好地预测新文档中的词语分布；一致性用于衡量主题之间的语义相关性，一致性得分越高，表示主题模型生成的主题具有更高的语义相关性。从图 6 可以看出，主题数为 3 时，困惑度和一致性指标表现较好。因此，选择“3”作为“中立”评论集的最优聚类主题数。按照此流程，分别对剩余不同情感评论集进行主题数寻优，并将聚类结果整理成表。

表 8 持“赞扬”和“惊喜”态度评价文本中的潜在主题

Table 8 Potential themes in the text with "praise" or "surprise" attitude

主题一	主题二	主题三
表达能力	生产力	革新
理解	高效	突破
准确	实用	前沿
逻辑性	好	高科技
多样	便捷	功能
直观	智能	创新
规范	自动化	还行
细腻	功能	优异

表 8 为微博用户对 ChatGPT 持“赞扬”和“惊喜”态度的评价文本中的潜在主题。主题一中“理解”“表达能力”“逻辑性”“准确”“直观”表现出微博用户认为

ChatGPT 的交互体验和功能表现较好，能够提供准确、逻辑性强、表达清晰的回答；主题二中“生产力”“实用”“高效”“智能”表明一些用户赞扬它的实用性和效率，主题三中“前沿”“革新”“突破”等主题词展现了微博用户对 ChatGPT 大语言模型技术性和创新性的认可。

表 9 持“中立”态度评价文本中的潜在主题

Table 9 Potential themes in the text with a "neutral" attitude

主题一	主题二	主题三
真实	技术替代	国产
靠谱	就业	创新
验证	机会	优势
误导	失业	技巧
客观	技能	功能
来源	职业发展	规范
可信度	矛盾	革新
查证	版权	发展趋势

表 9 为微博用户对 ChatGPT 持“中立”态度的评价文本中的潜在主题。主题一中“真实”“可信度”“验证”“来源”“查证”等关键词反映了用户对 ChatGPT 所生成信息的真实性和可靠性的关注；主题二中“技术替代”“失业”“矛盾”等关键词表明用户对技术替代影响劳动力需求和就业机会的关注；主题三中“国产”“创新”“功能”“规范”“发展趋势”等词表明，ChatGPT 虽然在技术性和创新性方面具有一定的优势，但随着越来越多国产大语言模型的发布，更加先进、智能和高效的算法也会应运而生，并因此形成激烈的市场竞争，部分用户目前处于观望态势。

表 10 持“担忧”态度评价文本中的潜在主题

Table 10 Potential themes in the text with a "worried" attitude

主题一	主题二	主题三
泄露	伦理	失业
保障	造谣	平等
隐私	诚信	就业
权限	责任	劳动力
监管	真实性	矛盾
政策	偏见	感觉
风险	差劲	稳定
意识	版权	贫富差距

表 10 为微博用户对 ChatGPT 持“担忧”态度的评价文本中的潜在主题。主题一中“泄露”“保障”“隐私”“权限”“监管”“政策”“风险”“意识”等词语表明用户对于 ChatGPT 使用过程中个人数据安全和隐私保护存在着担忧,关注数据泄露、权限控制和监管机制的问题;主题二中“伦理”“造谣”“诚信”“责任”“真实性”“偏见”“版权”等词语体现了用户对于 ChatGPT 所生成的信息可能引发伦理和道德问题的忧虑,包括信息真实性、内容操纵、造谣行为等;主题三中“失业”“平等”“就业”“劳动力”“矛盾”“稳定”“贫富差距”等词语反映了用户对于 ChatGPT 可能对就业市场和社会经济带来的影响的担忧。

表 11 持“愤怒”和“失望”态度评价文本中的潜在主题

Table 11 Potential themes in the text with an attitude of "anger" or "disappointment"

主题一	主题二	主题三
昂贵	响应	深度
繁琐	冗长	专业
访问限制	重复	限制
注册	逻辑	局限性
兼容	崩溃	狭窄
门槛	答非所问	复杂
限制	差	细节
开放	矛盾	外行

表 11 为微博用户对 ChatGPT 大语言模型持“愤怒”和“失望”态度的评价文本中的潜在主题。主题一中“昂贵”“繁琐”“访问限制”“注册”“门槛”“开放”等主题词表现出微博用户对该模型的可用性和使用成本的不满;主题二中“响应”“冗长”“重复”“逻辑”“崩溃”“答非所问”等主题词揭示了部分微博用户在认为该模型的解答效率、逻辑性以及信息准确性等方面还有待提升;主题三中“深度”“专业”、“限制”“局限性”“狭窄”“复杂”“细节”“外行”体现了部分微博用户对 ChatGPT 大语言模型处理复杂问题、专业问题、细节问题和知识范围方面存在局限和限制的失望。

4.2 主题演变可视化

将评论文本经时间序列处理后再进行的 LDA 主题分析能够提供主题随时间变化的动态信息和趋势,能够帮助发现不同主题在不同时期的流行度和重要性变化,从而更好地理解话题的演变过程。

加载已分类的 36 677 条数据集，首先，通过主题数综合评价指标确定主题数为 3；其次，将对应的时间字段转换为以天为单位的时间索引，然后在上文 LDA 模型构建完毕的基础上计算每篇文档的主题分布情况，如表 12 所示。同时使用指数加权移动平均（Exponential Weighted Moving Average，EWMA）对每个主题的时间序列进行平滑处理；最后，使用 matplotlib 库绘制时间序列图，将每个主题的平滑后的时间序列绘制在同一张图上。最终结果如图 7 所示。

