第二章 文献综述

第一节 投资者情绪研究综述

1. 投资者情绪的定义与测度

投资者情绪是行为金融学中的核心概念，反映了投资者对未来市场走势的预期和认知偏差。从理论界定角度看，Brown和Cliff（2006）认为投资者情绪本质上反映了市场参与者的期望，是导致市场波动的重要因素。Peng和Xiong（2005）则从认知资源角度提出，投资者情绪是一种稀缺的认知资源，有限的投资者关注会导致存在类别学习的现象，从而影响投资决策。Baker和Stein（2003）从市场流动性视角出发，指出在流动性强的市场中，价格主要由非理性的投资者主导，高流动性反映了投资者的积极情绪。

传统的投资者情绪测度方法主要依赖于间接的金融市场指标作为代理变量。Lee等（1991）率先使用封闭式基金的折现率作为投资者情绪的代理变量，发现该代理变量有效，基金的风险直接受到投资者情绪影响，从而影响资产组合的价格。Baker和Wurgler（2006）通过构建了包括IPO数量、股权、换手率、首日平均收益率和股息溢价等在内的复合指标来度量投资者情绪，发现投资者情绪像跷跷板一样决定了投资者如何选择投资标的。Huang等（2015）指出了利用主成分分析方法存在的缺陷，并使用偏最小二乘法来估计投资者情绪。Aboody等（2018）则创新性地使用隔夜收益率来衡量投资者情绪，发现投资者情绪对市场短期收益率有着显著的正向作用，同时低隔夜收益率的股票长期表现更好。

随着互联网技术和大数据分析方法的快速发展，基于互联网数据的新测度方法逐渐兴起，主要包括社交媒体数据、金融论坛数据和互联网搜索数据三大类。在社交媒体方面，Zhang（2011）等人通过收集六个月的Twitter信息并分析其情绪倾向，发现情绪化的Twitter与道琼斯、纳斯达克和标普500指数显著负相关。黄润鹏等（2015）利用新浪微博平台的接口分析微博情绪与上证综指的相关性，发现微博情绪指标的加入可以有效提升模型的准确率。在金融论坛数据方面，段江娇等（2017）构建了股票论坛关注度、投资者情绪一致性等指标，结果表明股票收益受到网络论坛关注和情绪影响，而股票交易量受到情绪一致性影响。在互联网搜索数据方面，郑瑶（2016）将投资者情绪与百度指数构建回归模型，证明了搜索指数有助于提高股票市场变量预测的精度。赵妍妍等（2017）则利用点互信息（PMI）衡量对应情感倾向的相关性分数，并借此构建大规模情感词典，提高了情绪测度的准确性。

二、投资者情绪与资本市场表现关系研究

投资者情绪对股票收益的影响是学术界关注的焦点。来自中国市场的研究表明，张强等（2007）发现相比于几乎不造成影响的个人投资者，机构投资者情绪是影响中国股市的系统因素。杨永伟（2018）研究中国权威媒体发布的新闻后发现，当新闻中蕴含强烈的积极情绪时，会使得股价大幅度上涨。国际市场研究方面，Lin等（2018）通过研究投资者情绪与美国股指期货市场及其对应的现货标的价格的关系，发现投资者情绪与期货市场表现呈现明显的负相关关系。Cedric等（2019）发现投资者关注度对投资者情绪有因果关系，并且投资者情绪对于大盘股而言较为短暂，对于小盘股而言情绪效应较为持久。Aboody等（2018）的研究表明投资者情绪对市场短期收益率有着显著的正向作用，而低投资者情绪时期的股票在长期表现更好，这反映了投资者情绪的短期推动与长期反转特征。

投资者情绪也对市场交易量产生显著影响。Baker和Stein（2003）的研究表明，高流动性反映了投资者的积极情绪，在流动性强的市场中，价格主要由非理性的投资者主导。段江娇等（2017）通过对股票论坛数据的分析发现，股票交易量主要受到情绪一致性的影响，而非单纯的情绪水平。李思龙等（2018）采用东方财富股吧论坛帖子考察投资者情绪，发现投资者的互动可以增加企业股东数量并且提高流动性，同时降低信息不对称性。这些研究表明投资者情绪不仅直接影响价格，还通过交易行为影响市场微观结构。

在牛熊市中，投资者情绪表现出显著的差异化影响。项辛怡（2018）的研究显示，投资者情绪对股票收益产生负向影响，并在牛市阶段和熊市阶段各具特点：投资者情绪对股票收益的影响在牛市比熊市更为强烈。这可能是因为牛市中投资者乐观情绪更容易形成羊群效应，导致市场反应更为剧烈。王聪等（2018）的研究结果表明上证指数涨跌幅与投资者情感之间同时存在线性与非线性关系，且日区间联动幅度较大；在构建模型时应综合考虑两者关系和投资者情感的异方差性。此外，网络新闻舆情也表现出对投资者情绪的差异化影响。Yang和Yan（2011）发现新闻舆情导致的高情绪在超过某个临界值后会带来负的超额收益，低于该临界值则超额收益为正。这种非线性关系表明，理解投资者情绪与市场表现的关系需要考虑市场环境和情绪临界点的综合影响。

