

分类号：
UDC：

密级：
学号： 412700220291

南昌大学专业学位硕士研究生
学位论文

ChatGPT 大语言模型的评论主题特征与情感分析研究

Research on Comment Theme Features and Sentiment Analysis of the ChatGPT
Large Language Model

孙逸宁

培养单位（院、系）：公共政策与管理学院

指导教师姓名、职称：朱益平 教授

专业学位种类：图书情报硕士

专业领域名称：

论文答辩日期：2024年05月26日

答辩委员会主席： 邹慧

评阅人：

2024年6月1日

摘 要

以 ChatGPT 为代表的大语言模型在赋能诸多行业并展现巨大潜力的同时，也带来了一系列不容忽视的问题与挑战。深入剖析用户对于 ChatGPT 大语言模型的情感反馈，不仅能为人工智能企业和从业者更好地推进 AIGC 技术的开发和应用提供一定参考价值，还能够为政府制定相关技术的监管策略和风险应对措施提供有益借鉴。

在爬取微博用户相关评论并完成数据预处理的基础上，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，全方位揭示公众对 ChatGPT 大语言模型的关注焦点和态度倾向，进而明确大语言模型所带来风险挑战的治理路径。结果表明，积极、中立和消极情感的评论分布相对均衡，大众对以 ChatGPT 为代表的大语言模型有着良好预期。与此同时，用户的主要关注点呈现复杂化的特点，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，又涉及到对外部社会环境的影响。

在此情况下，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。政府层面，建设人工智能法制体系和战略生态；企业层面，需在追求经济效益的同时，承担起数据治理与合规运营的重任，强化技术创新；个人层面，提高数字素养与信息安全意识，与时俱进地进行职业规划与技能升级，积极参与社会监督。三者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。

关键词：ChatGPT，大语言模型，机器学习，情感分析，治理路径

ABSTRACT

The emergence of large language models, epitomized by ChatGPT, has not only demonstrated remarkable potential across multiple industries but also brought forth a series of significant issues and challenges. A thorough analysis of users' emotional feedback towards the ChatGPT large language model can provide valuable insights for AI companies and professionals to advance the development and application of AIGC technology, as well as offer instructive references for governments in formulating regulatory strategies and risk response measures pertinent to such technologies.

Upon collecting and preprocessing relevant comments from Weibo users, a Bert model was constructed for sentiment classification, complemented with Latent Dirichlet Allocation (LDA) topic clustering and Autoregressive Integrated Moving Average (ARIMA) time series modeling. This comprehensive approach illuminated the public's focal points and attitude inclinations toward the ChatGPT large language model. The results showed a relatively balanced distribution among positive, neutral, and negative sentiments, indicating an overall optimistic anticipation of large language models exemplified by ChatGPT. Concurrently, user concerns manifested a complex nature, revealing a multifaceted array of challenges and risks encountered by ChatGPT in its widespread applications. These ranged from inherent limitations within its operational mechanisms to broader impacts on external socio-environmental contexts.

Under these circumstances, it is imperative for government entities, businesses, and individuals—each being critical influencers and drivers of AI technology development—to adopt comprehensive and multi-layered risk management strategies. At the government level, the establishment of a legal framework and strategic ecosystem for artificial intelligence is called for. At the corporate level, while pursuing economic benefits, enterprises must undertake the responsibility of data governance and compliance operations, reinforcing technological innovation. At the individual level, enhancing digital literacy and cybersecurity awareness, proactively updating career plans and skill sets, and participating actively in social

supervision are necessary. All three parties should forge an effective collaborative mechanism, collectively promoting the secure and healthy advancement of AI technology. They should strive to fully unleash its potential while adeptly managing associated risks and challenges, ultimately maximizing the shared benefits of technological dividends and societal welfare.

Key Words: ChatGPT, Large Language Model, Machine Learning, Sentiment Analysis, Governance Pathways.

目 录

第 1 章 绪论.....	1
1.1 研究背景.....	1
1.2 研究意义.....	3
1.3 文献综述.....	4
1.3.1 文本情感分析.....	4
1.3.2 网络热点事件情感分析.....	7
1.3.3 信息资源管理学科在 AIGC 领域的研究.....	9
1.3.4 文献述评.....	11
1.4 研究思路、方法、技术路线图.....	12
1.4.1 研究思路.....	12
1.4.2 研究方法.....	12
1.4.3 技术路线图.....	14
1.5 创新点.....	15
第 2 章 方法与模型.....	16
2.1 情感分类模型.....	16
2.1.1 Bert 预训练模型.....	16
2.1.2 支持向量机.....	17
2.1.3 随机森林.....	19
2.2 LDA 主题模型.....	20
2.3 ARIMA 时间序列模型.....	21
第 3 章 数据准备.....	23
3.1 数据来源.....	23
3.2 原始数据概览.....	23
3.3 评论情感倾向性标注.....	26
3.4 数据预处理.....	28
第 4 章 基于机器学习的评论情感分类.....	30
4.1 基于 Bert 预训练模型的情感分类.....	30
4.2 其他情感分类模型.....	33
4.2.1 基于支持向量机的情感分类.....	35
4.2.2 基于随机森林模型的情感分类.....	35
4.3 模型评价.....	36
4.3.1 混淆矩阵.....	36
4.3.2 ROC 曲线和 AUC 值.....	38
4.4 基于 BERT 模型的情感分类预测.....	40

第 5 章 基于 LDA 的评论主题聚类	42
5.1 模型构建.....	42
5.2 主题聚类结果.....	43
5.3 主题演变可视化.....	45
5.4 章节小结.....	47
第 6 章 基于时间序列的情感预测	49
6.1 时间序列预处理.....	49
6.2 时间序列模型构建.....	52
6.3 预测结果分析.....	54
第 7 章 对策建议分析	56
7.1 政府层面的策略与措施.....	57
7.1.1 建设人工智能法制体系.....	57
7.1.2 战略生态构建与实施引导.....	60
7.2 企业层面的行动方案.....	62
7.2.1 数据治理与合规运营.....	62
7.2.2 技术创新与社会责任担当.....	63
7.2.3 推进 AI 可持续发展与生态共建	64
7.3 个人层面的应对与自我调适	64
7.3.1 提高数字素养与信息安全意识.....	64
7.3.2 职业规划与技能升级.....	64
7.3.3 社会参与与监督.....	65
7.4 对策与建议小结.....	65
第 8 章 研究总结与展望	67
8.1 研究总结.....	67
8.2 研究不足与展望.....	68
参考文献.....	69

第 1 章 绪论

1.1 研究背景

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正在发生深刻的变革。在众多人工智能技术方向中，生成式人工智能(Artificial Intelligence Generated Content,AIGC)作为其前沿应用之一，在过去十年中从传统的基于统计语言模型向基于深度学习模型演进^[1]，通过神经网络建模，实现更准确地理解和分析语言、图像和音频等不同形式的输入，并生成更加自然和准确的输出。AIGC是指由人工智能生成的，包括文字、图片、音频和视频等各种形式的内容。在这些内容中，机器人可以通过学习大量数据或通过模仿人类思维过程来创作出独特的、具有创意性的成分，如新闻报道、小说、诗歌、音乐、绘画、动画等。这其中，最具有代表性的自然语言处理产品便是 OpenAI 于 2022 年 11 月 30 日发布的 ChatGPT-3.5 和一季度后推出的 GPT-4。

兴起于近年，由神经网络、深度学习等人工智能技术所推动的 GPT 并没有很长的历史，2015 年 OpenAI 成立专门研究实验室开展大语言模型(Large Language Models,LLM)研究，自 2018 年起发布初代 GPT 到 2022 年的 GPT-3.5，短短四年的时间，大语言模型的参数由 1.17 亿个扩大到 1750 亿个，预训练容量更是提升了九千多倍，达到了 45TB^[2]。训练环境的高度优化和改进使得 ChatGPT 获得了有目共睹的强大语言理解和生成能力，这些优秀的特性和性能，使得 ChatGPT 在发布后不到一周就拥有了超过 100 万的用户量，并且在 2023 年 4 月单月取得了全球访问量 17.6 亿次的惊人成绩，充分证明了它在人工智能领域重要的应用价值和市场竞争力。

近年来，中国政府也在人工智能领域发布了一系列政策，力求推动AI产业稳中向好发展。国务院于2017年发布《“十三五”国家战略性新兴产业发展规划》^[3]，将AI技术的发展上升至国家战略布局。同年7月，工信部发布《新一代人工智能发展规划》^[4]，提出了一系列重点任务和发展目标，目的是加快推动我国人工智能领域的发展和应用，并将其打造成为世界领先的人工智能创新中心。2020年7月，国家标准化管理委员会、中央网信办等五部门共同印发了《国家新一代

人工智能标准体系建设指南》^[5]，对AI领域标准体系研制的各项内容进行了详细描述，旨在推动中国人工智能标准化工作的发展，为新一代人工智能技术的研发、应用提供标准支撑，也为中国人工智能产业的提升提供了重要保障。2022年7月，科技部、教育部、工信部等六部门发布《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》^[6]，着眼于推动人工智能在重要领域和应用场景的深度融合，促进经济高质量发展。

总体来说，中国政府在人工智能领域的投入和支持力度不断增加，这也使得我国在人工智能技术研究和应用方面取得了一系列重要进展。在人工智能生成内容领域，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、数理推算、中文理解、多模态生成等方面具有广泛的应用价值^[7]；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。

在当前的网络环境和法律限制等多方面原因影响下，ChatGPT并未向中国内地和中国香港用户开放。但是，许多国内的开发者和机构已经将ChatGPT模型中文化，并开发了相应的中文镜像版GPT模型，提供给中国境内用户使用。这些自然语言处理产品为用户带来解放生产力的便利，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。国家新一代人工智能治理专业委员会于2019年6月17日正式发布《新一代人工智能治理原则——发展负责任的人工智能》，强调在推进新一代人工智能发展的过程中，各利益相关方应坚守道德伦理底线，确保人工智能技术的安全、可靠与可控，从而促进经济社会生态的可持续发展。为了加强对互联网信息服务深度合成技术应用的管理和监管，保障公众的合法权益和国家安全，中华人民共和国工业和信息化部、国家网信办、公安部于2022年11月25日发布了《互联网信息服务深度合成管理规定》^[8]；此外，在2023年4月11日，国家网信办也起草了《生成式人工智能服务管理办法（征求意见稿）》^[9]，向社会公开征求意见，体现了国家对人工智能技术发展过程中可能带来的风险的高度重视。与此同时，互联网上关于人工智能生成内容技术所带来的风险的争论愈演愈烈。有些人持积极态度，认为AIGC技术将会带来更加智能

的语音交互、自然语言处理等^[10]；而另一些人则持消极态度，担心 AIGC 技术会引发隐私泄露、数据造假、就业困难等问题**错误！未定义书签。**。

微博作为社交媒体平台，以其操作便捷、内容多元、信息扩散高效及更新频率高等特质，充分展现了大数据环境下的关键属性，并因此被普遍视为大数据时代极具代表性的媒介载体^[11]。尤其值得关注的是，新浪微博庞大用户基数日常生成的海量数据资源，已经成为学术界开展社会科学研究的重要资料来源。在上述背景下，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，并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，以期为国家政府在人工智能生成内容领域的风险应对提供一定思路，助力中国国内人工智能技术更长远、更稳健地发展。

1.2 研究意义

理论意义方面，舆情对于政策决策和社会发展具有重要的影响作用，对于舆情的分析和研究一直是学术界和实践领域的重要任务。人工智能生成内容技术的最具代表性之一是语言模型，而 ChatGPT 大语言模型则是目前最为出色和著名的语言模型之一，其发展与应用已经引起了广泛的关注和讨论，但是目前对于这一技术的舆论情感研究还较为有限。本文对新浪微博用户评论进行数据挖掘，深入探究大众对于 ChatGPT 大语言模型的关注点，为扩充人工智能生成内容技术的舆情研究提供了一定的参考价值。同时，本文采用了自然语言处理领域的前沿方法——Bert 模型进行情感分类，提高了研究结果的精度和有效性，将其与 LDA 主题聚类模型和时间序列模型相结合，进行多角度的综合分析，可以更加准确地理解和评估 ChatGPT 大语言模型在大众中的舆论表现和影响，有助于为同类型的研究提供新思路。

现实意义方面，对国家和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，同时为广大用户了解 AIGC 技术发展趋势和成长空间提供有益参考。通过对新浪微博用户评论的数据挖掘和情感极性分类，结合 LDA 主题聚类模型和时间序列模型，可以深入分析大众对于 ChatGPT 大语言模型的态度、关注点和未来预期，探究 ChatGPT 大语言模型的发展趋势和潜在风险，

以期为国家更加科学、全面、精准地监管 AIGC 技术，使其发挥最大的社会效益和经济价值提供可能的参考。此外，本研究还可以为人工智能企业和从业者提供有价值的信息，帮助他们了解市场趋势、用户需求和行业变化，为更好地推进 AIGC 技术的开发和应用提供一定参考。

1.3 文献综述

本文从文本情感分析、基于网络热点事件的情感分析和信息资源管理学科在 AIGC 领域的研究三个方面对研究现状进行梳理。

1.3.1 文本情感分析

传统的情感分析方法主要采用基于规则或词典的方法进行处理。情感词典是在语言学和心理学研究的基础上，根据人们在语言中表达的情感来建立的一种词库，包含了各种情感相关的词汇。在进行情感分析时，可以将文本中的情感词与情感词典进行匹配，从而确定该文本的情感极性。Liu^[12]在 2012 年提出情感词典 SentiWordNet，对情感词的极性进行了标注，并将文本的情感分析转化为文本中情感词的极性计算问题。王敏^[13]等使用英文的 SentiWordNet 和中文的知网词典进行情感词典构建，并制定了一系列规则，以帮助模型确定评论和字幕中单词或短语的情感极性，进行了视频弹幕和字幕的情感分析，最终发现弹幕的情感强度普遍高于字幕。这种于规则或词典的方法强调对情感词汇的掌握和利用，但是难以适应多样化和复杂化的情感表达方式。因此，不少学者将目光聚焦到针对文本分类任务的特征选择方法——SO-PMI 算法，该算法可以快速地从大规模的文本语料库中提取和筛选出与情感相关的单词，不仅大大缩短了情感词典构建的时间，也使其能够灵活地应用于不同的情感分类和情感分析任务。张云中^[14]等人通过文本挖掘和文献阅读获取国际关系智库文章和种子词，运用 SO-PMI 算法构建国际关系情感词典作为情感分析的支撑工具，依托情感趋同原理得到媒体型智库在国际关系话题上的舆论引导力排名。姜伶俐^[15]等人针对当前公共情感词典在微博语料覆盖度不足的现状，选取了 HowNet 以及大连理工大学情感词典作为微博基础情感词典，并创新性地提出了运用 Good-Turing 平滑技术优化的情感倾向点

互信息 (SO-PMI) 算法, 旨在精确识别并判定近年来大量涌现的网络情感词汇之情感极性, 得到了有效性较高的实验结果

然而, 情感词典方法的先天性不足限制了其适用范围和准确度。钟佳娃^[16]等人认为, 情感词的判别与选择取决于先验知识与实验设计, 不仅跨领域情感分析的分类效果不佳, 对于存在歧义或多义性的语料或长文本, 使用情感词典方法进行情感分析的效果也不尽人意。同时, 随着网络用语不断涌现, 情感词典方法的准确度必然会趋于下降。因此, 在传统情感分析领域, 基于规则和词典的方法逐渐被机器学习方法所替代。

基于机器学习的情感分析方法是目前应用最广泛的方法之一。这种方法通过使用大量的标注数据进行训练, 自动学习情感特征和模式, 从而可以更好地适应各种语言、领域和场景, 同时具有更高的准确性和泛化能力。基于机器学习的情感分类方法主要分为三类: 无监督学习、有监督学习以及半监督学习途径。

无监督学习不依赖预先标记的情感类别标签, 而是通过发掘文本内部以及文本之间的结构和模式来进行分类, 这种方法适合于报纸、学术文章等经过精细编辑校对后的结构化文本, 而在情感分析领域, 更多的是碎片化的、语义模糊的、缺乏上下文语境的、包涵各式各样表情符、缩略语甚至错别字的社交媒体语言。Turney^[17]率先实验性运用无监督学习对汽车、电影、银行和旅游目的地的评论进行情感倾向性分类, 将评论分类为推荐或不推荐。钟智锦^[18]等人认为, 结构话题建模、概率潜在语义分析等无监督算法常常难以识别人类在虚拟空间中发布的文字的真实含义, 正因如此, 有监督的机器学习应运而生。

在有监督学习中, 借助一组已知情感极性的训练样本集, 能够有效地对各类情感类别进行区分与划分。然而, 此类方法高度依赖于高质量、充足且精准标注的数据样本, 这意味着在实际应用中有监督方法通常要求投入大量的时间和人力成本用于人工标记和预处理数据集, 以确保模型训练的有效性和准确性^[19]。Pan g^[20]等人, 在其研究中首次运用了有监督学习技术对电影评论进行情感分类, 并对支持向量机 (SVM)、最大熵模型 (Maximum Entropy Model) 以及朴素贝叶斯分类器 (Naive Bayes Classifier) 这三种机器学习模型的表现进行了详尽比较, 有力证明了基于机器学习的分类技术在精度和效率上显著优于传统人工分类方法。徐琳宏^[21]等人选用手工标注的 Awas 共 8736 条引文信息做为数据集, 利用引文中情感表达的引文标识位置指引和情感词汇等特征, 采用支持向量机模型构建了引文情感的自动识别系统。唐慧丰^[22]等人曾开展对中文文本的情感分类实验,

使用了几种常见的机器学习方法,涉及到不同的特征选择方法、文本特征表示方法、特征选择机制、数据集规模和文本特征数量等方面的影响,发现针对短文本和多级情感分类问题,有监督的机器学习方法也能够取得不错的效果。杨爽^[23]等人创新性地提出了一种基于支持向量机的多层次情感分类方法论,此方法能够在情感表达、词汇词性以及语义内涵等多个特征层面进行深度挖掘,并成功实现了对情感的精细化五级分类。

半监督学习方法在情感分类任务中发挥着重要作用,特别是在缺乏大量标记数据的情况下,通过对未标注文本执行特征提取过程,能够有效提升情感分类的整体性能。Liu^[24]等人结合有监督与半监督机器学习技术,成功应用于微博用户立场的探测工作中。Jiang^[25]等人则提出了基于表情符号的情绪空间模型(ESM, Emotion Space Model)的半监督情感分类方法。相较于传统基准方法,ESM利用表情符号信息从未经标注的数据集中构建词语的向量表示,在情感分类任务上的表现取得了显著提升。

值得注意的是,传统机器学习方法中的分类特征一般只能针对特定问题,推广能力有限。相对于传统机器学习方法中使用的分类特征,深度学习模型具有更强大的表达能力和泛化能力,能够通过从数据中自动学习特征和模式,更好地捕捉文本与情感之间的复杂关系^[26],提高情感分析的准确性和效果,因此逐渐成为情感分析领域的主流方法。

