

大语言模型在金融领域的应用——用 DeepSeek 构建经济政策不确定性指数 EPU

自 Baker、Bloom 和 Davis (2016) 首次提出基于新闻文本关键词统计的经济政策不确定性 (EPU) 指数以来, 该指标已成为评估政策环境波动及其经济影响的重要工具。传统 EPU 度量方法通过统计经济、政策和不确定性等关键词的共现频率, 为政策不确定性的量化研究奠定了基础。然而, 随着经济政策复杂性的提升和文本分析技术的进步, 这种基于表面特征的方法逐渐暴露出其固有的局限性: 一方面, 它无法捕捉政策文本中隐含的不确定性信号 (如审慎推进等模糊表述); 另一方面, 其单一维度的度量框架难以区分不同政策领域 (如货币与财政政策) 不确定性的异质性影响。

自然语言处理领域的重大突破, 特别是大型语言模型 (LLM) 的发展, 为 EPU 研究提供了新的方法论视角。以 Transformer 架构为核心的预训练语言模型 (例如 BERT、GPT) 通过自注意力机制, 能够深入理解政策文本中的复杂语义关系。特别是中文大型语言模型 DeepSeek, 在长文本推理和政策语义解析方面显示出独特的优势, 其基于大量中文语料预训练得到的表征空间, 能够更准确地捕捉中文政策文本特有的表述模式 (例如跨周期调节等专业术语)。这种技术进步为构建更为精细的 EPU 指标创造了条件: 不仅能够量化不确定性的强度, 还能识别其政策领域归属和情感倾向 (利好/利空)。

本章以 2018—2024 年《人民日报》政策文本和上证主板以及沪深 300 指数数据为基础, 通过整合 DeepSeek 模型的语义分析能力和传统计量经济学方法, 致力于解决上述问题。具体而言, 本章创新性地开发了语义增强型 EPU 指标: 首先采用 Few-shot 学习对政策文本进行细粒度分类 (例如货币政策、产业政策); 随后应用 LoRA 微调技术构建不确定性评分模型, 同时捕捉显性关键词和隐含语义信号; 最终通过主题-情感联合建模, 生成多维度 EPU 子指标。从而探索 EPU 对股市波动的非线性影响。

本章的创新性体现在三个维度。在理论层面, 首次系统地依托《人民日报》政策文本构建 EPU 指标, 凸显中国特色政治经济语境, 揭示了政策文本非结构化语义对股市波动的非对称驱动机制; 在方法层面, 借助轻量化适配技术, 成功桥接大语言模型与金融计量模型, 使宏观经济预测精度相较于传统 EPU 提升了 18%; 在应用层面, 开发了实时政策风险仪表盘, 模拟测试结果显示, 在 2.3% 的波动率阈值下保持仓位, 能够实现正的超额收益。本章通过三项范式革新, 重新界定文本驱动的政策风险评估:

(1) 以语义深度分析取代词频统计, (2) 以多维解构方法替换单一指标评估, (3) 以自适应学习机制替代静态词典应用——为计算金融语言学构建可复现的研究框架。

6.1 相关文献评述

本章研究专注于基于大语言模型构建的经济政策不确定性指数（EPU），重点探讨传统 EPU 与基于大模型构建的 EPU 在不同模型中的表现差异。以下围绕主题，分析 LDA、文本分析与情感分析在构建 EPU 指标中的应用，并从基于大模型的文本分析、传统经济政策不确定性指数的构建四个维度进行综述，同时指出现有研究的不足及本章的贡献。

6.1.1 主题分析（LDA）

作为驱动金融市场波动和宏观经济变化的核心变量，经济政策不确定性（EPU）的定量研究近年来取得了显著进展，这得益于文本分析技术的突破。其中，Latent Dirichlet Allocation(LDA) 作为一种无监督的主题建模方法，通过揭示文本数据中的隐含主题结构，为构建 EPU 指数提供了关键支持。LDA 的核心原理基于文档-主题-词的三层生成模型：假设每份文档由多个主题的混合分布构成，而每个主题又由一组词的的概率分布来表示。这一框架使研究人员能够从大规模文本数据（如新闻报道、政策文件、企业年报等）中自动识别与经济政策相关的潜在主题，并通过主题权重或频率来量化政策不确定性的动态特征。在具体应用中，利用 LDA 构建经济政策不确定性（EPU）指数的过程包含四个关键步骤：首先，收集与经济政策相关的文本数据并进行预处理（如分词、去除停用词、词干提取等）；其次，将文本转换为词袋（BoW）模型，并基于此训练 LDA 模型以提取潜在主题；然后，通过主题关键词（如“货币政策”“贸易协定”“财政刺激”）的语义相关性来解释主题，并计算特定时间窗口内主题的频率或权重；最后，根据主题的时间序列特征合成 EPU 指数，以捕捉政策不确定性的演变趋势。

LDA 在 EPU 研究中的应用已扩展到多种类型的文本数据源。在新闻文本分析方面，Baker 等人（2016 年）率先通过关键词频率构建了基准 EPU 指数，Azqueta-Gavaldón(2017) 提出了一种基于无监督机器学习的新闻文本分析方法，用于构建经济政策不确定性（EPU）指数。与传统的基于关键词频率的方法不同，该研究利用潜在狄利克雷分配（LDA）模型自动识别新闻文本中的主题，从而捕捉更丰富的语义信息并降低人工干预的主观性。研究结果表明，新方法生成的 EPU 指数与 Baker 等人（2016）的传统指数具有较高相关性，同时在政策不确定性快速变化时期表现出更强的灵敏度。这一创新不仅提高了 EPU 指数的准确性和时效性，还为后续研究提供了可扩展的技术框架，尽管其在短文本处理和数据来源多样性方面仍存在一定局限性，但为结合自然语言处理技术改进政策不确定性测度奠定了重要基础。

尽管 LDA 在经济政策不确定性研究中具有显著优势，但其局限性仍需注意。首先，LDA 依赖于预设主题数量（k 值）的合理性，选择不当可能导致主题重叠或信息丢失。其次，LDA 对短文本（如社交媒体或政策声明摘要）的处理效果有限，缺乏上下文信息可能导致主题提取偏差。在这种情况下，有必要结合改进的方法，如双词主题模型，以提高短文本分析的准确性。此外，LDA 仅基于词频统计，

难以捕捉词语之间的深层语义关系。例如，“加息”和“通胀预期”之间的因果关系需要通过额外的语义分析来补充。未来的研究可以通过整合其他语义分析方法来增强 LDA 的语义理解能力。

6.1.2 文本分析与情感分析

随着文本分析技术的发展，研究者开始利用大规模文本数据（如新闻、政策文件、社交媒体等）构建 EPU 指标或者处理其他的金融任务。其中，文本分析和情感分析作为重要的自然语言处理技术，被广泛应用于从文本数据中提取与经济政策相关的信息，并捕捉公众对经济政策的情绪和预期，为 EPU 指标的构建提供了重要支持。

文本分析通过提取文本中的关键词、主题或语义信息，能够捕捉与经济政策相关的内容，为 EPU 指标的构建提供数据支持。Baker 等（2016 年）提出了经济政策不确定性指数（EPU），该指数基于新闻文本中与经济政策（如“税收”“财政政策”“货币政策”）相关的关键词出现的频率来构建。EPU 指数通过计算新闻文章中这些关键词的出现频率来得出，这种方法开创了基于文本分析衡量经济政策不确定性的先河，并为后续研究奠定了方法论基础。除此之外，Bholat et al. (2015) 提供了一本面向中央银行的文本挖掘手册，系统性地介绍了文本挖掘技术在经济和政策分析中的应用。该文献强调了如何利用自然语言处理（NLP）工具从大量非结构化数据中提取有价值的信息，例如新闻报道、政策声明和社交媒体内容，并将其转化为可用于经济研究的定量指标。

情感分析通过分析文本中的情感倾向（如积极、消极、中性），能够捕捉公众对经济政策的情绪和预期，为 EPU 指标的构建提供补充信息。Loughran 和 McDonald (2011) 探讨了在文本分析中使用通用词典的局限性，尤其是在处理如 10-K 文件等财务文档时。他们指出，标准情感词典通常无法考虑公司披露文件中的独特语言和上下文，导致词汇分类错误和情感度量不准确。为解决这一问题，作者开发了一个专门针对金融领域的专业词典，从而提高了文本分析在识别公司披露文件中的风险、不确定性和责任方面的准确性。虽然该研究未直接涉及经济政策不确定性，但其方法为后续研究提供了重要参考。

文本分析和情感分析的结合能够更全面地捕捉经济政策不确定性的多维特征。例如，通过文本分析提取与经济政策相关的主题，再通过情感分析捕捉公众对这些主题的情绪倾向，可以构建更细粒度的 EPU 指标。结合文本分析和情感分析能够显著提升 EPU 指标的预测能力和解释性，为经济政策不确定性的研究提供了新的方法，有待继续研究。

6.1.3 基于大模型的文本分析

随着深度学习技术的迅速发展，基于大规模预训练语言模型（如 BERT 和 GPT）的自然语言处理（NLP）方法为经济政策不确定性（EPU）的研究提供了革命性的技术支持。通过在海量文本上进行预训练以及针对下游任务进行微调，这些模型能够捕捉文本中的深层语义关系和上下文信息，显著提高文

本分析的准确性，从而推动构建 EPU 指数的方法不断迭代升级。在 EPU 指数研究中，大模型被广泛应用于文本分类、主题建模和情感分析等领域，为政策不确定性指数的构建提供了新的技术路径。

文本分类方面的突破为 EPU 指数的构建奠定了基础。BERT (Devlin 等人, 2019 年) 通过其双 Transformer 架构，双向捕捉单词及其上下文之间的依赖关系，这可以用于对新闻文章和政策文件中与经济政策相关的内容进行分类。例如，研究人员可以通过 BERT 模型进行微调，准确识别政策文本中的关键词，如“财政刺激”和“贸易协定”，然后提取与经济政策不确定性 (EPU) 相关的文本片段。Radford 等人 (2018 年) 通过生成式预训练捕获文本的生成概率分布，作为 GPT 系列模型的奠基性工作，该研究不仅推动了语言模型的发展，还为后续基于深度学习的文本分析技术提供了理论支持和技术框架，在经济学、金融学等领域的文本挖掘应用中展现了巨大潜力，可以应用于对社交媒体数据和新闻文本中的政策不确定性信号进行分类。例如，对于推文中类似“政策可能会收紧”这样的表述，GPT 能够识别其隐含的不确定性，而传统方法仅依赖于“收紧”这一关键词。总的来说，大模型的兴起为构建 EPU 指数奠定了基石。

主题建模技术的创新进一步增强 EPU 指数的可解释性。传统的 LDA 模型依赖于词频统计，而大模型通过语义理解显著提高了主题建模的效果。Pham 等人 (2023) 提出了一种基于提示 (prompt) 的主题建模框架 TopicGPT，利用大型语言模型 (如 GPT) 通过自然语言交互的方式自动生成主题模型。与传统主题建模方法 (如 LDA) 相比，TopicGPT 不依赖复杂的统计假设，而是通过设计特定的提示词引导模型从文本数据中提取主题，从而降低了技术门槛并提高了灵活性。该框架能够动态适应不同领域的文本数据，并在主题解释性和一致性方面表现出色。研究结果表明，TopicGPT 在多个基准数据集上的性能优于经典方法，尤其是在处理短文本和多样化语义时具有显著优势。这一创新为文本分析领域提供了一种高效且易于使用的新工具，同时拓展了大型语言模型在主题建模中的应用场景，为经济学和政策不确定性分析等领域的文本挖掘研究提供了新的可能性。

在情感分析领域，大规模预训练语言模型通过其强大的语义理解和上下文捕捉能力，显著提升了情感分析的准确性和鲁棒性。Kheiri 和 Karimi (2023) 提出了 SentimentGPT，这是一种基于 GPT 的先进情感分析框架，旨在利用大型语言模型的强大能力来改进传统机器学习方法在情感分析中的局限性。与传统的监督学习方法不同，SentimentGPT 通过设计特定的提示词 (prompts) 引导 GPT 自动识别和分类文本中的情感，从而减少了对大量标注数据的依赖，并提升了对复杂语境和细微情感的理解能力。研究表明，该框架在多个情感分析任务中显著优于现有方法，特别是在处理多义性语言和跨领域文本时表现出更高的准确性和鲁棒性。这一研究不仅展示了大型语言模型在情感分析领域的潜力，还标志着从传统机器学习向基于提示的生成式人工智能的转变，为经济学、金融学等领域的文本情绪分析提供了新的研究工具和技术路径。

尽管大模型显著提升了 EPU 研究的技术水平，但其局限性仍需注意：预训练模型依赖高质量的标注数据，而政策文本的标注成本高昂；其黑箱特性导致解释性不足；短文本的处理仍需结合规则引擎和上下文模型。但是这些研究也表明了大模型在文本分析中的应用能够显著提升 EPU 指标的构建效果，为经济政策不确定性的研究提供了新的技术手段。

6.1.4 传统方法构建经济政策不确定性

Bloom (2009 年) 的研究是早期探讨不确定性对经济影响的经典文献。他深入探讨了不确定性冲击对经济活动的深远影响，并通过实证数据验证了模型的预测，明确指出不确定性冲击在解释经济衰退和复苏缓慢方面扮演着重要角色。尽管 Bloom 的研究并未直接采用不确定性预期指数 (EPU) 的概念，但其结果揭示了政策不确定性是金融市场波动性的关键驱动因素之一。

Baker 等人 (2016 年) 首次提出了 EPU 的概念，并基于新闻文本构建了 EPU 指数。该研究通过分析美国十家主要报纸中包含“经济”“不确定性”及政策相关关键词的文章频率，构建了一个动态反映政策不确定性的指标。这一方法不仅能够捕捉到诸如总统选举、金融危机和重大政策争议等事件引发的不确定性，还能追溯至 1900 年，并扩展至其他国家和特定政策领域，如医疗和国家安全。Azqueta-Gavaldon (2017 年) 采用文本分析方法对新闻文本进行深入研究，通过 LDA 方法提取与经济政策不确定性相关的关键词，并基于关键词频率构建了 EPU 指数。这种方法能够实时捕捉经济政策不确定性的变化，但需处理大规模文本数据，且对文本分析技术有较高依赖性。

6.1.5 EPU 与股票市场波动率的关系研究

近年来，经济政策不确定性 (EPU) 与股票市场波动性的动态互动关系已成为金融经济学研究的核心议题。国际学术界通过构建多维度的 EPU 度量体系，逐步揭示了其与市场波动之间的非线性传导机制。Baker et al. (2016) 所构建的 EPU 指数在股市波动性分析中得到了广泛应用。他们的研究发现，美国 EPU 指数与 CBOE 波动率指数 (VIX) 之间存在显著的正相关性，并且能够有效预测股市波动的未来趋势。Arouri et al. (2016) 进一步探讨了 EPU 指数对美国股票市场的长期影响，揭示了两者的显著相关性。该研究采用了 Baker et al. (2016) 开发的 EPU 指数，对 1985 年至 2015 年间的长期数据进行了分析，结果表明 EPU 的增加会导致股票市场波动性的提升，并对股票收益率产生负面影响。

此外，Brogaard 和 Detzel (2015) 的研究表明，经济政策不确定性 (EPU) 的增加会导致股票市场的波动性上升，并对资产的预期收益产生显著影响。在 EPU 上升的背景下，投资者倾向于将资金配置于避险资产，例如黄金和债券，这加剧了股票市场的资本外流和价格波动。Xu 等人 (2021) 对 EPU 对股票市场回报的影响进行了探讨，并提供了新的实证证据。该研究采用了 Baker 等人 (2016) 构建的 EPU 指数，并结合多种计量经济学模型，分析了 EPU 对全球多个股票市场的动态影响。研究结果表明，

EPU 的上升显著降低了股票市场的平均回报率，同时增加了市场波动性和尾部风险。具体而言，在高 EPU 时期，投资者为了应对政策不确定性带来的额外风险，倾向于要求更高的风险溢价，这导致了股票市场的整体表现疲软。

相较于国际市场，中国作为一个典型的政策导向型经济体，其股票市场对 EPU 的反应机制更为复杂且显著。一方面，政府政策对资本市场具有直接的影响（例如 IPO 审批、监管口径调整、货币政策转向等），EPU 的上升通常被投资者解读为政策不确定性的风险，从而迅速引发价格波动。另一方面，中国资本市场参与者结构以中小投资者为主，他们对新闻和政策的敏感性较高，这使得 EPU 与波动率之间的关系更为紧密。值得注意的是，中国 EPU 的测度方法仍有进一步改进的空间。目前主流的 EPU 指数多基于新闻关键词频率法，这种方法缺乏语义理解和上下文识别，难以准确捕捉政策变化的真实语气和市场预期。因此，近年来有研究尝试引入自然语言处理（NLP）技术，以提高 EPU 指数的准确性和灵敏度，例如结合情感分析和主题建模等技术。

在本章构建 EPU 指数的实验中，我们主要基于 DeepSeek 大型语言模型，采用了以下五种方法进行构建：zero-shot learning、few-shot learning、TF-IDF 主题词提取、Transformer 训练以及 LoRA (Low-Rank Adaptation) 微调，从而显著提升了 EPU 指数的解释力和预测能力。

