

基于 CiteSpace 的财经新闻外文文献计量分析（2021–2025）

马颖¹ 常乐¹ 吴明海²

1. 沈阳理工大学外国语学院 辽宁沈阳 110000

2. 沈阳建筑大学外国语学院 辽宁沈阳 110000

摘要：本研究以中国知网 2021 至 2025 年收录的财经新闻外文文献摘要为分析文本，通过 CiteSpace 工具绘制其研究主题词突现图谱与聚类图谱，揭示该领域技术驱动下研究主题的动态演进路径。研究发现：主题聚类呈现“技术驱动”与“场景聚焦”的双重特征，涵盖“股票市场预测”“LSTM 模型创新”；技术方法从机器学习的初步引入（2021–2022）逐步深化至 LSTM 等深度学习模型的应用（2023–2025），股票价格预测成为核心场景；未来研究可向多模态数据融合方向转向。本研究采用可视化手段揭示 4 年来财经新闻外文文献研究热点及其演进特征，旨在为该领域学术与实践研究提供参考与启示。

关键词：财经新闻；外文文献；CiteSpace；突现分析；聚类分析

1. 研究背景

财经新闻是经济信息中具有新闻价值的信息^[1]。作为金融市场核心信息的传播媒介，对投资决策、政策解读与风险管理具有重要影响。金融领域和计算机领域的众多研究者关注股票预测问题^[2]，随着数据规模的激增和市场复杂性的提升，结合自然语言处理（NLP）与机器学习（ML）等技术，研究财经新闻已成为揭示市场动态与提升预测精度的关键途径，研究者通过文本挖掘技术提取财经网站信息，结合百度指数的搜索热度与算法模型，分析市场关键词对投资策略与收益率的影响^[3]。通过 CiteSpace 工具绘制研究主题词知识图谱工具，用可视化手段呈现该领域研究主题的演变趋势，厘清领域内研究主题聚类特征，揭示财经新闻对金融市场的深远影响，还为未来的跨学科创新研究提供了方向性指导。本研究聚焦 2021–2025 年中国知网收录的财经新闻英文文献，通过关键词突现及聚类分析系统探索财经新闻研究热点及其技术方法维度的演变，并进一步预判潜在的未来研究热点与创新机遇，为该领域的未来发展提供系统性指导。

2. 研究设计

在数据来源与筛选方面，本研究选择中国知网外文数据库为数据源，将时间跨度设定为 2021–2025 年，以“financial news”为检索词进行全库关键词，共检索词 112 篇英文文献（截止 2025 年 3 月 3 日）。

在分析工具与参数设置方面，本研究采用 CiteSpace 6.1 版本进行计量分析，该软件分析出来的图形称为“科学知

识图谱”，为科学计量学界的研究提供了有效路径^[4]。主题词聚类分析和突现时间切片参数设置为 2021–2025 年，切片长度 4 年。路径裁剪阈值（LRF=2.5, LBY=5），保留高频关键词与强关联节点。聚类算法为 LLR，模块性（Modularity=0.0561）验证聚类合理性。

在分析方法方面，突现分析是通过突现强度（Burst Strength）识别特定时间段内研究热点；聚类分析是结合模块性（Q>0.3 为显著聚类）与轮廓值（Silhouette>0.7 为高一致性），提取核心主题簇，揭示核心研究主题；网络结构分析是通过密度（Density=0.0231）与节点中心性（Centrality）来评估研究主题此词间的关联强度。

3. 主题词聚类分析

通过 LLR 算法生成 22 个聚类，筛选出 6 个高显著性聚类（Silhouette>0.9），其内容与特征如下：

技术驱动型聚类主题为 #2 lstm-based novel model，关键词包括 LSTM, attention mechanism, hybrid model，聚焦改进型 LSTM 框架，如结合注意力机制以增强时序特征提取能力。典型案例包括利用 LSTM-ATT 模型预测加密货币价格（Wang et al., 2023）。此外，主题聚类词 #4 stock price prediction，关键词包括 machine learning, volatility, time series，涵盖从传统 ARIMA 到深度学习的多模型比较，强调特征工程（如波动率指标、新闻情感得分）对预测精度的提升。

场景导向型聚类 #3 stock market index return，关键词包括 index tracking, risk premium, macroeconomic factors，主要分

