

人工智能在股票价格预测中的应用研究

陈姝琪

福州外语外贸学院，中国·福建 福州 350200

【摘要】股票价格预测在金融市场中至关重要，对投资者决策和金融机构运营影响深远。传统预测方法在处理复杂市场数据时存在局限，而人工智能技术凭借强大的数据处理和模式识别能力，为股票价格预测带来新契机。本文深入剖析人工智能在股票价格预测中的应用，涵盖多种人工智能模型的运用、多源数据融合处理、实际应用案例分析、面临挑战及未来发展趋势等，旨在为投资者和金融从业者提供参考，推动该领域发展。

【关键词】人工智能；股票价格预测；机器学习；深度学习

引言

作为金融市场核心组成部分的股票市场，其价格波动受宏观经济形势、行业竞争格局、企业财务状况、政策法规调整、投资者心理预期及突发重大事件等众多复杂因素影响，精准预测股票价格走势以助投资者把握投资时机、获取收益且对金融机构优化资产配置、防范风险意义重大；传统股票价格预测方法如基本面分析、技术分析和时间序列分析等虽有一定作用，但在应对股票市场的高度复杂性、非线性和海量数据时预测精度和效率受限；发展迅猛的人工智能技术，以机器学习、深度学习为代表的人工智能算法在处理复杂数据和模式识别上优势显著，为股票价格预测提供新思路和方法并成为研究热点。

一、股票价格预测的复杂性与传统方法的局限

(一) 股票价格波动的影响因素

在宏观经济层面，经济增长、通货膨胀、利率变动、汇率波动等以广泛影响股票价格的方式存在，其中经济增长强劲致企业盈利预期上升，使股价往往上涨，利率上升令企业融资成本增加、股票吸引力下降，致股价可能下跌；在行业层面，行业生命周期、竞争态势、技术创新等因素以影响行业内企业发展进而影响股价的态势存在，新兴行业企业若技术突破或市场份额扩大，股价可能大幅上涨，而传统行业企业若面临激烈竞争或技术替代，股价可能受挫；在企业微观层面，财务状况、经营战略、管理层能力、公司治理等因素以决定企业在价值进而影响股价的情形存在，企业财务报表显示盈利良好、资产质量高时，股价通常有支撑；此外，投资者心理和市场情绪以对股价有重要影响的状态存在，如市场恐慌或过度乐观时，股价

可能偏离基本面。

(二) 传统预测方法的局限性

通过研究宏观经济数据、行业报告、企业财务报表等以评估股票内在价值及判断股价合理性的基本面分析，存在需大量数据收集和分析、对宏观经济和行业发展趋势判断依赖大以及数据滞后性影响预测及时性而难捕捉短期股价波动等情况。基于股票历史价格和成交量数据并运用图表形态、技术指标等预测未来走势的技术分析，因市场变化快、历史规律未必重演，技术分析信号易受噪音干扰、缺乏严格理论基础且预测主观性强。如ARIMA等假设数据有稳定统计特征和趋势并通过分析历史数据预测未来的时间序列分析，由于股票市场受多种复杂因素影响，数据非线性和非平稳性突出，模型对突发事件和结构变化敏感，致使预测效果不理想。

二、人工智能技术在股票价格预测中的应用

(一) 机器学习模型在股票价格预测中的应用

1、线性回归模型

结合股票价格与影响因素间存在线性关系的假设，通过最小化误差平方和确定模型参数的线性回归，其中简单线性回归仅考虑如股价与企业盈利关系这类单一自变量，多元线性回归则综合如宏观经济指标、企业财务指标等多个自变量以预测股价，然而实际股票市场非线性强，致使线性回归模型困难以准确反映复杂关系而应用受限。

2、决策树与随机森林模型

将数据分类或预测问题转化为树形结构决策过程、基于特征值对样本分类且每个内部节点为属性上的测试条件、分支为测试输出、叶节点为类别或预测值但易过拟合的决

策树模型，与通过构建多个决策树、对训练数据有放回抽样、各决策树独立训练且最终预测结果由所有决策树投票或平均得出，能有效降低过拟合风险、提高预测准确性和稳定性，在股票价格预测中可综合考虑多种因素并分析其对股价影响的随机森林模型^[1]。