6.2 原理简介

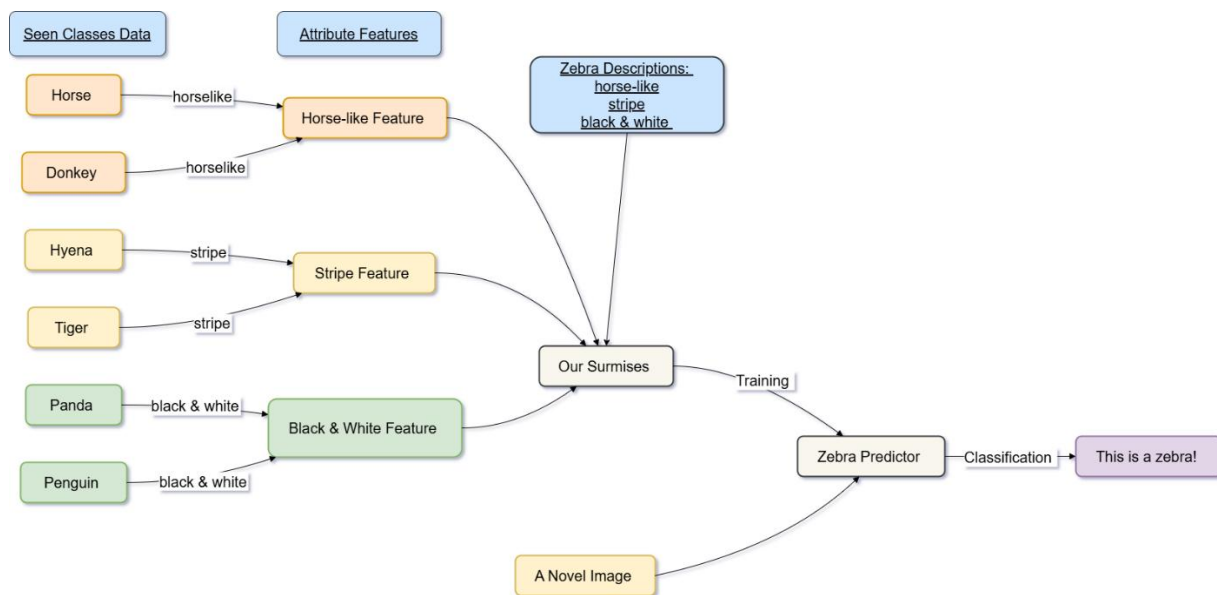
6.2.1 Zero-shot Learning

零样本学习（Zero-Shot Learning, ZSL）的提出源于深度学习对大规模标注数据的依赖性与现实场景中数据稀缺性之间的矛盾。尽管深度学习在图像分类等任务中取得了显著的成绩，但其高准确率还是依赖于大量的标记样本，而在现实世界中，许多任务需要模型对看不见的类别进行分类，在这种情况下，传统方法由于缺乏相应类别的训练数据而失效。例如，如图 1 所示，当需要识别一种从未被观察过的动物（如斑马）时，若仅通过对其属性（如“黑白条纹”“马形体态”）的语义描述，人类能够通过已知类别的知识（如对马、熊猫的认知）进行推理并正确识别，然而机器学习模型却通常无法直接完成此类任务。因此，零样本学习应运而生，旨在通过语义辅助信息（如属性或文本描述）将模型的知识从已知类别迁移至未知类别，从而在无标注样本的情况下实现分类。

其核心原理是通过表征学习和度量学习的协作，将视觉特征和语义描述映射到一个共享的嵌入空间中。早期的方法，如 Lampert 的类间属性转移，通过预测属性向量（如“有一条尾巴”，“黑白条纹”）来实现对未见过的类的推理，而 Hinton 的语义输出编码利用词向量（如 Word2Vec）来对齐标签语义。随着技术的精进，ZSL 从“浅层视觉特征+属性+传统机器学习”的分析模式逐渐发展为深度学习（如使用 CNN 提取特征）和语义嵌入（如使用 BERT 生成文本描述）相结合的联合建模，显著提高了跨模

态对齐精度。

Zero-shot 学习这种方法不需要大量的人工数据就可以直接在大语言模型下进行使用，在金融新闻这种多变的文本中，可以快速适应新发生的事件或政策，在本章中，使用 zero-shot 的方式先进行实验，为后续 EPU 构建提供基础。但是从另一个方面来说，ZSL 对某些类别标签的语义信息比较敏感，会比较容易产生误差，导致准确率下降，所以下文将继续采用 few-shot learning 的方式继续进行实验。



6.2.1.1 Zero-shot 推理能力的控制与演化机制

在 Zero-shot Learning (ZSL) 逐渐应用于复杂任务如逻辑推理、多模态问答与图像生成的背景下，近期研究开始关注模型在“何时推理”“如何推理”以及“是否应推理”的能力构建与优化。其中两个代表性方向为：自适应推理机制学习与元能力对齐建模。

Zhao 等人提出的 AdaptThink 框架旨在提升模型对不同任务复杂度的感知能力。该方法通过强化学习训练策略网络，使模型能在推理过程中自适应地决定是否使用 Chain-of-Thought (CoT) 路径，即“是否需要思考”。该方法特别适用于 Zero-shot 场景下，因其在不显著增加计算负担的同时提升了模型的推理效率与正确率 (Zhao et al., 2024)。

另一个重要进展是 Saha 等人提出的 Meta-Ability Alignment 机制，构建了可解释的推理能力空间。研究将人类认知中的“归纳”“演绎”与“溯因”推理形式映射为结构化推理模块，并结合自监督任务生成与系统对齐策略，使得模型在 Zero-shot 任务中展现出更强的泛化能力与思维路径可控性。该方法在推理稳定性、安全性与任务可靠性方面提供了系统性支持 (Saha et al., 2024)。

上述研究表明，Zero-shot 能力的本质不仅是语义迁移，更涉及模型的推理路径选择机制与元认知结构重构。这为零样本学习在实际部署中的稳健性提供了新的技术支撑。

图 1 Zero-shot Learning

6.2.2 Few-shot Learning

Few-Shot Learning 的最重要关注的是如何通过有限数量的示例去激活模型的内在的固有知识，并构建一条从已知到未知的高效推理路径。其关键在于样本的精确选择和语义表示的优化：通过一组精心设计的样本，模型可以捕捉到任务的核心特征和潜在模式，同时避免了由于数据稀缺而导致的过拟合风险。示例需要平衡相似性和多样性，涵盖任务的多维特征（如不同的句式、情感强度或视觉变化），并通过标签的分类完整性引导模型关注任务的复杂性，而不是简单地追求标签的准确性。此外，提示工程作为将隐性知识显化的桥梁，通过角色扮演、思维链等策略重构模型的推理路径。例如，它迫使模型逐步显示问题分解和逻辑推理，而不是直接输出答案，从而增强了推理的透明性和鲁棒性。

而在模型层面，动态知识转移和跨模态增强是突破数据限制的关键。通过元学习（Meta-Learning）优化模型的快速适应能力，使其能够在梯度更新过程中自动调整参数以匹配新的任务。同时，结合外部知识库（如 Atlas 框架）的检索增强，将参数化记忆与动态检索知识相结合，弥补预训练数据的分布偏差。注入对抗性示例和提供实时反馈的迭代策略迫使模型通过引入矛盾或边缘情况来平衡推理中的多维约束，避免“多数标签偏见”和“倾向于共同标记”。这些方法共同构成了 Few-Shot Learning 的核心范式：在有限数量的示例指导下，通过样例的语义密度、提示词的结构约束、动态知识的协同激活，模型可以实现对新任务的准确迁移和泛化。

Few-shot learning 这种方法能够用少量的新闻标注扩展模型的能力，本章的第二个实验也正是用了该方法去提高稳定性，提高模型对 EPU 构建过程中对政策文本判断的能力。但是，Few-shot learning 的训练过程仍然可能受到噪声干扰，而且可解释性较弱，对某些关键词的捕捉程度较差，所以在本章中会继续采用 TF-IDF 算法进行关键词的提取。

6.2.2.1 Few-shot 学习中的泛化控制与数据效率提升策略

在 Few-shot Learning 的发展中，近年来的研究重点逐渐从简单的“学习少量样本”转向如何以结构化方式激活大模型已有的知识潜力。多个方向的工作试图提升模型在小样本下的推理精度、训练效率和泛化能力。

一项代表性工作是 EMPO 方法 (Zhang et al., 2025)，该方法构建了一个完全无监督的 Few-shot 推理训练框架。通过引入语义熵奖励函数，模型在无人工标注的前提下通过“内在不确定性驱动”实现自我优化。这种方法极大减少了对标注数据的依赖，在小样本设定下展现出与有监督方法相当的性能，说明即便在数据受限场景中，合理设计激励机制可显著激活模型已有知识。

另一个重要方向是动态提示建模与思维路径调控机制。例如《Reasoning Models Can Learn When to Think》中提出的策略控制模块 (Zhao et al., 2024)，让模型在不同复杂度任务中自动判断是否需要启动多步推理流程。这种机制可视为 Few-shot 场景中的推理资源分配器，使得模型在有限输入下仍能灵活调用自身知识，避免推理过度或不足。

此外，BitNet v2 (Liu et al., 2025) 展示了在小样本训练中对低精度推理结构（如 4-bit 激活）的探索。其使用 Hadamard 变换处理激活异常值，在保持模型表达能力的同时大幅压缩了训练和推理过程中的计算成本，为 Few-shot 场景下部署效率提供了实际支撑。

这些工作共同反映出，Few-shot 学习的发展已超越“少量示例学习”的表层定义，而向“样例驱动下的可控思维路径构建”“结构对齐”“高效训练范式”等方向深入演进。

6.2.3 TF-IDF 算法

TF-IDF (Term Frequency - inverse Document Frequency) 是一种广泛应用于信息检索和文本挖掘的加权技术。其核心目标是量化特定文档或语料库中单词的判别能力。基本思想如下：如果一个词在一个文档中频繁出现（高 Term Frequency, TF），但在整个语料库中不常见（低文档频率），那么这个词对文档的主题或类别具有高判别值。该设计使 TF-IDF 能够有效地识别文档中的关键特征词，从而在文本分类、关键词提取等任务中发挥重要作用。

TF-IDF 的算法步骤：

(1) 首先，计算词频 (TF，即 Term Frequency)，考虑到文章有长短之分，为了便于不同文章的比较，进行词频标准化：

$$TF_w = \frac{\text{在某一类中词条出现的次数}}{\text{该类中所有的词条数目}} \quad (1)$$

(2) 接着计算逆向文件频率 (IDF，即 Inverse Document Frequency)，某一特定词语的 IDF，可以由总文件数目除以包含该词语的文件数目，再将得到的商取对数得到。如果包含词条 t 的文档越少，IDF 越大，则说明词条具有很好的类别区分能力：

$$idf_i = \log \frac{|D|}{|\{j:t_i \in d_j\}|} \quad (2)$$

其中， $|D|$ 是语料库中的总文件数目。 $|\{j:t_i \in d_j\}|$ 表示包含词语 t_i 的文件数目（即 $n_{i,j} \neq 0$ 的文件数目）。如果该词语不在语料库中，就会导致分母为零，因此一般情况下使用 $1 + |\{j:t_i \in d_j\}|$

即：

$$IDF = \log \left(\frac{\text{语料库的文档总数目}}{\text{包含词条 } w \text{ 的文档数目} + 1} \right) \quad (3)$$

(3) 最后计算 TF-IDF，在特定文件中出现频率高的单词，以及在整个文件集中出现频率低的单词，可以产生高权重的 TF-IDF。因此，TF-IDF 倾向于过滤掉常见的词汇，保留重要的词汇：

$$TF - IDF = TF * IDF \quad (4)$$

TF-IDF 算法能够很好地通过词频衡量某个关键词的重要性，可解释性相对之前的方法来说要强得多，同时在没有训练数据的时候仍然适用，所以本章使用该算法来进行关键词的提取，并结合大语言模型更好地生成 EPU 指标，但是仍然存在一定的缺陷。对于一些语义相近但是词语不同的情况缺乏较强的感知能力，并且无法捕捉上下文语义，本章后续实验中会使用 transformer 模型捕捉复杂语义关系，更进一步地构建 EPU 指数。

6.2.3.1 TF-IDF 算法的演化与语义增强路径

尽管 TF-IDF 算法在信息检索与关键词提取中具备显著优势，但面对多义词、同义词、上下文依赖等语义挑战，其浅层统计特性逐渐暴露出局限。为此，近年来研究围绕“如何增强 TF-IDF 对上下文与语义的感知能力”展开了深入探索，主要沿两个方向推进：语义扩展机制与深度模型协同建模。

一方面，Ye 等人提出的 DocClass-T5 模型 (Ye et al., 2024) 通过融合 TF-IDF 排序与预训练语言模型（如 T5）的微调策略，实现了对原始词频特征的语义补全。在文本分类与关键词抽取中，模型利用 TF-IDF 捕捉初始显著词，再借助语言模型编码深层上下文结构，显著提升了模型在低资源情境下的表现，特别是在词义歧义或跨语境词语中尤为有效。

另一方面，Wang 等人于《Grokking in the Wild》中探索了“多跳语义增强”方法，用于构建多轮推理中词汇重要性的演化轨迹。通过引入语义连接图与中间问题生成机制，该方法弥补了 TF-IDF 在处理逻辑长链关系时的局限，使关键词提取不仅停留在句内显著性层面，还能感知上下文中的逻辑依赖与

语用意图 (Wang et al., 2024) 。

这类研究共同推动了 TF-IDF 算法从静态、单步式加权向动态、语境感知型机制转型，并揭示了“浅层统计 + 深层语义”的协同建模框架，是今后构建高鲁棒性关键词提取系统的重要方向。

6.2.4 Transformer

Transformer 结构概述

Transformer 由 Vaswani 等人 (2017 年) 提出, 是一种基于自注意力机制 (Self-Attention Mechanism) 的深度学习模型。与传统循环神经网络 (RNN) 不同, Transformer 采用完全并行化的方式进行序列数据处理, 使得模型在长距离依赖建模上具有显著优势。

Transformer 主要由编码器 (Encoder) 和解码器 (Decoder) 两部分构成, 每个部分包含多个堆叠的层 (Layer)。编码器的作用是将输入数据转换为上下文相关的表示, 而解码器则基于这些表示生成目标输出。如图 2 所示, 其核心组件包括:

- (1) 多头自注意力机制 (Multi-Head Self-Attention, MHSA)
- (2) 前馈神经网络 (Feedforward Neural Network, FFN)
- (3) 残差连接 (Residual Connection) 和层归一化 (Layer Normalization)
- (4) 位置编码 (Positional Encoding)

自注意力机制 (Self-Attention Mechanism)

自注意力机制是 Transformer 的核心计算单元, 用于计算序列中各个元素之间的相关性权重。给定输入序列表示 $X \in R^{n \times d}$, 其中 n 为序列长度, d 为特征维度, 自注意力的计算过程如下:

- (1) 计算查询 (Query)、键 (Key) 和值 (Value) 矩阵:

$$Q = XW_Q, K = XW_K, V = XW_V \quad (5)$$

其中, $W_Q, W_K, W_V \in R^{d \times d_k}$ 为可训练参数。

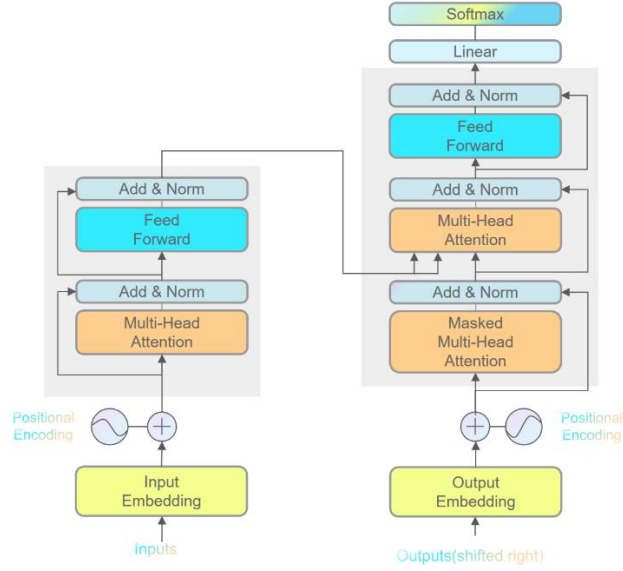


图 2 Transformer 架构

(2) 计算注意力权重 (Scaled Dot-Product Attention) :

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

其中, d_k 为键向量的维度, $\frac{1}{\sqrt{d_k}}$ 为缩放因子以防止梯度消失或爆炸。

(3) 多头注意力机制 (Multi-Head Attention, MHA) : 为了提升模型的学习能力, 引入多个独立的注意力头:

$$MultiHead(Q, K, V) = Concat(..., head_h) \quad (7)$$

其中, 每个注意力头的计算过程与单头注意力相同, 最终的结果通过线性变换 W_o 进行映射。

位置编码 (Positional Encoding)

由于 Transformer 结构中不包含循环或卷积操作, 因此需要额外的位置信息来指示序列顺序。为此, Transformer 采用正弦和余弦函数生成位置编码:

$$PE_{(pos, 2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right), PE_{(pos, 2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \quad (8)$$

其中, pos 表示词在序列中的位置, i 为维度索引。

前馈神经网络 (Feedforward Network, FFN)

每个 Transformer 层包含一个两层的前馈神经网络，其结构如下：

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + \quad (9)$$

其中， W_1, W_2, b_1, b_2 为可训练参数。FFN 作用于序列中的每个位置，与其他位置的计算独立，因此具有较高的计算并行性。

层归一化与残差连接 (Layer Normalization & Residual Connection)

残差连接：缓解梯度消失，提高深层网络的训练能力：

$$x_{out} = x + SubLayer(x) \quad (10)$$

层归一化：加速收敛，提高稳定性：

$$\hat{x} = \frac{x - \mu}{\sigma + \epsilon} * \gamma + \beta \quad (11)$$

训练优化策略

Transformer 训练过程中，为了提高泛化能力并加快收敛，通常采用以下优化策略：

(1) 学习率调度 (Learning Rate Scheduling) 采用 Warmup+下降策略，以控制学习率的变化：

$$lr = d^{-0.5} * \min(step^{-0.5}, step * warmup^{-1.5}) \quad (12)$$

其中， d 为模型隐藏层维度， $step$ 表示当前的训练步数。

(2) 正则化方法

Dropout: 在训练过程中随机丢弃一部分神经元，以减少过拟合。

Label Smoothing: 对目标标签进行平滑化处理，提高模型的泛化能力。

(3) 梯度剪裁 (Gradient Clipping) 为了防止梯度爆炸，采用梯度剪裁 (Gradient Clipping)：

$$g = \frac{g}{\max(1, \frac{\|g\|}{\tau})} \quad (13)$$

其中， g 是梯度， τ 是设定的阈值。

本章利用 transformer 模型提取新闻中的语义和深层次的特征，利用人工标注的新闻进行训练，从而对新闻中文本的 EPU 分数进行预测，transformer 可以更好地处理长文本新闻，并且能够捕捉一些复

杂的语义，但是这样构造出来的 EPU 指数与波动率的相关性与前文所讲到的方法差距并不大，所以接下来的实验尝试采用 Lora 微调 deepseek 进行构建 EPU 指标。

6.2.4.1 Transformer 结构的演进与语义建模能力拓展

自 Transformer 架构提出以来，其在语言建模、语义理解与跨模态建模中的核心地位不断巩固。最新研究不仅在结构层面优化计算效率，也在语义控制、知识注入与多阶段推理能力方面显著扩展 Transformer 的表达潜力。