2003年,Bengio^[27]等人提出使用一个三层前馈神经网络的语言模型进行建模,拉开了神经网络在自然语言处理领域的序幕。该模型成为了后来的深度学习自然语言处理模型的基础,经过不断演化,目前典型的神经网络学习方法有:卷积神经网络(Con-volutional Neural Network, CNN)、递归神经网络(Recurrent Neural Network, RNN)、长短时记忆(LongShort-Term Memory, LSTM)网络等。Li^[28]等人设计了一种以卷积神经网络为核心架构的中文微博意见摘要生成算法,引入了混合排序函数以量化和建模特征间复杂的语义关联性。实验证明,该方法在衡量情感分析性能的各项重要指标——准确率、召回率、精度以及AUC上均表现出卓越的性能水平,超过了传统的机器学习分类方法,这种方法为中文微博情感分析提供了一种新的参考。罗帆^[29]等人结合了循环神经网络和卷积神经网络的优势,构建了一个名为H-RNN-CNN的多层次深度学习模型。在该模型框架下,首先运用双层递归结构的RNN对文本序列进行深层次的上下文捕捉与建模,随后引入针对句子层级的处理机制,有效地解决了长文本情感分类中的复

杂依赖性和全局语境理解问题,从而提升了对长文本情感倾向识别的能力和准确性,为解决情感分析中面临的挑战提供了新思路。

2006年,Hinton^[30]等人提出了深度学习理念,构建了基于神经网络的模型体系,通过深入学习和解析数据的核心特征,有效提升了模型的学习能力。该方法利用深层网络模型对数据进行多层次的抽象和表示,实现了更加准确和高效的学习。近年来,随着深度学习技术的发展,基于深度学习的情感分析方法得到了更为广泛的应用。2017年,Google Brain 团队^[31]首度推出了 Transformer 架构,该架构摒弃了传统的循环神经网络机制,转而采用 Attention 机制构建全新的机器翻译模型框架,并在此基础上创新性地引入了多头注意力(Multi-head Attention)模块。顾军华^[32]等人设计了一种结合卷积注意力机制的模型,即 CNN_attention_LSTM,该模型旨在通过卷积运算捕获文本局部最优情感特征,并有效追踪文本情感极性转换过程中的语义变化信息。

在一系列自注意力机制(Self-Attention)的应用下,Google 团队于 2018 年 10 月发布为自然语言处理任务提供支持的通用的新型语言模型 BERT(Bidirectional Encoder Representation from Transformers)。它基于谷歌 2017 年发布的 Transformer 架构,使用海量的未标记数据进行预训练,从而使得模型具有了学习和理解文本数据的能力,并且可以通过不同任务间模型参数的共享和不断的微调来适应更加专业化和特定领域的任务,大大提高了模型的通用性和泛化能力^[33]。同时,Bert 模型采用双向上下文建模,即同时考虑上下文语境中的前后文信息,从而更好地捕捉文本中的关系和语义信息,在深度双向 Transformer 结构加持下,使得 Bert 模型能够以极高的有效性处理长文本序列^[34]。正是由于其优良性能和良好的泛化能力,本文使用 Bert 模型进行文本情感分类。

1.3.2 网络热点事件情感分析

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可靠和高效的技术支持。随着互联网的普及与社交媒体的发展,网络热点事件已经成为了当前社会各界广泛关注和研究的重要领域之一,了解公众对某个事件、话题或产品等的态度、情感和行为,在政府决策、商业市场、社

会舆情监测等领域都具有重要意义和价值。吴成斌^[35]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^[36]等人提出使用基于 CNN 的深度学习技术,设计了高校舆情设计系统框架图,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,获得了较好的分类性能。

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^[37]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^[38]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;此外,网络热点事件的传播数量在同一时间框架内也具有区域性。

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^[39]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,采用大数据技术对热点事件中网民的情绪进行实时客观测量,构建了国人网络情绪动态图谱。卢恒^[40]等人从社会学视角出发,结合情报领域定量方法,全面阐释和解构网络热点事件的社会本质和演进规律,采用了意见领袖影响力评价、LDA 主题模型和 SnowNLP 情感分析方法对重大疫情中网络热点事件意见领袖、主题分布和情感走势进行分析,从人群、内容和情绪三种社会属性揭示演化规律。李冉^[41]等人使用心理学理论从情绪、态度等方面对微博热点事件进行分析,从传播学角度探讨其传播特点和影响因素,深入剖析舆情背后公众的关注重点,最终得出结论:影响力权威媒体的官方账号影响力更大、公众更愿意相信官方消息、意见领袖的观点可以引导舆情。

1.3.3 信息资源管理学科在 AIGC 领域的研究

作为信息科学和计算机科学的交叉学科,信息管理和图书情报领域的学者也高度关注 ChatGPT 类生成式人工智能的发展和应用。李白杨^[42]等人立足于互联网的演化发展,从网络形态演进、内容生产、人机交互模式、网络资源组织四个层面梳理了 AIGC 发展的基础条件,以数据赋值、模型赋智、空间赋能三个维度为着力点,分别探讨了 AIGC 的技术特征、技术要素和发展阶段,最后认为,AIGC 与技术算法的融合应用,为信息组织、数据资产管理、用户研究和信息伦理四个角度的研究与实践带来了实质性的影响。詹希旒^[43]等人围绕生活场景、服务场景、文娱场景、科技场景和商业场景五大应用场景的自有特征,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,认为 AIGC 的发展机遇主要体现在基于群体智能的内容共生、基于数智融合的能力升维和基于增量市场的价值共创三个层面,同时提出 AIGC 的技术基础是数字科技的更新迭代,建构理念是创作空间的扩容增益,但其本质特征还是内容生态的创新发展。赵浜^[44]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,并采用专家打分法进行测评,详细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。

部分学者也对智慧图书馆领域的 AIGC 技术应用进行了探讨。蔡子凡^[45]等人在介绍 AIGC 技术的演进历程以及其在图书馆智慧服务中的应用场景后,通过对已有文献和案例进行分析总结,认为 AIGC 技术可以应用于图书馆智慧服务中,包括自动推荐适合购买的图书、提供便捷查询服务、收集并整合购书信息等;AIGC 技术可以提高图书馆决策咨询服务的精准度、可信度和卓越度;基于 AIGC 技术的信息素养教育模式可以提升学习效率和体验感。储节旺^[46]等人综合分析了相关文献,得出人工智能生成内容可以提高图书馆服务效率和质量、帮助个性化推荐阅读材料、数字化和索引实体收藏等结论。赵杨^[47]等人聚焦新一代人工智能背景下的智慧图书馆变革与发展趋势,从基础设施层、算法支撑层、数据资源层、业务功能层和服务应用层 5 个部分构建融合 AIGC 技术的智慧图书馆体系框架,进而从转型目标制定、服务体系重构、技术设备升级、内容矩阵建设和应用生态拓展等方面提出转型实现路径,通过深入剖析由 AIGC 技术引领的智慧图书馆转型框架、实施步骤及其面临的挑战,为数智化环境下智慧图书馆的转型升级与持

续发展提供了新的思路与解决方案。

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^[48]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 类技术应用正通过智能办公、自动文摘、自动综述和自动简报、机器翻译、媒体传播等对信息管理与传播产生颠覆性影响，同时也指出了这些技术应用存在的数据偏差问题、透明度问题、隐私问题、恶意使用问题等。陆伟^[49]等人从信息管理角度出发，认为 ChatGPT 等大型语言模型在信息加工、整合和生成方面具有强大能力，可以推动信息资源建设和 AI 技术升级，但需要考虑其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。游俊哲^[50]提出了行为规制模式和对研究者使用生成式 AI 的不同场景分配不同程度的义务，并建议对科研评价规则进行创新，以规避生成式 AI 可能带来的风险和问题。为了深入研究 AIGC 技术带来的机遇与挑战，Dwivedi^[51]等人汇集了计算机科学、市场营销、信息系统、教育、政策、酒店和旅游、管理、出版和护理等领域专家的 43 篇文章，在承认 ChatGPT 提高生产力的能力，并表示它可能会在银行、酒店和旅游以及信息技术行业带来重大收益，并加强管理和营销等商业活动的同时，也考虑到它的局限性、对实践的干扰、对隐私和安全的威胁，以及偏见、滥用和错误信息的后果。漆晨航^[52]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^[53]等人高度关注 ChatGPT 对知识生产秩序的挑战与重构，指出生成式 AI 通过深度学习和预训练模型颠覆了传统的知识生产和传播方式，在打破“知识资本”的垄断，并推动知识权力向多元、平衡发展的同时，也引发了知识膨胀、泡沫化问题，导致信息过载及有效信息处理难度增大。在关于人工智能应秉持“以技术为本”还是“以人为本”这一核心理念的争论中，梁少博更倾向于并坚定支持后者，他主张，推动“以人为本”的生成式 AI 发展，关键在于强化对底层理论及伦理价值体系的探索、深化对提升模型可解释性和公平性相关技术的研究，并且重视对其应用场景与管控机制的深入剖析^[54]。

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^[55]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，并运用解释现象学的方法对访谈结果进行分析，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，同时，在认知维度，对 ChatGPT 的认知主要包括技术

认知、社会认知、学科认知以及学术认知 4 个方面，以此为基础，提出了加强数据管理建设、提高人工智能素养及优化信息资源管理学科体系等应对以 ChatGPT 为代表的人工智能生成技术负面效应的策略与建议。张海^[56]等人以 ChatGPT 用户为研究对象，采用访谈的方式收集数据，共收集到 23 份访谈资料，分析后构建了 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，最后提出了加强技术本土化和信息安全保护、增加社会支持、引导用户正确使用等建议，以促进 ChatGPT 的本土化和提高用户的使用体验。

1.3.4 文献述评

通过梳理文献可以发现，随着时间推进，文本情感分析方法不断创新和完善，热点事件情感分析研究也在方法和领域上不断拓展和深化，为目前相关领域的研究提供了有效支持和引领；同时，学术界也聚焦 AIGC 技术的应用进行了一系列研究。但不足之处在于，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 自发布后引发轰动至今仅有半年时间，且其应用范围和形式多种多样，相关研究和分析相对较少；另一方面，图情学科对 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，仅有部分学者使用访谈、问卷等方法研究用户对 ChatGPT 的态度和认知，围绕该技术的舆论情感研究较为分散和局限。此外，在热点事件舆论情感研究中，预测情感趋势常见以从心理学、社会学等学科的角度出发进行定性研究，使用预测算法和模型进行预测分析的定量研究有待更全面、更完善的发展。

因此，本文采用自然语言处理领域的前沿方法——Bert 模型进行情感分类，将其与 LDA 主题聚类模型和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。

1.4 研究思路、方法、技术路线图

1.4.1 研究思路

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，编写爬虫代码获取有关“ChatGPT 大语言模型”的微博用户评论，通过文本预处理后，构造情感分类模型。为提高情感分析结果的准确性，本文基于自然语言处理领域先进研究成果 Bert 预训练模型构建情感分类模型，结合随机森林、支持向量机模型进行对比实验，确定最佳情感分类器。

其次，在利用训练完毕的 Bert 模型对爬取的微博用户评论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。

1.4.2 研究方法

(1) 文献分析法

充分利用中国知网、万方、Web of Science、百度、Google 等数据库或搜索引擎对国内外机器学习、情感分析和 GPT 类大语言模型的相关研究进行检索、阅读和整理，了解该领域研究成果，结合相关理论基础，确定本文的研究方向和内容。

(2) 文本分析法

从非结构化的在线评论中提炼出有价值的信息，将网络评论文本转化为结构化数据，进行深度文本挖掘，以揭示其中隐含的有效内容。运用 Python 爬虫技术，针对微博平台搜集了围绕“ChatGPT”话题的用户评论文本，并对这些数据进行了预处理。依托 Bert 预训练模型搭建了情感分类模型，并与随机森林及支

持向量机模型进行对比测试，通过多维度模型评价指标确定性能更优的分类模型，将其应用于微博评论的情感极性分类，并利用 LDA 主题模型挖掘用户对以 ChatGPT 为代表的大语言模型的主要关注点。

（3）时间序列分析法

时间序列分析是研究时间数列间关系及其内部规律的一种方法，主要是对随时间变化而有规律地变化的现象进行统计模型建立和预测，可用于探索现象的规律性、估计趋势与季节性。本文以评论情感分类为基础，构建 ARIMA 时间序列模型进行情感趋势预测。

1.4.3 技术路线图

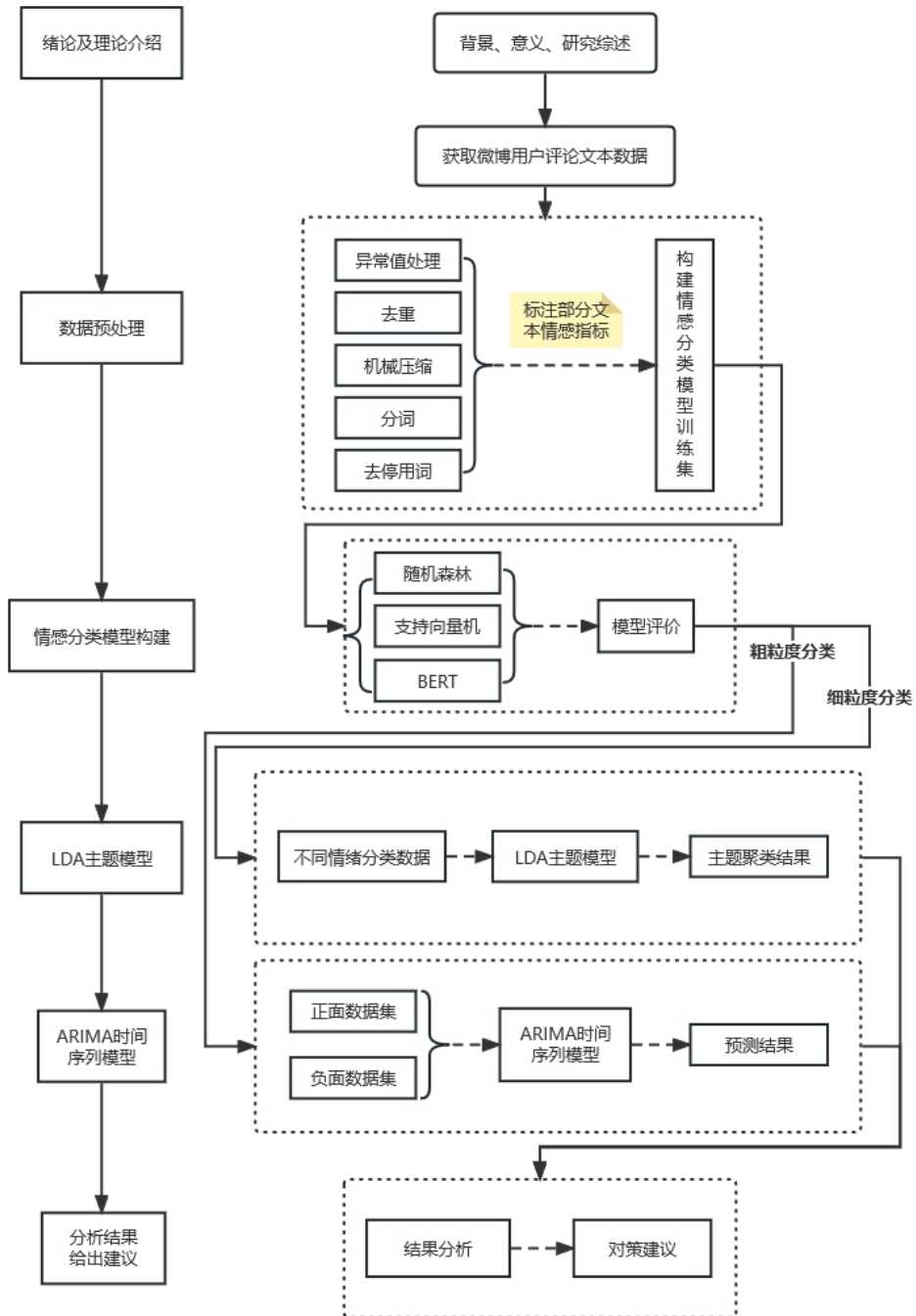


图 1.1 技术路线图

1.5 创新点

本文的创新点主要包括以下两个方面：

其一，本文在传统粗粒度情感类别的基础上进一步细分出细粒度指标，更加适应用户复杂情绪的评论分析任务，提高了情感识别的精度和深度。同时，多种情感分类模型（包括传统的随机森林和支持向量机等模型，以及 BERT 预训练模型）被运用到分类任务中，通过比较不同模型在多个评价指标上的表现，确定了本场景下最佳的情感分类模型，取得了较好的分类效果。

其二，在继承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，打破了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，相较于传统的情感分析而言，分析结果更加完善。

第 2 章 方法与模型

2.1 情感分类模型

2.1.1 Bert 预训练模型

BERT(Bidirectional Encoder Representation from Transformers), 是 2018 年 10 月由 Google AI 研究院提出的一种预训练模型, 它基于谷歌 2017 年发布的 Transformer 架构, 通过预训练来学习无标注数据中的深度双向表示, 是 NLP 发展史上里程碑式的模型成就。BERT 整体框架包含预训练和微调两个阶段。

BERT 的预训练阶段基于 Transformer 架构, 对输入序列进行并行处理, 并捕捉全局依赖关系。在这之中, BERT 使用了更具有先进性的 Masked Language Model (MLM)和 Next Sentence Prediction (NSP)预训练任务。

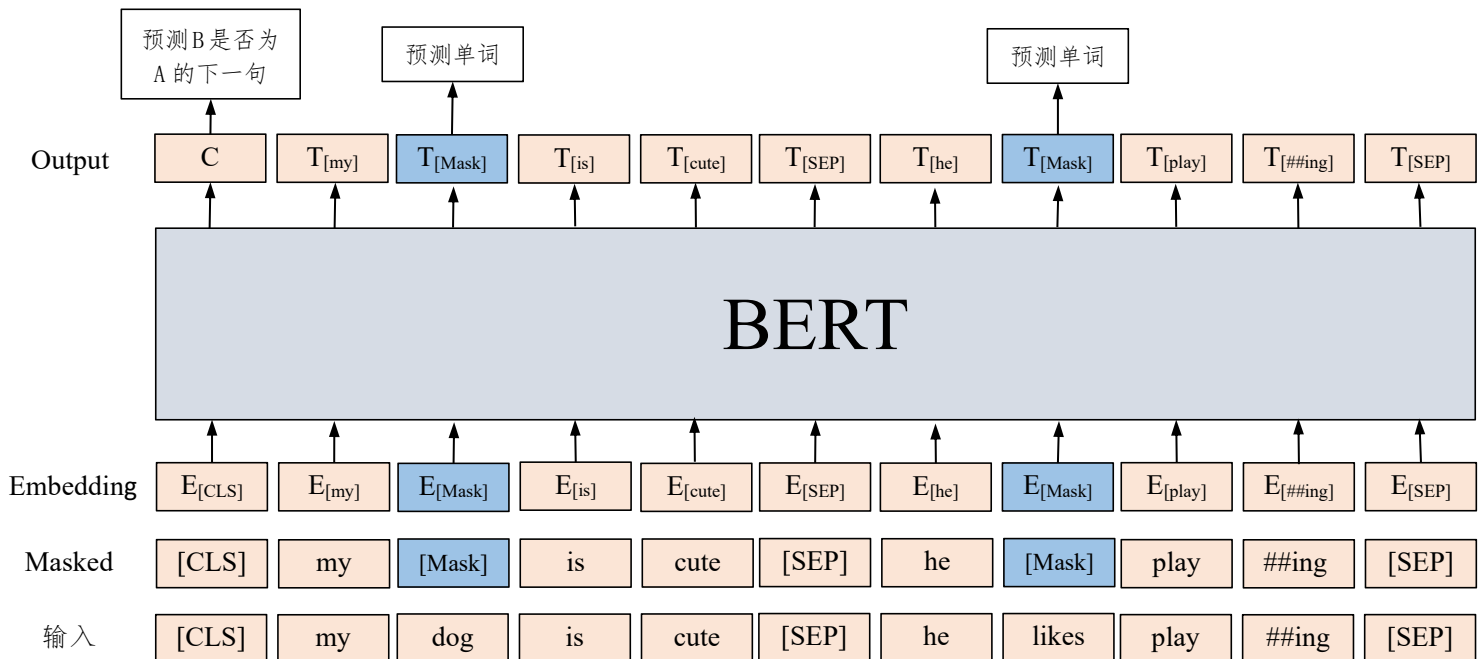


图 2.1 Bert 模型结构图

在 MLM 任务中, BERT 输入序列中的一些随机选择的词会被遮盖(用[MASK]符号替换), 模型需要通过上下文来预测这些遮盖的词。在 NSP 任务中, 模型