析市场指数回报的驱动因素，如美联储政策变动对 S&P 500 指数的影响 (Baker et al., 2022)。主题聚类词 #5 financial information，关键词包括 text mining, event extraction, earnings call，主要涉及非结构化文本（如财报电话会议记录）的情感分析与事件信号挖掘，用于企业财务风险预警。

方法论对比聚类 #3 comparative study，关键词包括 LSTM vs. GRU, model robustness, overfitting，对比不同深度学习架构（如 LSTM 与 GRU）在相同数据集下的表现，探讨模型过拟合与泛化能力。

聚类演进趋势呈现出技术深化和数据融合特征。从早期通用机器学习 (2021–2022) 转向特定模型优化 (2023–2025)，如 LSTM 与 Transformer 的混合架构，体现出技术深化特征；2023 年后，“#5 financial information”聚类中“multimodal data”关键词频次增加，表明多源数据整合成为新趋势，体现出数据融合特征。

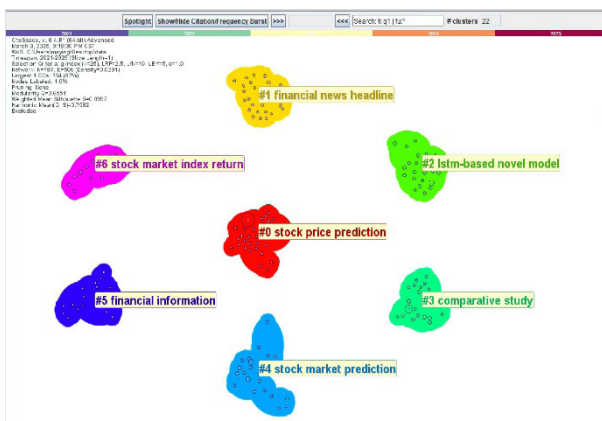


图 1 财经新闻研究主题词聚类图谱

Top 25 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2021 - 2025
financial markets	2021	2.39	2021	2022	
media sentiment	2021	1.12	2021	2021	
economic news	2021	1.12	2021	2021	
deep learning	2021	0.97	2021	2022	
machine learning	2022	1.22	2022	2022	
financial news article	2022	0.95	2022	2023	
financial news report	2022	0.87	2022	2022	
learning method	2022	0.87	2022	2022	
multiple perspectives	2022	0.87	2022	2022	
short-term memory	2023	1.83	2023	2023	
stock price	2023	1.36	2023	2023	
long short-term memory	2023	0.91	2023	2023	
news sentiment	2023	0.91	2023	2023	
textual data	2023	0.91	2023	2023	
incorporating financial news	2023	0.91	2023	2023	
machine learning model	2023	0.91	2023	2023	
prediction model	2023	0.91	2023	2023	
significant correlation	2023	0.91	2023	2023	
positive impact	2023	0.91	2023	2023	
learning approach	2023	0.91	2023	2023	
stock prediction	2023	0.91	2023	2023	
superior performance	2023	0.91	2023	2023	
stock feature	2023	0.91	2023	2023	
stock price prediction	2024	1.65	2024	2025	
sentiment analysis	2021	0.93	2024	2025	

图 2 财经新闻研究主题词突现图谱

4. 主题词突现分析

如图 2 所示，研究主题词突现图谱从技术方法维度解

释了该领域研究热点的阶段性转移特征。

2021 年为技术工具的初步引入和探索阶段，研究主题以传统金融分析为主，如：主题突现强度最强的主题词为“financial markets”（突现强度 2.39）和“economic news”（突现强度 1.12），表明该时期的财经新闻研究以基础性财经市场分析为核心。此外，突现词“deep learning”（突现强度 0.97）和“sentiment analysis”（突现强度 0.93）表明，财经新闻研究领域开始尝试引入深度学习和情感分析等技术方法开展研究，如通过情感分析辅助解读经济新闻对市场的影响，此时技术应用较为分散，尚未形成系统性框架。

2022 年为机器学习主导阶段，“machine learning”（突现强度 1.22）突现强度高于其他主题词而成为年度核心关键词，这表明该领域研究重心向数据驱动的技术方法转移。此外，进一步分析“financial news report”（突现强度 0.95）、“multiple perspectives”（突现强度 0.87）、“learning method”（突现强度 0.87）可见，该阶段财经新闻报道更关注利用机器学习方法探究多源数据的整合（如新闻文本、社交媒体、市场交易数据）、跨学科视角的融合，如：通过机器学习模型分析不同渠道的财经新闻对市场波动的差异化影响。