3、支持向量机模型

支持向量机旨在通过寻找能最大化分类间隔的最优分类超平面来解决分类问题，在股票价格预测中，将股价走势分为上涨、下跌或持平类别以训练模型，对于非线性问题，借助核函数把低维数据映射到高维空间从而找到线性可分超平面，虽在小样本、非线性及高维数据处理上表现良好，但因参数选择和核函数确定对预测结果影响大而需精细调优。

（二）深度学习模型在股票价格预测中的应用

1、人工神经网络模型

结合生物神经网络结构和功能模拟且由大量神经元节点组成、通过神经元间连接权重传递和处理信息的人工神经网络，其中多层感知机作为简单前馈神经网络包含输入层、隐藏层和输出层且层间神经元全连接；在股票价格预测中，输入层接收股价历史数据、成交量、宏观经济指标等信息，隐藏层对信息进行非线性变换，输出层输出预测结果，然而传统人工神经网络存在梯度消失或梯度爆炸致使训练困难的问题。

2、循环神经网络及其变体

结合循环神经网络能处理序列数据且神经元输出可反馈到输入以捕捉数据时间序列信息，在股票价格预测中能有效处理股价随时间变化的序列特征；长短期记忆网络（LSTM）和门控循环单元（GRU）作为 RNN 的改进变体，解决了 RNN 的梯度消失和长期依赖问题，其中 LSTM 通过输入门、遗忘门和输出门控制信息流动来选择性记忆和遗忘历史信息，GRU 简化 LSTM 结构以提高计算效率，它们在股票价格预测中广泛应用以更好捕捉股价长期趋势和短期波动^[2]。

3、卷积神经网络模型

卷积神经网络最初用于图像识别，通过卷积层、池化层和全连接层提取数据特征，在股票价格预测中，将股价数据视为特殊“图像”，让卷积层的卷积核扫描数据以提取局部特征，使池化层对特征降维以减少计算量，最后由全

连接层输出预测结果，其能自动提取数据特征、减少人工特征工程工作量，且对数据平移、旋转等变换有一定不变性，在处理股价数据的时空特征上具备优势。

4、人工智能模型在股票价格预测中的优势

人工智能模型凭借处理海量、高维、复杂数据并自动提取有效特征、挖掘数据间隐藏关系和模式的能力，弥补传统方法不足；机器学习模型基于数据驱动，凭借不依赖严格理论假设且适应性强的特点，通过大量历史数据训练学习股价变化规律以提高预测准确性；深度学习模型凭借具有强大非线性拟合能力的多层神经网络结构，更精准刻画股票市场的复杂非线性关系，有效捕捉股价波动特征，在长期趋势和短期波动预测上表现出色；此外，人工智能模型在训练完成后，凭借快速处理新数据并实时更新预测结果的能力，为投资者提供及时决策支持。

三、多源数据融合与人工智能预测模型的构建

（一）数据来源与数据预处理

股票价格预测需包含股票历史交易数据（开盘价、收盘价、最高价、最低价、成交量等）以反映股价波动和市场交易活跃程度、企业财务数据（营业收入、净利润、资产负债表等）以体现企业经营状况和财务健康、宏观经济数据（GDP、利率、通货膨胀率等）以影响股票市场整体走势、行业数据（行业增长率、市场份额等）以反映行业发展趋势和竞争格局，以及新闻资讯、社交媒体数据等蕴含市场情绪和投资者预期信息的多源数据。数据预处理作为关键步骤，鉴于原始数据可能存在影响模型训练效果的缺失值、异常值和噪声，常用的处理方法包括缺失值处理，如删除含缺失值样本、均值填充、插值法或基于模型预测填充；异常值检测和处理，通过统计方法或机器学习算法识别异常值并采用修正或删除处理；数据标准化，将数据归一化到特定范围如 [0, 1] 或 [-1, 1] 以消除量纲影响并提高模型收敛速度和稳定性。

（二）多源数据融合方法

为充分利用多源数据信息而需的有效融合方法中，特征拼接是将不同数据源特征（如股票交易数据特征、财务数据特征和宏观经济数据特征）按顺序拼接成作为模型输入的新特征向量的简单直观方法，然而该方法未考虑不同数据间关联；基于模型融合的方法是先分别用不同数据源（如分别用股价历史数据训练LSTM模型，用财务数据训练决