通过预测上下句之间是否连贯来得出句子之间的关系，对于每个句子对，有 50% 的概率是正确的连贯上下句配对，而另外 50% 的概率是随机选择的不连贯上下句配对。MLM 和 NSP 任务在 BERT 的预训练阶段起到关键作用。MLM 任务促使 BERT 学会利用上下文信息进行词预测，以获得更准确的词向量表示；而 NSP 任务则有助于 BERT 学习句子级别的语义关系，提高对句子级任务的迁移能力。通过这两个任务的结合，BERT 模型能够学习得到更丰富、更全面的语言表示，为下游任务的微调提供更好的基础^[57]。

在预训练后，BERT 模型通过微调来适应各种下游任务。微调阶段通常包括任务特定的网络结构和少量的任务特定参数，通过在下游任务上进行有监督的训练，BERT 可以根据具体任务进行精细调整，并实现更好的性能^[58]。

在本文的分类任务中，不对 BERT 预训练模型进行微调，首先使用预训练的 BERT 作为特征提取器，加载预训练分词器 Bert-Base-Chinese，然后在 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。将数据集划分为训练集、验证集和测试集，选择 AdamOptimizer 作为优化器，并使用交叉熵损失函数作为目标函数。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：

$$Loss = -(y \log P + (1-y) \log(1-P)) \quad (2.1)$$

在模型内加载训练集的评论文本，根据预测结果和真实标签计算损失，并使用反向传播算法更新模型参数，重复这个过程，直到模型收敛或达到设定的训练轮数。最后，使用验证集评估模型的性能，计算模型在验证集上的准确率、精确率、召回率等指标，并根据结果进行模型调优，如调整超参数、调整网络结构等，最终训练出最优表现的情感分类模型。

2.1.2 支持向量机

支持向量机是由 V.N.Vapnik 等人于 1968 年提出的一种有监督机器学习方法，主要思想是通过找到一个最优的决策边界（超平面）来将不同类别的数据样本分开，这个决策边界使得距离两个类别最近的样本点到该边界的距离最大化，从而具有更好的泛化能力^[59]。对于线性 SVM，可以直接使用原始特征；对于非线性 SVM，需要通过核函数将数据映射到高维特征空间来实现。

由图 2.2 所示，针对两类待分类文本，存在一个能有效区分这两类文本的最优超平面 H，以及两个与最优超平面 H 平行且距离最近的辅助超平面 H1、H2，它们之间的间隔大小为 $2/\|w\|$ 。因此，在选择最佳超平面时，应确保其既能准确划分两类文本数据集，又能最大化间隔值 $2/\|w\|$ 。这种算法在面临统计样本数量较少的情况下，仍能有效地发掘出良好的统计规律性。

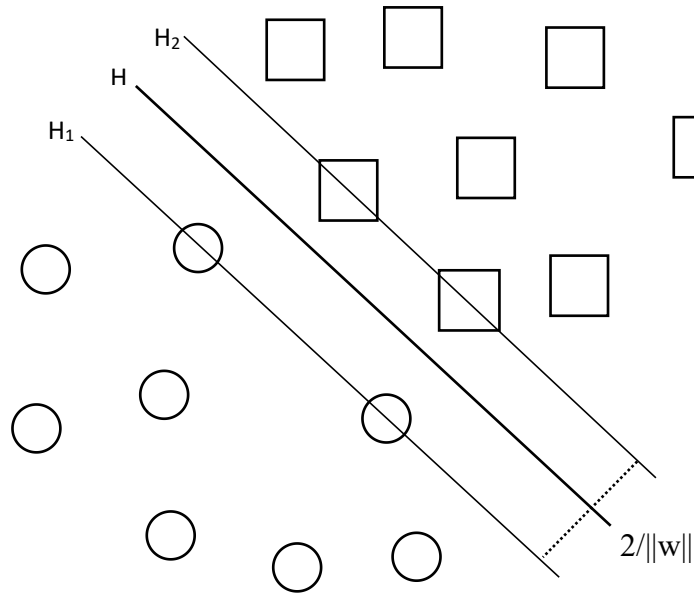


图 2.2 支持向量机模型最优分类超平面

假设样本集 (x_i) ，其中， $i \in [1, n]$ ， d 代表特征的数量或维度，每个样本 (x_i) 都有 d 个特征， $y_i \in \{-1, 1\}$ ，线性空间分类函数为 $g(x) = wx + b$ ，则分类超平面满足：

$$y_i(w * x_i + b) - 1 \geq 0 \quad (2.2)$$

理想的分类超平面应当确保上述公式恒成立，并在此基础上尽可能使 $\|w\|$ 的值最小化，这样的超平面即被视为最优分类边界。通过运用最优化理论求解得到 w 的理想取值，随后结合支持向量的概念来确定偏移量 b 的最优解，从而构建出最佳的分类函数模型：

$$f(x) = \begin{cases} 1 & w * x + b * > 0 \\ 0 & w * x + b * = 0 \\ -1 & w * x + b * < 0 \end{cases} \quad (2.3)$$

支持向量机算法在解决各类问题时表现优秀，特别是在小样本、非线性分类场景下展现出了强大的适应能力。当应用于文本类别划分任务时，该算法能够实现较高的分类准确度和召回率，有效提升文本分类效果。

2.1.3 随机森林

随机森林 (Random Forest)是一种用于分类和回归的强大的集成学习算法，其核心原理是通过构建多个决策树，并将它们的结果组合来进行分类^[60]。

构建决策树主要包括以下步骤：首先，从训练数据集中采用有放回抽样的方式随机选择一个子集，用于构建决策树；其次，从根节点开始，根据某个优化准则（如信息增益、基尼系数）选择最佳的特征和切分点，将数据集分割成两个子集，然后，递归地在每个子集上重复这个过程，直到达到最大深度或节点中的样本数小于阈值为止，最终得到决策树的叶节点，每个叶节点代表一个类别或一个回归值。值得一提的是，随机森林模型引入了随机选择特征和随机抽样数据两个随机因素，较为有效地增加了模型的多样性和准确性。

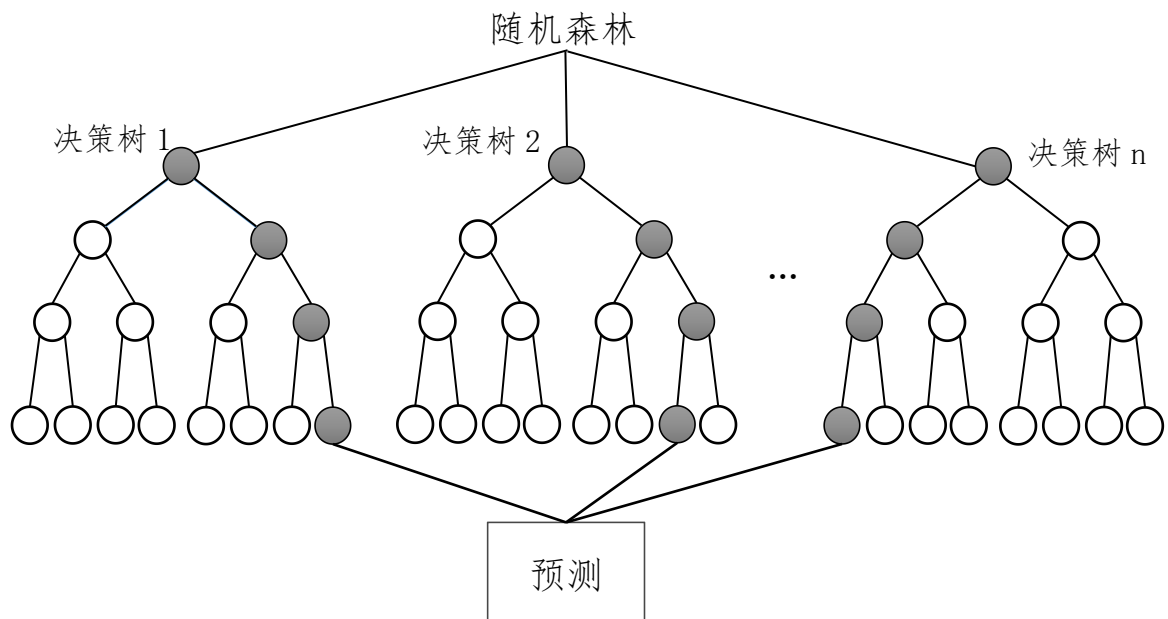


图 2.3 随机森林模型示意图

当需要对一个新的样本进行分类时，随机森林模型会将该样本输入到每个决策树中进行分类，采用多数表决的方式，即选择得票最多的类别作为最终的预测

结果。

2.2 LDA 主题模型

LDA (Latent Dirichlet Allocation) 模型是一种用于主题建模的概率生成模型, 由 D. M. Blei^[61]等人在 2003 年提出。作为一种无监督学习算法, 它能够从大规模文本语料库中发现隐藏的主题结构。

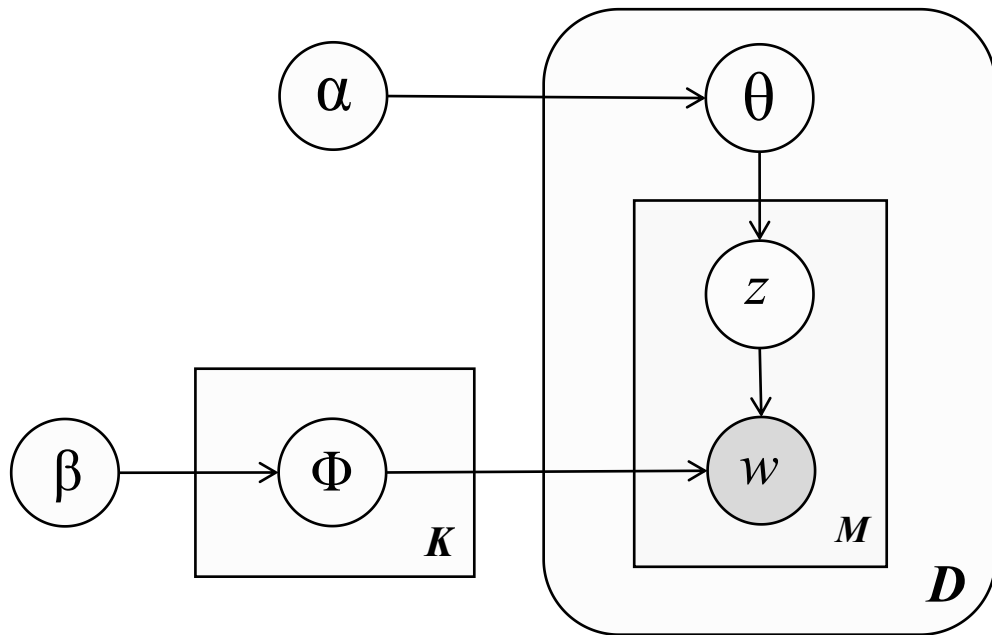


图 2.4 LDA 模型示意图

LDA 模型生成过程如图 2.4 所示, 其中, α 和 β 为分布超参数, Φ 为主题随机采样词分布, K 为设置的主题数, θ 为随机采样主题分布, D 表示训练语料的文档数, M 表示一篇文档中的词语量, z 和 w 分别表示某个主题和词语。首先, 对于每篇文档, 从主题分布中随机选择一个主题; 然后, 根据选定的主题, 从该主题的词汇分布中选择一个词汇; 重复这个过程, 直到为文档中的每个位置生成一个词汇。通过这种生成过程, LDA 模型可以模拟文档的生成过程, 并通过观察到的文本数据推断出每篇文档的主题分布和每个主题的词汇分布。

在 LDA 模型中, 主题分布和词汇分布都是通过概率分布进行建模的, 主题分布服从狄利克雷分布, 而词汇分布则是基于主题的多项式分布。通过使用 Gibbs 采样等方法, 可以对 LDA 模型进行参数估计和推断, 得到每篇文档的主题分布

和每个主题的词汇分布,可以帮助我们从小量文本中抽取有用的信息,并支持进一步的数据分析和决策制定^[62]。

2.3 ARIMA 时间序列模型

自回归差分移动平均模型 (Autoregressive Integrated Moving Average model, ARIMA) 是时间序列预测分析方法之一,该模型基于前期的数据拟合来预测未来的数值变化^[63]。其中, AR 表示自回归 (Auto Regression), I 表示单整阶数 (Integration), MA 表示移动平均模型 (Moving Average)。

在时间序列分析中,平稳性是指经过多次差分后得到的序列能够在未来一段时间内保持原有的形态和惯性,只有为平稳序列时才能够对时间序列进行分析建模,一般来说,验证平稳性通常采用观察法和单位根检验法。观察法是指通过直观地观察时间序列的趋势图,当变化曲线均匀分布在零刻度线上下两侧时,可以初步认为序列平稳;单位根检验法则是使用例如 Augmented Dickey-Fuller test (ADF 检验) 等方法来进一步确定时间序列的平稳性。

单位根过程的存在会导致序列不平稳,并使回归分析中出现伪回归,通过 ADF 检验中的 ADF 值和 p 值,可以判断序列是否满足平稳性要求。一方面,根据不同程度拒绝原假设的统计值 (如 1%、5%、10%),当 ADF 值越小,越倾向于拒绝原假设,说明序列不存在单位根,即时间序列是平稳的;反之,如果 ADF 值较大,则表明时间序列是不平稳的,需要进行差分转化为平稳序列;另一方面,当 p 值小于 0.01 时,也可以认为序列是平稳的。

在 ARIMA 模型中, p 、 d 和 q 是三个重要的参数: p 表示模型中自回归部分的阶数,即当前观测值与前面 p 个观测值之间的关系; d 表示使时间序列平稳所需进行的差分次数; q 表示模型中移动平均部分的阶数,即当前观测值与前面 q 个误差 (残差) 的线性组合之间的关系。

表 2.1 ARMA 模型阶数的判断

模型 (序列)	自相关函数	偏自相关函数
AR(p)	拖尾	第 p 个后截尾
MA(q)	第 q 个后截尾	拖尾
ARMA(p , q)	拖尾	拖尾

如表 2.1 所示,在平稳时间序列中,通过分析自相关函数 (ACF) 和偏自相

关函数 (PACF)，可以初步判断 ARMA 模型的阶数 p 和 q 。其中，拖尾指序列以指数率单调递减或震荡衰减，而在大于某个常数 k 后，序列快速趋于 0 的情况则称为 k 阶截尾^[64]。在得到最终的 p 、 d 和 q 三个参数后，构建 ARIMA 模型并进行模型检验。

第3章 数据准备

3.1 数据来源

网页爬虫的流程可以简单理解为发送 HTTP 请求、接收响应数据、解析 HTML 源代码、提取并存储数据、循环遍历五个步骤^[65]，本文对微博用户发布的有关 ChatGPT 大语言模型的评论进行爬取，使用 requests 库发送 HTTP 请求，获取相应内容。具体爬取步骤如下：

第一步，导入 requests、pandas 和 fake useragent 等库，使用字典形式设置请求头 headers，并定义 cookie_list，其中包含多个 Cookie 字符串，用于随机选择一个作为请求的 Cookie；

第二步，分别定义函数 weibo_LongText_data 和 weibo_data，前者接收一个微博的 mid 参数，通过发送 HTTP 请求，请求该微博的详情页面，然后从返回的 JSON 数据中解析出全文内容并返回，用于获取长微博的全文内容；后者接收关键字 keyword 和最大页数 max_page 两个参数，用于获取普通短微博的内容，在函数内部，通过循环发送 HTTP 请求，获取搜索到的、JSON 格式的不同页数结果，然后在返回的 JSON 数据中解析出的用户昵称、发布时间、微博内容、IP 国家属地和 IP 省份属地等信息，并保存到列表中。

3.2 原始数据概览

本文以“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 48679 条数据，去重后剩余 44474 条。对评论数据文本长度作描述性统计分析，最短的一条评论是 5 个字符，最长的一条评论是 6080 个字符，平均每条评论文本长度为 137 个字符，评论长度中位数为 67，评论长度标准差为 232；文本长度在区间上的分布如图 3.1 所示，数据整体呈现出右偏分布，字数特别多的评论占少数。由表 3.1，长度小于 200 的评论数据达到整体数据量的 82.5%。

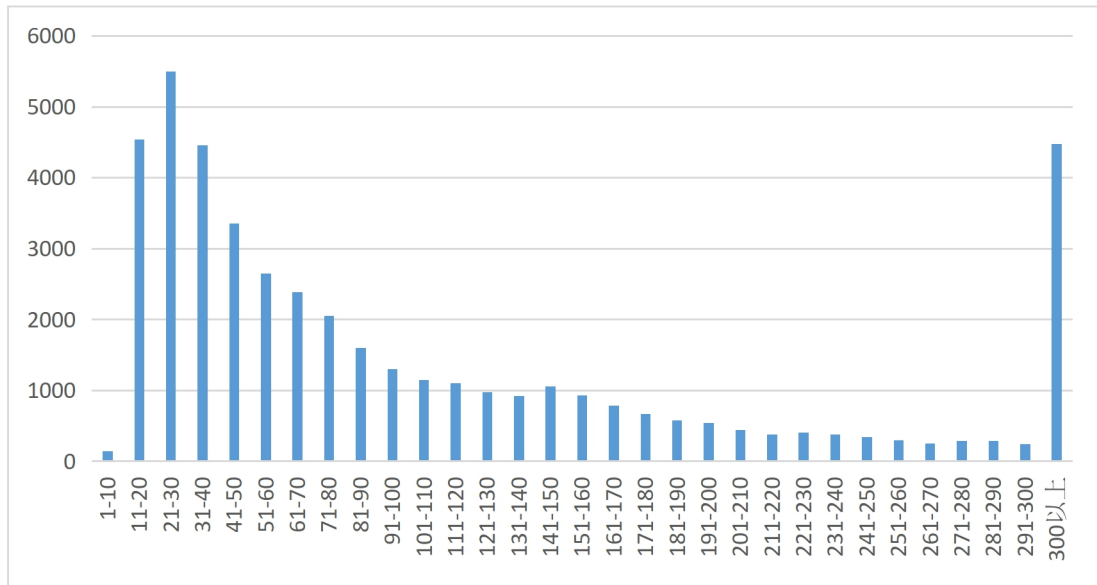


图 3.1 评论文本数据文本长度在区间上的分布图

表 3.1 评论文本数据字符数统计表

评论文本字符数	数据条数	频率	累计条数	累计频率
1-10	146	0.33%	146	0.33%
11-20	4538	10.20%	4684	10.53%
21-30	5496	12.36%	10180	22.89%
31-40	4460	10.03%	14640	32.92%
41-50	3354	7.54%	17994	40.46%
51-60	2652	5.96%	20646	46.42%
61-70	2387	5.37%	23033	51.79%
71-80	2055	4.62%	25088	56.41%
81-90	1600	3.60%	26688	60.01%
91-100	1298	2.92%	27986	62.93%
101-110	1151	2.59%	29137	65.51%
111-120	1102	2.48%	30239	67.99%
121-130	977	2.20%	31216	70.19%
131-140	924	2.08%	32140	72.27%
141-150	1053	2.37%	33193	74.63%
151-160	931	2.09%	34124	76.73%
161-170	784	1.76%	34908	78.49%
171-180	672	1.51%	35580	80.00%
181-190	574	1.29%	36154	81.29%
191-200	537	1.21%	36691	82.50%