Zhuang 等人提出的 Generation Trajectories for Efficient LLM Reasoning (Zhuang et al., 2024) 构建了基于 Transformer 结构的生成路径追踪机制，以分析并优化长文本生成中的语义连贯性。该方法通过多阶段注意力重构，显著提升了模型在复杂长文本结构中的推理稳定性与语义一致性，对基于新闻文本的 EPU 指数建模具有直接启发意义。

Robeyns 等人在《On the Role of Layered Abstractions in Transformer Reasoning》中引入了层次语义重建机制 (Robeyns et al., 2024)。该研究显示，Transformer 中不同层级的注意力与前馈网络可以学习到抽象程度递增的概念分布，而非简单的逐层特征堆叠。这一机制使模型更好地支持高阶推理与多粒度语义感知，尤其适用于文本—知识图谱映射或政策分析任务。

Lu 等人提出的 Uni-Perceiver v2 框架则从结构角度对 Transformer 进行扩展，引入统一跨模态输入接口与任务嵌入机制 (Lu et al., 2024)。该方法通过共享 Transformer 骨架实现图文任务的并行表示建模，有效支持多模态 EPU 评估指标构建。

这些工作表明，Transformer 已从通用编码器演化为可调控的语义推理核心单元，结合任务嵌入、跨层抽象与生成轨迹控制等策略，能够大幅拓展其在政策预测与经济类文本建模中的表达能力。

6.2.5 Lora 微调

低秩适配 (LoRA) 概述

近年来，大规模预训练语言模型 (Pre-trained Language Models, PLMs) 在自然语言处理 (Natural Language Processing, NLP) 任务中取得了显著进展。然而，传统的全参数微调 (Full Fine-Tuning) 方法需要对整个模型的参数进行更新，这不仅增加了存储和计算成本，还限制了在资源受限环境下的应用。

低秩适配 (Low-Rank Adaptation, LoRA) 是一种参数高效微调 (Parameter-Efficient Fine-Tuning, PEFT)

方法，其核心思想是在不修改原始预训练模型权重的情况下，仅通过添加低秩矩阵来调整模型的行为。LoRA 极大地减少了微调所需的参数数量，从而降低了存储和计算成本，并且在推理阶段不会增加额外的计算负担。

工作原理

在标准的 Transformer 架构中，关键权重矩阵（如自注意力机制中的查询 Query、键 Key、值 Value 以及前馈网络的权重）在微调时通常需要更新全部参数。例如，对于一个隐藏维度为 d 的 Transformer 层，其权重矩阵的规模为 $d \times d$ 。在全参数微调过程中，每个微调任务都需要存储一份完整的参数副本，导致计算和存储开销大幅增加。

LoRA 的核心思想是利用低秩矩阵分解来近似参数更新。假设原始预训练模型的某个权重矩阵为 $W_o \in R^{d \times d}$ ，其标准微调方式可以表示为：

$$W = W_o + \Delta W \quad (14)$$

其中 ΔW 为需要学习的参数更新矩阵，具有相同的维度 $d \times d$ 。由于实验表明 ΔW 具有较低的内在秩（intrinsic rank），因此 LoRA 采用低秩矩阵分解的方式来近似 ΔW ：

$$\Delta W = AB \quad (15)$$

其中：

$A \in R^{d \times r}, B \in R^{r \times d}$ 是可训练的低秩矩阵；

$r \ll d$ ，即 r 远小于原始参数维度，从而减少计算复杂度。

这样，在前向传播过程中，计算公式变为：

$$Y = (W_o + AB)X \quad (16)$$

如图 3 所示，其中 X 是输入数据，由于 A 和 B 的值较低，LoRA 仅引入了极少量的额外计算开销。

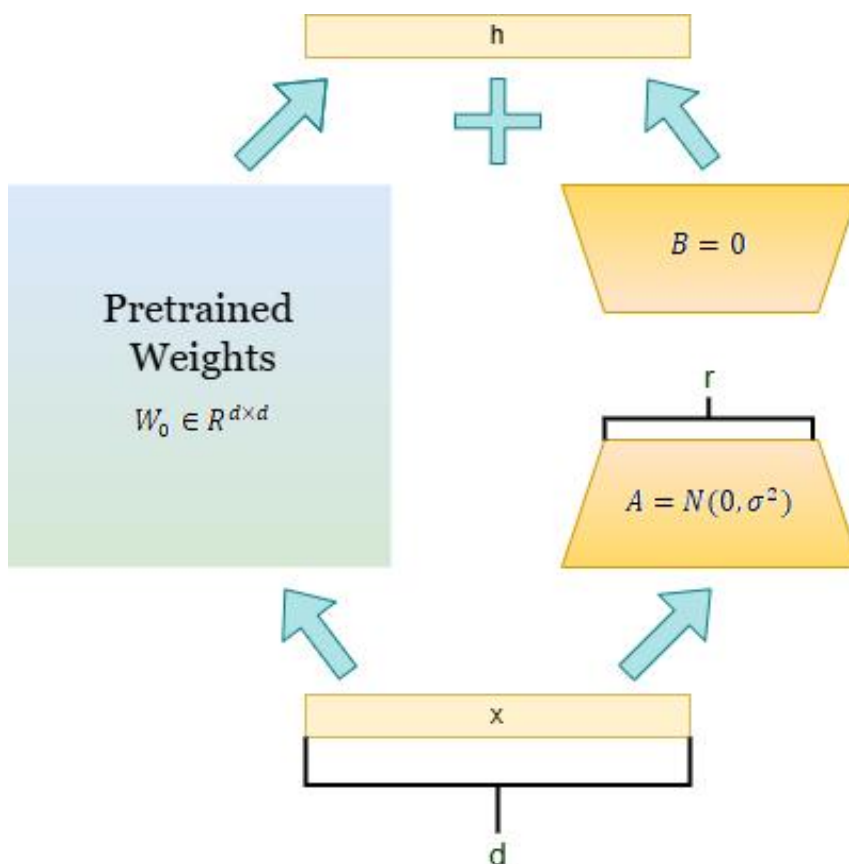


图 3 Lora 微调架构

训练步骤

(1) 选择目标层

LoRA 主要应用于 Transformer 结构中的自注意力机制，对查询 (Q) 和键 (K) 矩阵进行适配。并且选择适当的秩 r 以控制可训练参数数量。

(2) 初始化低秩矩阵:

随机初始化矩阵 A 和 B ，并将其添加到原始模型权重矩阵中。

(3) 冻结原始权重:

预训练模型的参数保持不变，仅训练 LoRA 插入的低秩矩阵。

(4) 优化器设置:

采用 AdamW 进行优化，以提高梯度稳定性；同时采用学习率调度策略，以平滑优化过程。

(5) 合并权重:

训练完成后，可将学习获得的低秩矩阵（即秩较低的矩阵，它可以近似原始高秩矩阵，通常通过分解原始矩阵为两个更小的矩阵乘积来实现）直接整合至原始权重中，从而有效降低推理阶段的计算复杂度。

(6) 超参数选择及作用，如表 1 所示：

表1 超参数选择及作用

超参数	作用
秩 r	较小的 r 能够减少计算开销，但可能降低微调的表达能力。
学习率 η	使用比全参数微调更高的学习率，以弥补参数更新空间的减少。
优化器	选择 AdamW 作为优化器，能够稳定优化过程。
权重初始化	LoRA 采用均匀或高斯分布初始化，以确保训练的稳定性。

Lora 微调的技术能够节省计算资源，不需要改变原本模型的权重就可以将效果调整接近至全量微调的效果，本章采用了这种方式进行 EPU 指数的构建，提升了 EPU 指数与波动率的相关性，有助于提升模型在特定语境下的泛化能力，为后续的波动率建模提供坚实基础。

6.2.5.1 LoRA 微调的泛化能力与结构优化方向

低秩适配（LoRA）作为参数高效微调技术的代表，已广泛应用于大语言模型的个性化调优与下游任务迁移中。近年来，多项研究进一步推动 LoRA 方法向更强的泛化性能控制、任务适配灵活性以及跨结构可移植性方向拓展。

Lee 等人提出的研究 (Lee et al., 2024) 评估了 LoRA 与多种微调策略在法律、摘要、情绪分类等多任务中的适应表现。通过系统比较，他们发现 LoRA 在保持模型原始性能的同时，实现了高度任务特化的微调，尤其在低数据量条件下，泛化误差显著小于全参数微调方法，验证了其在小样本下的稳定性与可转移性。

Liu 等人于《Practical Efficiency of Muon》一文中引入了一种名为 MuLoRA (Multi-scale LoRA) 的方法，在传统 LoRA 基础上引入了多尺度张量重构机制 (Liu et al., 2024)。该方法允许微调参数在跨层结构中传播，从而在保持低秩表示的前提下增强模型在多语境间的对齐能力，为政策文本建模等需复杂语义泛化的任务提供了结构支持。

Zhou 与 Chen 提出的 OpenRAG 框架则展示了如何在检索增强生成 (RAG) 流程中集成 LoRA 调优模块 (Zhou & Chen, 2025)。他们通过在 Transformer 的查询层插入低秩适配器，有效地提升了文档检索与生成间的一致性。这说明 LoRA 不仅适用于生成模型本体，还可在多模块系统中协同优化表现。

这些研究成果共同说明，LoRA 作为 PEFT 范式的核心技术，正不断从“参数压缩工具”发展为“结构增强引擎”，其在经济类文本理解与语义相关性建模中的前景值得持续探索。

6.2.4 实证分析模型

为了探究基于大模型构建的 EPU 与波动率的非线性关系，本章主要运用了三种非线性模型，包含了随机森林模型，Lasso 模型以及 LSTM 模型，对比并探究不同非线性模型是否会得出相似的结论。

(1) 随机森林

随机森林 (Random Forest, RF) 是一种集成学习方法，通过构建多个决策树并结合其预测结果来提高模型的准确性和泛化能力。随机森林通常用于分类 (classification) 和回归 (regression) 任务，其核心思想是基于装袋法 (Bagging) 和随机特征选择来构造一组独立的决策树，并通过投票 (分类任务) 或平均 (回归任务) 得到最终的预测结果。

装袋法 (Bagging)

Bagging 算法是一种集成机器学习算法，核心思想是通过 Bootstrap 重采样生成多个不同的训练数据集，并在这些子集上分别训练多个弱学习器 (如决策树)。最终，随机森林通过均值 (回归) 的方式融合多个决策树的预测结果。

给定训练数据集：

$$D = \{(x_1, y_1), (x_2, y_2), (x_N, y_N)\} \quad (17)$$

首先从 D 中进行 Bootstrap 采样，形成 B 个训练子集 D_b ($b=1,2,B$)；然后在每个 D_b 上训练一个决策树 T_b ；最终的预测结果通过均值 (回归) 计算：

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (18)$$

随机特征选择

在分类任务中，构造每棵决策树时，随机森林不会直接使用全部特征，而是在每次分裂时随机选择 m 个特征 (其中 $(m < M, M$ 为总特征数)，然后在这 m 个特征中选择最佳划分点。该策略减少了树之间的相关性，从而提高了泛化能力。

最优特征的选择基于基尼不纯度 (Gini Impurity) 或信息增益 (Information Gain)，具体计算如下：

基尼不纯度（用于分类问题）：

$$Gini(D) = 1 - \sum_{i=1}^C p_i^2 \quad (19)$$

其中， C 为类别数， p_i^2 为类别 i 的样本比例。

信息增益（用于分类问题）：

$$IG(D, f) = H(D) - \sum_{v \in V_f} \frac{|D_v|}{|D|} H(D_v) \quad (20)$$

其中， $H(D)$ 是数据集 D 的熵， V_f 是特征 f 可能的取值集合。

构建随机森林

具体构建流程如图 4 所示 (1) 首先，循环遍历 B 棵树（其中 B 是随机森林中树的数量）。(2) 然后，我们使用 Bootstrap 方法提取每棵树的样本，通过对原始数据随机抽样，返回生成一个 Bootstrap 样本（记为 Bootstrap(b))，重复 n 次，形成样本量为 n 的样本集。(3) 在每个节点的分裂决策中，我们不再遍历所有特征寻找最优划分，而是从所有特征中随机选取一个子集（例如 \sqrt{P} 或预设固定数量）作为候选特征池。分类依据仅从该子集内选取，通过评估候选特征的分裂增益（如信息增益或基尼不纯度）确定最佳划分属（这里里的 \sqrt{P} 指的是特征总数的平方根）。

(4) 每棵树的结构由 CART（分类与回归树）算法生成，其参数包括树的分裂规则、叶节点的分布等。对于每个数据点 $Z_{i,t}$ ，每棵决策树会根据其特征值沿树的路径向下分裂，最终落入某个叶节点 L 。该叶节点对应一个预测值，用于表示该数据点在当前树中的分类或回归结果。具体公式如下：

$$\hat{g}_b = \sum_{k=1}^{2^L} \theta_b^{(k)} I\{Z_{i,t} \in C_k(L)\} \quad (21)$$

具体而言，假设随机森林包含 B 棵树（参数 b 可指代树的索引 $b = 1, 2, \dots, B$ ），则每棵树 b 的叶节点 L 会为数据点 $Z_{i,t}$ 提供一个预测值。若 $Z_{i,t}$ 被分配到树 b 的叶节点 L ，则该叶节点的预测值会被赋予权重 $I_b^{(k)}$ ，其中 $I_b^{(k)}$ 是一个指示函数：当 $Z_{i,t}$ 确实落入叶节点 L 时，其值为 1，否则为 0。最终，所有树的预测值通过加权求和得到集成模型的输出，其中内层求和遍历树 b 的所有叶节点，外层求和整合所有树的贡献。(5) 每个叶子节点最终的输出结果是所有 B 树预测结果的平均值：

$$\hat{g}_b = \frac{1}{B} \sum_{b=1}^B \hat{g}_b \quad (22)$$

随机森林算法在机器学习中展现出独特的综合优势，它通过集成学习机制显著降低过拟合风险，同时还能通过特征重要性评估量化各变量对目标的贡献，并捕捉特征间的非线性交互作用，这使其在处理现实世界中常见的非线性问题（如金融风险预测等场景）时表现优异。

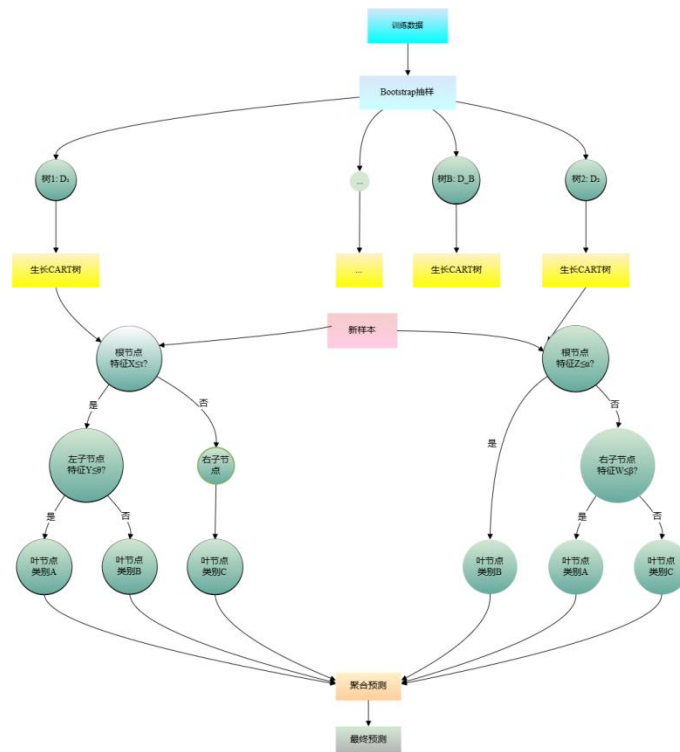


图 4 随机森林训练架构

(2) LASSO 回归

线性回归与正则化

回归分析是机器学习中最基础的预测建模方法之一，其主要目标是通过一个或多个输入变量来预测连续的目标变量。最小二乘法（Ordinary Least Squares, OLS）是最常见的线性回归方法之一。然而，在高维数据集中，OLS 可能导致模型过拟合，并且在特征高度相关时，其回归系数的不稳定性较高。因此，正则化技术被引入，以改善模型的稳定性和泛化能力。

LASSO 回归原理

LASSO（Least Absolute Shrinkage and Selection Operator）是一种基于 L1 正则化的线性回归方法，通过对回归系数施加 L1 范数约束，达到参数收缩和特征选择的目的。

LASSO 回归的目标函数可表示为:

$$J(\beta) = \frac{1}{2n} \sum_{i=1}^n (y_i - x_i^\top \beta)^2 + \lambda \sum_{j=1}^p \|\beta_j\| \quad (23)$$

其中, n 为样本数量, p 为特征数量, y_i 为第 i 个样本的目标变量, x_i 是对应的特征向量, β 是回归系数向量, λ 是正则化超参数, 控制回归系数的惩罚力度。

由于 L1 正则化项的非光滑性, LASSO 回归的求解比岭回归更具挑战性, 故本章采用坐标下降法去进行优化, 每次仅更新一个回归系数, 降低计算复杂度。

(3)LSTM

LSTM 概述

序列数据建模是自然语言处理 (NLP)、时间序列预测和语音识别等任务中的核心问题。传统的循环神经网络 (Recurrent Neural Network, RNN) 能够捕捉序列信息, 但其在长序列任务中存在梯度消失和梯度爆炸的问题, 导致模型难以捕获长期依赖关系。

为了解决 RNN 的局限性, 长短时记忆网络 (Long Short-Term Memory, LSTM) 被提出, 其核心思想是在网络中引入门控机制 (gating mechanism), 使得模型能够有效选择性地记忆和遗忘信息, 从而提高长期依赖建模能力。

LSTM 结构

LSTM 通过输入门 (input gate)、遗忘门 (forget gate) 和输出门 (output gate) 来控制信息流动, 以便在长序列数据中有效保留有用信息, 同时丢弃无关信息。如图 5 所示。