表 12 整体文本潜在主题

Table 12 Underlying themes of the overall text

主题一	主题二	主题三
真实性	技术替代	国产
伦理	就业	创新
误导	革新	革新
诚信	失业	突破
客观	技能	表达能力
来源	劳动力	还行
规范	自动化	理解
查证	结构失衡	昂贵

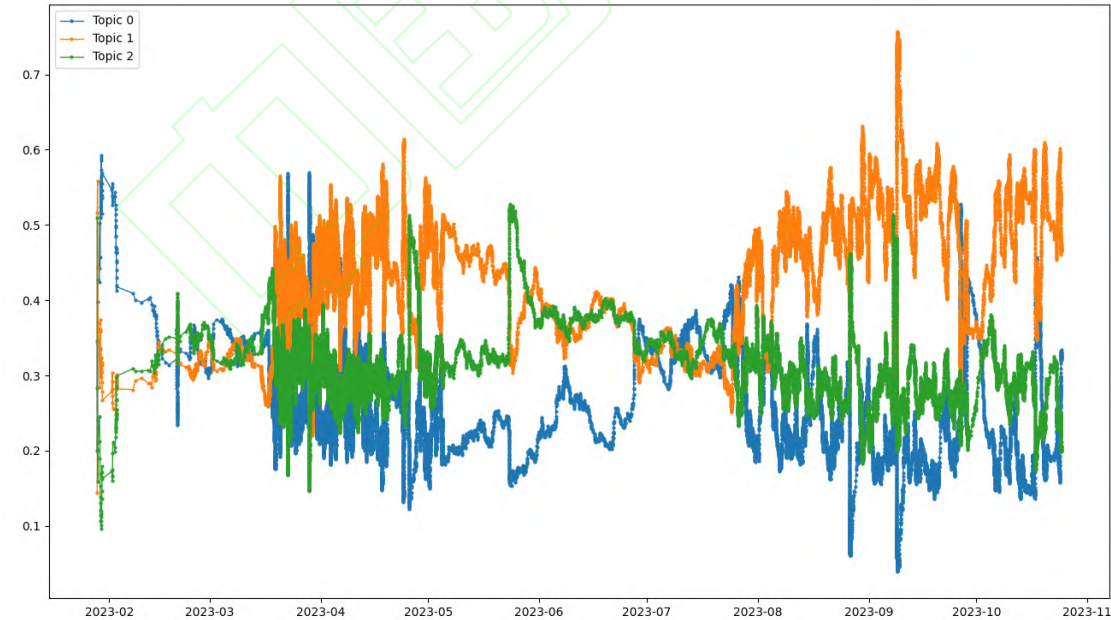


图 7 LDA 主题演变可视化

Figure 7 LDA topics evolution

如表 12 所示，主题一涉及的内容是关于用户对 ChatGPT 在信息真实性、伦



理标准、信息可靠性的关注，以及对其生成内容进行查证的需求。用户关注的焦点在于模型生成的文本是否准确、可信，并且希望可以验证信息的真实性。主题一的关注度呈“较高-较低-中等”的趋势。初始因用户对 ChatGPT 新技术的好奇而高，但随着不准确信息的出现，用户对其真实性和可信度产生担忧。GPT4 发布后，信息更加准确具备时效性，但误导性信息风险仍存，如 AI 绘图和未经验证信息，这也是用户对 ChatGPT 生成内容真实性进行质疑的原因之一。

主题二涉及了关于 ChatGPT 对就业市场的影响，包括技术替代和对就业机会、劳动力需求造成变化。这一话题始终保持较高讨论热度，并经历了两次明显的关注度提升。原因包括 2023 年大语言模型技术的显著进步和广泛应用，以及全球经济和就业格局的深刻变化，引发了对人工智能如何影响未来就业机会和劳动力市场的广泛关注。

主题三涉及的内容主要是关于 ChatGPT 国产化和相关技术的创新、优势和发展趋势的讨论。用户关注的焦点在于国内是否有能力开发类似的大语言模型，以及该模型在实际应用中的功能和创新表现，也表现出来对人工智能产品的价格方面的考量。主题三在整个时间序列中的关注度总体上保持平稳，用户对国产化和相关技术的讨论较为均衡，没有出现非常明显的高峰或低谷。

基于 LDA 主题聚类分析，ChatGPT 在广泛应用中所遭遇的挑战和风险呈现多元化特征，既涵盖其内部运作机制的局限性，如对训练数据质量的高度依赖、在专业领域应用的局限性以及缺乏透明度和可解释性，又涉及对外部社会环境的影响，表现为隐私保护难题、数据霸权隐患、诱发违法犯罪行为的可能性以及对学术诚信和自主学习能力的潜在冲击，甚至还包括因技术进步带来的结构性失业风险。鉴于此类挑战和风险展现出的复杂交织与广泛影响特性，制定并实施全方位、多维度的风险防控策略显得尤为迫切且必要。

## 5 基于时间序列的情感预测

### 5.1 时间序列预处理

本部分内容以上文评论情感分类为基础，将情感分类的结果转换为时间序列模型的输入，构建时间序列模型进行情感趋势预测。通过参考金融学领域学者对投资者情感量化的方法，本文综合整理出“简单情感指数”“看涨指数”和“情感差