评论文本字符数	数据条数	频率	累计条数	累计频率
201-210	439	0.99%	37130	83.49%
211-220	381	0.86%	37511	84.34%
221-230	402	0.90%	37913	85.25%
231-240	383	0.86%	38296	86.11%
241-250	339	0.76%	38635	86.87%
251-260	293	0.66%	38928	87.53%
261-270	252	0.57%	39180	88.10%
271-280	284	0.64%	39464	88.73%
281-290	289	0.65%	39753	89.38%
291-300	243	0.55%	39996	89.93%
300 以上	4478	10.07%	44474	100.00%

经过对评论数据的观察，发现一些评论内容过于繁杂、词汇重复单一，缺乏意义，这种情况可能是由于用户在进行评论时并没有真正想要表达评论的内容，也有可能是刷评的水军。为了确保评论内容的有效性、节省计算资源以及提高实验数据的准确性，本文限定单个评论文本长度为 0-200 个字符，筛选去重后共计 36677 条，汇总作为实验数据，存储格式为 CSV 文件。部分实验数据格式如表 3.2 所示：

表 3.2 部分原始数据示例

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
CemeteryAdministrator	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...聪明到有点儿...吓人，它是不是已经有自我意识了啊????	中国	江苏
NinesevenTy	2023/2/19 23:20	我寻思着这 ChatGpt 也不行啊，我一个问题就给他干懵逼了都	中国	河南

用户昵称	发布时间	微博内容	IP 国家属地	IP 省份属地
天津股侠	2023/2/23 12:45	@周鸿祎: ChatGPT 可能两三年内就会产生自我意识威胁人类	中国	天津
他叫她小豆包耶	2023/3/5 12:07	ChatGPT 离大谱各种胡编乱造是吧	中国	河南
白鱼 Fiasili	2023/3/8 6:05	chatGPT 查文献太快了, 让我怀疑我这个研究生白读了 哈哈哈哈哈 哈哈哈哈哈	中国	台湾
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	

36677 条实验数据中, 按国家地区划分, 用户发文排名前十的国家分别是中国、美国、英国、日本、澳大利亚、新加坡、加拿大、德国、韩国、荷兰。在中国范围内, 用户发文排名前十的省份分别是北京、广东、上海、浙江、江苏、山东、四川、湖北、福建、河南。

3.3 评论情感倾向性标注

原始数据经过筛选整合后精简了数据数量, 但进行情感分析还缺乏必要的情感倾向信息, 因此需要进行人工标注。在 36677 条实验数据中进行随机抽样, 得到 10000 条人工情感标注目标数据。邀请两位从事自然语言处理相关研究的硕士生, 基于对 ChatGPT 在微博上的表现和用户反馈的观察和分析, 以及对 ChatGPT 本身的了解和认识, 展开了深入的讨论, 最终选取“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, 邀请了一位情报领域的专家参与情感分类指标的构建工作。以粗粒度指标作为基础, 逐一讨论并总结出用户对 ChatGPT 情感体验的关键词和表达方式。经过多轮讨论和反复确认, 我们逐步细分出了赞扬、惊喜、中立、担忧、愤怒和失望等六个情感次类。

在情感分类指标的构建过程中, 专家和硕士生们进行了多次讨论和交流, 对数据集中的评论进行了独立统一的情感分类, 最终达成了共识。这些情感主类和次类被认为能够全面涵盖微博用户对 ChatGPT 的情感体验, 具有较高的有效性、

可靠性和客观性。

表 3.3 情感分类指标信息

情感倾向编号	情感主类	情感次类	含义	特征词
0	积极	赞扬	用户对 ChatGPT 的优点和价值肯定、赞美、赞同、的认可	肯定、赞美、赞同、欢乐、兴奋等
1		惊喜	与 ChatGPT 有关的的某方面超出了用户的预期	惊喜、惊奇、震撼等
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
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
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
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

在遵循三个情感主类、六个情感次类的含义和特征的基础上，结合评论数据的背景和上下文进行情感倾向标注，尽可能避免主观偏见，确保标注的一致性和准确性。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 3.3 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。

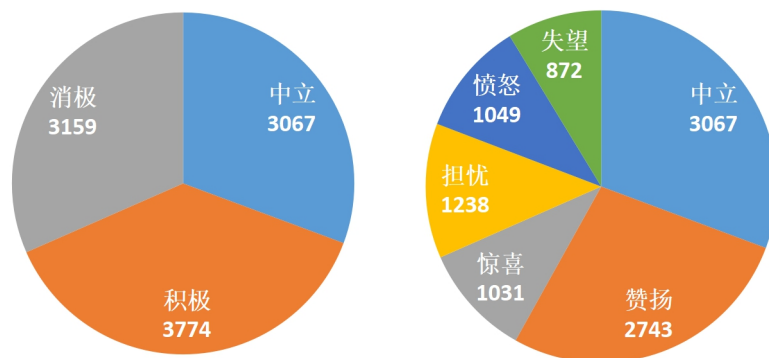


图 3.2 已标注粗粒度和细粒度评论基本信息

为尽可能保证各情感评论数量的平衡，分别在六种细粒度情感评论里按照帕累托法则采取 8: 2 的比例随机划分，将数量多的部分合并为训练集，余下部分合并为测试集，最终得到训练集包含为 8000 条数据，测试集包含 2000 条数据。部分训练集数据如表 3.4 所示。

表 3.4 部分数据标注结果示例

用户名	发布时间	评论内容	IP 省份	IP 国家	情感倾向
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	湖南	中国	0
彦凌	2023/2/14 9:49	这分析推理能力，继续训练下去一定还会有更多让人类震惊的对话	北京	中国	1
李大锤同学	2023/3/17 18:32	用 ChatGPT 搜到的十篇文献都是他自己编的，甚至还给我编了个 doi 出来，气死我了	北京	中国	4
陈老湿说	2023/3/18 0:27	难怪这几天霸占全网热搜 难怪让一些科技巨头感到害怕 太让人毛骨悚然的 AI 回答了 微软新 Bing 支持 Copilot 之后 搜索的意义不再是更快更准确呈现结果 而是感觉真的有人跟你在交流 #GPT4##New Bing##AI#	北京	中国	3
垒土成台	2023/3/25 22:44	百度李彦宏：文心一言和 ChatGPT 的水平差了 2 个月，但可以追赶	湖北	中国	2
莫白 27	2023/3/31 13:57	短短一天对 ChatGPT 彻底失望。不懂请不说，更不要乱说	加拿大	加拿大	5

3.4 数据预处理

未经过数据预处理，实验数据集中存在许多特殊符号，例如重复的标点符号、非文本字符如“#”、“@”等，以及其他无关信息。本文调用 Python 中的 re 模块来过滤这些无效字符和无用信息。在去除无用信息后。分词是将文本分割成单独的词语或标记的过程，通过分词，可以将一篇文本划分为众多独立的词汇单位，以便于后续的处理和分析^[66]。在众多分词方法中，jieba 具有高准确率、快

速、支持多种分词模式、支持繁体中文和开源免费等优势，在自然语言处理等相关领域被广泛应用^[67]，因此，本文调用 Python 中的 jieba 包进行文本分词操作。加载原始数据 CSV 文件，读取“微博内容”列，jieba 参数设置 cut_all=False，HMM=True，使用空格符进行分词操作。

同时，去停用词也是文本预处理的重要步骤之一。停用词是指那些在文本中频繁出现但对于文本整体含义贡献不大的常见词语，例如“的”、“是”、“在”等，这些词语通常不包含明确的语义信息，而是起到连接词或结构性作用。去除这些停用词，一方面可以减少文本中的噪音和冗余信息，使得后续的文本分析更加关注有意义的词汇；另一方面可以减少数据量，提高模型处理效率^[68]。本文采用开源停用词表“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，调用 Python 中的 NLTK (Natural Language Toolkit) 包进行停用词的过滤操作。

表 3.5 部分分词及去停用词结果示例

微博内容	分词结果	去停用词结果
小试 ChatGPT，还是被惊艳到了... 后面直接把审核员的邮件内容复制给他回复...还可以，虽然是应声虫级的，但也算得体了...	小试 ChatGPT， 还是 被 惊艳 到了 后面 直接 把 审核 员 的 邮件 内容 复制 给 他 回复 还 可 以 ， 虽 然 是 应 声 虫 级 的 ， 但 也 算 得 体 了 ...	小试 ChatGPT，惊 艳 审核员 邮件 内容 复制 回复 应声虫 级 算 得 体 了 ...
ChatGPT 出问题了，或者说只是问题暴露了。它撒谎——其实不能，就是胡说八道不负责任！本来就只是一个 NLP 模型——自然语言处理，natural language processing。	ChatGPT 出 问 题 了 ， 或 者 说 只 是 问 题 暴 露 了 。 它 撒 谎 —— 其 实 不 能 ， 就 是 胡 说 八 道 不 负 责 任 ！ 本 来 就 只 是 一 个 NLP 模 型 —— 自 然 语 言 处 理 ， natrual language processing 。	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
不想找论文了 让 chatgpt 帮我找了十篇，没有一篇是真的 事实证明 我要失业了 ~~~#chatgpt#	不 想 找 论 文 了 让 chatgpt 帮 我 找 了 十 篇 ， 没 有 一 篇 是 真 的 事 实 证 明 我 要 失 业 了 ~ ~ ~ # chatgpt #	不 想 找 论 文 chatgpt 帮 找 十 篇 一 篇 真 的 的 事 实 证 明 我 要 失 业

第4章 基于机器学习的评论情感分类

4.1 基于 Bert 预训练模型的情感分类

BERT 预训练模型包含一些特定的参数配置，例如，BERT 预训练模型的基本版本 BERT-Base-Uncased，拥有 12 层的 Transformer 编码器，隐藏层大小为 768，注意力头数为 12，总参数量为 110M。

首先，Transformer 是一种基于自注意力机制的深度学习模型，用于在序列数据中进行编码和解码，BERT 模型使用了 Transformer 架构作为其核心组件，只使用了 Transformer 的编码器部分，用于从输入句子中提取语义信息；该编码器由多个层组成，每个层都包含一组自注意力子层和前馈神经网络子层，较大的 BERT 模型通常有更多的层，以便更好地捕捉输入句子中的上下文信息。其次，隐藏层大小是指 Transformer 编码器中隐藏状态的维度，在 BERT 模型中通常为 768 或 1024，即编码器输出的每个单词都表示为一个 768 维或 1024 维的向量。隐藏层过大会导致模型参数增多，增加训练和推理的计算成本，并容易出现过拟合问题。再次，BERT 模型使用多头自注意力机制，其中每个头都可以关注输入序列的不同部分。头数指的是自注意力机制中注意力头的数量，较大的 BERT 模型通常具有更多的注意力头，可以更好地捕捉输入句子中的不同关系。最后，总参数量是指所有模型参数的数量，它取决于模型的大小和架构，可以作为衡量模型规模和复杂度的指标^[69]。

除了提供基本版本 BERT-Base-Uncased 外，BERT 官方还提供了多个适用于不同的任务、语言处理需求和算力大小的预训练模型版本，如 BERT-Base-Multilingual-Cased（适用跨多种语言的任务）、BERT-Base-Chinese（针对中文的 BERT 模型）等。本文选择中文 BERT 模型 BERT-Base-Chinese，该预训练模型拥有 12 层的 Transformer 编码器，隐藏层大小为 768，注意力头数为 12，总参数量为 110M。

深度学习框架的选择方面，本文选择 PyTorch 框架进行 BERT 模型构建。虽然 PyTorch 和 TensorFlow 都是深度学习领域主流且极受欢迎的框架，但相较于 TensorFlow 使用静态计算图，需要先定义完整的计算图而言，PyTorch 使用动态计算图，可以在运行时动态地构建、修改和调试计算图，使得代码更加灵活和易

于调试^[70]。

在 Python 中导入 torch、transformers 等相关包，加载已下载好的 BERT-Base-Chinese 预训练模型，使用 tokenizer_class.from_pretrained 从模型中加载分词器；使用 pandas 导入文本数据，调用 tokenizer.encode_plus 进行文本编码，生成输入编码序列（input_ids）和注意力掩码（attention_mask），随后按照 65%、20%、15%比例划分训练集、验证集、测试集，并使用 TensorDataset 对训练集、验证集和测试集进行打包，调用 DataLoader 进行数据加载。定义 Bert_Model，加载预训练模型权重，通过 BERT 模型和全连接层进行分类，使用 AdamW 优化器对模型参数进行优化，并使用余弦退火学习率调度器进行学习率的动态调整。

接下来，定义 train_and_eval 函数，用于训练和验证模型，设定损失函数为交叉熵损失，根据设定的轮数进行训练和验证，每轮先将模型设置为训练模式，然后逐个批次地计算模型输出并计算损失、反向传播、优化参数和调整学习率。在每个训练步骤的几个时间点打印训练进度信息。随后，将模型设置为评估模式，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。定义 evaluate 和 predict 函数，分别用于评估性能和进行预测，将模型设置为评估模式（model.eval()），然后使用无梯度计算环境（torch.no_grad()）进行推断，区别主要在于：评估过程最终使用 accuracy_score 函数计算并返回准确率；预测过程则是将预测结果直接保存到文件。

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。需要注意的是，train_and_eval 函数内有一定数量的训练参数，如 Max_Seq_Length（输入序列的最大长度）、Learning_Rate（模型训练时的学习率大小）、EPOCHS（模型训练过程中数据集遍历的轮数）和 Batch_Size（每个训练批次中的样本数量）等，这些超参数大小受到如数据集大小和特点、计算资源的可用性以及任务需求等多种因素影响。通过修改模型结构和调整超参数大小，在特定任务的数据集上对模型进行有监督的训练，能够最终使得模型适应特定的任务^[71]，这个阶段也被称为微调（Fine-tuning）阶段。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。

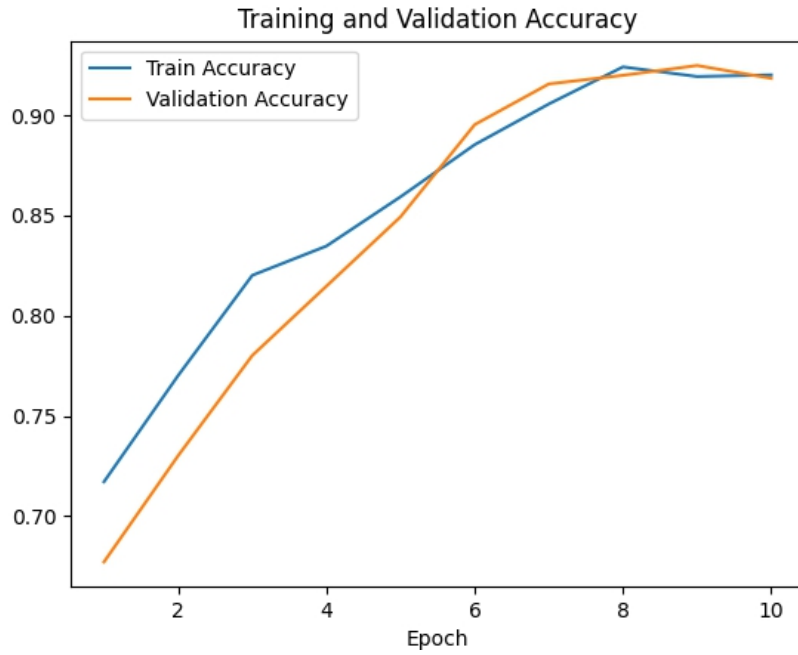


图 4.1 准确率曲线图

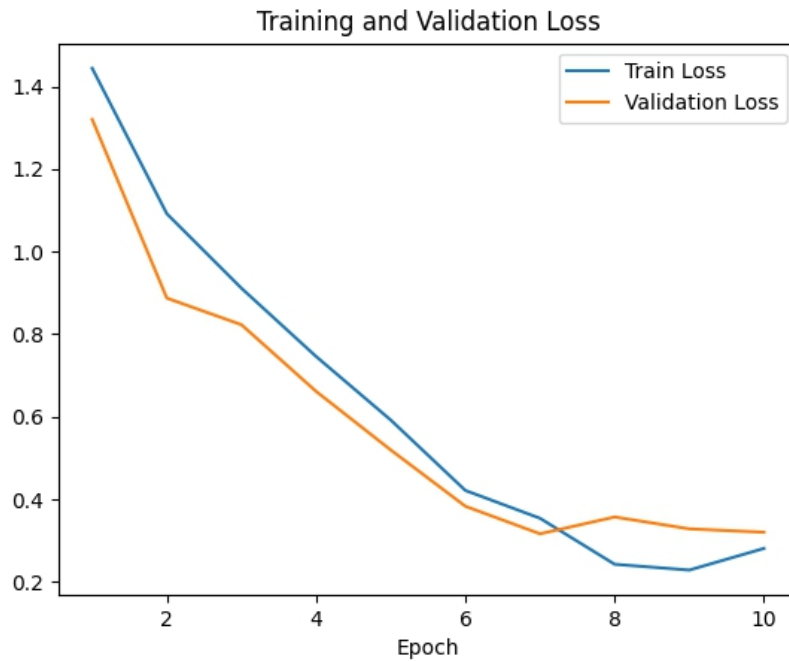


图 4.2 损失值曲线图

由上图可知，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，这说明模型具有很好的泛化能力。

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，表现出了良好的性能和泛化能力，因此可以用于实际应用场景中。

4.2 其他情感分类模型

本小节介绍使用支持向量机和随机森林两种传统机器学习算法进行情感分类。在自然语言处理中，我们需要将文本表示为数值形式，以便计算机计算和分析。不同于BERT使用深度神经网络对大量文本数据进行预训练，将数据编码生成对应于文本内容的、具有语义信息的高维句向量表示，传统机器学习算法（如支持向量机和随机森林）并不直接利用文本的语义信息。因此，在加载传统机器学习模型进行训练之前，首先需要对评论文本进行向量化处理^[72]。

传统的向量化方法包括独热编码和词袋模型。独热编码将每个单词映射为一个唯一的向量，所有词汇的向量互斥且正交，表示某个词是否出现；词袋模型（如CountVectorizer和TfidfVectorizer方法）则将文本表示为一个向量，其中每个维度对应于一个词汇，并记录该词汇在文本中的出现次数或权重。在这之中，相比于CountVectorizer，TfidfVectorizer方法综合考虑了一个词语在当前文档中的重要程度以及在整个文档集合中的普遍重要性，能够降低普遍出现的常见词语的权重，提高少见但有区分性的词语的权重，因此更有助于捕捉关键信息，过滤噪音词语^[73]。在Python中创建TfidfVectorizer对象，并设置max_features参数为1000，分别调用vectorizer.fit_transform和vectorizer.get_feature_names来将文本数据转换为TF-IDF矩阵，并获取特征词列表。

在传统的向量化方法的同时，本文引入词嵌入方法进行模型表现的横向对比。词嵌入是一种更先进的向量化方法，它使用神经网络模型学习词汇之间的语义关系和上下文信息，将语义相近的单词映射到相似的向量空间附近，不仅可以对单词进行向量化，还可以应用于句子和文档级别的向量化^[74]。在词嵌入中，单词的含义和语义特征可以通过向量的距离和相似度来表示。常见的词嵌入方法包

括 Word2Vec、GloVe 和 FastText 等。

GloVe (Global Vectors for Word Representation) [75]采用了全局共现矩阵来学习词向量,可以更好地捕捉词与词之间的全局语义信息,从而得到更好的全局词汇表示,另外, GloVe 在训练过程中可以使用预先计算好的共现矩阵来节省计算时间和存储空间,同时, GloVe 中的词向量可以通过简单地进行加减运算来计算两个词之间的语义相关性[76],相较于 Word2Vec 方法采用局部上下文窗口,以及需要在训练过程中不断扫描数据集[77], GloVe 方法有着更好的全局词汇表示、更好的稳定性和更好的词向量加减效果。

FastText[78]在 2016 年由 Facebook Research 开发并开源,它使用了一个浅层神经网络模型,将词表示为子词的集合,并利用子词的信息来推测未知词的特征,通过聚合子词的向量来表示整个词的语义信息[79]。相对于 Word2Vec 和 GloVe 来说, FastText 学习速度更快,在处理稀有词和低频词时表现更好,并且能够处理歧义词、复杂语义词和词典外的陌生词,更重要的是,除英语词向量模型外, FastText 官方还提供了多达 157 种语言的词向量模型,使其在处理中文语料的任务中较 GloVe 方法有优的表现。因此,本文使用 FastText 方法进行文本向量化。

进入 FastText 官方网站(fasttext.cc),下载中文预训练词向量模型 cc.zh.300.bin 文件,该模型维度为 300。在 Python 中使用 Pandas 库加载基于 2.3 小节经过情感标注后的训练集和测试集,并使用 fasttext.load_model 进行词向量模型调用,然后定义一个 get_average_vector 函数接受文本作为输入,函数内部将文本分割成单词,并逐个检查该词是否存在于 cc.zh.300.bin 模型中,如果存在,则获取该单词对应的词向量,并将其添加到一个向量列表中;如果不存在,那么该词将被忽略,不进行任何处理。最后,如果向量列表不为空,则通过 np.mean 函数计算向量列表的平均值,并作为结果返回;如果向量列表为空,则返回一个全为 0 的 300 维零向量。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。

表 4.1 部分句向量示例

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	[-1.11070052e-02 1.50568234e-02
时间 整理 资料 写 memo 写 文件 改 合	4.19823229e-01 -7.93736354e-02...]