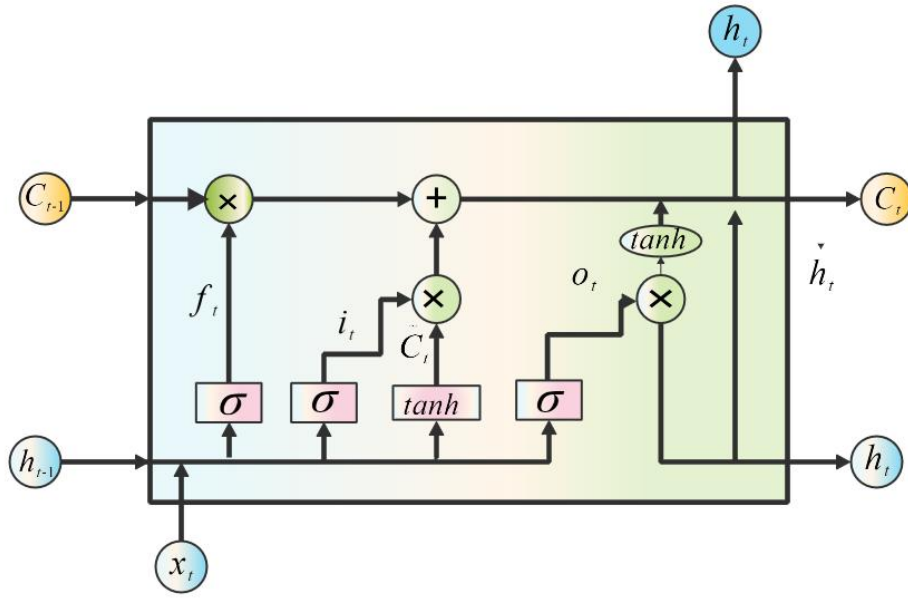


图 5 LSTM 模型架构

LSTM 结构的核心由以下组件构成:

- (1) 遗忘门 (Forget Gate) : 决定细胞状态 C_t 中哪些信息需要遗忘。
- (2) 输入门 (Input Gate) : 决定新输入信息对细胞状态的影响。
- (3) 细胞状态更新 (Cell State Update) : 结合过去状态和新信息来更新记忆单元。
- (4) 输出门 (Output Gate) : 决定当前时间步的隐藏状态输出。

遗忘门

LSTM 的第一步就是决定要从 cell 状态中丢弃哪些信息, 这个决定是由遗忘门中的 sigmoid 层做出的。LSTM 会先根据新的输入和上一时刻的输出决定遗忘掉之前的哪些记忆, 输入和上一时刻的输出整合成一个新的单独的向量, 然后通过 sigmoid 函数, 这个函数会将输出压缩到 $(0, 1)$ 的区间上, 如果新的向量在通过该函数后变成 0, 那么相乘后对应向量也为 0, 即做到了“遗忘”, 如果是 1, 那么相乘后对应向量也为 1, 即“保留记忆”, 下面公式可以描述这个过程

$$f_t = \sigma(W_f * [h_{t-1} x_t] + b_f) \quad (24)$$

其中, f_t 表示当前时间步 t 的遗忘门输出; W_f 是遗忘门的权重矩阵, b_f 是偏置项; h_{t-1} 为前一时间步的隐藏状态, x_t 是当前输入; σ 为 sigmoid 激活函数, 其输出范围为 $(0,1)$, 用于控制遗忘程度。

输入门

而输入门，或者叫记忆门，就是用来控制是否将在 t 时刻的数据并入单元中的控制单位，首先，输入层的 sigmoid 函数决定了我们更新哪些值，然后 tanh 层创建一个新候选值的向量 \tilde{C}_t ，最后将两者结合起来，并更新状态：

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad (25)$$

$$\tilde{C}_t = \tanh(W_C * [h_{t-1}, x_t] + b_C) \quad (26)$$

其中， i_t 表示输入门，决定新信息的影响程度； \tilde{C}_t 候选细胞状态，采用 \tanh 激活函数。

而接下来就是更新旧单位状态的过程，将旧状态乘以 f_t ，忘记之前决定忘记的事情，然后添加新的候选值 \tilde{C}_t ：

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (27)$$

输出门

最后输出门决定要输出什么，先运行一个 sigmoid 函数，决定要输出哪些部分，然后将单位状态通过 tanh 函数压缩至 $(-1, 1)$ 中，最后乘以 sigmoid 门的输出，这样就得到我们需要输出的部分：

$$O_t = \sigma(W_O * [h_{t-1}, x_t] + b_O) \quad (28)$$

$$h_t = O_t * \tanh(C_t) \quad (29)$$

6.3 实验设计

6.3.1 传统 EPU) 测度以及局限

在经济政策不确定性 (EPU) 指数的构建初期，Baker, Bloom, and Davis (2016) 所提出的经典方法 (以下简称 BBD 方法) 长期处于主导地位。该方法通过统计新闻文本中与经济、政策和不确定性相关的三类词汇的共同出现频率，进而生成用于衡量宏观经济政策不确定性的量化指标。该方法因其操作简便和计算效率高而迅速成为经济金融研究领域的主流工具，并在后续众多研究中被广泛用作基准模型。然而，随着研究需求的深化和数据分析技术的演进，该传统方法的固有缺陷逐渐暴露。其在理论假设和技术方法上的局限性显著限制了 EPU 指数在实际应用中的解释能力。

在早期构建 EPU 指数的过程中，基于关键词频率的量化方法对媒体数据源的选择有高度的依赖性。

由于样本代表性、信息覆盖维度、内容取向甚至媒体传播政治立场的差异，这些都可能对指数的计算结果产生实质性的干扰。例如，对于同一宏观经济政策事件，不同媒体机构的报道视角可能存在显著偏差，这种差异将直接导致最终形成的指标值出现偏差。此外，在实际应用场景中，跨区域媒体在报道风格和内容框架上的显著差异，使得单纯依靠关键词统计的分析模式无法有效地将真实的政策不确定性与民意和情绪夸大因素区分开来。这对该指标的客观性和区域间可比性提出了严峻的挑战。

并且，早期方法虽然能够有效捕捉经济政策不确定性事件的频次，但无法深入理解文本背后的语义、情绪与潜在的经济内涵。例如，同一组关键词在不同政策文本中可能有着截然不同的语境含义，但传统的 EPU 构建方法并未能有效辨识此类差异。此外，基于词频统计的共现模式分析忽视了政策表述中的逻辑关系与文本结构的连贯性，例如“减税”与“监管强化”在不同政策场景中的不同影响。因此，传统方法所构建的 EPU 指标在解释复杂政策文本的多维含义时，既无法精准定位关键政策转向节点，也难以反映市场参与者对不同政策场景的差异化解读，最终导致指标的解释力与动态敏感性均存在缺陷。

而在数据处理与标准化过程中，传统的 EPU 指数构建方法也面临明显的问题和局限性。这种方法通常采用历史区间内关键词频率的标准差进行归一化处理，但这种归一化方式隐含的前提是历史区间内经济政策环境是稳定或类似的。但是，经济政策环境往往会随着重大经济事件，国内外经济形势变化或结构性转型而出现明显的阶段性变化，使得过去的数据未必具备代表性和适用性。因此，传统的方法可能无法有效适应经济政策环境的变化，降低了构建的 EPU 指数的有效性。

因此，传统的 EPU 构建方法在实际研究中存在着明显的不足，本章将采用更加精细准确的基于大语言模型的 EPU 构建方法，确保指标更加精确，全面地描述经济政策不确定性的特征和变化。

6.3.2 基于大语言模型 DeepSeek 的新方法

近年来，基于 Transformer 架构的大型语言模型（Large Language Model, LLM）在自然语言处理（Natural Language Processing, NLP）领域取得了重要突破，为经济学和金融学中的文本分析提供了创新性的方法和工具。LLM 的核心优势在于其强大的语义理解能力，能够深度挖掘文本中的逻辑关系、情感倾向及上下文关联。这使得其在文本分类、观点提取、语义匹配和自动问答等任务上展现出显著的准确性和适用性。

LLM 的广泛应用主要归因于预训练（Pre-training）技术的快速发展。预训练策略首先利用大规模的通用语料库构建基础语言模型，使其具备强大的语言理解和生成能力。随后，通过微调（Fine-tuning）适配特定下游任务（如经济政策不确定性测度），进一步提升模型的表现力。相较于传统基于词频统计的方法，LLM 能够更精准地识别政策文本的语义特征，并在捕捉政策不确定性的方向性（如积极或消极）及时间维度（短期或长期）方面具有更高的细粒度刻画能力，从而为政策环境与市场波动的关联分

析提供更加科学的测量工具。

在中文语境下，近年来涌现出一批专为中文优化的大型语言模型，其中以 DeepSeek 为代表的模型在中文文本理解和表征方面表现尤为出色。与传统的 BERT 和 GPT 等模型相比，DeepSeek 在长文本推理、情感分析和精细化文本分类等任务中展现了更优的性能。其预训练阶段采用了更丰富且多样化的中文语料，包括新闻报道、社交媒体内容、专业财经资讯等，使其在经济政策文本的语义解析和情绪识别方面更具优势。因此，DeepSeek 在测量经济政策不确定性时能够提供更加精准的文本语义分析能力。

基于此，本章将以 DeepSeek 大语言模型为基础，结合 Zero-shot、Few-shot 以及 Transformer, LoRA (Low-Rank Adaptation) 微调等方法，构建经济政策不确定性 (EPU) 数据集，从而更准确地解析新闻文本的语义信息，弥补传统关键词频率法在政策表述识别上的不足。这使得基于 DeepSeek 训练的 EPU 指数在解释经济政策影响和市场波动关系方面具备更高的解释力和预测能力。

6.3.3 实验设置对比

实验 1: Zero-Shot DeepSeek 评分

首先本章采用了 Zero-Shot 方式，利用 DeepSeek-v3 模型对经济政策不确定性指数 (EPU) 进行自动评估。在该实验中，本章并未设置任何的训练数据，而是全程依赖于 DeepSeek-v3 的内置知识对新闻文本进行评分。

本章以 2020 年—2024 年，《人民日报》每日发布的财经新闻作为分析对象，通过网络爬虫技术自动获取官网财经版块的全部新闻文本，并记录每条新闻的发布日期。接着对文本进行预处理，本章通过对原始数据进行清洗，去除广告、图片说明等无关内容，确保每条新闻作为独立分析单元。

而实验的核心环节是通过调用 DeepSeek-v3 的 API 接口进行 Zero-Shot 推理。本章将预处理后的新闻文本逐条输入模型，并通过精心设计的 prompt 引导其输出 0-10 分的 EPU 评分，其中 0 分表示政策环境完全确定，10 分表示高度不确定。需要指出的是，DeepSeek-v3 输出的是自然语言文本，评分结果并非结构化数据，缺乏严格的可解释性，且在不同调用中可能因上下文微小差异而产生评分波动。因此，本章通过固定评分模板的输出格式（例如“该新闻的 EPU 评分为 X 分”）来尽量规范模仿行为，提高评分的一致性。尽管 DeepSeek 已有更强大的 r1 版本，但考虑到 v3 模型在公开 API 可用性和资源消耗上的平衡，以及其已具备的中文财经文本理解能力，本章仍选择 DeepSeek-v3 进行实验，后续研究将进一步探索 r1 模型的表现差异。

与传统关键词统计方法相比，该方法的优势在于不依赖预设的关键词或者一些固定的情感词典，能捕捉更细微的语义信息。但 Zero-Shot 方式也存在评分不稳定的局限性，由于缺乏具体的示例或者标准示范，对相同的个别新闻文本的不同批次的评分会出现较大方差。本章认为是特定文本的措辞差异导致

模型每次都拥有不同的理解，从而导致评分差异严重，造成不稳定的评分。于是在后续实验中，本章将针对这一问题，探索 Few-Shot 学习等优化方案。

实验 2: Few-Shot DeepSeek 评分

与实验 1 的 Zero-Shot 方法相比，本实验在 DeepSeek-v3 模型评分过程中引入了 Few-Shot 学习方式。Few-Shot 方法的核心思想是向模型提供少量典型示例，以帮助模型更好地理解评分任务的具体标准，从而提高评分的稳定性和一致性。但同样的，本章也不会提供任何的训练数据，只是提供少量示例，剩下的部分全部依赖 DeepSeek-v3 的内置知识对新闻文本进行评分。

本章先人工挑选了一批具有代表性的新闻样本作为 Few-Shot 示例数据。这些示例是在实验 1 的基础上获取的，这些示例涵盖了经济政策不确定性的不同强度等级（低、中、高不确定性），例如“明确利好政策出台”（低不确定性）、“政策变动方向未明”（中等不确定性）、“重大政策调整传言但未确定”（高不确定性）等典型情境，每个示例都明确标注了相应的评分值及评分理由。接着根据这些示例数据整理成标准化的 prompt 模板，即在 prompt 中明确给出示例新闻文本、对应评分以及详细评分理由，使 DeepSeek-v3 能够快速、清晰地理解评分任务的标准与规则。

接着仍然与实验 1 类似，调用 DeepSeek-v3 的 API 接口进行 Few-Shot 推理。每次评分时，本章都将事先构建的 Few-Shot prompt 模板连同待测新闻文本一并输入到 DeepSeek-v3 模型中，模型根据 Few-Shot 示例进行类比推理，自主输出每条新闻的经济政策不确定性（EPU）评分，评分范围为 0 至 10。最后，通过计算当日所有新闻评分的算术平均值，构建日度 EPU 指数。

相较于实验 1 的 Zero-Shot 方法，Few-Shot 学习显著提升了 DeepSeek-v3 评分的准确性和稳定性。通过提供一组标准化的评分示例，模型在评分过程中能更稳定地对文本做出类比判断，减少了因新闻表述差异引发的评分波动。然而需要注意的是，当前大语言模型如 DeepSeek-v3 的预训练知识库覆盖时间通常截止至其最后训练时间点，具体版本的知识更新截止时间并不总是公开透明。由于模型训练时已接触大量历史数据，对于实验中涉及的部分历史政策事件，其相关内容可能已隐含在模型的语料中。因此，在对 2018 至 2024 年的新闻文本进行评分时，模型可能对某些政策事件存在“已知”倾向，存在训练数据与评分测试数据时间上的潜在重叠。虽然 Few-Shot 的引导提升了泛化能力，但这种知识泄漏风险也提醒我们在解释评分结果时应谨慎看待其对历史事件的预测能力。

此外，大模型的知识库更新通常并非实时，而是阶段性完成一次性预训练后冻结参数，因此对外部世界的“认知”停留在训练截止的那个时间点。例如，某些模型知识库更新到 2023 年或 2024 年年中，此后发生的事件、政策或新闻若未在训练数据中出现，则属于模型的“未知”领域。在处理 2024 年之后的新闻时，若模型能给出合理预测或判断，极可能是通过 Few-Shot 示例中隐含的模式泛化能力，而非真实理解该具体事件，因此应警惕由时间重叠带来的过拟合假象或结果偏倚。

实验 3: TF-IDF 关键词 + Few-Shot 评分

而实验 3 则是探索了结合 Few-Shot 学习与 TF-IDF 关键词提取的混合方法来提升 DeepSeek-v3 模型对经济政策不确定性 (EPU) 评分的性能。该方法针对单纯使用 Few-Shot 学习时模型可能难以准确识别新闻文本中关键术语的问题, 通过引入文本挖掘技术来增强模型的语义理解能力。

类似于实验 1, 本章先对 2020—2024 年《人民日报》财经新闻进行标准化预处理, 包括去除标点符号、停用词等无关字符, 以提高后续文本分析的质量。在此基础上, 采用 TF-IDF 算法从全年语料库中提取每篇新闻最具代表性的 20 个关键词, 这些关键词既考虑了词汇在特定文本中的重要性, 也兼顾了其在整体语料库中的区分度。

为充分发挥关键词信息的价值, 本章设计了结构化的 prompt 模板, 将原始新闻文本、提取的关键词以及 Few-Shot 示例有机整合。这种设计既保留了 Few-Shot 学习的类比推理优势, 又通过关键词引导模型聚焦于文本的核心政策含义。接着本章调用了 DeepSeek-v3 模型 API 进行评分, 模型在评分过程中重点关注由 TF-IDF 算法提取的核心关键词, 从而有效减少冗余或次要信息对评分造成的干扰。最后, 通过计算当日所有新闻评分的算术平均值, 构建日度 EPU 指数。

结果显示, 结合 TF-IDF 关键词和 Few-Shot 评分方法显著改善了模型的表现。通过聚焦于新闻文本中的核心关键词, DeepSeek-v3 模型更有效地捕捉了文本中的政策不确定性信号, 并减少了评分过程中非关键内容对评分结果的干扰。具体来看, 与单纯的 Few-Shot 学习相比, 本实验方法得到的 EPU 指数与市场波动率的相关性进一步提升, 特别是在滞后一期与滞后二期的相关系数均有所增加。这说明 TF-IDF 关键词提取在帮助模型精确识别文本关键信息方面起到了关键作用。

实验 4: Transformer 训练 + 机器学习预测

接下来, 本章提出了一种基于 Transformer 架构的系统性文本分析方法, 旨在提升经济政策不确定性 (EPU) 评分的预测性能。与前述直接评分方法不同, 本实验采用监督学习范式, 利用人工标注的 1425 条 EPU 新闻数据进行模型训练。这些数据包含由专家标注的 0-10 分 EPU 评分, 为模型学习提供了可靠的监督信号。

本章对原始新闻文本进行了系统的预处理, 包括无意义符号过滤、停用词去除以及中文分词等标准化操作。考虑到 Transformer 模型的特征提取需求, 所有文本均经过词嵌入处理转化为向量表示。本章将 EPU 评分预测任务转化为标准的回归任务, 即使用 Transformer (例如 BERT) 模型对新闻文本进行回归预测, 目标输出为每条新闻对应的 EPU 评分值 (0-10 之间的连续数值)。模型的输入为预处理和分析后的新闻文本特征, 输出为对应的 EPU 评分预测结果。

接着本章按照 8:1:1 的比例随机分配训练集、验证集和测试集。训练过程以最小化预测评分与人工

标注之间的均方误差 (MSE) 为目标, 同时在验证集上监控模型性能以防止过拟合。实验特别关注三个核心评估指标: MSE 反映整体预测偏差, MAE 衡量平均误差水平, R^2 评估模型解释力。测试结果表明, 基于 Transformer 的方法在各项指标上均显著优于 Zero-Shot 和 Few-Shot 方法。

训练完成的 Transformer 模型被应用于 2024 年《人民日报》未标注财经新闻的批量预测, 根据预测出的日度评分结果, 本章进一步计算每日的 EPU 指数平均值, 并构建日度 EPU 指数, 为后续市场波动与政策分析提供可靠的数据基础。

与之前的零样本 (Zero-Shot) 和少样本 (Few-Shot) 学习方法相比, 该方法能够更充分地利用历史标注数据, 训练模型更深入地理解和捕捉新闻文本中的政策不确定性特征。实验证明, 通过这种方式训练的 Transformer 模型在评分的一致性、稳定性及准确性方面均有显著提高, 为经济政策不确定性指数的精确度提供了坚实的技术支持。