异指数”三种情感量化计算公式。

简单情感指数（Simple Sentiment Index, SSI）。SSI 的取值范围为任意整数，当某时段积极情感的评论数量大于消极情感的评论数量时，SSI 值为正，反之则为负。简单情感指数可以较为直观地反映出网民在这一时段的情感倾向。

$$SSI=i^{pos} - i^{neg} \tag{6}$$

看涨指数（Bullish Index, BI）。参考 Antweiler 和 Frank<sup>[36]</sup>的方法，BI 的取值范围为[0,1]，看涨指数越高，说明越多的网民持积极的态度，事件的走向趋好的可能性就越大。

$$BI = \ln\left(\frac{1+i^{pos}}{1+i^{neg}}\right) \tag{7}$$

情感差异指数<sup>[37]</sup>（Sentiment Discrepancy Index, SDI）。SDI 是用来衡量不同群体或样本之间情感表达的差异程度的指标。SDI 的取值范围为[0,1]，情感差异指数越大，说明不同群体之间的情感差异越明显，反之则情感表现更加统一。

$$SDI = \left|1 - \frac{i^{neg} - i^{pos}}{i^{neg} + i^{pos}}\right| \tag{8}$$

在上述三项表达式中，指标  $i^{pos}$  代表积极情感文本的量化统计量，而  $i^{neg}$  则对应消极情感文本的量化统计量。

前文提到，数据收集阶段爬取了 2023 年 1 月 2 日至 2023 年 10 月 29 日的评论数据，基于上文的粗粒度分类结果，为了减少数据中的随机噪声，提高预测的准确性，本次实验以周为单位将数据进行合并，以期更好地捕捉潜在的趋势和季节性。将 2023 年 1 月 2 日至 2023 年 10 月 29 日共 301 天的数据合并成 43 周的数据，以周为单位，按照三种情感量化计算公式分别进行计算，汇总后得到如表 13 所示的部分周情感指数计算结果，并绘制如图 8 所示的三种情感指数随时间的变化趋势图。

表 13 部分周情感指数统计结果

Table 13 Statistical results of partial weekly emotion index

周数	SSI	BI	SDI
1	-7	-0.062242309	0.968609865

2	6	0.0588405	0.97029703
3	-9	-0.084557388	0.957345972
4	34	0.287682072	0.855932203
5	33	0.142790718	0.928416486
.....	.....	.....	.....
39	104	0.254404666	0.873170732
40	191	0.488802712	0.759748428
41	90	0.279584862	0.860681115
42	61	0.158981607	0.920469361
43	121	0.310757157	0.845466156

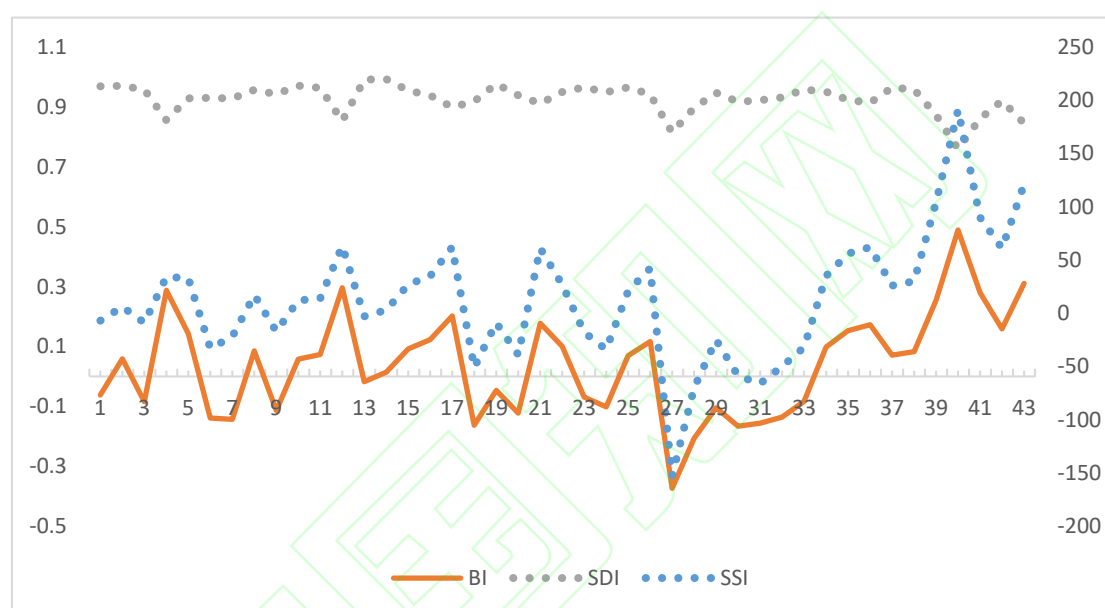


图 8 三种情感指数随时间变化趋势图

Figure 8 The trend of the three emotion indexes over time

从图 8 中可以看出，看涨指数 BI 的波动较大，表明网民情绪态度倾向变化大；情感差异指数 SDI 在 0.7 至 1 之间小幅波动，表明用户情感分歧显著。BI 是根据积极和消极态度评论数量的比例计算得出的指标，能够直接反映情感倾向，便于发现周期性的情感波动和长期趋势。SSI 的取值受到积极和消极评论数量的影响较大，如果某一时段的评论数量较少，即使其中存在着强烈的情感倾向，SSI 可能无法准确反映这种情感；SDI 更适用于比较不同文本或不同人群之间的情感差异，在反映单个文本或事件的情感倾向上有所欠缺。