策树模型)训练多个模型,再将各模型预测结果通过加权平均、投票等方式融合,此方法虽能发挥不同模型优势但计算成本高;构建统一特征空间这一重要方法则是通过深度学习模型等(如用自编码器将文本形式的新闻数据和数值形式的财务数据)将不同数据源数据映射到同一特征空间以挖掘数据间潜在关系,进而进行融合和预测^[3]。

(三) 人工智能预测模型的构建与训练

构建人工智能预测模型时,需根据数据特点和预测任务,以深度学习模型为例,选择合适模型结构(确定网络层数、每层神经元数量、激活函数类型等)和参数,选择合适损失函数(如均方误差用于回归任务,交叉熵用于分类任务)衡量模型预测结果与真实值差距;模型训练采用训练数据集,通过常用优化算法(随机梯度下降及其变体Adagrad、Adadelta、RMSProp、Adam等)调整模型参数使损失函数最小化,训练过程中,为防止过拟合,采用正则化方法(如L1和L2正则化、Dropout技术等)并通过早停法避免模型在训练集上过拟合;训练完成后,用验证数据集评估模型性能并调整参数优化模型,最后用测试数据集测试模型泛化能力,评估预测准确性和可靠性。

四、人工智能在股票价格预测中面临的挑战与应对策略

(一) 数据质量与数据安全问题

结合股票市场数据量大、来源复杂且数据质量参差不齐,缺失值、错误值和噪声数据影响模型预测准确性,以及金融数据含敏感信息致数据安全和隐私保护至关重要的情况,为提高数据质量,建立严格的数据采集和清洗流程,采用多种数据验证和纠错方法并结合人工审核以确保数据准确性和完整性;在数据安全方面,加强数据加密、访问控制、数据脱敏等技术应用并制定完善的数据安全管理以防止数据泄露和滥用。

(二) 模型的过拟合与欠拟合问题

过拟合时模型呈现出在训练集上表现良好却在测试集与新数据上泛化能力差的情况,欠拟合时模型出现无法充分学习数据特征与规律且预测准确率低的状况;为解决过拟合问题而采用正则化方法、增加训练数据量、调整模型结构和参数等手段,诸如使用L1和L2正则化约束模型复杂度以及运用Dropout技术随机丢弃部分神经元来防止过拟合;

针对欠拟合则通过增加模型复杂度(如增加神经网络层数或神经元数量,或采用更复杂模型结构),同时优化数据预处理和特征工程以提高数据质量和特征有效性^[5]。

五、结论与展望

(一) 研究结论总结

结合人工智能技术为股票价格预测提供有效手段、机器学习和深度学习模型在处理复杂股票市场数据上优势明显且能挖掘传统方法难以发现的规律和模式以提高预测准确性和效率、多源数据融合可综合利用多方面信息来提升预测模型性能,以及应用中面临数据质量、模型过拟合与欠拟合、市场环境变化等挑战且需相应策略应对等内容。

(二) 未来发展趋势展望

展望未来,结合随着人工智能技术发展和金融市场数字化转型,人工智能在股票价格预测领域将有更广阔应用前景,包括新人工智能算法和模型将不断涌现如结合强化学习和深度学习的方法使模型能在动态市场环境中自主学习和优化策略,大数据、云计算、物联网等技术发展将提供更丰富数据资源和强大计算能力以进一步提升预测模型精度和效率,以及跨学科研究将加强,经济学、金融学、计算机科学等领域专家合作深入理解股票市场运行机制与人工智能技术结合点,推动该领域理论和实践发展并为投资者和金融市场提供更精准、智能的股票价格预测服务等方面。

参考文献:

- [1] 刘思含. Shapley-NSGA-III-CatBoost 算法在股票价格预测及交易策略中的应用 [D]. 内蒙古大学, 2023.
- [2] 李辰宇. 基于多源数据融合的股票价格预测模型及应用研究 [D]. 山东财经大学, 2023.
- [3] 张琳. 基于人工智能的中国股票价格预测与异质性研究 [J]. 暨南学报(哲学社会科学版), 2023, 45 (03): 123–132.
- [4] 周家润. 基于改进粒子群优化神经网络对股票价格预测研究 [D]. 长江大学, 2023.
- [5] 王国兰. 人工智能算法在股票价格波动规律预测中的应用 [J]. 长春大学学报, 2022, 32 (02): 20–23+38.

作者简介:

陈姝琪 (2004.11-) , 女, 汉族, 福建永泰人, 本科。