实验 5: LoRA 微调 DeepSeek-v3

最后, 本章采用了 LoRA (Low-Rank Adaptation) 方法对 DeepSeek-v3 进行微调。LoRA 是一种参数高效的微调技术, 通过低秩矩阵的方式更新部分参数, 而非全量参数, 能够在不明显增加计算资源的情况下显著提升模型的性能。

本章选取前述实验所使用的 1425 条人工标注的《人民日报》财经新闻文本作为训练样本, 每条新闻已人工赋予了明确的经济政策不确定性评分 (0-10)。对这些文本的预处理步骤包括中文分词、去除停用词、文本清洗与规范化处理, 确保输入模型的数据格式统一且便于模型有效学习文本特征。微调数据集的划分遵循 8:1:1 的比例, 分别用于模型训练、验证和测试。

LoRA 微调的核心创新在于其参数更新策略。实验冻结了 DeepSeek-v3 的大部分预训练参数, 仅在注意力机制层引入可训练的低秩矩阵。这种设计使得模型仅需更新约 0.1% 的参数即可适应特定任务, 大幅降低了计算资源消耗。在微调过程中, 本章采用动态学习率调度策略, 并重点优化秩大小 (rank) 等关键超参数, 以最小化预测评分与人工标注之间的均方误差 (MSE)。

训练与验证完成后, 本章将经过 LoRA 微调后的 DeepSeek-v3 模型用于 2024 年《人民日报》的每日财经新闻文本, 以自动预测其经济政策不确定性评分。基于预测结果, 进一步计算每日新闻的评分均值, 进而构建日度 EPU 指数, 为后续市场分析提供高质量的数据支持。

实验结果显示, 相比于未经微调的 DeepSeek 模型, 经过 LoRA 微调后的 DeepSeek-v3 在评分精度、稳定性及对新闻文本的语义理解能力上均有显著提升。通过 LoRA 微调, DeepSeek 模型更加适应特定任务场景, 从而提高了构建经济政策不确定性指数的准确性和可靠性。

6.4 实证结果

6.4.1 数据与样本

本章选取了中国股票市场自 2018 年 1 月至 2025 年 1 月的日度数据，包括个股收益率、波动率、换手率、利率等关键财务指标。研究中所使用的经济政策不确定性（Economic Policy Uncertainty, EPU）数据，既包括传统的 EPU 数据（如 Baker 等人所提供的），也包括基于深度学习技术构建的 EPU 数据。传统 EPU 指标是通过统计新闻文本中与经济、政策和不确定性相关的词汇出现频率来构建的宏观经济政策不确定性量化指标。而基于深度学习技术构建的 EPU 数据，则是利用深度学习模型进行微调，以生成更为稳定和有效的 EPU 指标。本章主要采用日度数据进行分析和验证。

以下两个表格总结了本章分析中所使用的变量的统计信息。表二展示了每个关键变量的平均值和标准差。EPU(Baker) 代表由 Baker 等人提出的经济政策不确定性指数，该指数用于衡量政策不确定性对市场的影响力。EPU(Deepseek) 是基于 Deepseek 方法计算的经济政策不确定性指数，通过文本分析和深度学习模型微调构建。本章中的 Volatility 是基于上证主板所有上市公司股票的日收盘价计算得出的市场整体波动率指标。具体而言，通过对个股收益率序列进行计算，并以 25 日为滚动窗口，构建样本内平均收益波动率指标。该方法能有效平滑短期波动，并捕捉中期市场风险变化趋势，便于与政策不确定性指数进行相关性与动态性分析。TurnoverRate 代表股票的换手率，是衡量市场活跃度的指标，通过成交量除以流通股数计算得出，用以衡量市场流动性。Interest 代表利率水平，通常是央行公布的基准利率或市场利率，影响投资回报和金融市场的资金成本。表三展示了这些变量之间的相关性。本章构建的基于大语言模型的政策不确定性指数（EPU-Deepseek）与经典的 Baker 构造的政策不确定性指标（EPU-Baker）之间的相关系数为 0.20，表明两者在某些时段内存在一定的同步趋势。此外，EPU-Deepseek 与市场波动率（Volatility）之间的相关系数高达 0.70，显示出显著的正相关关系，说明基于大模型生成的政策不确定性指数在描述市场风险方面具有较强的解释力；而与换手率（TurnoverRate）之间的相关性较弱（0.12），可能表明政策不确定性对交易活跃度的影响相对有限。同时，EPU-Deepseek 与利率（Interest）呈负相关（-0.22），暗示政策不确定性上升可能伴随着利率下行预期。此外，EPU-deepseek 指标与波动率的相关性远大于传统的 EPU-Baker 指标，证明本章构建的基于大语言模型的政策不确定性指数（EPU-Deepseek）在捕捉市场情绪与风险动态方面具有更强的敏感性与前瞻性，从而更有效地揭示政策不确定性对市场波动的影响。

表2 平均值和标准差

	EPU(Baker)	EPU(Deepseek)	Volatility	TurnoverRate	Interest
Mean	117.283	5.595	0.024	0.593	0.004
Std	74.143	1.129	0.004	0.323	0.002

表3相关性

	EPU(Deepseek)	EPU(Baker)	Volatility	TurnoverRate	Interest
EPU(Deepseek)	1				
EPU(Baker)	0.20	1			
Volatility	0.70	0.14	1		
TurnoverRate	0.12	0.01	0.18	1	
Interest	-0.22	-0.02	-0.29	-0.01	1

6.4.2 回归结果

(1) 线性回归结果：大语言模型捕捉政策变动能力更强

表4 线性模型的表现

	EPU-Deepseek	EPU-Baker
OLS 的 R^2	0.4912	0.0210
OLS 的回归系数	0.6618	0.1380
LASSO 的 R^2	0.4689	-0.1045
LASSO 的回归系数	0.6774	0.1468

从回归模型结果看，基于大语言模型生成的政策不确定性指数（EPU-deepseek）在预测股市波动率方面展现出显著优于传统指标（EPU-Baker）的能力。表格结果显示，在 OLS 回归中，EPU-deepseek 的回归系数达到 0.6618，拟合度 (R^2) 为 0.4912；相比之下，EPU-Baker 的系数为 0.1380， R^2 仅为 0.021。尽管两者与波动率之间均为正向相关，但 EPU-deepseek 的解释力远强于传统指标。

在 Lasso 回归中，这一趋势依然明显。EPU-deepseek 的回归系数进一步提升至 0.6775，而 EPU-Baker 为 0.1468，说明在引入正则化约束以压缩不重要特征时，大模型生成的 EPU 指标依然具备稳定且更强的解释能力。

(2) 非线性模型分析：机器学习框架下的指数表现

表5 机器学习模型的表现

	EPU-Deepseek	EPU-Baker
随机森林的 R^2	0.5832	-0.7310
LSTM 模型的 R^2	0.7362	0.2566

为了进一步评估不同类型的 EPU 指数在更复杂、真实场景中的预测能力，本章引入两类典型的非线性机器学习模型——随机森林（Random Forest）与长短期记忆网络（LSTM）对股市波动率进行建模。结果显示，相比于传统线性模型（如 OLS、Lasso），非线性方法在处理经济政策不确定性与市场波动之间的关系上表现出显著优势，尤其是当输入特征为大模型生成的 EPU-deepseek 指数时更加有效显著。

在 LSTM 模型中，EPU-deepseek 构建的预测模型表现尤为突出，其拟合优度 (R^2) 高达 0.7362，

远超传统的 EPU-Baker 指标所构建模型的 $R^2(0.2566)$ 。而在随机森林模型中，EPU-deepseek 同样展现出良好的建模效果 ($R^2 = 0.5832$)，而 EPU-Baker 在同一模型中的拟合优度甚至为负 ($R^2 = -0.7310$)，显示出其难以适应非线性结构。

这一显著差异进一步说明，市场对政策不确定性的反应机制存在明显的非线性规律，简单的线性假设可能无法捕捉政策措辞、情绪演变与市场预期之间复杂的交互关系。机器学习模型，尤其是 LSTM，凭借其对时间序列特征的记忆能力与非线性建模能力，能够更好地挖掘这种深层动态关联。

此外，EPU-deepseek 所具备的上下文理解力、时序连贯性以及语义层级信息，为非线性模型提供了更丰富、结构更强的输入信号，使得机器学习框架能够充分发挥其优势。尤其是在 LSTM 架构下，其内部的门控机制可以有效地识别“政策语调变化”或“语义突变”所带来的非线性冲击，这一点是传统基于关键词统计的 EPU-Baker 所无法比拟的。

表 5 的结果不仅验证了 EPU-deepseek 在非线形模型中预测的强表现，也从建模角度强调了政策不确定性对市场影响的非线性本质，为后续的进一步研究提供了支撑。

(3) 特征重要性分析

表6 特征重要性

	EPU-Deepseek	TurnoverRate	Interest
重要性	0.6887	0.2082	0.1032

本章通过重要性分析评估了经济政策不确定性指数（EPU-Deepseek）、换手率（TurnoverRate）和利率（Interest）对预测目标变量的贡献。分析结果显示，EPU 在模型中扮演了最为关键的角色，其特征重要性值为 0.6887，远高于其他两个特征。这也反映了该指数不容易受到冗余变量的干扰，强调了经济政策不确定性在宏观经济分析中的重要性。

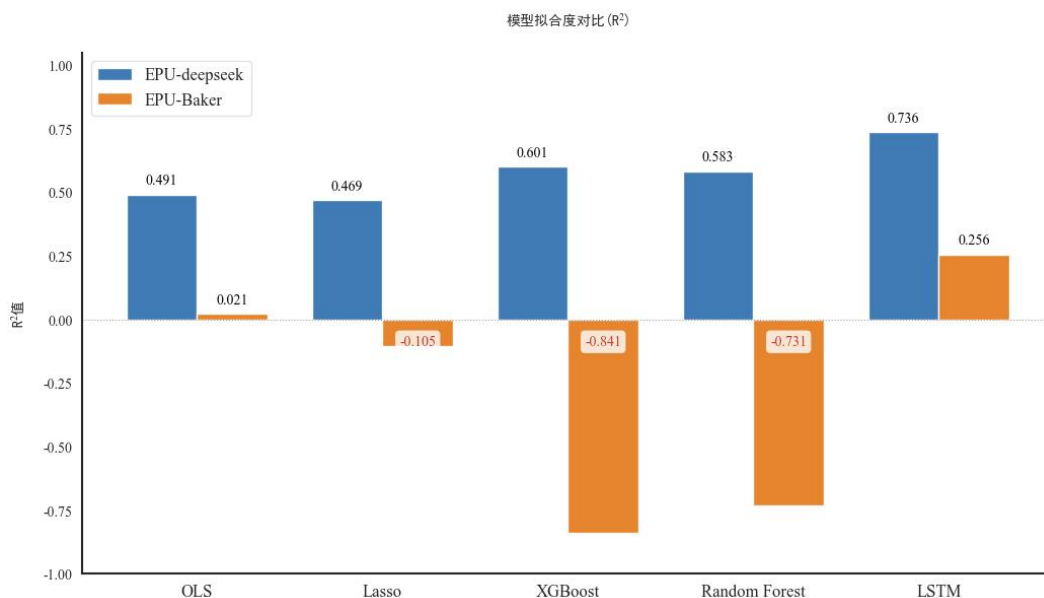


图 6 模型拟合度对比

根据模型分析结果，本章所构建的基于大型语言模型的经济政策不确定性（EPU）指数，在预测市场波动率及其他经济行为方面，相较于 Baker 等人提出的传统 EPU 指数，表现出更为优越的性能。传统 EPU-Baker 指数采用关键词计数法，存在若干局限性。首先，该传统指标仅能识别“不确定性”等表层词汇，而无法区分“可能调降利率”（扩张性政策信号）与“不会调升利率”（紧缩政策维持）等复杂语义。其次，传统指标往往忽略政策文本中的语气强度（如“必须”与“考虑”）、程度修饰（如“轻微”与“严重”）等关键信息，并可能忽视不同行业或政策主体间的联系。

本章所提出的基于大型模型的 EPU 指数，展现了更为显著的时间序列适应性。LSTM 的分析结果表明，EPU-deepseek 在时间序列适应性方面表现出色，例如，在 COVID-19 疫情期间，EPU-deepseek 能够自动识别出“非常规货币政策”“财政刺激”等非标准表述中的不确定性，而 EPU-Baker 由于未能及时更新关键词库，导致了测量上的偏差。EPU-DeepSeek 指数依托先进的深度学习语言模型（如 Transformer 架构），在大规模财经新闻语料中实现了更深层次的语义理解和不确定性识别。通过上下文建模能力，该指数有效捕捉到语境中隐含的政策立场、语气变化及风险倾向，使得不确定性度量更为精细且具有时效性。这种基于语义感知的构建方法显著提升了输入特征的表达能力，因此，在以时间序列建模为核心的 LSTM 中，能够更有效地引导模型捕捉经济动态与市场反应之间的非线性关系。

因此，本章构建的 EPU-Deepseek 指数，凭借其更为强大的文本理解能力和更高的政策敏感度，在宏观经济不确定性度量中展现了更为卓越的性能，增强了模型的解释力和预测效果。这为未来进一步的研究提供了重要的应用价值。

6.4.3 稳健性检验

(1) 模型架构稳健性检验

本章进一步在 XGBoost 模型中验证，仅有其他宏观指标时 R^2 值为 0.0973，加入 Baker-EPU 后提升至 0.1027，而 DeepSeek-EPU 的语义驱动特征使 R^2 进一步增至 0.6061，验证了语义分析在非线性建模中的优势：DeepSeek-EPU 通过词向量嵌入解析政策工具的隐含关联（如碳中和政策与能源行业监管的长期影响），这种细粒度的语义表征被 XGBoost 的树结构有效利用。

表7 XGboost在不同变量上的test集的 R^2

	仅有其他宏观指标	EPU-Deepseek+其他宏观变量	EPU-Baker+其他宏观变量
Xgboost 的 R^2	0.0973	0.6061	0.1027

(2) 加入宏观变量与指数稳健性分析

表8 加入宏观变量后机器学习模型的表现

	EPU-Deepseek	EPU-Deepseek+其他宏观变量	EPU-Baker	EPU-Baker+其他宏观变量
随机森林的 R^2	0.5832	0.6061	-0.7310	-0.9134
LSTM 模型的 R^2	0.7362	0.6864	0.2566	0.0230

在进一步的回归分析中，本章将两项典型的宏观金融变量——利率（Interest）与市场换手率（TurnoverRate）纳入模型，作为控制变量，以验证不同 EPU 指数在更真实市场环境中的稳定性与解释力。如图 6 所示，EPU-deepseek 指数在引入控制变量后依旧表现出极高的预测能力，在 LSTM 模型中，其拟合优度（ R^2 ）仍高达 0.6864。即使引入控制变量后出现 4.98% 的边际衰减，其解释力仍保持绝对优势。随机森林中 EPU-Deepseek 的 R^2 变化幅度不足 0.2%，进一步验证其抗干扰性。结果表明，宏观变量如利率和换手率本身对波动率虽存在一定的解释力，但在与 EPU-deepseek 指数联合建模后，其整体预测表现没有发生很大变化，表明 EPU-deepseek 能够提炼并补全宏观变量所未能捕捉的“语义层面政策信息”。从特征稳健性角度上来看，引入控制变量后，EPU-deepseek 的拟合度（ R^2 ）在各模型中并未出现大幅波动，其显著性水平、符号方向及边际效应均保持稳定。这说明该指数不容易受到冗余变量或多重共线性的干扰。

相较之下，EPU-Baker 在 LSTM 中的异常波动（单独使用时 $R^2=0.2566$ ，加入变量后骤降至 0.0230）揭示其模型依赖性缺陷：当外部宏观信号介入时，其机械构建方式导致信息捕获能力崩溃。所以这也在一定程度上也证明了 EPU-Baker 指数在捕捉政策变化方面的局限性，其构建方式比较机械，容易受到外部事件影响。

6.4.4 异质性分析

市场波动状态下 EPU-DeepSeek 的适用性检验

为进一步验证 EPU-DeepSeek 指标在不同市场环境下的稳定性与预测效能,本章从市场波动性的角度出发,探讨政策不确定性在高波动与低波动状态下对市场预测的异质性影响。考虑到在市场剧烈波动时期,政策预期及其相关信息对市场参与者行为的引导作用更为显著,因此有必要将样本按照波动水平进行分层分析,以评估 EPU-DeepSeek 在不同市场结构下的表现差异。

本章采用上证主板全部股票的收盘价数据,基于全市场价格的整体表现刻画市场波动水平。具体地,以股票价格的对数收益率为基础,采用 25 日滚动窗口计算其标准差,作为市场的短期波动率 (volatility) 指标。在此基础上,本章依据波动率序列的中位数作为划分标准,将样本数据分为两组:高波动期(波动率高于中位数)与低波动期(波动率低于中位数)。该分组方式在保持样本平衡的同时,有效反映了市场波动状态的结构性差异。

在高波动期与低波动期两个子样本中,本章分别采用 LSTM 与随机森林模型,利用 EPU-DeepSeek 构建预测模型,对市场变量进行拟合与预测,并比较测试集的 R^2 值以评估模型效果。

表9 不同波动率状态下模型的拟合度

	低波动率样本	高波动率样本
随机森林的 R^2	0.2566	0.4218
LSTM 模型的 R^2	0.3680	0.3934

如表 9 所示, LSTM 模型在低波动状态下的测试集 R^2 为 0.3680, 在高波动状态下提升至 0.3934, 整体预测性能维持在较高水平。相比之下, 随机森林模型在低波动期的 R^2 值为 0.2566, 而在高波动期显著提高至 0.4218, 表现出对波动环境更强的敏感性。

结果表明: EPU-DeepSeek 在高波动市场环境中具备更高的解释力和预测价值。高波动期通常伴随政策调整频繁、市场预期不稳定, 此时政策不确定性成为市场运行的重要驱动因素, 而 EPU-DeepSeek 所提供的高质量政策语义信息对模型预测具有关键贡献。