因此，综合研究内容和数据特点，本文选择看涨指数 BI 作为原始时间序列。

## 5.2 时间序列模型构建

首先对原始序列进行平稳性检验。本文采用单位根检验方法（Augmented

Dickey-Fuller, ADF), 用于判断时间序列数据是否具有单位根, 从而确定序列的平稳性。

表 14 原始时间序列 ADF 检验结果

Table 14 Results of ADF test of original time series

ADF 检验结果	返回值
检验统计值	-0.7897
P 值	0.8221
置信度 1%	-3.6104
置信度 5%	-2.9391
置信度 10%	-2.6081

对原始时间序列的 ADF 检验结果如表 14 所示。其中 P 值为 0.8221, 显著大于 0.05, 为非平稳时间序列, 需要进行差分操作。一阶差分公式如下公式所示。

$$C(x)=x_t-x_{t-1} \tag{9}$$

其中,  $x_t$  为当前单位时间的值,  $x_{t-1}$  为上一个单位时间的值。经一阶差分后的时间序列如图 9 所示。

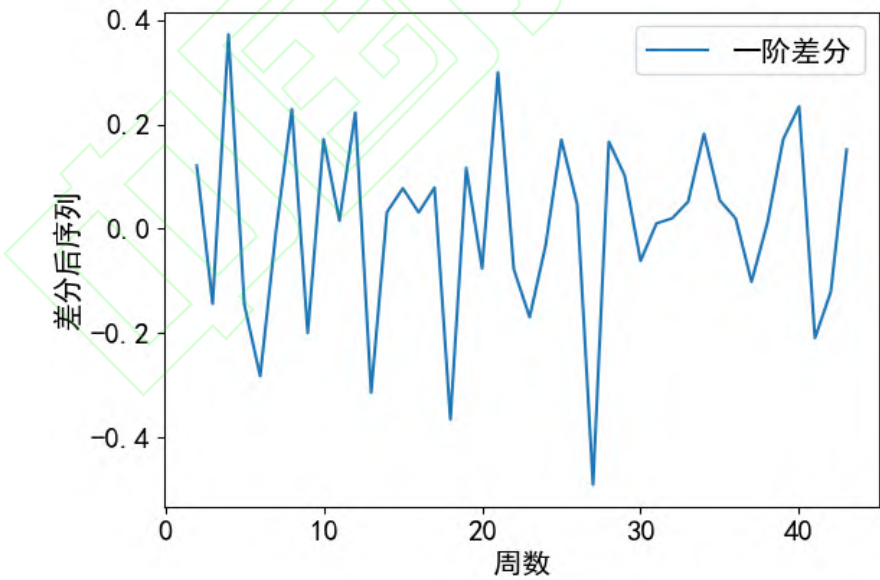


图 9 一阶差分后的时间序列

Figure 9 Time series after first-order difference

对一阶差分后的时间序列再进行 ADF 检验, 结果如表 15 所示。

表 15 一阶差分后的时间序列 ADF 检验结果

Table 15 ADF test results of time series after first-order difference

ADF 检验结果	返回值
检验统计值	-7.1450
P 值	0.0000
置信度 1%	-3.6104
置信度 5%	-2.9391
置信度 10%	-2.6081

在 ADF 检验过程中，发现 P 值低于 0.01，并且所有相关的统计临界值均满足标准，拒绝了原假设并通过了平稳性检验。

ADF 检验完毕后，继续对时间序列进行非白噪声校验。对于本研究的时间序列，其一阶差分序列的校验结果如表 16 所示。

表 16 白噪声检验结果

Table 16 Test results of white noise

白噪声检验结果	LB 统计量	统计量 P 值
原始时间序列	7.9751	0.0047
一阶差分时间序列	4.1176	0.0424

上述校验数据显示，一阶差分序列对应的 P 值低于 0.05 阈值，这证实了其已具备非白噪声属性。据此，接下来绘制一阶差分序列的自相关图和偏自相关图来确定模型参数 p 和 q 的合适取值。

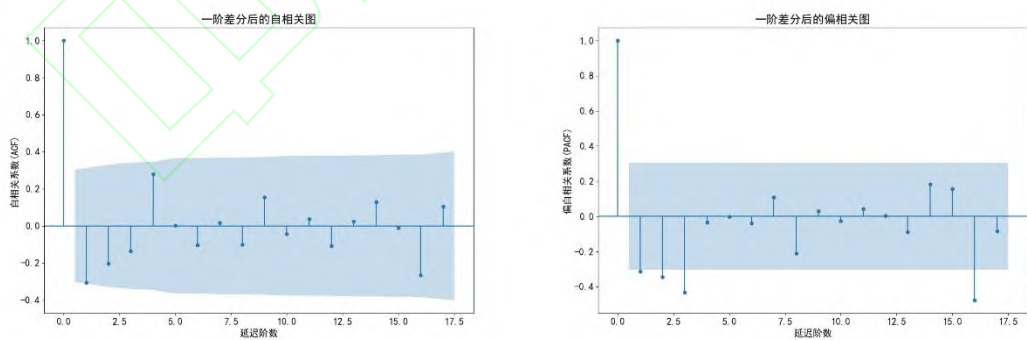


图 10 自相关图和偏自相关图

Figure 10 Autocorrelation graph and Partial correlogram

在图 10 的结果上，进一步分析最小信息准则和贝叶斯信息准则，最终确定了 p、q 分别为 0 和 1 的取值。结合早先在平稳性验证阶段所得到的结果，即差分阶数 d 的有效估计为 1，从而构建 ARIMA(0,1,1) 模型来对该时间序列进行分