文本内容	句向量表示
同 写 稿子 饭 嘴里	
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 natrua l 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...]

4.2.1 基于支持向量机的情感分类

在 Python 中导入 sklearn 库内支持向量机模块的 SVC 类，并导入 sklearn 库中 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，最终找到最佳的模型超参数组合^[80]。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，可能容忍更多的错误分类，有防止过拟合的倾向，而较大的 C 值则会追求更高的训练准确率，并尽量避免错误分类，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^[81]。

表 4.2 支持向量机超参及候选值

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。

4.2.2 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，增加决策树的数量可以减少

随机误差，提高模型的稳定性和表现，但也会增加计算成本，过大的决策树数量也可能导致过拟合；`max_depth` 表示决策树最大深度，控制决策树的深度有助于防止过拟合、节省计算资源，较小的深度可以增加模型的泛化能力，但也可能降低其在训练集上的拟合能力；`min_samples_split` 表示节点分裂所需最小样本数，如果节点的样本数小于这个值，就不会进行分裂；`min_samples_leaf` 表示叶节点所需最小样本数，当分裂操作导致叶节点的样本数小于该值时，便停止分裂，增加这个值可以防止模型过度细分，防止过拟合情况出现^[82]。

表 4.3 支持向量机超参及候选值

参数	候选值
<code>n_estimators</code>	100, 300, 500, 800, 1000
<code>max_depth</code>	5, 10, 15, 20
<code>min_samples_split</code>	2, 4, 6, 8, 10
<code>min_samples_leaf</code>	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。

4.3 模型评价

4.3.1 混淆矩阵

对于多分类问题，混淆矩阵是一种常见的评估模型性能的方法。混淆矩阵^[83]可以展示模型在不同类别上的预测结果与真实标签之间的对应关系，从而提供了更详细的分类结果信息。混淆矩阵的每一行表示真实标签，每一列表示预测标签。矩阵中的每个元素表示模型将样本预测为某一类别的数量。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^[84]。

表 4.4 混淆矩阵

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。

$$\text{Acc} = \frac{\text{TP}_{11} + \text{TP}_{22} + \text{TP}_{33} + \dots + \text{TP}_{NN}}{\text{总样本数}} \quad (4.1)$$

(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 :

$$R = \frac{\text{TP}_{ii}}{\text{TP}_{i1} + \text{TP}_{i2} + \text{TP}_{i3} + \dots + \text{TP}_{iN}} \quad (4.2)$$

(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 :

$$P = \frac{\text{TP}_{ii}}{\text{TP}_{i1} + \text{TP}_{i2} + \text{TP}_{i3} + \dots + \text{TP}_{iN}} \quad (4.3)$$

(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。

$$F1 = \frac{2PR}{P + R} \quad (4.4)$$

经过综合计算评估, 汇总不同模型各项评价指标如表 4.5 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 4.5 不同模型各项评价指标

模型	准确率	精确率	召回率	F1-Score
BERT	0.9048	0.9197	0.9012	0.9104

模型	准确率	精确率	召回率	F1-Score
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

4.3.2 ROC 曲线和 AUC 值

接受者操作特性曲线^[85] (Receiver Operating Characteristic Curve, 简称 ROC 曲线) 是机器学习中可视化分类模型性能的常用工具。它以二分类模型为基础, 将真阳性率 (True Positive Rate, TPR) 作为纵轴, 假阳性率 (False Positive Rate, FPR) 作为横轴绘制出的曲线。ROC 曲线通过改变分类模型的阈值来绘制。当阈值设定为最大或最小值时, 模型的预测结果会全部被划分为正例或负例, 此时 TPR 和 FPR 的值分别为 1 和 0。而当阈值设定为介于最大和最小值之间时, TPR 和 FPR 的值会随之改变。

除了绘制 ROC 曲线外, 通过计算曲线下面积也是评估模型性能的常见方法, 这个面积被称为 AUC (Area Under the Curve)。AUC 的取值范围在 0.5 到 1 之间, 越接近 1 表示模型的性能越好, 越接近 0.5 表示模型的性能越差。

以 BERT 分类模型为例。首先, 使用训练好的 BERT 模型对测试集进行预测, 并计算预测概率。使用 `test_loader` 迭代测试集中的每个样本, 将其传递给模型, 然后使用 `softmax` 函数计算概率, 并将结果附加到 `probabilities` 列表中。

接下来, 将原始标签转换为二进制形式。因为 ROC 曲线只能用于二分类问题, 所以需要将多分类问题转换为多个二分类问题, 每个问题都是将某个类别与其他所有类别区分开来。使用 `label_binarize()` 函数将标签转换为二进制形式。具体地, 将长度为 `n_samples` 的一维数组转换为一个大小为 `(n_samples, n_classes)` 的二维数组, 其中每列都是一个类别的二进制标记, 对于每个类别, 计算其 ROC 曲线, 并绘制在同一张图表上。同时, 使用 `roc_curve()` 函数计算每个类别的 FPR、TPR 和阈值, 并使用 `auc()` 函数计算 AUC 值。

除了绘制 ROC 曲线并计算 AUC 值外, 本次实验还使用了宏平均^[86] (macro-average) 来计算接受者操作特性曲线。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 ROC 提供了对整个多类别分类模型性能的总体评估，它将每个类别视为同等重要，不考虑类别之间的样本不平衡情况，因此在处理类别不平衡的问题时有着优良的表现。首先，将所有类别的 FPR 进行汇总，并使用 `np.unique()` 函数去除重复值；然后，对于每个类别，使用 `interp()` 函数在这些点上插值它们的 TPR；最后，求平均值并计算 AUC，然后使用 `plt.plot()` 函数绘制曲线。结果如图 4.3、图 4.4 所示。

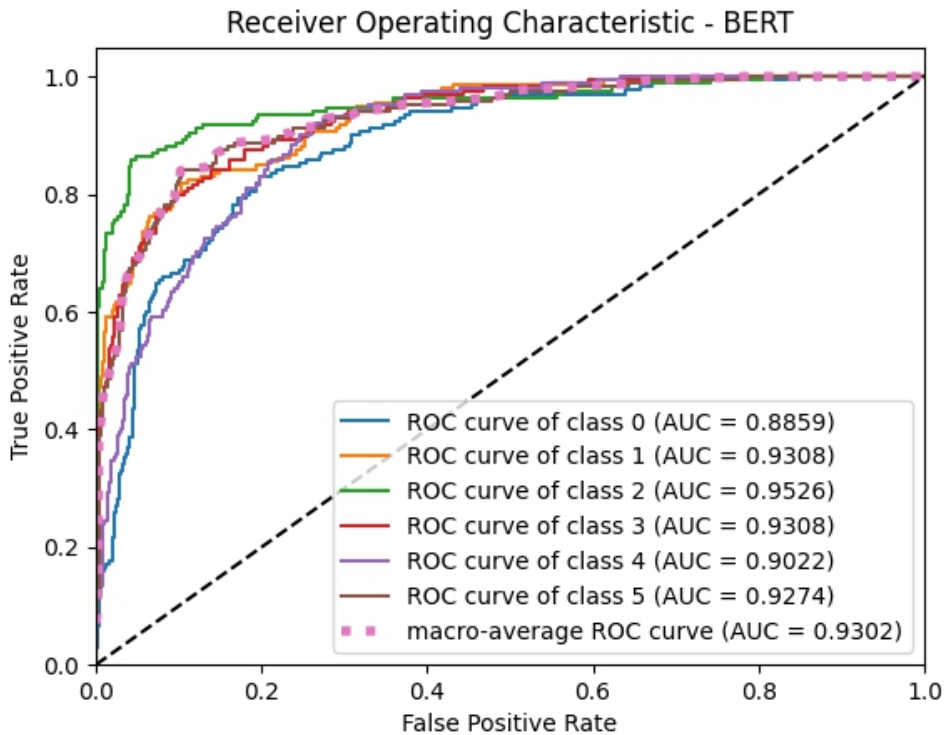


图 4.3 ROC 曲线及 AUC - BERT

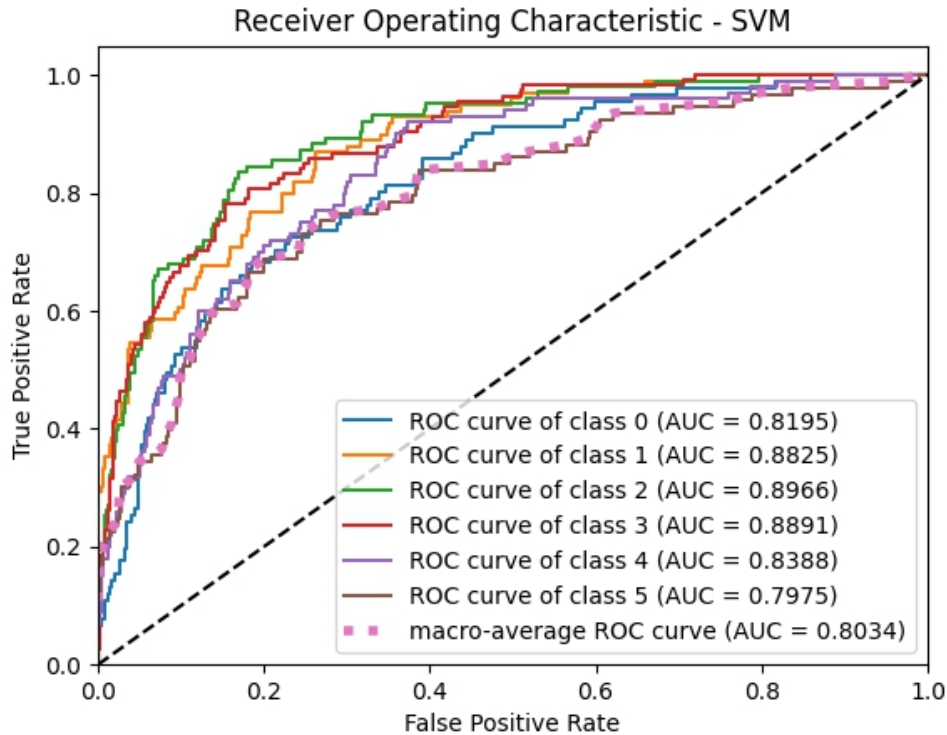


图 4.4 ROC 曲线及 AUC - SVM+FastText

从宏观平均 ROC 曲线的位置来看,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,这意味着模型能够有效地识别这些类别的正例,分类的准确性能够得到保障;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可能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。

4.4 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 3.2 小节已训练好的 `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，得到 26677 条评论情感倾向结果。为了实现资源的最大化利用、提高研究的整体效率，便于进一步分析，在数据有限的情况下，本文选择结合人工标注的精准性与模型预测的大规模处理能力，将人工标注的数据和模型预测的结果进行整合，最终得到 36677 条评论及相应情感倾向数据。

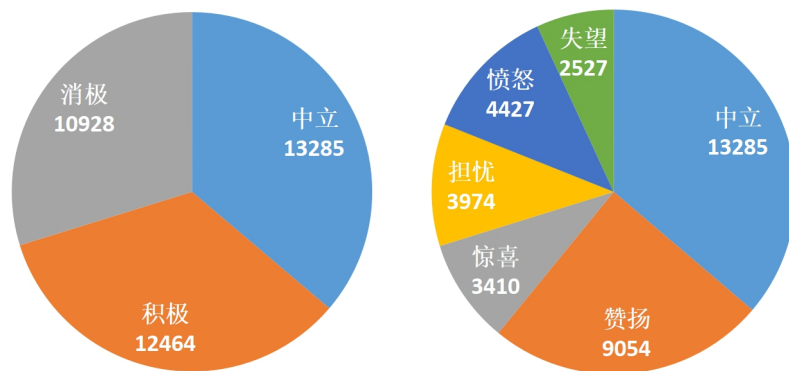


图 4.5 整体评论集各情感类别评论数量

各情感类别的评论数量结果如图 4.5 所示。整体评论集共有 36677 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 13285 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 的反响呈现复杂且多元化的特点。这种态度的形成，其原因是人们对新兴技术的期待与担忧并存，亦反映出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，也对制定政策和指导未来的发展方向具有不可忽视的价值。

第 5 章 基于 LDA 的评论主题聚类

5.1 模型构建

在 Python 中导入 LatentDirichletAllocation 库, 加载不同情感倾向类型的评论集, 按照 3.3 小节中 FastText 词嵌入方法的流程分别进行向量化, 获得文本的平均向量表示。使用 `np.vstack(data.values)` 获取文档-词矩阵, 并将文档-词矩阵传递给 LDA 模型进行拟合, 之后调用 `lda.transform` 获取文本在 LDA 主题空间内的表示。LDA 模型内有多个超参数供手动调整, 其中, 主题数 K 通常被认为最能直接影响模型的最终效果: 较小的主题数可能导致生成的主题过于笼统, 无法很好地解释数据; 而较大的主题数可能会使主题过于细分, 导致主题之间的区分变得困难。因此, 在拟合模型之前, 我们需要通过计算不同主题数下的困惑度、一致性、余弦相似度指标以及可视化方法综合确定最佳主题数。

创建两个空列表 `perplexity_scores` 和 `coherence_scores`, 用于存储困惑度与一致性得分。该函数使用 FastText 模型 `fasttext_model.wv` 来训练主题模型, 迭代次数设置为 5 次, 并在每次迭代中设置主题数为当前的数值。分别调用 `log_perplexity` 和 `coherence_model_lda.get_coherence` 计算当前主题数下的困惑度与一致性, 并将其添加到相应的列表中。调用 `matplotlib` 进行可视化展示, 以情感类别为“中立”的共 13285 条评论集为例, 两种指标迭代结果如下:

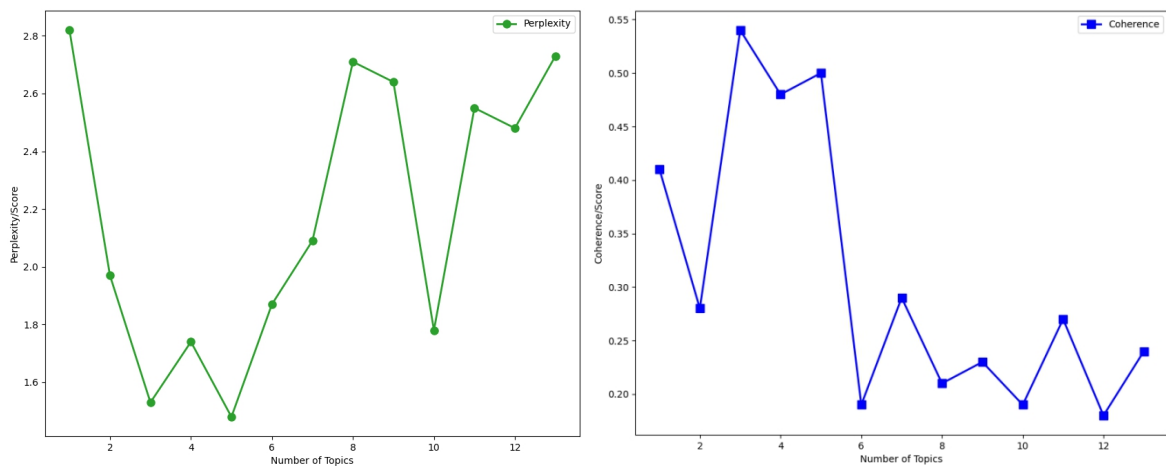


图 5.1 LDA 困惑度 (左) 与一致性 (右)

模型评价指标中，困惑度（Perplexity）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（Coherence）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。从图中可以看出，主题数为3时，困惑度和一致性指标表现较好，可以认为，生成的文本相对准确和连贯，主题之间的关联性也较强。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。

5.2 主题聚类结果

表 5.1 持“赞扬”和“惊喜”态度评价文本中的潜在主题

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 5.1 为微博用户对 ChatGPT 大语言模型持“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”表现出微博用户由 ChatGPT 大语言模型通过学习人类语言的方式来生成流畅、自然的文本，如文章、新闻等，而对其语言能力、表达能力的惊叹；主题二中“生产力”、“实用”、“高效”、“智能”表明，一些用户可能会赞扬它的实用性和效率，认为它可以在很多场景下提高工作效率和生活品质；主题三中“前沿”、“革新”、“突破”等主题词展现了微博用户对代表人工智能技术的前沿进展——ChatGPT 大语言模型技术性和创新性的认可。

表 5.2 持“中立”态度评价文本中的潜在主题

主题一	主题二	主题三
真实	技术替代	国产
靠谱	就业	创新

主题一	主题二	主题三
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 5.2 为微博用户对 ChatGPT 大语言模型持“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 大语言模型所生成信息的真实性和可靠性的关注，可能认为需要对信息来源进行查证和验证以确保其准确性；主题二中“技术替代”、“失业”、“矛盾”等关键词表明技术替代对劳动力需求和就业机会的影响，以及由此可能引发的社会矛盾深受大众所关注；主题三中“国产”、“创新”、“功能”、“规范”、“发展趋势”等词表明，ChatGPT 大语言模型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 5.3 持“担忧”态度评价文本中的潜在主题