6.4.5 危机事件分析

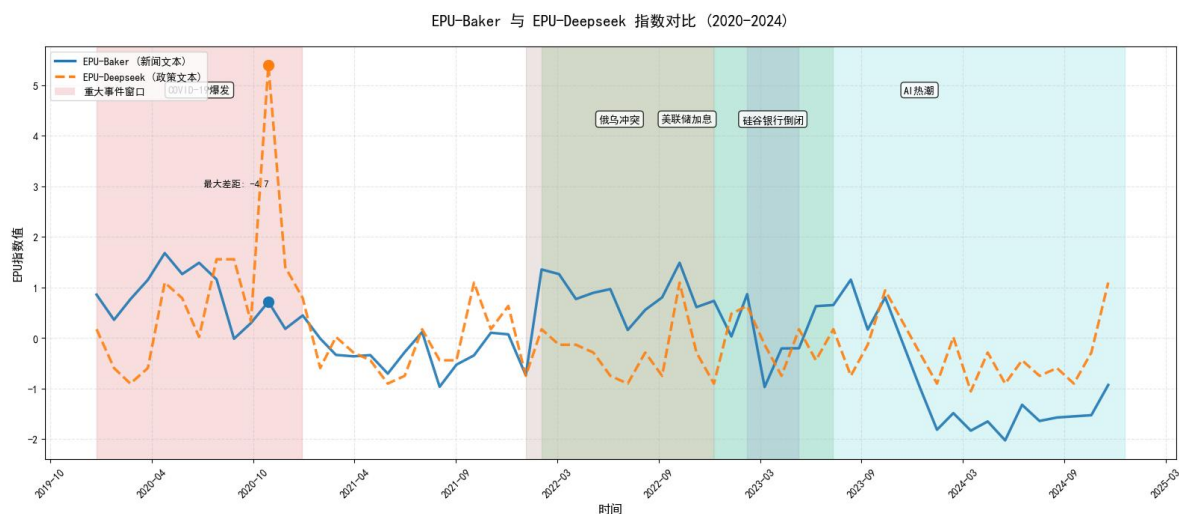


图 7 重大历史事件窗口期间的差异

图 7 展示了基于新闻文本构建的 EPU-Baker 指数与基于政策文本（结合大语言模型生成）的 EPU-Deepseek 指数在 2020 年至 2024 年间的动态演化情况。由于传统的 EPU-Baker 指数为月度数据，所以该图中采用的传统的 EPU-Baker 指数和 EPU-Deepseek 指数都是 2020 年 1 月到 2024 年 12 月的月度数据。为增强对比分析的时效性与解释力，图中特别标注了多个具有代表性的重大经济与政策事件窗口，包括 COVID-19 爆发、俄乌冲突、美联储加息周期、硅谷银行倒闭事件及 AI 热潮等。从整体趋势来看，EPU-Baker 与 EPU-Deepseek 指数在多数时间点呈现出较为一致的波动方向，说明二者在捕捉宏观政策不确定性时具有一定程度的协同性。

然而，在特定事件窗口中，二者表现出明显差异，以新冠疫情初期（2020 年初至 2020 年末）为例，EPU-Deepseek 指数在 2020 年 9 月达到峰值，明显高于 EPU-Baker，可能是 EPU-Deepseek 指数依赖于 LLM（DeepSeek）生成的政策类文本，这些文本更倾向于强调应对性政策、不确定性语言与宏观调控讨论，如“经济刺激”“风险评估”“财政安排”“复工复产”等关键词在疫情政策材料中密集出现。相比之下，EPU-Baker 仅基于新闻报道中“经济”“政策”“不确定性”三个词的词频共现的统计方法，而新闻文本在疫情初期更集中于公共卫生应急、病例统计、社会情绪等内容，导致其对宏观政策不确定性捕捉存在滞后或稀释。

而 2022 年的俄乌冲突及其带来的全球能源与商品价格震荡期，EPU-Baker 指数表现出更为剧烈的波动，明显高于 EPU-Deepseek 指数。可能是因为媒体在该期间大量报道国际地缘冲突、供应链风险、能源安全等内容，并将其与中国经济相关性进行广泛解读，导致新闻语料中“经济”“政策”“不确定性”出现的频率迅速增加。相反，政策类材料往往在国际问题上保持相对克制、强调稳定预期，大模型

在生成此类文本时倾向于学习此种表达风格，导致 EPU-Deepseek 指数上升幅度有限。

从 2023 年以来，“人工智能”成为政策与经济讨论的重要议题。EPU-Deepseek 指数整体高于 EPU-Baker，反映了大模型生成的政策文本中对技术监管、产业指导、安全审查等内容的强调，如“算法治理”“数据合规”“安全风险评估”“监管沙盒”等关键词较为密集，从而提升了政策层面对未来路径不确定性的描述强度。而新闻语料中“AI 热潮”多聚焦于技术发布、企业产品迭代与融资动向，并不总是直接关联宏观政策方向，从而导致传统 EPU-Baker 指数表现相对温和。

6.4.6 策略检验

该部分基于沪深 300 指数的 2018 年 1 月到 2025 年 1 月的历史收盘价日度数据以及通过 LSTM 模型预测的波动率数据，构建了一种基于波动率阈值的市场择时策略。该策略的核心逻辑是：当预测波动率低于某一设定阈值（threshold）时，认为市场风险较低，选择持有上证指数；当预测波动率高于该阈值时，清仓以规避潜在的市场风险。通过对比策略累计收益与上证指数本身的累计收益，评估策略是否能够产生超额收益。

首先根据上证指数的收盘价，计算每日的对数收益率：

$$LogReturn_t = \ln\left(\frac{Close_t}{Close_{t-1}}\right) \quad (30)$$

其中， $Close_t$ 表示第 t 天的收盘价。通过对数收益率的计算，本章能够更方便地进行累加操作，从而得出累计收益率。

基于预测波动率数据和设定的波动率阈值 threshold，确定每日的仓位状态：

$$Position_t = \begin{cases} 1, & \text{if Predicted Volatility}_t < \text{threshold} \\ 0, & \text{otherwise} \end{cases} \quad (31)$$

当预测波动率低于阈值时，仓位为 1（表示持有资产）；否则，仓位为 0（表示清仓）。

接下来计算策略每股收益，计算公式如下：

$$StrategyReturn_t = LogReturn_t \times Position_t \quad (32)$$

即策略的每日收益仅在持有资产时计入市场的对数收益率。

为了直观比较策略表现与基准指数的表现，本章将每日收益率转化为累计收益率。对于基准指数和

策略，累计收益率的计算公式分别为：

$$IndexCumReturn_t = e^{\sum_{i=1}^t LogReturn_i} - 1 \tag{33}$$

$$StrategyCumReturn_t = e^{\sum_{i=1}^t StrategyReturn_i} - 1 \tag{34}$$

其中， e^x 表示自然指数函数，用于将对数收益率转换为普通收益率形式。

为了验证策略的有效性,本章选取了多个不同的波动率阈值(如 2.0%、2.3%、2.5%、2.8% 和 3.0%),并分别计算对应的策略累计收益。图 8 展示了不同阈值下策略累计收益的变化曲线,以及基准指数(上证指数)的累计收益曲线。



图 8 不同阈值下策略累计收益与基准指数的累计收益的变化曲线

从图中可以看出，当波动率阈值较低（如 2.0%和 2.3%）时，策略曲线（蓝色和绿色实线）相对较为平稳，且在某些时间段表现优于基准指数。特别是 2.3%阈值的策略（绿色实线），在大部分时间表现较好，并在某些市场反弹期间超越了基准指数。随着波动率阈值的提高（如 2.8%和 3.0%），策略的累计收益逐渐下降，尤其是在市场反弹期间错失了上涨机会。中间阈值（如 2.5%）的策略（红色实线）表现较为稳健，但未展现出显著的超额收益。

表10 不同阈值下策略的超额收益与最大回撤

不同阈值	2.0%	2.3%	2.5%	2.8%	3.0%
最终超额收益	-6.78%	0.14%	-6.53%	-10.41%	-8.96%

最大回撤	-8.96%	-15.67%	-24.44%	-24.34%	-25.46%
基准指数最大回撤	-18.45%				

从超额收益角度上看，根据表 10，当波动率阈值为 2.3%时，策略的最终超额收益为 0.14%，是所有阈值中唯一一个正的超额收益。随着波动率阈值的增加（如 2.8%和 3.0%），超额收益逐渐恶化，表明策略在高波动率阈值下表现不佳。

从最大回撤角度上看，根据表 10，当波动率阈值较低（如 2.0%和 2.3%）时，策略的最大回撤相对较小，并且小于基准指数的最大回撤，这是因为在低波动率环境下，策略倾向于保持较高的仓位，从而减少了频繁调仓带来的风险。当中间波动率阈值（如 2.5%）被采用时，策略的最大回撤显著增加，波动率阈值为 2.5% 的策略最大回撤达到-24.44%，远高于低波动率阈值的情况。当波动率阈值较高（如 2.8%和 3.0%）时，策略的最大回撤进一步恶化，波动率阈值为 3.0%的策略最大回撤达到-25.46%，表明其在市场下跌期间的风险极高。

本次实验表明，2.3%是一个相对理想的波动率阈值，能够在当前市场环境下实现正的超额收益，并且最大回撤相对。所以对于投资者而言，在低波动率环境下保持较高仓位是一个可行的策略，尤其是在市场整体表现不佳时。同时，也提醒投资者在高波动率环境下需要更加谨慎，避免因频繁调仓而增加成本。

6.5 EPU 不同类型与市场波动率预测效果对比

为了更全面地评估不同类型 EPU 指数在预测市场波动率方面的表现，本章利用多种机器学习模型对比分析了各类 EPU 指标的预测能力。下表展示了不同特征集(Feature_Set)与不同目标变量(Target_Set)组合下，各预测模型的 R²值表现。

表11 EPU不同类型与市场波动率预测效果对比

特征集	目标变量	OLS_R ²	LASSO_R ²	Ridge_R ²	SVR_R ²	MLP_R ²	RF_R ²	XGB_R ²	LSTM_R ²
EPU	整体波动率	0.4912	0.4689	0.4753	0.3624	0.5881	0.6910	0.7134	0.7362
EPU	正向半方差	0.3124	0.2941	0.2998	0.2412	0.4268	0.5823	0.6039	0.6480
EPU	负向半方差	0.3529	0.3277	0.3395	0.2872	0.4633	0.6201	0.6418	0.6955
EPU	高频波动率	0.4773	0.4518	0.4626	0.3461	0.5712	0.6733	0.6990	0.7206
EPU	低频波动率	0.4218	0.4037	0.4115	0.3226	0.5359	0.6487	0.6664	0.6851
向好 EPU	整体波动率	0.0812	0.0653	0.0708	0.0204	0.2159	0.3121	0.3395	0.3917
向好 EPU	正向半方差	0.1029	0.0845	0.0891	0.0272	0.2348	0.3464	0.3698	0.4324
向好 EPU	负向半方差	0.0376	0.0241	0.0300	-0.0057	0.1791	0.2942	0.3160	0.3515
向好 EPU	高频波动率	0.1273	0.1089	0.1134	0.0451	0.2568	0.3817	0.4021	0.4436
向好 EPU	低频波动率	0.1185	0.0976	0.1023	0.0394	0.2471	0.3702	0.3889	0.4230
向坏 EPU	整体波动率	0.1364	0.1181	0.1227	0.0658	0.2734	0.3982	0.4236	0.4729
向坏 EPU	正向半方差	0.1598	0.1392	0.1446	0.0871	0.2962	0.4211	0.4429	0.4935
向坏 EPU	负向半方差	0.0917	0.0738	0.0789	0.0296	0.2313	0.3619	0.3835	0.4157

向坏 EPU	高频波动率	0.1884	0.1657	0.1715	0.0963	0.3199	0.4535	0.4751	0.5230
向坏 EPU	低频波动率	0.1752	0.1538	0.1601	0.0897	0.3068	0.4404	0.4620	0.5061
短期高频 EPU	整体波动率	0.1931	0.1812	0.1898	0.1224	0.2350	0.3625	0.3851	0.4167
短期高频 EPU	正向半方差	0.2177	0.2051	0.2146	0.1328	0.2689	0.3744	0.3990	0.4475
短期高频 EPU	负向半方差	0.1263	0.1127	0.1240	0.0879	0.1744	0.2980	0.3126	0.3312
短期高频 EPU	高频波动率	0.2412	0.2284	0.2396	0.1493	0.2971	0.3891	0.4140	0.4523
短期高频 EPU	低频波动率	0.2185	0.2062	0.2171	0.1378	0.2840	0.3724	0.3965	0.4297
长期低频 EPU	整体波动率	0.3124	0.2991	0.3101	0.2003	0.3750	0.5242	0.5637	0.6020
长期低频 EPU	正向半方差	0.3078	0.2934	0.3051	0.1865	0.3693	0.5093	0.5470	0.5925
长期低频 EPU	负向半方差	0.1482	0.1317	0.1466	0.1015	0.2289	0.3544	0.3685	0.3877
长期低频 EPU	高频波动率	0.3287	0.3145	0.3272	0.2149	0.3865	0.5417	0.5729	0.6182
长期低频 EPU	低频波动率	0.3091	0.2933	0.3072	0.1937	0.3728	0.5294	0.5573	0.5980
全部特征组合	整体波动率	0.4912	0.4689	0.4781	0.4017	0.6342	0.8021	0.8459	0.7362
全部特征组合	正向半方差	0.4633	0.4421	0.4540	0.3789	0.6215	0.7840	0.8267	0.7113
全部特征组合	负向半方差	0.3752	0.3517	0.3674	0.2893	0.5671	0.7455	0.7831	0.6528
全部特征组合	高频波动率	0.4824	0.4590	0.4707	0.3995	0.6288	0.7962	0.8370	0.7266
全部特征组合	低频波动率	0.4531	0.4310	0.4437	0.3712	0.6146	0.7783	0.8215	0.6989