析。

5.3 预测结果分析

将前 43 周的数据作为模型训练集，加载 ARIMA(0,1,1) 模型，对自 2023 年 10 月 29 日起未来 15 周的微博评论情感走向进行预测。得到如图 11 所示的预测结果。同时，绘制这 43 周的积极、中立和消极三种情感的评论数据的时间分布图，如图 12 所示。

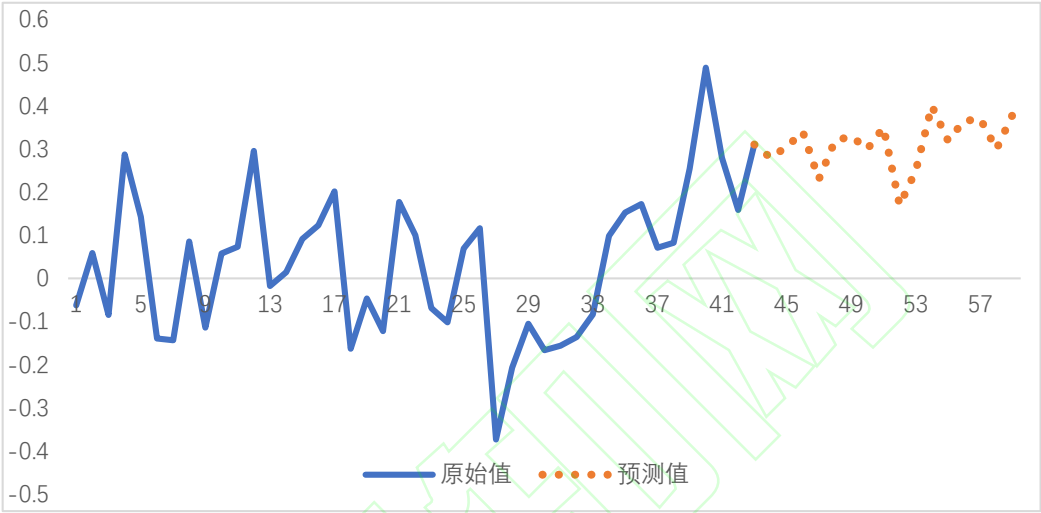


图 11 情感预测曲线图

Figure 11 Emotion prediction curve

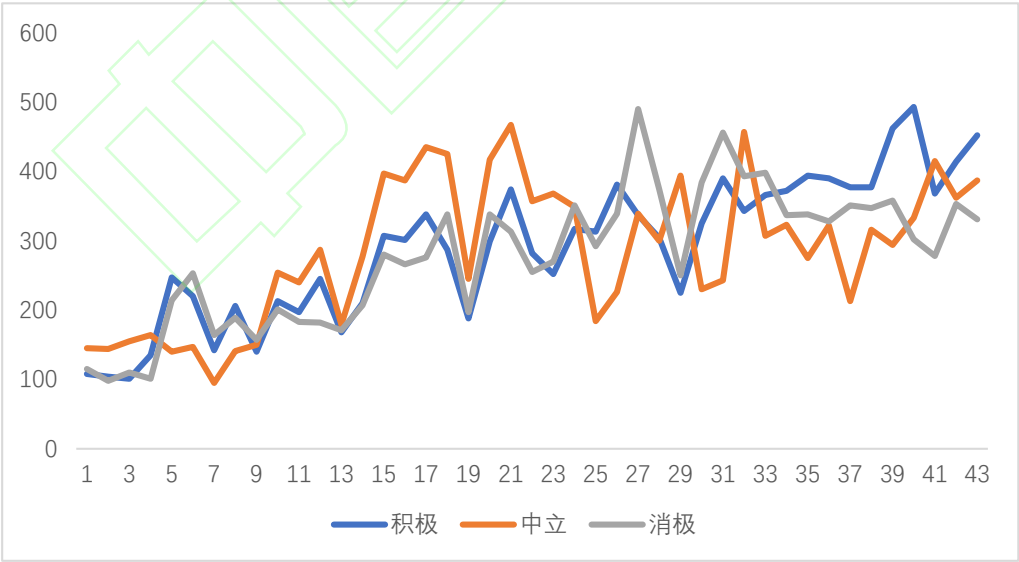


图 12 三种情感主类评论数据的时间分布图

Figure 12 The temporal distribution of the comment data of three affective main classes

在选取的 43 周内，BI 指数起伏变化较大，且显著地受到突发因素的影响。

三类情感评论数量均呈现波动性变化。整体而言，积极情感评论通常处于主导地位，在多次波动后保持较高水平；中立情感评论在整个过程中波动较为剧烈，在多个阶段呈现上升趋势；消极情感评论尽管整体数量较少，但在某些阶段如第 13 周至 27 周显著增多。