主题一	主题二	主题三
隐私	伦理	失业
保障	造谣	平等
泄露	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 5.3 为微博用户对 ChatGPT 大语言模型持“担忧”态度的评价文本中的潜在主题。主题一中“隐私”、“泄露”、“保障”、“权限”、“监管”、“政策”、“风险”、“意识”等词语表明用户对于 ChatGPT 大语言模型使用过程中个人数据安全和隐私保护存在一定担忧，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 大语言模型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 大语言模型可能对就业市场和社会经济带来的影响的担忧，包括科技发展这把双刃剑可能带来的失业风险、结构失衡、不稳定等社会问题。

表 5.4 持“愤怒”和“失望”态度评价文本中的潜在主题

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 5.4 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对该模型的使用和在回答问题时不足之处的不满，在解答效率、逻辑性以及信息准确性等方面获得的体验可能较差；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不满和失望。

5.3 主题演变可视化

相较于传统 LDA 主题聚类的静态结果，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，可以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；

同时，通过绘制主题随时间的变化图表，可以直观地展示不同主题的发展趋势，使分析结果更易于理解和传达。

加载 4.4 小节下的整体数据集共 36677 条已分类评论，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 5.1 小节 LDA 模型构建完毕的基础上计算每篇文档的主题分布情况，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 5.2 所示。

表 5.5 整体文本潜在主题

主题一	主题二	主题三
真实	技术替代	国产
伦理	就业	创新
隐私	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

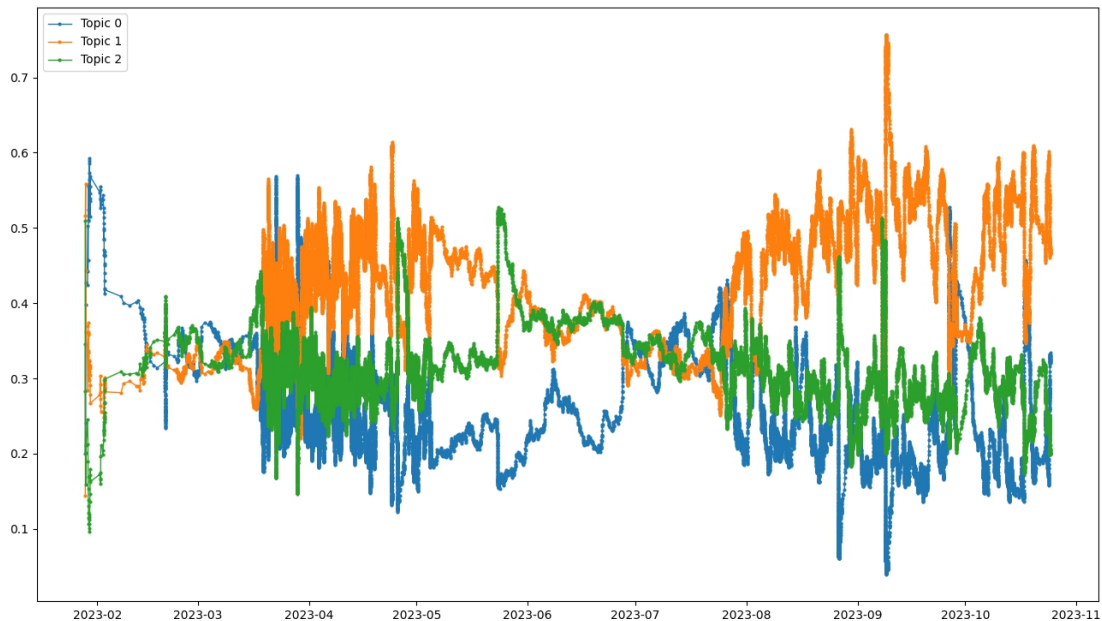


图 5.2 LDA 主题演变可视化

主题一涉及到的内容可能是关于 ChatGPT 大语言模型在使用中可能引发的伦理问题，包括用户隐私泄露、生成内容时的真实性和可靠性，以及用户对其生成内容进行验证和查证的需求。用户关注的焦点在于模型生成的文本是否准确、可信，并且希望能够找到有效的方法来验证和核实信息的真实性。主题一在整个时间序列中的关注度呈“较高-较低-中等”的趋势，原因可能是由于在初始阶段，用户对 ChatGPT 的新技术充满了好奇，同时随着使用次数的增加，一些不准确或虚假的信息被产生出来，引发了用户对以 ChatGPT 大语言模型为代表的人工智能生成的内容真实性和可信度的担忧。GPT4 发布后，ChatGPT 大语言模型的智能程度大幅提升，一方面，它能够提供更加时效的信息和分析，同时在处理当前的话题和问题时能够展现出更高的准确性和相关性；另一方面，虽然大语言模型在生成内容时尽力保持客观和准确，但仍然有可能存在误导性信息的情况；AI 绘图的风靡、未经验证信息等内容的持续存在也是用户对 ChatGPT 大语言模型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 大语言模型对就业市场的影响和技术替代的讨论：是否会取代传统工作岗位，从而对就业机会和劳动力需求造成变化。这一话题在讨论过程中始终保持较高热度，并经历了两次明显的关注度跃升。这背后的原因，一方面可能是由于 2023 年大语言模型技术取得了显著的进步，应用范围更加广泛；另一方面，全球的经济状况与就业格局也在经历深刻变革，随着这些变化的持续发酵，越来越多的人开始关注人工智能技术是否及如何影响到未来的就业机会分配和整体劳动力市场需求。

主题三涉及到的内容主要是关于 ChatGPT 大语言模型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自主开发类似的大语言模型，以及该模型在实际应用中的功能、规范和创新表现。同时，用户也对类 ChatGPT 人工智能的使用成本保持了持续关注。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。

5.4 章节小结

从 LDA 主题聚类结果可以看出，用户对 ChatGPT 广泛应用的挑战和风险关注点呈现多元化特征，结合“中立”、“担忧”、“愤怒”和“失望”评论集聚

类主题词进行分析，主要可以分为“自身内部隐患”和“可能引发的外部社会问题”。

自身内部问题方面，ChatGPT 大语言模型在广泛应用时面临多重挑战。首先，很大一部分对 AI 大模型持观望态度的用户认为，GPT 系列模型选择闭源，大多国产模型基于 LLaMA，国产大模型架构自主化之路道阻且长。其次，其训练数据源自未经筛选和事实验证的互联网文本集合，可能导致模型提供错误、滞后或带偏见的信息，在处理权威性、时效性和精确度要求高的问题时尤为明显。再次，尽管具有广泛的知识覆盖面，但在专业领域如法律咨询、编程、金融分析及医学诊断中，由于缺乏实时更新和实践经验，ChatGPT 可能无法提供深度精准的专业见解。人们同样关注，作为深度学习模型，ChatGPT 存在透明度和可解释性不足的问题，其决策过程如同“黑盒”，难以追溯并解释具体依据，这不仅增加了算法歧视与误判的风险，也使得合规性审查变得复杂。

在可能引发的外部社会问题方面，隐私保护是首要关注点，一方面，ChatGPT 在收集和处理用户交互数据过程中，涉及敏感信息和个人隐私，即使企业采取安全措施，仍存在潜在的数据泄露风险；另一方面，AI 大模型企业积累大量用户数据以优化模型性能，易陷入隐私权保护与模型优化的“两难”境地。其次，大型语言模型背后的企业如 OpenAI 等形成的潜在数据霸权和垄断态势极有可能加剧数字鸿沟，影响市场竞争和社会分配正义。另外，ChatGPT 还可能被恶意利用以开发恶意代码、传播虚假信息甚至侵犯版权，从而诱发违法犯罪行为。在教育学术领域，过度依赖 ChatGPT 可能导致学术诚信受损、学生自主学习能力退化，并催生快餐式学习文化。同样引发关注的是，随着 ChatGPT 等智能工具逐步取代人力完成内容创作、客户服务等工作，结构性失业风险加大，可能引发就业结构变化和社会收入差距扩大等一系列社会经济问题。

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。

第6章 基于时间序列的情感预测

6.1 时间序列预处理

本章节以第四章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。

$$SSI = i^{pos} - i^{neg} \quad (6.1)$$

看涨指数(Bullish Index, BI)。参考 Antweiler 和 Frank^[87]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。

$$BI = \ln \left(\frac{1 + i^{pos}}{1 + i^{neg}} \right) \quad (6.2)$$

情感差异指数^[88](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。它可以用于比较不同文本、不同时段或不同人群之间的情感偏好或情感分布差异。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。

$$SDI = \left| 1 - \frac{i^{neg} - i^{pos}}{i^{neg} + i^{pos}} \right| \quad (6.3)$$

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 4.4 小节的粗粒度分类结果，以日为单位将每一天的数据进行合并，部分统计结果如表 6.1 所示

表 6.1 部分统计结果

日期	积极情感	消极情感	中性情感	日期	积极情感	消极情感	中性情感
2023/1/2	25	22	21	2023/6/1	39	40	59
2023/1/3	19	24	32	2023/6/2	45	35	60
2023/1/4	15	12	20	2023/6/3	54	58	72
		
2023/5/18	59	74	93	2023/10/16	81	41	64
2023/5/19	19	21	20	2023/10/17	59	40	33
2023/5/20	26	22	28	2023/10/18	79	47	36
2023/5/21	55	50	65	2023/10/19	54	37	81
2023/5/22	66	63	89	2023/10/20	10	92	38
2023/5/23	44	45	69	2023/10/21	66	58	46
2023/5/24	70	62	92	2023/10/22	65	38	64
2023/5/25	69	58	87	2023/10/23	78	62	59
2023/5/26	23	18	21	2023/10/24	76	48	51
2023/5/27	71	43	74	2023/10/25	61	50	72
2023/5/28	31	24	35	2023/10/26	57	46	53
2023/5/29	28	18	26	2023/10/27	70	51	22
2023/5/30	56	45	61	2023/10/28	46	43	58
2023/5/31	29	23	32	2023/10/29	64	31	72

从表中可以看出，每日的相关情感评论数量相对较少，在以日为单位的情况下，较小的样本量可能导致统计分析出现更多的随机波动。因此，为了减少数据中的随机噪声，提高预测的准确性，本次实验将每日数据合并为每周数据，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，将2023年1月2日至2023年1月8日定为第一周，以此类推，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6.2所示的部分周情感指数计算结果，并绘制如图6.1所示的三种情感指数随时间的变化趋势图。

表 6.2 部分周情感指数统计结果

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

周数	SSI	BI	SDI
6	-33	-0.139171566	0.930232558
7	-22	-0.143100844	0.928104575
.....			
37	26	0.07126302	0.964285714
38	30	0.082691716	0.958563536
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

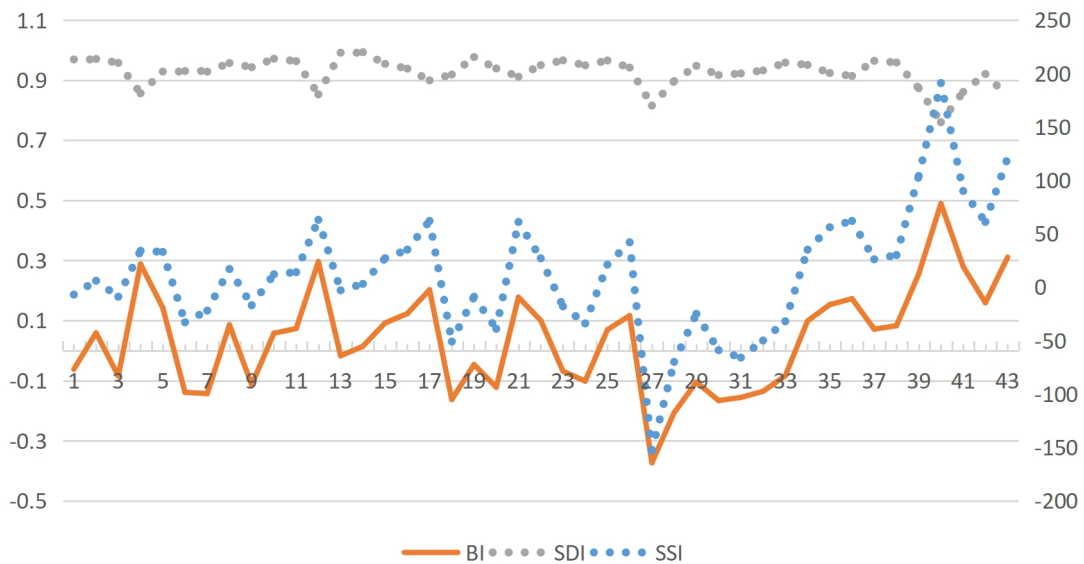


图 6.1 三种情感指数随时间变化趋势图

从图中可以看出，看涨指数 BI 在相应坐标轴上下的波动较大，表明网民情绪态度倾向变化大；情感差异指数 SDI 基本在 0.7 至 1 区间小范围波动，表明网民情感分歧较大。一方面，看涨指数是根据积极和消极态度评论数量的比例计算得出的指标，能够最直接反映情感倾向，发现周期性的情感波动，观察长期的情感变化趋势；另一方面，看涨指数取值范围在 0 到 1 之间，容易理解，便于解释。由于简单情感指数 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；情感差异指数 SDI 更适用于比较不同文本、不同时段或不同人群之间的情感偏好或情感分布差异，在反映单个文本或事件的情感倾向上有所欠缺。因

此, 参考研究内容, 并结合数据特点, 综合对比后选择看涨指数 BI 作为原始时间序列。

6.2 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)。ADF 检验是一种常用的单位根检验方法, 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 6.3 原始时间序列 ADF 检验结果

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 6.3 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 4.6 所示。

$$C(x) = x_t - x_{t-1} \quad (6.4)$$

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 6.2 所示。

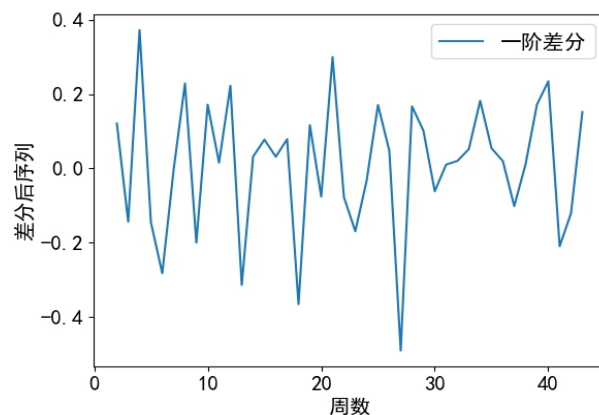


图 6.2 一阶差分后的时间序列

对一阶差分后的时间序列再进行 ADF 检验，结果如表 6.4 所示：

表 6.4 一阶差分后的时间序列 ADF 检验结果

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 值显著低于阈值 0.01，并且所有相关的统计临界值均满足标准，从而显著地拒绝了原假设并通过了平稳性检验。这意味着该时间序列具备平稳性。

ADF 检验完毕后，继续对时间序列进行非白噪声校验。在这一环节中，如果经过检验后的 P 值小于 0.05，则可以断定该时间序列并非白噪声序列，相反则表明为白噪声序列，不具备时间相关性，故而不适合进行预测分析。对于本研究的时间序列，其一阶差分序列的校验结果已记录于表 6.5 中。

表 6.5 白噪声检验结果

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 值确实低于 0.05 阈值，这证实了其已具备非白噪声属性。据此，接下来绘制了一阶差分序列的自相关图和偏自相关图，通过对两图中截尾与拖尾形态的观察来确定模型参数 p 和 q 的合适取值。

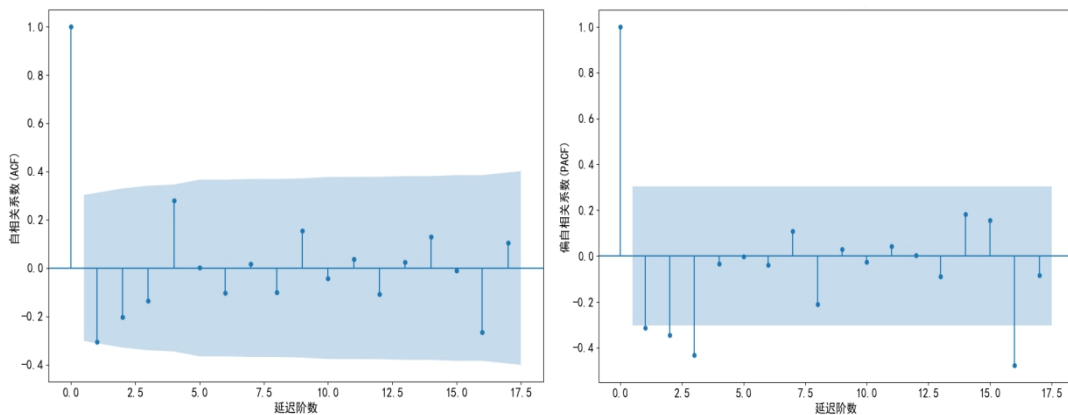


图 6.3 一阶差分后的自相关图（左）与偏相关图（右）

在自相关图和偏自相关图的结果上，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。

6.3 预测结果分析

取前 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 6.4 所示的预测结果。

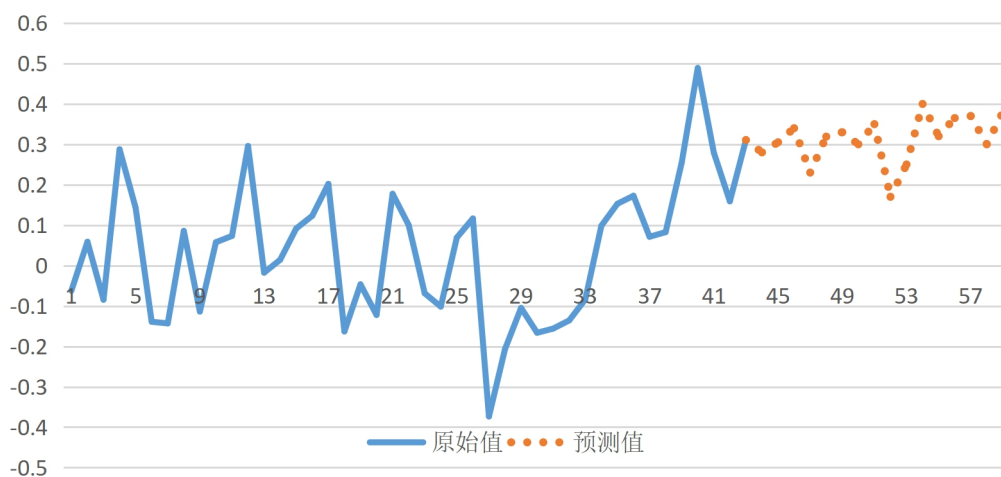


图 6.4 情感预测曲线图

在 2023 年 1 月 2 日至 2023 年 10 月 29 日共 43 周的时间内，BI 指数起伏变化较大，且显著地受到突发因素的影响。2023 年 2 月初，OpenAI 推出 ChatGPT Plus，月度订阅价格为 20 美元；次月，推出 GPT-4，仅限 ChatGPT Plus 会员使用。自此，ChatGPT 正式踏上商业变现之路，引发网民激烈讨论。2023 年 3 月 20 日，ChatGPT 宕机 12 小时，全球 ChatGPT 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，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 年年底至 2024 年年初期间内对 ChatGPT 大语言模型进行了包括但不限于新增多模态功能、免费开发语音功能、降低会员费用和完善生态链等的多项改进，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 模型的局限性也应当被重视：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，真实世界中的数据往往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^[89]。2023 年 11 月初，OpenAI 发布了 GPT-4 Turbo 和 GPT-s，对 ChatGPT 大语言模型进行了重要强化，并调整了使用价格，利好普通用户。而预测结果里相应时间段的情感值却大幅波动，迅速下降。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，导致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，并将其应用于日常生活和工作中。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，展现出了令人满意的效果。通过预测网民的情感态度，用户可以更好地了解社会热点、参与讨论；政府可以更好地了解 and 把握社会舆论的走向，更加精准地了解 and 回应群众的需求，制定出针对不同情况的有效策略和措施；社会组织和媒体也能够更好地回应公众关切、引导舆论发展。从企业角度来看，企业可以了解公众对于此类大语言模型技术的接受程度，更全面地发掘用户对自身品牌、产品或服务的态度和偏好，帮助企业调整品牌策略，提高竞争力和可持续发展能力。