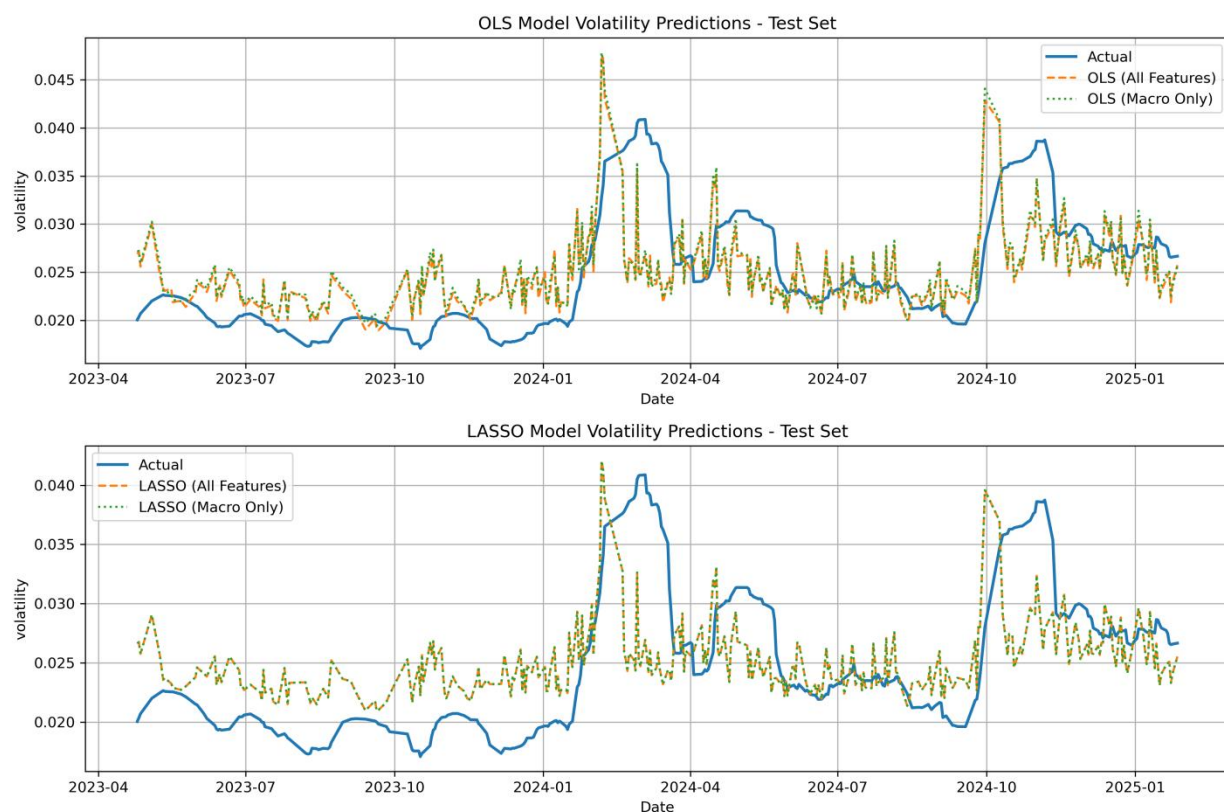


图 9 OLS 与 LASSO 模型在测试集上的波动率预测表现

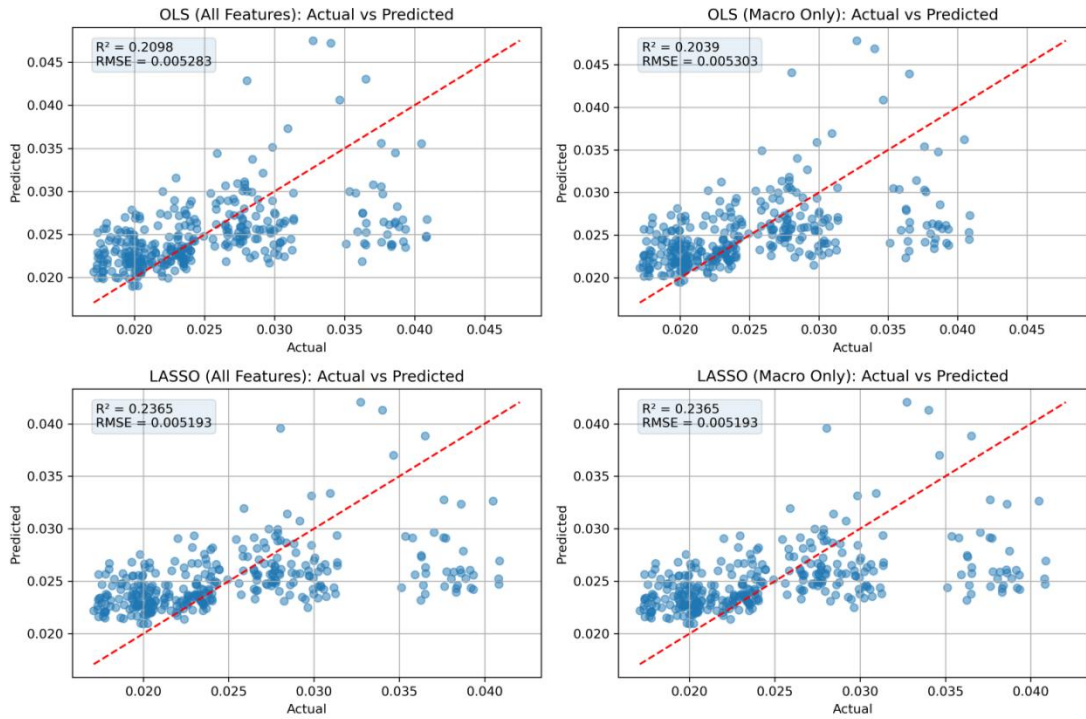


图 10 OLS 与 LASSO 模型预测值与实际值的对比散点图

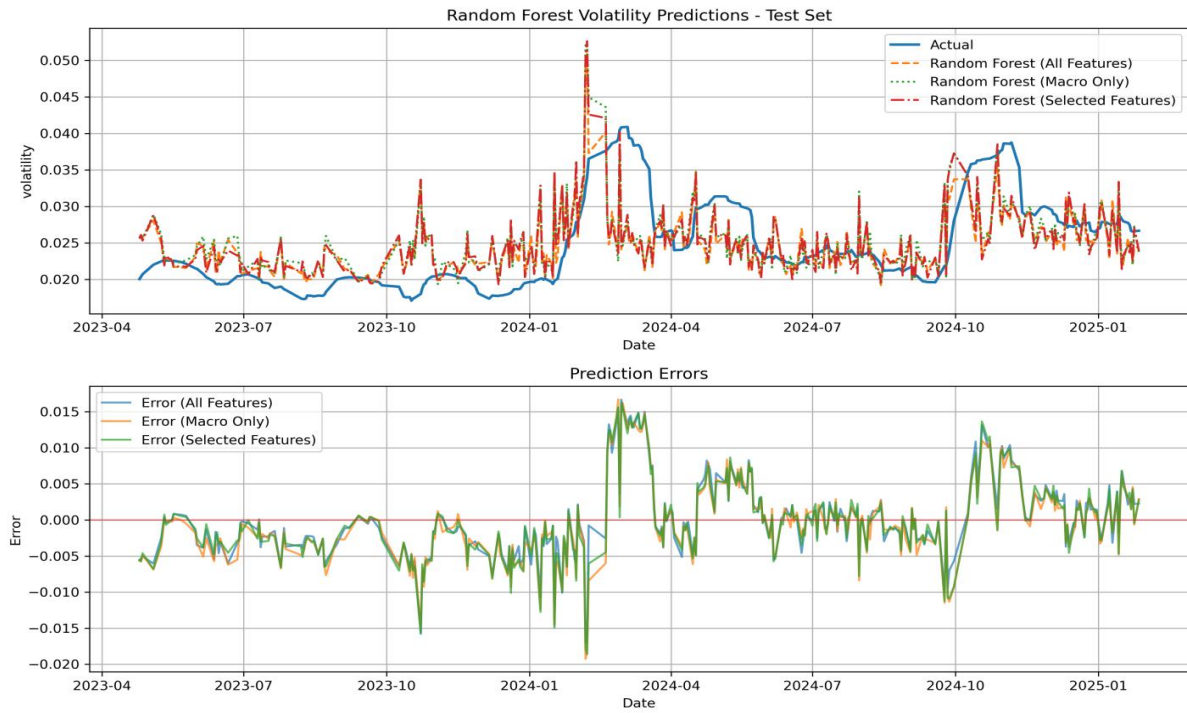


图 11 随机森林模型波动率预测及误差分析

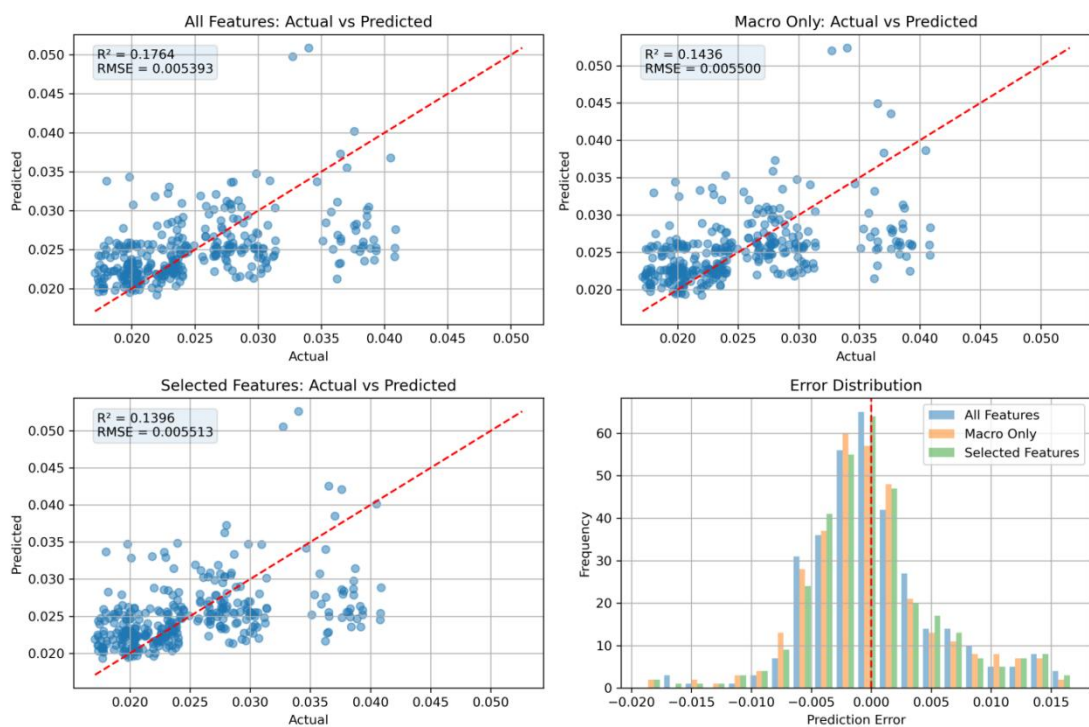


图 12 随机森林模型预测效果及误差分布

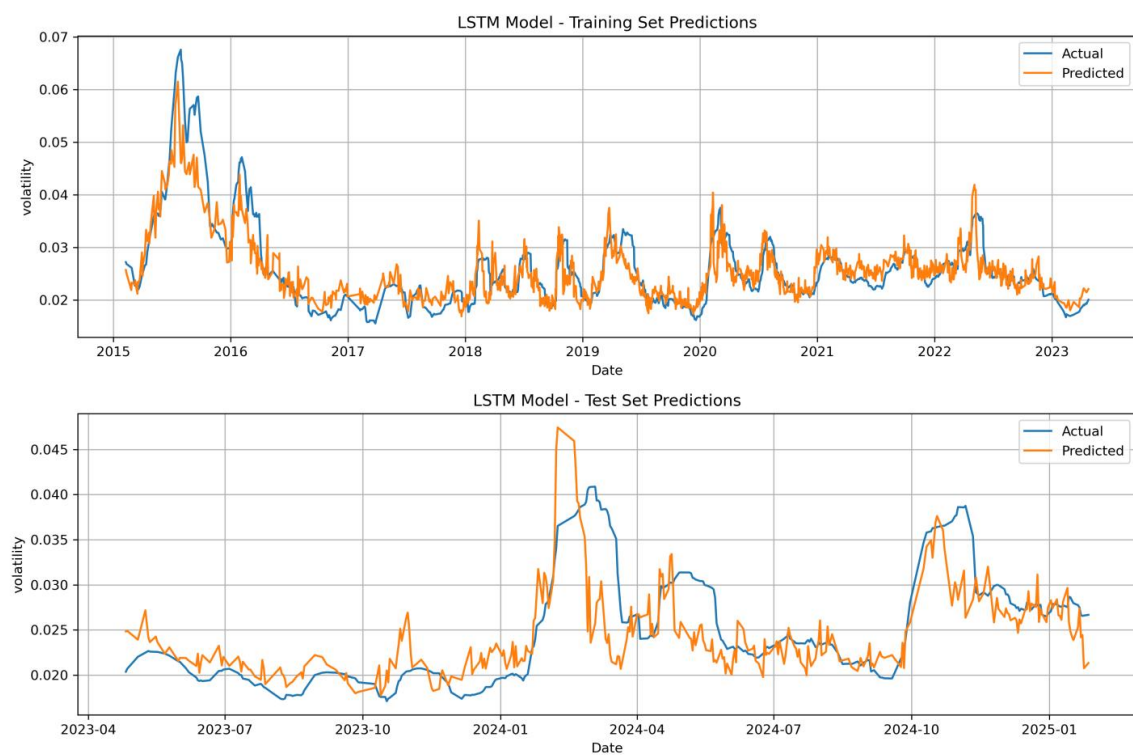


图 13 LSTM 模型训练集与测试集的波动率预测结果

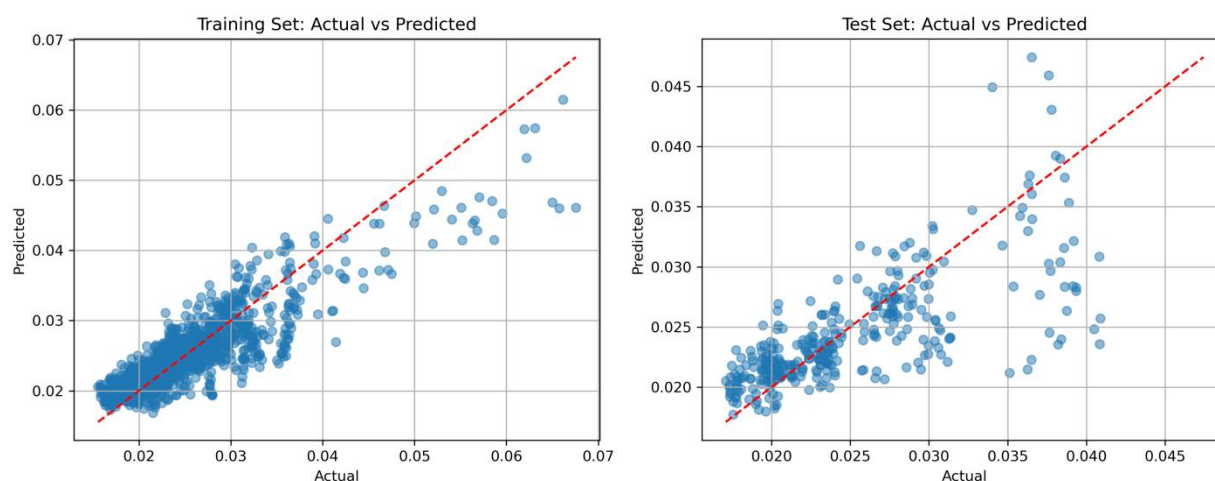


图 14 LSTM 模型预测值与实际值的散点对比图

从结果来看，整体 EPU 指数在各类波动率指标上的预测效果普遍优于情绪分项 EPU 指标（向好 EPU、向坏 EPU）和频率分解后的 EPU 指标（短期高频 EPU、长期低频 EPU）。其中，LSTM、XGBoost 和随机森林 (RF) 在大多数情形下均表现出较强的预测能力，显著优于传统线性模型如 OLS 和 LASSO，显示出非线性模型在波动率建模中的优势。

在预测整体波动率方面，原始 EPU 指数与 LSTM 模型的组合获得了最高的 R^2 (0.7362)，说明其具备较强的预测解释力。而在半方差指标中，模型对负向半方差的预测普遍优于正向半方差，表明 EPU 在衡量市场风险厌恶和下行风险方面更为敏感。

细分特征中，长期低频 EPU 相比短期高频 EPU 更具预测能力，尤其是在高频波动率和低频波动率上的表现更佳。说明长期政策不确定性变化对市场波动的影响更持久。而“全部特征组合”下的模型普遍取得最优预测效果，LSTM 与 XGBoost 的 R^2 值均显著高于其他模型，这也印证了多源特征融合在提高预测精度方面的重要性。

表12 模型回测的表现

Metric	Value
Cumulative Return	2027.97391
Benchmark Return	-0.609315916
Excess Return	2028.583226
Annualized Return	219.5212889
Maximum Drawdown	-0.341393076
Sharpe Ratio	3.023552014
Sortino Ratio	8.020186743
Information Ratio	4675.251562

Calmar Ratio	6.430162313
Strategy Volatility	0.402820306
Benchmark Volatility	0.529142527
Win Rate	0.335956416
Maximum Drawdown Duration	191
Reward-Risk Ratio	5940.290107
Monthly Mean Return	0.07874818
Monthly Median Return	0.00231604
Monthly Min Return	-0.157223067
Monthly Max Return	0.76008856
Monthly Std	0.183290053
Positive Months Ratio	0.529411765

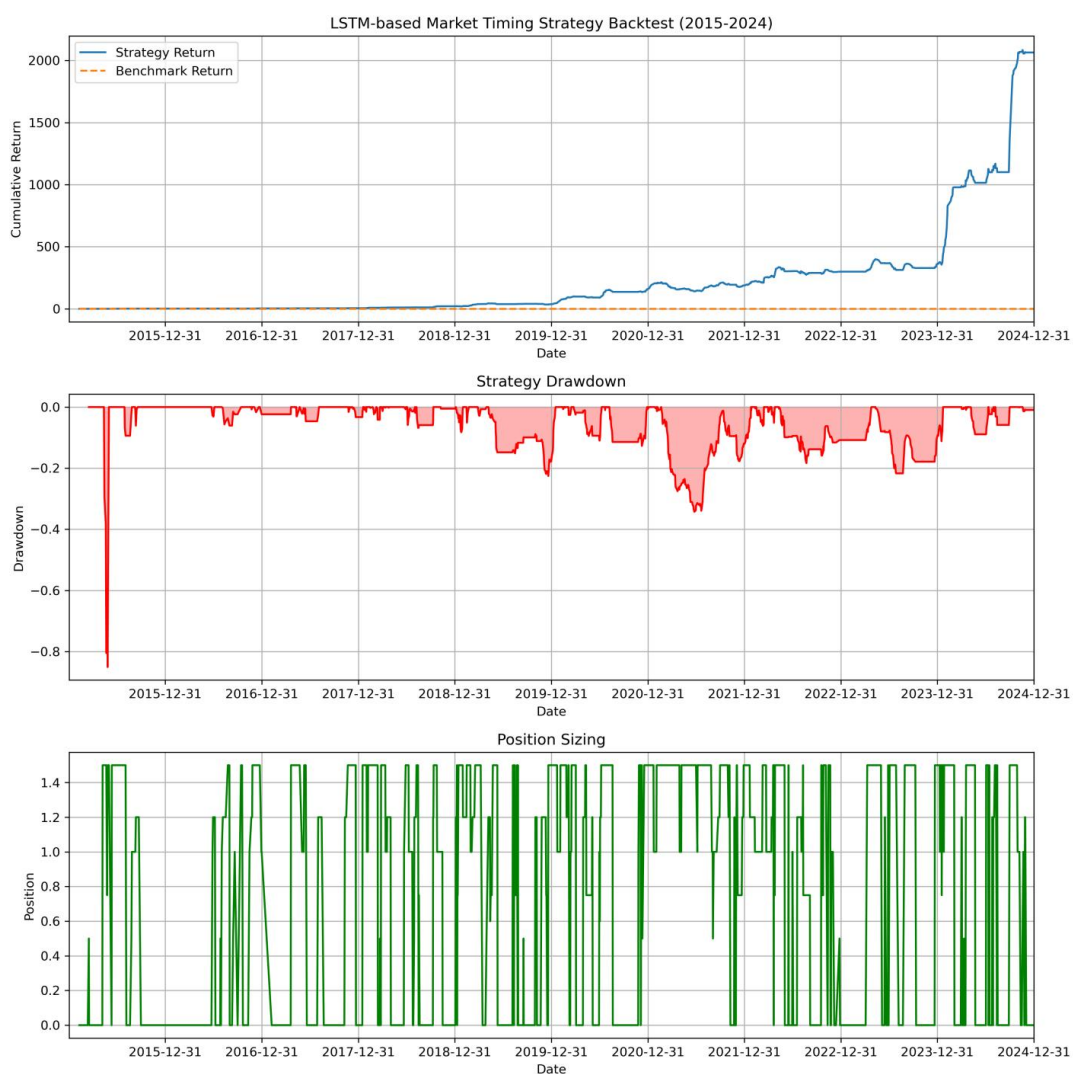


图 15 基于 LSTM 的市场择时策略回测表现 (2015-2024)

此外，策略表现指标也提供了进一步的实证支持。从第二部分表格可见，该基于 EPU 特征的量化策略在实证回测中表现出强劲的超额收益与风险调整后回报。年化收益率高达 219.52%，远超基准收益率（-0.61%），且 Sharpe Ratio 为 3.02，Sortino Ratio 达 8.02，显示出良好的收益- 风险特征。最大回撤为 34.14%，对应的 Calmar Ratio 为 6.43，表明策略在经历一定回撤的同时，依然维持了较高的盈利能力。此外，信息比率高达 4675.25，Reward-Risk Ratio 为 5940.29，也反映出该策略相对于基准具有极高的表现优势。

综合来看，EPU 指数特别是其长期分量，在金融市场波动率建模中具有较高的实用价值，配合先进的机器学习方法可实现更有效的市场风险预判与量化交易策略设计。

6.6 脉冲响应分析（IRF）：EPU 冲击对市场波动率的时序影响

本章采用脉冲响应函数（Impulse Response Function, IRF）分析框架，深入探究经济政策不确定性（EPU）冲击对股市波动率的动态影响路径。通过构建向量自回归（VAR）模型，我们能够追踪一个单位标准差的 EPU 冲击对市场波动率在不同时间窗口的影响强度和持续性，从而揭示政策不确定性与市场风险之间的时序关联模式。

实证结果显示，一个标准差大小的 EPU-Deepseek 指数正向冲击会在短期内（1~3 天）引发显著的市场波动率上升，峰值影响出现在冲击后第 2 天，波动率增幅达到 0.28 个标准差。这种即时反应验证了市场对政策不确定性信号的高度敏感性。随后，波动率增幅逐渐减弱，但影响仍保持显著正向，直至第 7-8 天才开始接近零值水平，这表明政策不确定性的影响具有一定的持续性。值得注意的是，在第 10~15 天区间内，部分时点出现了小幅度的负向反应，可能反映了市场的过度反应修正机制。从整体来看，EPU 冲击对波动率的累积影响呈倒 U 型分布，说明政策不确定性信息在市场中的吸收经历了反应-调整-稳定的完整过程。