在所选取的 43 周时间内，2023 年 2 月初，OpenAI 推出 ChatGPT Plus，不久后推出 GPT-4，限制会员使用，开始踏上商业变现之路，引发了激烈讨论，三类情感评论数量均开始上升；2023 年 3 月 20 日，ChatGPT 宕机 12 小时，全球用户受到影响；2023 年 5 月 27 日，中国大陆 ChatGPT 用户遭遇大面积封号，受此负面影响，消极情感评论数量增加，BI 指数相应下降；2023 年 7 月 4 日，ChatGPT 浏览器联网模式下线，13 周后，联网功能回归，这一期间用户体验受到影响，BI 指数较低，联网功能恢复后 BI 指数增长迅猛；10 月 17 日，“文心大模型 4.0”正式发布，用户反响热烈，BI 指数表现起伏波动。

预测部分共 15 周（2023 年 10 月 29 日至 2024 年 2 月 11 日），结合实际情况，OpenAI 在 2023 年底至 2024 年初期间内对 ChatGPT 大语言模型进行了新增多模态功能、免费开发语音功能、降低会员费用等多项改进，相应地，预测的情感值同步提升，表明预测模型具有一定的准确性。但考虑到样本较少、时间跨度短等原因，预测模型仍有很大的提升空间。

ARIMA 模型是一种线性模型，它假设数据呈现线性趋势和固定的周期性，然而，真实世界中的数据具有复杂的非线性关系和非固定的周期性，对于长期的趋势和更复杂的季节性模式，ARIMA 模型往往表现出较差的拟合效果<sup>[38]</sup>。2023 年 11 月初，OpenAI 发布了 GPT-4 Turbo 和 GPT-s，优化了 ChatGPT 并下调了价格，但预测结果显示情感值在此期间大幅波动并下降。由此可见，ARIMA 模型无法很好地处理由突发事件引发的情感倾向变化，同时，随着时间进程的不断延伸，模型预测误差逐渐呈现出一定程度的增长，预测准确度有所下降，这一现象也符合时间序列预测模型的基本特征。

观察 BI 指数预测结果，未来用户对 ChatGPT 大语言模型的情感评价呈现出积极的趋势，更多用户可能会愿意尝试使用这一技术。可以认为，用户对 ChatGPT 大语言模型的积极情感评价受到技术发展、用户体验改善和社会认知提升等多个

方面因素的影响,这些因素共同推动了 ChatGPT 在用户中的普及和接受度增加。

总体来说,该情感预测模型在对微博用户评论情感走向进行预测的任务中,展现出了较为满意的效果。

## 6 相关结论与政策启示

### 6.1 相关结论

以 ChatGPT 为代表的大语言模型的广泛应用,在深度解放生产力、赋能诸多行业领域发展的同时,也带来了风险与挑战。本文通过基于 Bert 预训练模型构建情感分类模型,分类研究结果显示积极、中立和消极情感的评论分布相对均衡,分别为 34.0%、36.2%和 29.8%,赞扬、愤怒、担忧、惊喜和失望情感类别依次占总评论数的 24.7%、12.1%、10.8%、9.3%、6.9%。这表明在社交媒体上,用户对“ChatGPT 大语言模型”的态度是多元化的,既有对技术的期望和认可,也存在担忧和质疑。用户基于个人经验、期望以及价值观念的差异,对 ChatGPT 产生了复杂且多样化的反应。这种对新兴技术既抱有期待又存在担忧的复杂心理,也反映了用户在接受新技术过程中的理性考量。

在情感分类的基础上,使用 LDA 主题模型分别对细粒度分类评论集进行主题聚类,通过主题聚类,研究发现以 ChatGPT 为代表的大语言模型在广泛应用中所遭遇的挑战和风险呈现多元化特征,既涵盖其内部运作机制的局限性,又涉及对外部社会环境的影响,主要关注点呈现复杂化、多元化的特征。

通过建立情感时间序列,构建时间序列模型进行情感预测。预测结果发现用户对 ChatGPT 等大型语言模型的情感反馈正朝着积极方向发展,这表明随着时间的推移,此类技术在用户中的接受度和满意度会逐渐提升。这种积极趋势可能源于 ChatGPT 技术的持续改进、应用场景的不断拓展以及用户对其价值认识的深入,同时也表明 ChatGPT 等大型语言模型在未来有着广阔的市场潜力。

作为当前自然语言处理领域的一项重大突破,以 ChatGPT 为代表的大语言模型在其广泛应用的过程中,也随之带来了风险与挑战,政府、企业和个人作为影响和推动 AI 技术发展的重要参与者,采取全面、多层次的风险应对措施势在必行。

### 6.2 政策启示

在应对 ChatGPT 大语言模型广泛应用所带来的挑战与机遇时，政府、企业和个人各自扮演着不可或缺的角色，彼此间存在着紧密的联系与互动。政府、企业和个人三个层面应形成有效的联动机制，共同推动人工智能技术的安全、健康发展，充分释放其潜能，同时妥善应对风险挑战，以实现科技红利与社会福祉的最大化共享。

### （1）政府构建人工智能法制体系，建设战略生态引领实践

人工智能法制体系的建设包含着多方面的内容。首先，政府亟需建立一套健全的数据合规制度，详尽规定大型语言模型在数据全生命周期中各个环节的标准操作程序，设立独立公正的第三方数据审计机构；其次，政府应当加强对人工智能产业的反垄断执法力度，并制定完善 AI 生成内容的数据产权法律制度；此外，推进网络安全法制与监管机制建设，将利用 AI 工具从事诈骗、侵犯隐私权、制造及传播不良信息等活动列为违法行为，并确立相应的处罚标准。在网络空间治理层面，着重强化对 AI 生成内容的识别与审核机制，对非法获取和使用未经授权 AI 服务及其产出物的产业链条的查处力度，切断源头供给和技术支持。