第二节 自然语言处理技术研究进展

1. 文本情感分析方法演进

自然语言处理（Natural Language Processing, NLP）作为人工智能领域的重要分支，主要研究通过计算机对自然语言进行分析、理解和识别，广泛应用于机器翻译、文本分类等多个场景。在文本情感分析领域，技术方法经历了从基于词典到机器学习再到深度学习的演进过程。

基于词典的传统方法是最早的文本情感分析方法，其核心是通过构建情感词典，对文本进行分词后进行情感匹配和分类。这种方法的主要步骤包括词干提取、分词、词形还原和词性标注等基础处理。然而，基于词典的方法存在局限性，主要体现在需要大量人力维护词典，且难以准确处理网络新词和缩略词。

随着机器学习技术的发展，研究者开始采用更为智能的文本分析方法。这类方法首先需要将文本数据按比例划分为训练集和测试集，通过观察损失函数的变化来优化模型性能。相比词典方法，机器学习方法具有更强的自适应能力，但仍需要大量人工标注的训练数据。

近年来，深度学习方法在文本情感分析领域取得了突破性进展。以卷积神经网络(CNN)、循环神经网络(RNN)、长短期记忆网络(LSTM)和Transformer为代表的深度学习算法，在文本情感分类任务中展现出强大的性能。目前多款专用于深度学习的程序框架，如Google公司的TensorFlow和Facebook公司的PyTorch，这些工具为深度学习研究提供了极大便利。

二、自然语言处理技术在金融领域的应用

金融文本处理具有其特殊性，主要体现在专业性强、词义模糊度高等方面。传统情感分析方法在金融领域的应用面临着专业词汇识别困难、上下文语义理解不准确等挑战。为解决这些问题，研究者开始构建专门的金融领域词典，如CFSD中文金融情感词典，并通过持续更新和扩充来提高分析效果。

在不同模型的应用效果比较方面，实证研究表明深度学习模型表现最为优异。这主要得益于深度学习模型能够自动学习特征，适应金融文本的复杂性。特别是在处理时间序列信息方面，LSTM等模型展现出独特优势。然而，金融领域的文本分析仍需要结合专业知识，单纯依靠算法难以获得理想效果。

第三节 研究评述

近年来,投资者情绪与股市表现的关系研究取得了显著进展。现有研究从理论和实证两个层面较为系统地探讨了投资者情绪的本质内涵、测度方法及其对资本市场的影响机制。在情绪测度方面,研究方法经历了从传统金融指标到互联网大数据的转变,特别是基于社交媒体、金融论坛和搜索数据的新型测度方法极大地丰富了研究手段。在影响机制方面,研究发现投资者情绪不仅直接影响股票收益和交易量,还在不同市场环境下呈现出显著的差异化特征。与此同时,自然语言处理技术的快速发展为投资者情绪研究提供了有力的技术支持,从基于词典的传统方法到深度学习算法的应用,极大地提升了文本情感分析的准确性和效率。

现有研究仍存在若干不足之处：首先，现有的情绪指标在时效性和准确性方面还有待提高，难以完全捕捉市场情绪的快速变化。其次,多数研究关注投资者情绪的整体水平,而对情绪的结构性特征研究不足,如情绪的分布特征、波动特征等缺乏深入探讨。其次,在采用互联网数据测度情绪时,往往忽视了不同信息源的可靠性差异,且未充分考虑信息传播过程中的时滞效应。再次,现有的自然语言处理模型在处理金融专业文本时仍面临着专业词汇识别困难、上下文语义理解不准确等技术瓶颈。此外,研究多集中于探讨投资者情绪对市场的单向影响,对二者之间可能存在的动态反馈机制关注不足。最后，对情绪影响市场的传导机制研究不够深入，缺乏对微观层面的细致分析。

未来研究进行改进和深化的方向：一是构建更加实时和精准的情绪指标体系，充分利用高频数据捕捉市场情绪变化；二是构建多源异构数据融合框架,综合考虑不同信息源的特点和权重，建立更为完善的情绪测度体系；三是针对金融领域特点开发专门的自然语言处理模型,提高文本分析的准确性；四是建立投资者情绪与市场表现的双向互动模型,更好地揭示二者的动态关系。五是加强对市场微观结构的研究，深入分析投资者情绪对个股和不同类型投资者的差异化影响。这些改进将有助于更全面地理解投资者情绪对股市的影响机制。