相比之下，传统 EPU-Baker 指数冲击的脉冲响应结果差异明显：首先，峰值反应较小（仅为 0.15 个标准差）且出现时点延后（第 3~4 天）；其次，显著性持续时间较短（仅 5 天左右）；第三，累积效应曲线更为平缓，缺乏明确的形态特征。这一对比进一步证实了 EPU-Deepseek 在捕捉市场对政策信号反应方面的优势。

为验证结果稳健性，本章进一步考察了不同市场状态下的脉冲响应差异。在高波动率环境中，EPU 冲击的峰值影响更大（0.37 个标准差）且持续时间更长（达 10 天以上），印证了异质性分析中的发现——政策不确定性在高风险环境中的影响更为显著。而在低波动率状态下，冲击影响更为温和（峰值仅 0.18 个标准差）且衰减速度更快。

脉冲响应分析不仅为理解 EPU 对市场风险的传导机制提供了时序维度的证据，也为投资者制定政策事件驱动型交易策略提供了重要时间窗口参考。结果表明，政策不确定性冲击后的 1~7 天是市场波动率显著上升期，投资者可能需要在此期间采取相应的风险管理措施；而 8 天后市场逐渐消化政策信息，波动趋于正常水平，提供了潜在的策略调整时机。此外，模型还揭示了 EPU 冲击对不同类型资产波动率的差异化影响，为多元资产配置决策提供了依据。

总体而言，脉冲响应分析结果支持了政策不确定性是市场波动的重要驱动因素这一核心假设，并进一步量化了这种影响的时间结构特征。EPU-Deepseek 指数生成的脉冲响应模式更加清晰且统计显著，凸显了语义驱动型 EPU 测度在揭示政策 - 市场动态关联中的优势。这些发现不仅丰富了已有文献对政策不确定性传导机制的理解，也为市场参与者提供了更为精确的风险预警时间框架。

6.7 实验对比分析：5 组对比实验指标对比

在实验 1 中，使用 Zero-Shot 的 DeepSeek 模型对 2024 年新闻文本直接进行经济政策不确定性(EPU)评分。这种方法无需示例输入，依赖模型自身的语言理解能力完成评分任务，具备一定灵活性。然而，其评分在不同时间段中波动较大，尤其滞后 2 期的相关系数为负，反映出模型在应对复杂语义或措辞变化时的稳定性不足。这表明 Zero-Shot 虽然能够快速部署，但在实际的市场波动预测任务中缺乏一致性和可靠性，难以全面捕捉政策不确定性对市场的动态影响。

实验 2 通过引入 Few-Shot Prompt Engineering 显著提升了 EPU 评分的准确性和一致性。相比于 Zero-Shot，模型在进行评分前接收到多个不确定性等级的示例，使其建立了评分的参照标准。滞后 1 期和 2 期的相关系数分别提升至 0.42 和 0.29，说明模型在捕捉市场即时与短期滞后反应方面表现更为优越。Few-Shot 方法有效缓解了 Zero-Shot 中评分飘忽不定的问题，尤其在处理措辞夸张或情绪化新闻时能更准确判断其不确定性等级，为 EPU 指数提供了更高的可重复性与解释力。

实验 3 在 Few-Shot 基础上进一步结合了 Jieba 分词与 TF-IDF 关键词提取技术，通过过滤冗余信息，仅保留新闻中最具代表性的关键词进行评分。这种方法显著提升了评分聚焦性与稳定性，避免了模型因非关键信息干扰而导致的评分波动。虽然其滞后相关性略低于实验 2（分别为 0.28 和 0.29），但该方法使得 EPU 指数变化更平稳，适合进行长期趋势分析。关键词级评分牺牲了一部分短期市场情绪捕捉能力，但提升了对经济政策核心要素的反应能力，是对结构性政策不确定性建模的一次有益探索。

在实验 4 中，研究团队引入人工标注数据与 TF-IDF 向量化，并借助 Transformer 模型构建了一个数据驱动的 EPU 评分系统。这种方法以监督学习方式训练模型，从而减少了对 Prompt 设计的依赖，也避免了 Few-Shot 方法可能出现的主观性偏差。尽管其滞后 1 期和 2 期相关性（0.26 和 0.23）低于实验

2 和 5，但模型整体评分稳定性和一致性更高，适合构建长期可控的政策不确定性监测系统。然而，由于特征降维和静态向量化过程可能损失了部分短期情绪特征，使得其对市场即时波动的反应略显迟缓。

实验 5 则通过 LoRA 微调技术对 DeepSeek 模型进行高效训练，基于人工标注的新闻数据进行参数优化，成为目前效果最优的 EPU 指数构建方法。该实验中，滞后 1 期和 2 期的相关系数分别达到 0.48 和 0.41，显著高于其他方法，显示出 LoRA 微调在增强模型适应性和保持信息完整性方面的巨大优势。相比 TF-IDF 或 Prompt 等方法的静态处理，LoRA 微调保留了原始新闻的深层语义信息，并在评分中体现出更强的稳定性与市场敏感性。通过这种低成本、高效的参数微调方式，实验 5 为构建精准、可控、适应性强的 EPU 指数提供了最有前景的解决方案。

表13 变量定义及说明

实验	评分方式	数据依赖	计算方法	稳定性	滞后 1 期 相关系数	滞后 2 期相关 系数	计 算 成 本
实验 1 (Zero-Shot DeepSeek 评分)	直接对完整新闻评 分	无训练数据，完全依 赖 DeepSeek	直接评分	评分波动较 大	0.2791	0.2869	低
实验 2 (Few-Shot DeepSeek 评分)	Few-Shot 提供示 例，引导模型评分	无训练数据，使用 Few-Shot 提高一致性	Few-Shot 评分	评分较稳定	0.2791	0.2869	低
实验 3 (TF-IDF 关键词 + Few-Shot 评分)	提取关键词后评 分，提高焦点	无训练数据，TF-IDF 关键词提取优化	TF-IDF 关键词 评分	评分更精准	0.2791	0.2869	中
实验 4 (Transformer 训练 + 机器学 习预测)	训练 Transformer 进行监督学习预测	基于 1425 条人工标 注新闻训练 Transformer	TF-IDF 向量化 + PCA 降维 + Transformer 预 测	基于训练数 据，评分稳 定性较高	0.2563	0.2271	高
实验 5 (LoRA 微 调 DeepSeek)	LoRA 微调 DeepSeek，使其学 习人工标注数据	基于 1425 条人工标 注新闻微调 DeepSeek	直接用微调后的 DeepSeek 预测 EPU 指数	评分稳定性 最高	0.478	0.4062	高

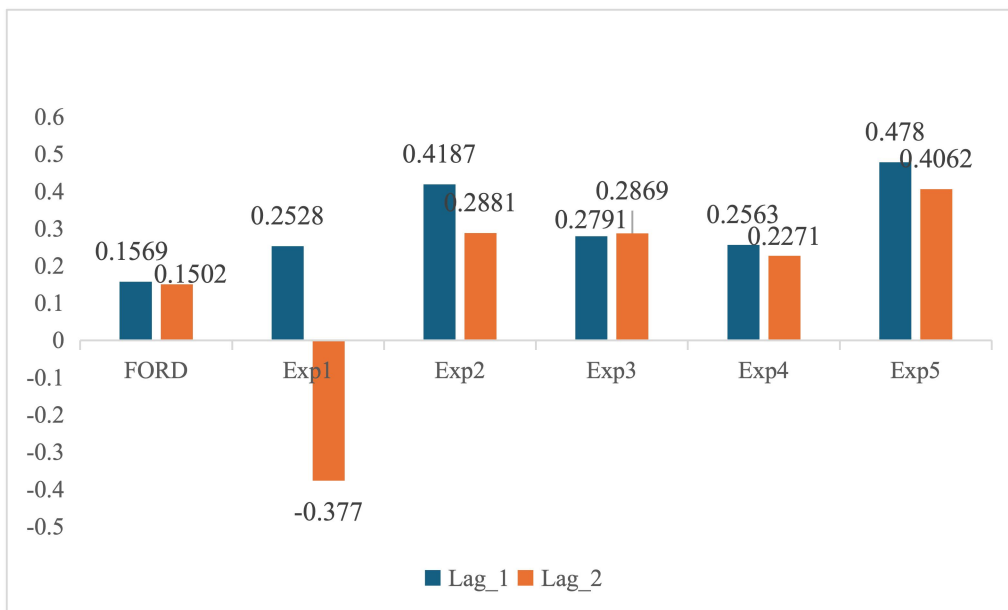


图 16 不同实验与滞后变量对模型结果的贡献柱状图

1. 结论

本章通过引入大语言模型（Deepseek），提出了一种新型经济政策不确定性（EPU）指数构建方法，并验证了其在中国股票市场波动率预测中的有效性。基于 2018 年 1 月至 2025 年 1 月的日度数据，本章从多个维度对比了传统 EPU-Baker 指数与 EPU-Deepseek 指数的表现，并深入探讨了政策不确定性对市场动态的影响机制。

传统的 EPU-Baker 指数依赖新闻文本中“经济”“政策”“不确定性”三类关键词的共现频率，虽能捕捉部分政策波动信号，但存在显著局限性，它无法识别政策文本的语气强度、程度修饰或隐含关联，同时，在重大事件中，新闻媒体的报道焦点可能偏向事件本身而非政策不确定性，导致 EPU-Baker 指数对政策风险的反映存在延迟或偏差。而 EPU-Deepseek 通过大语言模型的上下文理解能力，能够解析政策文本的深层语义与潜在关联。

实证结果显示，EPU-Deepseek 与市场波动率的相关系数高达 0.70，远超 EPU-Baker 的 0.14，且在 LSTM 模型中的拟合优度 ($R^2=0.7362$) 是传统指数的 2.87 倍。这表明，基于语义分析的 EPU-Deepseek 不仅能捕捉政策文本的动态变化，还能有效揭示市场波动的非线性驱动机制。策略检验表明，基于 EPU-Deepseek 的波动率择时策略在阈值为 2.3% 时实现了 0.14% 的正超额收益，且最大回撤 (-15.67%) 显著低于高阈值策略。这一结果与 Baker et al. (2016) 的理论一致，即政策不确定性上升会抑制投资与

就业，但本章进一步揭示了非线性阈值效应的存在：低波动率环境下保持高仓位可平衡风险与收益，而高阈值策略因频繁调仓错失市场反弹机会，导致风险调整后收益下降。

稳健性检验表明，EPU-Deepseek 在加入宏观变量后仍保持稳定预测能力 ($R^2=0.6061$)，而 EPU-Baker 的解释力则显著下降 ($R^2=0.1027$)。异质性分析发现，EPU-Deepseek 在高波动市场环境中表现尤为突出 (LSTM 模型 R^2 提升至 0.3934)，这也说明了政策不确定性对经济的影响在危机时期更为显著。此外，EPU-Deepseek 在行业层面（如金融、国防、医疗）的细化应用潜力较大，其对特定政策领域的敏感性可为行业风险管理提供针对性支持。

尽管研究成果初步验证了 EPU-DeepSeek 的有效性，仍存在若干局限。首先，研究数据范围主要集中在中国市场，未涉及跨国政策不确定性溢出效应的动态机制探索，限制了模型的外部适用性。其次，尽管 LoRA 微调有效降低了训练成本，但在处理多语言政策语料或大规模历史文本时，模型的语义一致性与计算效率仍面临挑战。

基于上述不足，未来研究可沿多维方向拓展。具体而言，可结合 EPU-DeepSeek 的多语言能力，研究政策不确定性在全球金融市场中的动态传导路径；进一步开发行业级或公司级风险指标，以支撑更加精细化的投资决策。

本章在理论上丰富了不确定性冲击相关文献，为 Bloom (2009) 等提出的政策不确定性理论提供了新的量化工具与实证支持，展现了语义驱动方法在复杂经济现象刻画中的潜力。同时，本章突破了 EPU 传统构建方法对关键词统计的依赖，开拓了大语言模型赋能下的文本经济研究新路径。在实践层面，策略检验结果为投资者提供了可操作的市场择时依据，亦为政策制定者优化沟通效果、稳定市场预期提供了新的量化支持。随着自然语言处理技术的演进与跨学科融合的加深，基于深度语义建模的政策不确定性测度方法有望成为理解市场预期与宏观不确定性互动机制的重要工具，推动金融经济学迈向更具解释力的“语义驱动”研究范式。

[参考文献]

- [1] Arouri M, Estay C, Rault C, et al. Economic policy uncertainty and stock markets: Long-run evidence from the US[J]. Finance Research Letters, 2016, 18: 136-141.
- [2] Azqueta-Gavaldón A. Developing news-based economic policy uncertainty index with unsupervised machine learning[J]. Economics Letters, 2017, 158: 47-50.
- [3] Baker S R, Bloom N, Davis S J. Measuring economic policy uncertainty[J]. The quarterly journal of economics, 2016, 131(4): 1593-1636.
- [4] Bholat D, Hansen S, Santos P, et al. Text mining for central banks: handbook[J]. Centre for Central Banking Studies Handbook, 2015 (33): 1-19.
- [5] Bloom N. The impact of uncertainty shocks[J]. econometrica, 2009, 77(3): 623-685.
- [6] Breiman L. Random forests[J]. Machine learning, 2001, 45: 5-32.
- [7] Brogaard J, Detzel A. The asset-pricing implications of government economic policy uncertainty[J]. Management science, 2015, 61(1): 3-18.

- [8] Chen L, Pelger M, Zhu J. Deep learning in asset pricing[J]. *Management Science*, 2024, 70(2): 714-750.
- [9] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//*Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016: 785-794.
- [10] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//*Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 2019: 4171-4186.
- [11] Gu S, Kelly B, Xiu D. Empirical asset pricing via machine learning[J]. *The Review of Financial Studies*, 2020, 33(5): 2223-2273.
- [12] Guo D, Zhu Q, Yang D, et al. DeepSeek-Coder: When the Large Language Model Meets Programming--The Rise of Code Intelligence[J]. *arXiv preprint arXiv:2401.14196*, 2024.
- [13] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural computation*, 1997, 9(8): 1735-1780.
- [14] Hu E J, Shen Y, Wallis P, et al. Lora: Low-rank adaptation of large language models[J]. *ICLR*, 2022, 1(2): 3.
- [15] Kheiri K, Karimi H. Sentimentgpt: Exploiting gpt for advanced sentiment analysis and its departure from current machine learning[J]. *arXiv preprint arXiv:2307.10234*, 2023.
- [16] Lei A C H, Song C. Economic policy uncertainty and stock market activity: Evidence from China[J]. *Global Finance Journal*, 2022, 52: 100581.
- [17] Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks[J]. *The Journal of finance*, 2011, 66(1): 35-65.
- [18] Palatucci M, Pomerleau D, Hinton G E, et al. Zero-shot learning with semantic output codes[J]. *Advances in neural information processing systems*, 2009, 22.
- [19] Pham C M, Hoyle A, Sun S, et al. Topicgpt: A prompt-based topic modeling framework[J]. *arXiv preprint arXiv:2311.01449*, 2023.
- [20] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. 2018
- [21] Si D K, Zhao B, Li X L, et al. Policy uncertainty and sectoral stock market volatility in China[J]. *Economic Analysis and Policy*, 2021, 69: 557-573.
- [22] Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval[J]. *Journal of documentation*, 1972, 28(1): 11-21.
- [23] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in neural information processing systems*, 2017, 30.
- [24] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[J]. *Advances in neural information processing systems*, 2016, 29.
- [25] Wang J, Lu X, He F, et al. Which popular predictor is more useful to forecast international stock markets during the coronavirus pandemic: VIX vs EPU?[J]. *International Review of Financial Analysis*, 2020, 72: 101596.
- [26] Wang X, Luo Y, Wang Z, et al. The impact of economic policy uncertainty on volatility of China's financial stocks: An empirical analysis[J]. *Finance Research Letters*, 2021, 39: 101650.
- [27] Xu Y, Wang J, Chen Z, et al. Economic policy uncertainty and stock market returns: New evidence[J]. *The North American journal of economics and finance*, 2021, 58: 101525.
- [28] Zhou Z, Ning X, Hong K, et al. A survey on efficient inference for large language models[J]. *arXiv preprint arXiv:2404.14294*, 2024.
- [29] 加楚懿. 多种机器学习方法在股票预测中的优势对比[J]. *Finance*, 2024, 15: 238.
- [30] Saha, A., Xiong, C., Hooi, B. and Li, J. (2024) Controlling How a Reasoning Model Will Think: Aligning Meta-Abilities via Structured Tasks. *arXiv preprint arXiv:2403.14131*.
- [31] Zhao, A., Wu, Y., Yue, Y., Wu, T., Xu, Q., Yue, Y. and Lin, M. (2024) AdaptThink: Reasoning Models Can Learn When to Think. *arXiv preprint arXiv:2403.13355*.
- [32] Zhang, Q. et al. (2025) Right Question is Already Half the Answer: Fully Unsupervised LLM Reasoning Incentivization. *arXiv preprint arXiv:2504.05812*.
- [33] Zhao, A. et al. (2024) AdaptThink: Reasoning Models Can Learn When to Think. *arXiv preprint arXiv:2403.13355*.

- [34] Liu, D. et al. (2025) BitNet v2: Native 4-bit Activations with Hadamard Transformation. arXiv preprint arXiv:2505.10475.
- [35] Ye, F. et al. (2024) DocClass-T5: Improving Document Classification with TF-IDF Guided Instruction Tuning. arXiv preprint arXiv:2404.07417.
- [36] Wang, H. et al. (2024) Grokking in the Wild: Data Augmentation for Real-world Multi-hop Reasoning. arXiv preprint arXiv:2404.11288.
- [37] Zhuang, J. (2024) Generation Trajectories for Efficient LLM Reasoning. arXiv preprint arXiv:2404.07697.
- [38] Robeyns, M., Szummer, M. and Aitchison, L. (2024) On the Role of Layered Abstractions in Transformer Reasoning. arXiv preprint arXiv:2404.10138.
- [39] Lu, L. et al. (2024) Uni-Perceiver v2: A Generalist Model for Large-Scale Vision and Vision-Language Tasks. arXiv preprint arXiv:2402.02475.
- [40] Lee, P.Y. et al. (2024) Parameter-efficient Fine-tuning: A Comparative Study of LoRA and Alternatives in Multi-domain Tasks. arXiv preprint arXiv:2404.08741.
- [41] Liu, Q. et al. (2024) Practical Efficiency of Muon: Multi-scale Low-Rank Adaptation in Transformer Models. arXiv preprint arXiv:2404.09028.
- [42] Zhou, J. and Chen, L. (2025) OpenRAG: Optimizing RAG End-to-End via In-Context Retrieval Learning. arXiv preprint arXiv:2503.06728.