在应对人工智能技术对社会伦理、教育体系和劳动力市场所带来的影响时，政府与教育部门还须前瞻布局教育政策，以确保人才培养适应未来发展的需求。同时，积极促进技术创新和市场竞争，推动行业健康发展。进一步完善科技成果奖励机制，设立专项奖项表彰该领域的突破性成果，加快相关技术从实验室到市场的转化速度。

### （2）企业重视数据治理和合规运营，技术创新与责任担当并行

在企业层面，构建全面的数据治理体系是确保 AI 语言模型开发与应用合规、安全的核心任务。这一体系应当紧密围绕隐私保护、数据质量控制以及内容安全性展开，并深度结合现行国际国内和地区性法规要求，实现从数据生命周期管理到模型训练及输出内容的全方位合规。企业要遵循严格的数据生命周期管理原则，提升用户对 AI 系统的信任度和接受度，重视用户需求。

其次，企业要深化和拓宽垂直领域专业知识，运用针对性的数据集进行强化训练，开发特定领域的垂类大模型，使其能够精准掌握并灵活运用各专业领域的基础理论、实践规范以及最新的研究成果。与此同时，主动承担社会责任，定期

发布翔实的营运报告来强化透明度和公信力。

### (3) 个人增强数字素养与安全意识，注重职业规划与技能升级

个人用户作为技术的直接使用者和受益者，要提升数字素养与信息安全意识，对日常网络活动中的各类敏感信息（如个人隐私、学术成果和工作数据等）采取高度警觉的态度和谨慎处理方式，避免信息泄露或滥用。对 AI 技术对各行各业产生的深刻影响，个人应当积极洞察行业动态、前瞻发展趋势，并主动对接市场需求，以确保自身的知识结构与核心竞争力与时俱进。树立自主学习意识，培养独立思考和批判性思维能力，紧跟时代步伐进行职业规划与技能升级，积极参与社会监督，推动人工智能技术与个人生活、工作深层次地融合。

## 7 结语

本文通过社交媒体大数据洞察用户情绪与观点，对关于“ChatGPT 大语言模型”的微博评论进行情感分析研究。研究发现，用户在对以 ChatGPT 为代表的大语言模型的积极、中立和消极情感相对平均，与此同时，用户对 ChatGPT 大语言模型的关注焦点呈现复杂化、多元化的特征；其次，通过 ARIMA 时间序列模型预测网民的情感态度，发现未来用户对 ChatGPT 大语言模型的情感评价呈现出积极的趋势。

本文也存在一些局限。首先，本研究选取微博作为主要的数据来源来洞察公众对 ChatGPT 大语言模型的情绪反应，虽然微博作为中国主流互联网社交平台之一具有较高的代表性，但其并不能完全反映整个社交网络环境中的舆论全貌。其次，本研究所抓取的微博评论数据的时间跨度仅为九个月，数据量较小，这在一定程度上限制了我们公众态度随时间演变趋势的深入理解。最后，尽管本研究采用了包括 Bert 在内的多种模型进行情感分析，并验证了 Bert 模型在该领域的优越性，但在比较模型效能上尚有一定的扩展空间。未来可以从纳入更多种类的主流社交平台数据、延长数据收集的时间段和引入更多的深度学习模型等方面开展进一步的研究。

## 参考文献



- [1] 任鹤云,王涛,冉晓琳.商业银行零售业务转型研究——以平安银行为例[J].中国集体经济,2024(2):108-111.
- [2] 苏洁.数字人将成 AI 领域新热点[N].中国银行保险报,2024-01-17(005).
- [3] 林建华,王蓉,李咏梅,等. “人工智能时代的未来高等教育研讨会”会议综述[J/OL].高等工程教育研究,1-7.
- [4] Stanford University Human-Centered Artificial Intelligence. THE AI INDEX REPORT[EB/OL].(2023-11-01)[2024.11.02]. <https://aiindex.stanford.edu/>.
- [5] 李白杨,白云,詹希旎,等.人工智能生成内容(AIGC)的技术特征与形态演进[J].图书情报知识,2023,40(1):66-74.
- [6] 中国信息通信研究院.人工智能生成内容(AIGC)白皮书(2022 年)[R/OL]. <http://www.caict.ac.cn/kxyj/qwfb/bps/202209/P020220902534520798735.pdf>.2024-03-23.
- [7] 王静静,叶鹰,王婉茹. GPT 类技术应用开启智能信息处理之颠覆性变革[J].图书馆杂志, 2023, 42 (5): 9-13.
- [8] 毛太田,汤淦,马家伟,等.人工智能生成内容(AIGC)用户采纳意愿影响因素识别研究——以 ChatGPT 为例[J/OL].情报科学,1-15[2024-11-03].<https://kns-cnki-net.wvpin.ncu.edu.cn/kcms/detail/22.1264.G2.20231103.1005.010.html>.
- [9] 李锋. 类 ChatGPT 人工智能背景下国家安全情报工作的机遇、挑战和应对[J].情报理论与实践, 2024, 47 (4): 98-104.
- [10] 吴成斌. 文本情感分类算法在舆情分析中的应用[D].江西:南昌大学,2021.
- [11] 黄萍,朱惠娟,陈琳琳. 基于深度学习的情感分类技术在高校舆情分析中的应用研究[J]. 软件工程, 2021, 24 (11): 59-62.
- [12] 王毅. 社交网络分析视阈下做好网络舆情管理的策略[J]. 兰台内外, 2020 (8): 34-35.
- [13] 赵爱,毕硕本,王军,等. 基于网络舆情的信息空间网络结构及特征研究[J]. 中国科技论坛, 2017 (11): 149-157.
- [14] 丁晓蔚.网民情绪分析及相应舆情风险管理研究——基于大数据热点事件[J].当代传播,2019(6):51-54.
- [15] LIANG D,DAI Z,WANG M, et al. Web celebrity shop assessment and improvement based on online review with probabilistic linguistic term sets by