2023 年为深度学习与模型深化阶段，“long short-term memory”（突现强度 0.91）、“textual data”（突现强度 0.91）和“machine learning model”（突现强度 0.91）等突现词的出现表明，该阶段研究的技术应用进一步得到了细化。此外，长短期记忆网络（LSTM）技术引用说明研究重点转向了提升时序数据处理能力，尤其适用于股票价格等具有时间依赖性的预测场景。主题词“textual data”与“news sentiment”的同步突现且强度相同，揭示了非结构化文本数据（如财经新闻）的价值挖掘趋势，如：通过自然语言处理技术对文本进行情感分析，通过提取和分析情感倾向来预测市场情绪。2023 年，关键词“stock price prediction”（突现强度 1.36–1.83）和“prediction model”（突现强度 0.91）成为焦点，突显研究向金融预测的精准化转型。例如，结合 LSTM 模型与新闻情感分析，构建动态股票价格预测框架。此外，“incorporating financial news”（突现强度 0.91）表明，研究者意识到单一数据源的局限性，开始探索多模态数据融合（如新闻文本 + 历史交易数据 + 宏观经济指标）以提升预测精度。

2024–2025 年为预测模型进一步优化阶段，“stock price

prediction” 突现强度达 1.65, 同时 “sentiment analysis” 重现突现 (突现强度 0.93), 反映情感分析与市场预测的长期协同效应, 如通过实时新闻情感监测动态调整预测参数。此外, 关键词 “superior performance” (突现强度 0.91) 的出现, 可能指向模型在回测或实际应用中表现出的显著优势, 推动研究向技术落地的纵深发展。

5. 结语与启示

通过对中国知网外文数据库收录的英文财经新闻文献进行聚类与突现分析, 本研究揭示了该领域的研究主题的技术方法维度的演化特征。近 4 年来, 研究重点逐步从基础金融分析转向以机器学习和深度学习为核心的技术驱动方法应用研究, 其中 LSTM、情感分析、新闻文本挖掘等创新方法在市场预测中取得了显著进展; 多模态数据的融合及实时性分析逐步成为了关注重点, 以应对金融市场的动态复杂性。

近年来, 全球金融市场在应对复杂国际形势中不断变革、融合创新, 并推动可持续发展转型^[5]。未来财经新闻研究可重点关注以下几个维度: 持续关注技术创新, 探索多模态数据融合, 整合新闻文本、历史交易数据和宏观经济因素, 以提升金融市场预测的精准度; 关注跨学科协作, 通过金融、自然语言处理 (NLP) 和机器学习领域的合作, 构建更智能化的财经新闻分析框架; 关注实用性提升, 聚焦模型的实际应用场景, 增强模型性能实时性、稳健性和可解释性, 为财经领域政策制定者和参与者提供决策支持工具。

综上所述, 财经新闻领域呈现出从技术优化向实际应用落地的转型趋势, 基于 CiteSpace 的研究热点及其演变分析对启示未来财经领域研究提供了参考依据。

参考文献:

- [1] 田静, 关众著. 财经新闻传播学 [M]. 北京: 中国经济出版社, 2022.33.
- [2] 熊浩博. 融合知识图谱和情感分析的股价预测模型 [D]. 华东师范大学, 2022.
- [3] 孟雪井, 杨亚飞, 赵新泉. 财经新闻与股市投资策略研究——基于财经网站的文本挖掘 [J]. 投资研究, 2016, 35 (08): 29-37.
- [4] 刘林沙, 陈秋雷编著. 传播学数据分析方法 [M]. 成都: 西南交通大学出版社, 2023.152.
- [5] 马忠富. 迎变革 促融合 助推金融市场高质量发展 [J]. 债券, 2023, (11): 39.

作者简介:

马颖 (1973—), 抚顺人, 沈阳理工大学外国语学院副教授, 语言与翻译中心主任, MTI 硕士生导师, 研究方向为科技英语笔译、经贸英语笔译。

常乐 (1972—), 沈阳人, 沈阳理工大学外国语学院教授, 外国语学院院长, MTI 硕士生导师, 研究方向为科技英语笔译、经贸英语笔译。

吴明海 (1979—), 营口人, 沈阳建筑大学外国语学院副教授, 研究方向为语料库翻译。