第7章 对策建议分析

回顾前文的研究，通过情感分类、主题聚类和情感预测方法对微博用户关于 ChatGPT 大语言模型的评论进行多维度解析，结果显示积极、中立和消极情感的评论分布相对均衡，用户对 ChatGPT 大语言模型的预期评价良好，反映出公众对于 ChatGPT 技术的认知和接纳程度总体上持较为开放和理性的态度，越来越多的用户可能将大语言模型更广泛地采纳和融入到日常活动与工作场景中。大语言模型的影响力和应用范围将进一步扩大，其带来的便利性和效率提升将更加深入人心。

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，带来了多维度、多层次的风险与挑战，既有技术层面的固有局限，又有伦理、法律和社会经济层面的深层次矛盾。在当前全球数字化进程以及特定地区的网络环境与法律法规约束背景下，尽管 ChatGPT 原版服务尚未在中国内地正式上线，但国内众多开发者和机构已着手将 ChatGPT 的核心技术理念实现本地化转换，推出了一系列基于大语言模型架构的中文版本产品，为国内用户提供高效便捷的自然语言处理解决方案。在前文的情感分析中，积极评论体现了用户对 ChatGPT 在提高效率、增强体验、拓宽知识边界等方面的期待和肯定；中立评论则体现出用户对于 ChatGPT 技术既有欣赏也有疑虑，表明用户对技术的效能和影响保持着冷静审视；而消极评论则主要集中在对潜在风险和伦理问题的担忧，如数据安全、隐私保护、教育诚信和职业稳定性等。这种复杂且多元化的反响不仅表明以 ChatGPT 为代表的大语言模型在实际应用中具有的深远社会影响和变革潜力，也揭示出随着广泛普及，其潜在的风险与挑战也将随用户基数的增长和应用场景的拓展而放大。在此情况下，采取全面、多层次的风险应对措施势在必行。

接下来的章节，将着重阐述政府、企业和个人三个层面的应对措施，通过制定健全的法规政策、强化企业责任和道德规范、提升公众数字素养与适应能力等方式，来抵御以 ChatGPT 为代表的大语言模型所带来的风险挑战，促进这一强大技术在健康有序的环境中发挥更大的正面作用，实现社会经济的可持续发展。

7.1 政府层面的策略与措施

7.1.1 建设人工智能法制体系

(1) 建立健全严格的数据合规制度

随着大数据与人工智能技术的深度融合，数据作为一种关键生产要素的重要性日益凸显，尤其在大型语言模型训练场景中，所涉及的数据规模庞大且往往涵盖用户的个人隐私信息，因此，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的合法采集、存储、处理和传输的标准操作程序。强调遵循数据最小化原则^[90]，仅允许基于必要性原则收集个人数据用于模型训练，并要求相关企业采用高标准的脱敏技术和加密算法确保数据的安全处理与存储。同时，有必要设立独立公正的第三方数据审计机构，定期对企业进行合规审查，监督其严格遵守相关法律法规，并及时发现并纠正在数据管理过程中的潜在安全隐患。

另一方面，政府应当强化和完善数据权益保障的法律制度建设，明确界定用户数据的所有权、使用权以及利益分配机制。例如，推动实施数据可携带权、数据开放共享等创新制度，鼓励数据资源在保护隐私的前提下合理流动与利用，同时设置严格的个人数据隐私保护标准，通过法律手段严格限制企业非法收集或滥用个人信息的行为，从而维护公众的信息权益。

(2) 打击垄断行为，营造公平竞争环境

AI 技术及大语言模型特殊的资源密集型发展模式可能加剧市场集中度，带来垄断风险，因此，政府应当加强对人工智能产业的反垄断执法力度，通过修订和完善现有法规，限制数据寡头的形成和发展，促进市场开放和公平竞争，鼓励中小型企业参与创新并获取合理市场份额。具体来讲，加强反垄断法规的执行力度，不断修订完善《反垄断法》以适应数字经济发展趋势，特别增设有关于大数据和 AI 领域市场支配地位的判定标准，如设定市场份额或数据规模阈值，防止数据寡头形成并利用市场力量实施歧视性策略，对市场竞争构成壁垒或阻碍创新。设立专门针对 AI 领域的反垄断指南，规范大语言模型开发公司在数据获取、算法开发及市场竞争中的行为，通过严格审查企业并购活动等方式来保障正常的、有活力的市场集中度。

(3) 加强人工智能生成内容的知识产权制度建设

大语言模型生成内容的法律地位及其权利归属是人工智能领域亟待解决的问题。鉴于 AI 创作的独特性，尤其是大语言模型如 ChatGPT、文心一言和通义千问等能够独立生成富有创意和独创性的文本作品，法律制度应与时俱进，明确此类内容的产权归属。

政府应尽快制定和完善 AI 生成内容的数据产权法律制度，明确界定 AI 创作作品的著作权、使用权和流转规则，平衡创作者、用户与公众的利益关系，确保 AI 生成内容在产生、使用、传播过程中的权益分配合理合法。例如，可采用“操作主义”原则，即在无明确法律规定情况下，将版权权益赋予实际操控 AI 工具的使用者或拥有该工具及其训练数据的所有权的组织实体^[91]，等等。

进一步来说，确保知识产权的有效管理与保护，构建针对 AI 生成作品的官方注册登记机制势在必行。注册登记机制的核心是对 AI 创作活动进行权威性和可追溯性的记录，当 AI 系统如大语言模型生成具备独创性或创造性特征的作品时，实际操作者、开发者或者声称拥有版权的相关方可以在指定的官方平台上提交详细的版权登记申请。在详尽描述作品生成背景，如所使用的 AI 模型类型、训练数据来源以及作品产生的具体流程后，由平台负责对提交信息进行全面审查，结合技术评估与法律考量，确认作品的原创性及版权归属，并将相关信息存档备案，形成具有法律效力的确权文件。通过正式的版权登记，赋予了 AI 生成作品明确的法律地位，不仅有助于降低交易成本，提高版权流转效率，也有利于净化 AI 创作市场的秩序，鼓励合法合规的商业行为，激发研发人员的创新热情和技术投入，树立我国在国际知识产权领域的良好形象。

实践中，法院通过审理涉及 AI 生成内容版权纠纷的案件，可以逐步积累判例，形成一套完整的司法解释体系，为今后相似案例提供指导。包括但不限于确定 AI 原创性的判定标准、侵权行为的具体形态以及赔偿额度的计算依据等关键问题。倡导法院在处理相关案件时积极适用新规定，通过案例分析和判决释义，逐步确立清晰且具有前瞻性的司法裁判规则。

(4) 加强网络安全法制与监管机制建设

在当前日益复杂的网络环境中，高级人工智能工具如 ChatGPT 等技术的广泛应用，既带来了生产力与创新力的显著提升，也催生了新的网络安全威胁^[92]。针对新型网络犯罪形态，政府应推动立法改革，制定针对性强、涵盖全面的法律法规，明确将利用先进 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。

网络空间治理层面，应着重强化对 AI 生成内容的识别与审核机制。在社交媒体平台和新闻传媒领域，一方面，要求相关企业落实主体责任，采取有效措施及时对 AI 产生的信息进行真实性核验，尤其是对于可能引发社会恐慌或误导公众认知的虚假新闻与言论，必须做到快速发现、准确甄别并予以纠正；另一方面，鼓励企业进行技术研发，产出能够精准检测 AI 生成虚假信息的对抗性技术，包括但不限于开发基于机器学习和自然语言处理算法的检测工具，用于实时监控网络信息流，识别出具有潜在危害性的 AI 生成内容，并迅速启动应急响应机制，协同相关部门进行有效处置。此外，加大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，通过切断源头供给和技术支持，尽可能地从根源上遏制此类网络犯罪。

(5) 建立和完善算法信息披露制度

构建全面的算法透明度机制是确保大语言模型在安全、可靠和公平的基础上运行的关键举措。政府通过制定与时俱进的法规，可以要求 AI 服务提供商在保护商业秘密的同时，公开其核心算法的基本原理和技术细节，包括所使用的机器学习模型、训练过程以及优化策略，同时详细解释算法的主要功能、适用范围以及潜在的风险与局限性，特别是数据偏见问题和决策过程中的不透明性。

为了增强公众对大语言模型的理解和信任，法律应规定开发者提供易于理解和操作的界面或报告工具，允许用户追踪并查询与其交互历史相关的输入输出内容，了解影响答案生成的具体因素和决策路径。这样既有助于用户监督模型行为，也有利于提升算法公正性和可解释性。

(6) 权威性伦理和技术规范的应用与推进

在人工智能对话系统的设计、开发与应用过程中，制定一套具有普适性和权威性的伦理和技术规范至关重要。在此背景下，政府应当采取一系列行动以切实推动和落实上述伦理和技术规范。政府通过积极参与制定相关法律法规，将阿西洛马原则和 IEEE 伦理标准的核心理念融入国家法律体系，来为业界提供明确的伦理指南和监管依据^[93]；建立有效的监管机制，设立专门机构负责监督审查 AI 对话系统的合规性，确保其遵循已有的伦理和技术标准；在全球化视野下，积极参与国际合作，共同探讨并推进全球统一的人工智能伦理和技术规范，从而有效应对跨国界的技术挑战和社会影响。与此同时，政府可通过财政支持、项目资助等方式鼓励产学研各界合作研发符合伦理要求的 AI 技术，加强伦理教育和公众宣传工作，提高公民数字素养和社会认知水平。

(7)调整和完善劳动市场法规政策

在面对以 ChatGPT 为代表的大规模语言模型等人工智能技术进步带来的劳动市场结构转型和社会不平等问题加剧的背景下,政府扮演着关键的引导和调节角色。立法部门应当强化相关法规建设,通过出台《人工智能技术影响下的职业培训与转岗促进法》等专项法律,明确政府将设立专项资金,针对受大语言模型技术冲击较大的高风险行业或岗位,提供定制化的职业技能培训项目,帮助员工及时更新知识技能,向更具发展潜力的新兴领域或附加值较高的岗位转移,从而降低因技术替代导致的失业风险。

同时,在社会保障政策方面,政府需构建灵活且富有韧性的社会安全防护网,针对可能因自动化技术而面临暂时性失业的群体,可增设临时生活补贴计划,为他们在寻找新工作期间提供基本的生活保障,缓解短期经济压力。此外,加大职业技能培训补贴力度,鼓励企业、教育机构以及研究部门紧密合作,共同开发适应新技术环境的课程内容和培训体系,使受影响的劳动者有机会提升自身能力,顺利融入新的就业市场;以及开展持续的职业咨询服务,帮助失业者了解新兴行业趋势,掌握市场需求,提升自我竞争力。

7.1.2 战略生态构建与实施引导

(1)教育与人才培养

在应对人工智能技术的快速发展及其对社会伦理、教育体系和劳动力市场所带来的深刻影响时,政府与教育部门必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。

人工智能伦理与数字素养教育在当今时代的重要性日益凸显,应当作为不可或缺的组成部分纳入各级各类学校的课程体系之中。基础教育阶段的教学应适时融入科技伦理内容,通过生活化、情境化的教学案例,引导学生深入探讨人工智能带来的伦理难题,在了解前沿科技、理解前沿科技的基础上,培养他们从道德和社会影响角度理解并应用科技成果的能力;高等教育阶段,设立专门的人工智能伦理学必修课程有助于培育学生的批判性思维和法律意识,使其能够对 AI 技术的发展和应用进行深入学习与思考,并承担起相应的社会责任。同时,政策层面应当倡导和组织多样化的科普活动及专题讲座,利用线上线下相结合的方式,提升全社会对于 AI 伦理问题的认知水平,强化公众对于 AI 系统中个人信息保

护的重视程度，辨别信息来源的可靠性，从而提高用户的自我防护能力，鼓励正确且负责任地使用智能工具，防范潜在的滥用风险。

值得重视的是，算法科技人员作为人工智能和数据驱动技术的开发者与实施者，其在技术向善伦理认同方面扮演着至关重要的角色。他们不仅需要具备扎实的专业技能和创新能力，更要在技术研发过程中秉持伦理原则和社会责任感，确保所开发的算法和技术能够服务于人类社会福祉，体现“科技向善”的核心价值观。因此，应将伦理道德、社会责任以及法律法规等相关教育与实践流程融入到算法科技人员学术研究和技术的方方面面，引导科技人员形成“以人为本、科技向善”的价值观。

教育实践中，教育部门应在教学设计与评价体系上进行调整，强调过程性评价和创新能力培养，限制 AI 生成内容在学术作业中的不当应用，确保学术诚信；倡导互动式学习，培养学生的信息检索、分析和整合能力，认可原创性和独立思考的价值。

(2) 科研投资与引导

政府应当积极促进技术创新和市场竞争，推动行业健康发展。通过公共资助项目和科研合作计划等形式，扶持中小型企业及个人开发者参与到大语言模型的研发中，并采取公开招标的方式吸引国内外优质团队参与竞争；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，鼓励大学、研究所与企业共同构建实验室，研发能够有效检测和防御由 AI 驱动的恶意软件的安全技术产品，共享研究成果，在加快从实验室到市场的转化速度的同时提高网络安全防护水平。

针对专业领域对高质量、安全性高的大语言模型的需求，政府可以设立“垂类大模型基金”，为致力于特定行业 AI 技术研发的企业提供包括低利率贷款、税收优惠在内的多元化政策支持。此外，政府还可通过优先采购符合国家信息安全规范且注重用户信息权益保护的大语言模型产品和服务，利用政府采购的市场导向作用，激发企业的研发热情，推动国内自主大模型包括垂类大语言模型的研发和广泛应用^[94]，从而实现整个行业的技术创新和产业升级。

7.2 企业层面的行动方案

7.2.1 数据治理与合规运营

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，如《中华人民共和国个人信息保护法》、《中华人民共和国数据安全法》、欧盟《一般数据保护条例（GDPR）》、《加利福尼亚消费者隐私法案（CCPA）》等，以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。

企业必须遵循严格的数据生命周期管理原则，在数据收集环节实施明确的用户知情同意机制，坚持数据最小化原则，仅采集对模型训练至关重要的必要数据，并采用先进的加密技术和匿名化手段进行存储，从而有效防止敏感信息泄露。此外，推动研发实践差分隐私、同态加密等技术，并结合联邦学习等分布式机器学习框架，使模型能够在不直接接触原始数据的情况下进行训练与优化，降低隐私风险^[95]。

为确保 AI 模型训练数据的公正性和准确性，引入多元化的数据集，构建一套融合先进机器学习算法与人工审核策略的内容审核系统势在必行。一方面，可以设立一支具备跨学科背景的专业数据审查团队，定期对训练集进行审计和筛选，剔除可能存在的歧视性元素、涉及隐私或错误信息等内容；另一方面，对接权威数据源，采用自动化工具结合人工复核的方式，严格筛查输入数据，在数据获取阶段加强筛选流程，确保模型学习素材能够实时更新且质量可靠。

在面对公众对 AI 决策过程日益增长的透明度需求时，企业必须积极推行可解释人工智能（XAI）实践，以提升用户对 AI 系统的信任度和接受度。通过将模型预测与决策过程进行可视化表达，并提供易于理解的解释性说明，能够有效地增强 AI 系统在社会应用中的公正性和可靠性认知^[96]。进一步而言，在产品的设计层面，ChatGPT 等对话式 AI 应明确告知用户其提供的信息可能存在时效性和权威性限制，强调在关键决策场景下需参照官方或其他可靠渠道的信息源，以此确保在算法运营全过程中严格遵守国际及地区法律法规。技术研发上，创新开发各类可解释性技术方案，如优化注意力机制的可视化展现，探索局部敏感性分析方法，以及结合规则或逻辑推理框架构建混合型 AI 模型，使复杂大语言模型在

保持高智能水平的同时，具备内在的自我解释能力，让每一个生成的回复都能有迹可循，促进用户更好地理解 and 信任 AI 决策的过程与结果。

7.2.2 技术创新与社会责任担当

模型优化方面，应注重深化和拓宽垂直领域专业知识的学习深度与广度。通过精心筛选并设计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，运用更为先进的自然语言理解和生成技术，除图像、音频、视频外，尽可能实现结合多模态数据输入（例如物联网传感数据、生物识别数据），使得大语言模型不仅能够理解基本概念，还能充分考虑具体情境下的个性化因素，进而提供更加精确和个性化的解答方案。在科技和医学等快速迭代发展的领域，构建动态知识更新机制是保障大语言模型与时俱进的关键环节。通过搭建高效整合权威信息源（如学术期刊 API、行业数据库等）的实时数据流管理系统，使得大语言模型能够及时捕获并吸收最新知识，紧跟技术发展前沿，并将这些更新融入到决策支持过程之中。

为实现 AI 产品的持续优化和社会价值的最大化，企业需要建立并维护一个有效的用户参与和反馈系统，鼓励广大用户积极参与，提供关于 AI 生成内容的真实情况反馈。对于用户反馈的问题，企业应有专门的技术团队进行细致研究分析，并基于此进行针对性的算法迭代和优化工作。企业可以通过多样化的激励机制如积分奖励、荣誉认证等等让用户感受到企业对自身意见的尊重与重视，进一步提高用户社群的活跃度和满意度。

社会责任方面，企业应当通过定期发布详实的营运报告来强化透明度和公信力。这些报告应当包含企业在数据治理领域的具体实践，如数据生命周期管理的完善、隐私保护措施的加强、信息安全等保障等方面的工作进展和实际成效，此外还应详细阐述企业在防范 AI 误用和滥用方面的努力，包括算法公平性审计、偏见消除机制、模型解释性和可追溯性的改进举措等。通过这种方式，企业不仅可以向公众展示其在负责任的 AI 发展道路上所取得的进步，也能够为行业树立起高标准的合规典范。企业同样应当投身于公益活动与公共教育事业，着力提升公众对 AI 技术的认知水平和数字素养。可以通过举办公益讲座、在线课程、工作坊等形式普及 AI 基础知识，让更多群体能够理解和适应智能科技的发展。

7.2.3 推进 AI 可持续发展与生态共建

人才培养与学术合作是强化行业根基的重要环节。企业应与高等教育机构紧密联手，设立联合培养项目，支持奖学金计划、实习基地建设和科研项目合作等形式，培养具备深厚专业技术知识和高伦理素养的 AI 人才。通过资助前沿课题研究、共享实验资源等方式，鼓励创新思维和技术突破，为 AI 领域输送源源不断的高质量人才资源。