- using sentiment analysis and fuzzy cognitive map[J]. Fuzzy Optimization and Decision Making, 2020, 19(4): 561–586.
- [16] 卢恒, 张向先, 闫伟. 重大疫情中网络舆情的多属性演化分析[J]. 情报科学, 2022, 40(1): 158-165, 192.
- [17] 詹希旎, 李白杨, 孙建军. 数智融合环境下 AIGC 的场景化应用与发展机遇[J]. 图书情报知识, 2023, 40 (1): 75-85, 55.
- [18] 赵浜, 曹树金. 国内外生成式 AI 大模型执行情报领域典型任务的测试分析[J]. 情报资料工作, 2023, 44 (5): 6-17.
- [19] 王静静, 叶鹰. 生成式 AI 及其 GPT 类技术应用对信息管理与传播的变革探析[J]. 中国图书馆学报, 2023, 49(6): 41-50.
- [20] 陆伟, 刘家伟, 马永强, 等. ChatGPT 为代表的大模型对信息资源管理的影响[J]. 图书情报知识, 2023, 40(2): 6-9, 70.
- [21] DWIVEDI K Y, HUGHES L, ISMAGILOVA E, et al. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy[J]. International Journal of Information Management, 2019, DOI:10.1016/j.ijinfomgt.2019.08.002.
- [22] 漆晨航. 生成式人工智能的虚假信息风险特征及其治理路径[J]. 情报理论与实践, 2024, 47 (3): 112-120.
- [23] 王婷瑜, 张新俏, 杜智涛. 权力的解构与强化: ChatGPT 重建知识生产秩序[J]. 数字出版研究, 2024, 3 (1): 43-50.
- [24] 段荟, 张海, 王东波. 信息资源管理领域科研人员对 ChatGPT 态度、认知及应对策略研究[J]. 情报理论与实践, 2023, 46(7): 17-24.
- [25] 张海, 刘畅, 王东波, 等. ChatGPT 用户使用意愿影响因素研究[J]. 情报理论与实践, 2023, 46(4): 15-22.
- [26] 宫承波. 新媒体概论[M]. 北京: 中国广播影视出版社: 2021.
- [27] 张冬, 魏俊斌. 情感驱动下主流媒体疫情信息数据分析与话语引导策略[J]. 图书情报工作, 2021, 65(14): 101-108.
- [28] AURELIEN GERON. 《Hands-On Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems》[M]. California: O'Reilly Media. 2019.
- [29] 丁晓蔚. 互联网金融全面风险情报体系构建研究——金融情报学的视角[J]. 情

报学报,2022,41(12):1280-1293.

- [30]汪海燕,黎建辉,杨风雷. 支持向量机理论及算法研究综述[J]. 计算机应用研究, 2014, 31 (5): 1281-1286.
- [31]刘勇,兴艳云.基于改进随机森林算法的文本分类研究与应用[J].计算机系统应用,2019,28(5):220-225.
- [32]郑翔,李明杰.中国古代书目提要结构功能识别研究——以《四库全书总目》著录的古代科技文献为例[J].图书馆杂志,2022,41(12):96-103.
- [33]DAVIS J.The Relationship Between Precision Recall and ROC Curves[C]//Proceedings of the 23th International Conference on Machine Learning, 2006:233-240.
- [34]FAWCETT T . An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006, 27 (8): 861-874.
- [35]欧石燕,陈嘉文.科学论文全文语步自动识别研究[J].现代情报,2021,41(11):3-11.
- [36]ANTWEILER W,FRANK Z M. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards[J]. The Journal of Finance, 2004, 59 (3): 1259-1294.
- [37]LUTZ B,PROLLOCHS N,NEUMANN D. Predicting sentence-level polarity labels of financial news using abnormal stock returns[J]. Expert Systems With Applications, 2020, DOI:10.1016/j.eswa.2020.113223.
- [38]刘杨,周浩月,陆锦琪,等.基于 Prophet 模型对流感样病例发病趋势的预测[J/OL].疾病监测:1-7.

**作者简介:** 朱益平 (1989-), 男, 江西南昌人, 博士, 教授, 主要从事政策文本分析、 用户信息研究, E-mail: zhuyipingnet@ncu.edu.cn

慕钰 (2000-), 女, 安徽六安人, 信息资源管理硕士研究生, 主要从事信息计量研究 , E-mail:muyu192540@163.com

孙逸宁 (2000-), 男, 湖北咸宁人, 图书情报硕士研究生, 主要从事政策文本分析、 数字图书馆研究,E-mail: tranquility0821@163.com