为了营造开放共赢的生态环境，企业可选择性地开放部分核心技术与资源，对外提供易于集成的 API 接口和服务，鼓励第三方开发者利用这些资源进行应用创新和增值服务开发。同时，构建并维护一个活跃的开发者社区，通过论坛、研讨会、黑客马拉松等活动加强交流与协作，监督合作伙伴对开放资源的使用情况，确保其符合伦理道德和法律法规要求，防止技术滥用或误用带来的潜在风险。

7.3 个人层面的应对与自我调适

7.3.1 提高数字素养与信息安全意识

个人需系统地学习并掌握个人信息保护的基本策略和技术手段，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，避免无意识的泄露或滥用。具体到 AI 工具的应用上，用户需要深刻认识到其双刃剑特性，特别是在学术研究、医疗法律等领域，不仅要辨明是非，审慎对待由 ChatGPT 等大语言模型生成的内容，更要遵循透明度原则，明确标注并引用 AI 在创作过程中的贡献，以维护学术诚信和知识产权，避免潜在的风险和不诚信行为。

7.3.2 职业规划与技能升级

在人工智能时代背景下，个体职业规划与技能升级成为了实现个人持续发展和适应未来工作环境的核心策略。面对 AI 技术的迅速崛起及其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。日常工作中，利用自动化和智能化手段提高工作效率与质量的同时，也需要警惕过分依赖 AI 所带来的潜在风险，在面对复杂问题时摆脱对 AI 工具的过度依赖，适时切换至人工判断模式，发挥人类特有的直觉、情感及伦理道德考量的优势。

在时间资源和经济条件允许下，参与各类职业技能培训项目，针对新兴的人工智能领域进行有针对性的学习，例如通过在线课程体系掌握数据分析、机器学习等关键技术，利用学术资源跟踪阅读最新研究论文，参与开源社区讨论，等等，不断更新知识结构、磨砺专业技能、强化自主学习能力，并在实践中探索人机协作的最佳平衡点，促进个人职业发展与社会进步相协调，充分挖掘和释放人工智能时代赋予个体的潜能与机遇。

7.3.3 社会参与与监督

在数字化社会的构建与发展中，公民作为网络空间中的主体，不仅应积极关注并投身于人工智能伦理规范、数据权益保护以及 AI 技术对社会影响等重大议题的公共讨论之中，而且要主动倡导和践行负责任的 AI 应用理念，通过自身的行动推动形成健康的数字文化观念。

个人应当利用社交媒体、网络平台等工具，发起或响应倡议活动，宣传普及有关 AI 使用的道德标准和社会责任，引导公众正确看待和使用智能工具。例如，针对 ChatGPT 等 AI 产品，可以分享如何合理、合规地运用这些工具，同时揭示潜在的风险和滥用现象，提高大众对于信息真实性和隐私权的认识。

此外，面对 AI 技术被滥用或对个人及公众权益构成侵害的现象，每一位公民都有责任行使自己的举报权，及时将问题反馈给相应的监管部门，共同携手政府和企业维护网络空间的公正和安全。同时，普通民众也可以积极加入到企业履行社会责任的行动中，通过日常消费选择、舆论力量等途径，倡导企业遵循伦理准则，在开发和应用 AI 技术时秉持道德底线，在追求经济效益的同时，更加重视社会公益和长远发展目标。

7.4 对策与建议小结

政府、企业和个人三个层面在应对 ChatGPT 大语言模型广泛应用所带来的挑

战与机遇时，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。政府层面通过建设人工智能法制体系和战略生态，为 AI 技术的发展奠定了坚实的法规基石和战略导向，确保大语言模型能够在规范化、可持续的轨道上前行；企业作为技术创新和应用的主力军，需在追求经济效益的同时，承担起数据治理与合规运营的重任，强化技术创新，同时积极响应社会责任，通过共建 AI 可持续发展生态，促进整个行业的健康发展，确保大语言模型的实际应用既能满足用户需求，又能有效规避潜在风险；个人用户作为技术的直接使用者和受益者，提升数字素养与信息安全意识，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，有助于构建良好的数字社会环境，有力推动人工智能技术与个人生活、工作更深层次地融合，并在一定程度上缓解和抵御技术发展带来的潜在负面影响。

政府、企业和个人三个层面在面对国内外大语言模型的应用与发展时，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。

第8章 研究总结与展望

8.1 研究总结

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，也带来了多维度、多层次的风险与挑战。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，试图探究用户对以 ChatGPT 为代表的大语言模型的主要关注点，分析其在实际应用中所暴露出的风险和问题，并为相关主体提出针对性的治理策略与应对措施。具体研究过程如下。

编写爬虫代码获取有关“ChatGPT 大语言模型”的微博用户评论，通过文本预处理后，人工标注部分评论数据的情感倾向，整合作为训练集。为提高情感分析结果的准确性，本文基于自然语言处理领域先进研究成果 Bert 预训练模型构建情感分类模型，结合随机森林、支持向量机模型进行对比实验，确定 Bert 预训练模型为最佳情感分类器，并进行情感分类。结果显示，在 36677 条评论数据中，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%，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%。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，分析大众对于“以 ChatGPT 为代表的大语言模型”的主要关注点；同时，按天数统计不同粗粒度情感的评论数量，建立情感时间序列，构建时间序列模型进行时间序列分析。

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，又涉及到对外部社会环境的影响。用户在对其有着良好预期的同时，主要关注点呈现复杂化、多元化的特征。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，本研究据此提出了相应的治理路径：政府层面，建设人工智能法制体系和战略生态；企业层面，需在追求经济效益的同时，承担起数据治理与合规运营的重任，强化技术创新；个人层面，提高数字素养与信息安全意识，与时俱进地进行职业规划与技能升级，积极参与社会监督。三者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。

8.2 研究不足与展望

尽管本研究通过深入分析社交媒体大数据并运用先进的自然语言处理技术揭示了用户对 ChatGPT 大语言模型的关注焦点并给出相关对策建议，但仍存在一些局限性和未来可进一步探索的方向：

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和观点，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。因此，未来的研究工作中，可以考虑纳入微信、百度贴吧、各类专业论坛以及博客等多种类型的国内主流社交平台数据，甚至延伸至国际社交平台如推特等，从而构建更为全面的社交媒体数据分析框架；同时，进行国内外公众情绪对比研究，探索全球化背景下不同文化和社会环境下公众对大语言模型的认知差异。

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。为了更好地把握用户情感变化的长期动态，未来研究有必要延长数据收集的时间窗口，增加样本容量，确保研究结论的稳定性和时效性。同时，随着 ChatGPT 及同类大语言模型在更多场景下的应用，持续追踪和分析不断涌现的新鲜舆情数据，也将有助于我们及时了解和评估这类新兴技术的社会影响及风险演变。

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。未来研究可以进一步引入更多的深度学习模型，比如 Transformer 系列的其他变体如 RoBERTa、ALBERT 或者 GPT 系列模型等，进行更加丰富和深入的对比实验，以期探寻在情感分析任务中性能更优且更具普适性的算法模型。

参考文献

- [1] Zhang, C., et al., One Small Step for Generative AI, One Giant Leap for AGI: A Complete Survey on ChatGPT in AIGC Era. arXiv e-prints, 2023: p. arXiv:2304.06488.
- [2] Brown, T.B., et al., Language Models are Few-Shot Learners. arXiv e-prints, 2020: p. arXiv:2005.14165.
- [3] 国家重点专项规划之——“十三五”国家战略性新兴产业发展规划（国发〔2016〕67号）[EB/OL].https://www.ndrc.gov.cn/fggz/fzzlgh/gjjzxgh/201705/t20170508_1196736.html. 2017-05-08.
- [4] 国务院印发《新一代人工智能发展规划》[EB/OL].http://www.gov.cn/xinwen/2017-07/20/content_5212064.htm. 2017-07-20.
- [5] 国家标准化管理委员会, 中央网信办, 国家发展改革委, 科技部, 工业和信息化部. 关于印发《国家新一代人工智能标准体系建设指南》的通知（国标委联〔2020〕35号）[EB/OL].http://www.gov.cn/zhengce/zhengceku/2020-08/09/content_5533454.htm. 2020-08-09.
- [6] 科技部等六部门关于印发《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》的通知（国科发规〔2022〕199号）[EB/OL].http://www.gov.cn/zhengce/zhengceku/2022-08/12/content_5705154.htm. 2022-08-12.
- [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(05): 9-13.
- [8] 国家互联网信息办公室 中华人民共和国工业和信息化部 中华人民共和国公安部 令《互联网信息服务深度合成管理规定》（第12号）[EB/OL].http://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm. 2022-12-11.
- [9] 国家互联网信息办公室关于《生成式人工智能服务管理办法（征求意见稿）》公开征求意见的通知[EB/OL].http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm. 2023-04-11.
- [10] 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: 1-12.
- [11] 宫承波. 新媒体概论[M]. 中国广播影视出版社:, 2021: 381.
- [12] Liu B. Sentiment analysis: Mining opinions, sentiments, and emotions[M]. Cambridge University Press, 2015.
- [13] 王敏, 徐健. 视频弹幕与字幕的情感分析与比较研究[J]. 图书情报知识, 2019(5): 109-119.
- [14] 张云中, 李佳书. 媒体型智库国际关系话题的情感分析: 原理解析、工具构建与应用探索[J/OL]. 情报科学, 2023: 1-20.
- [15] 姜伶俐, 何中市, 张航. 基于 Good-Turing 平滑 SO-PMI 算法构建微博情感词典方法的研究[J]. 现代计算机(专业版), 2018(10): 15-20.

- [16] 钟佳娃, 刘巍, 王思丽, 杨恒. 文本情感分析方法及应用综述[J]. 数据分析与知识发现, 2021, 5(6):1-13.
- [17] Turney P. D.. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews[C]// Meeting of the Association for Computational Linguistics. 2002: 417-424.
- [18] 钟智锦, 王童辰. 大数据文本挖掘技术在新闻传播学科的应用[J]. 当代传播, 2018(5):12-18.
- [19] 王婷, 杨文忠. 文本情感分析方法研究综述[J]. 计算机工程与应用, 2021, 57(12):11-24.
- [20] Pang B, Lee L, Vaithyanathan S. Thumbs up: sentiment classification using machine learning techniques[C]. Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002: 79-86.
- [21] 徐琳宏, 丁堃, 林原, 杨阳. 基于机器学习算法的引文情感自动识别研究——以自然语言处理领域为例[J]. 现代情报, 2020, 40(1):35-40+48.
- [22] 唐慧丰, 谭松波, 程学旗. 基于监督学习的中文情感分类技术比较研究[J]. 中文信息学报, 2007, 21(6):88-94
- [23] 杨爽, 陈芬. 基于 SVM 多特征融合的微博情感多级分类研究[J]. 数据分析与知识发现, 2017, 1(2):73-79.
- [24] Liu L, Feng S, Wang D, et al. An empirical study on Chinese microblog stance detection using supervised and semi-supervised machine learning methods[M]// Natural language understanding and intelligent applications. [S.l.]: Springer International Publishing, 2016.
- [25] Jiang F, Liu Y, Luan H, et al. Microblog sentiment analysis with emoticon space model[C]// Chinese National Conference on Social Media Processing, 2014.
- [26] 陈龙, 管子玉, 何金红, 彭进业. 情感分类研究进展[J]. 计算机研究与发展, 2017, 54(6):1150-1170.
- [27] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3: 1137-1155.
- [28] Li Q, Jin Z, Wang C, et al. Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems[J]. Knowledge-Based Systems, 2016, 107: 289-300.
- [29] 罗帆, 王厚峰. 结合 RNN 和 CNN 层次化网络的中文文本情感分类[J]. 北京大学学报(自然科学版), 2018, 54(3):459-465.
- [30] Hinton G E , Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [31] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. arXiv: 1706.03762, 2017.
- [32] 顾军华, 彭伟桃, 李娜娜, 董永峰. 基于卷积注意力机制的情感分类方法[J]. 计算机工程与设计, 2020, 41(1):95-99.
- [33] Devlin, J., et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in 2019 CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES (NAACL HLT 2019), VOL. 1. 2019: Conference of the

- North-American-Chapter of the Association-for-Computational-Linguistics-Human Language Technologies (NAACL-HLT). p. 4171-4186.
- [34] 李舟军, 范宇, 吴贤杰. 面向自然语言处理的预训练技术研究综述[J]. 计算机科学, 2020, 47(03):162-173.
- [35] 吴成斌. 文本情感分类算法在舆情分析中的应用[D]. 南昌大学, 2021.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] 李冉, 王佳英, 单菁等. 突发事件微博网络舆情分析及可视化——以“武汉封城”为例[J]. 情报探索, 2022, (2):67-72.
- [42] 李白杨, 白云, 詹希旎, 李纲. 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.
- [43] 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.
- [44] 赵滨, 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.
- [45] 蔡子凡, 蔚海燕. 人工智能生成内容(AIGC)的演进历程及其图书馆智慧服务应用场景[J]. 图书馆杂志, 2023, 42(4):34-43+135-136.
- [46] 储节旺, 杜秀秀, 李佳轩. 人工智能生成内容对智慧图书馆服务的冲击及应用展望[J]. 情报理论与实践, 2023, 46(5):6-13.
- [47] 赵杨, 张雪, 范圣悦. AIGC 驱动的智慧图书馆转型: 框架、路径与挑战[J/OL]. 情报理论与实践:1-10.
- [48] 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/OL]. 中国图书馆学报:1-12.
- [49] 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.
- [50] 游俊哲. ChatGPT 类生成式人工智能在科研场景中的应用风险与控制措施[J/OL]. 情报理论与实践:1-11.
- [51] Dwivedi, Y.K., et al., "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT, 2023. 71.
- [52] 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, 4

- 7(3):112-120.
- [53] 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.
- [54] 梁少博. 面向“以人为本”的生成式人工智能: 挑战与路径[J]. 图书馆建设, 2023(4):19-21.
- [55] 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/OL]. 情报理论与实践:1-11..
- [56] 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.
- [57] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [58] 段瑞雪, 巢文字, 张仰森. 预训练语言模型 BERT 在下游任务中的应用[J]. 北京信息科技大学学报(自然科学版), 2020, 35(6): 77-83.
- [59] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3):273-297.
- [60] 李航. 统计学习方法(第 2 版)[M]. 北京:清华大学出版, 2019.
- [61] Blei, D.M., A.Y. Ng and M.I. Jordan, Latent Dirichlet allocation. JOURNAL OF MACHINE LEARNING RESEARCH, 2003. 3(4-5): p. 993-1022.
- [62] Griffiths, T.L. and M. Steyvers, Finding scientific topics. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 2004. 101: p. 5228-5235.
- [63] Box G E P , Jenkins G M , Reinsel G C .Time series analysis forecasting and control - Rev. ed.[J].Journal of Time, 1976, 31(2):238-242.
- [64] 岳丽欣, 周晓英, 陈旖旎. 基于 ARIMA 模型的信息构建研究主题趋势预测研究[J]. 图书情报知识, 2019(5):54-63+72.
- [65] 杨健, 陈伟. 基于 Python 的三种网络爬虫技术研究[J]. 软件工程, 2023, 26(2):24-27+19.
- [66] 李学龙, 龚海刚. 大数据系统综述[J]. 中国科学:信息科学, 2015, 45(1):1-44.
- [67] 石凤贵. 基于 jieba 中文分词的中文文本语料预处理模块实现[J]. 电脑知识与技术, 2020, 16(14):248-251+257.
- [68] 杨欢. 文本情感分类预处理研究[J]. 电子技术与软件工程, 2016(10):187.
- [69] 史童月, 王中卿. 基于 Transformer 的自然语言处理预训练语言模型概述[J]. 信息与电脑(理论版), 2022, 34(10):52-56.
- [70] 黄玉萍, 梁炜萱, 肖祖环. 基于 TensorFlow 和 PyTorch 的深度学习框架对比分析[J]. 现代信息科技, 2020, 4(4):80-82+87.
- [71] 丁美荣, 刘鸿业, 徐马一等. 面向机器阅读理解的多任务层次微调模型[J]. 计算机系统应用, 2022, 31(3):212-219.
- [72] 张爱华, 荆继武, 向继. 中文文本分类中的文本表示因素比较[J]. 中国科学院研究生院学报, 2009, 26(3):400-407.
- [73] 柴玉梅, 王宇. 基于 TFIDF 的文本特征选择方法[J]. 微计算机信息, 2006(24):24-26.
- [74] 刘胜杰, 许亮. 基于词嵌入技术的文本表示研究现状综述[J]. 现代计算机, 2020(1):40-43.
- [75] Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation

- [C]//Conference on Empirical Methods in Natural Language Processing. 2014.
- [76] 陈珍锐, 丁治明. 基于 GloVe 模型的词向量改进方法[J]. 计算机系统应用, 2019, 28(1): 194-199.
- [77] 李心蕾, 王昊, 刘小敏等. 面向微博短文本分类的文本向量化方法比较研究[J]. 数据分析与知识发现, 2018, 2(8): 41-50.
- [78] Bojanowski P , Grave E , Joulin A , et al. Enriching Word Vectors with Subword Information[J]. Transactions of the Association for Computational Linguistics, 2017, 5: 135-146.
- [79] Joulin A , Grave E , Bojanowski P , et al. Bag of Tricks for Efficient Text Classification[J]. 2017.
- [80] 徐晓明. SVM 参数寻优及其在分类中的应用[D]. 大连海事大学, 2014.
- [81] 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.
- [82] 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.
- [83] 李荣陆. 文本分类及其相关技术研究[D]. 复旦大学, 2005.
- [84] Davis J . The Relationship Between Precision-Recall and ROC Curves[C]//Proceedings of the 23th International Conference on Machine Learning, 2006. 2006.
- [85] Fawcett Tom. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006.
- [86] 汪云云, 陈松灿. 基于 AUC 的分类器评价和设计综述[J]. 模式识别与人工智能, 2011, 24(1): 64-71.
- [87] Antweiler W, Frank M Z. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards[J]. Journal of Finance, 2004, 59(3): 1259-1294.
- [88] Lutz B, Prilochs N, Neumann D . Predicting sentence-level polarity labels of financial news using abnormal stock returns[J]. Expert Systems with Applications, 2020, 148: 113223.
- [89] 王立柱. 基于计算智能的时间序列模型及预测研究[D]. 大连理工大学, 2017.
- [90] 杨婕. 论个人信息保护中的必要原则: 法律理解、实践困境与化解思路[J]. 信息安全研究, 2023, 9(7): 675-680.
- [91] 郑少飞. 人工智能技术应用风险法律分配的制度逻辑[D]. 吉林大学, 2023.
- [92] 刘恒, 王伟, 靳文虎等. AI 时代网络信息安全问题及对策——以 ChatGPT 为例[J]. 网络安全技术与应用, 2023, (11): 160-162.
- [93] Stahl, B.C. and D. Eke, The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. International Journal of Information Management, 2024. 74: p. 102700.
- [94] 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, 2023: 1-8.
- [95] 缪昊洋, 高谭芮, 汤影. 基于生成模型的联邦学习隐私保护算法[J]. 电子设计工程, 2023, 31(24): 81-84+89.
- [96] 钟晓雯. 从算法“黑箱”走向算法透明: 基于“硬法—软法”的二元法治理模式[J]. 中国海商法研究, 2023, 34(4): 53-62.