

学校代码 10125

专业代码 081203

山西财经大学

# 硕士学位论文

题目 基于 RWKV 模型的股票预测  
与投资组合方法研究

姓 名 张宸  
专 业 计算机应用技术  
研究方向 大数据技术与应用  
所属学院 信息学院  
指导教师 郝剑龙

二〇二五年三月



University Code 10125

Major Code 081203

Shanxi University of Finance & Economics

# Thesis for Master's Degree

Title Research on Stock Prediction  
and Portfolio Methods Based on RWKV  
Model

Name Chen Zhang

Major Computer Applied Technology

Research Orientation Big Data Technology and Application

School school of information

Tutor Jianlong Hao

March, 2025



山西财经大学

## 学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究所做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日





## 学位论文版权使用授权书

本学位论文作者完全了解学校有关保管、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权山西财经大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于保密□，不保密□。在      年解密后适用本授权书。

（请在以上方框内打“√”）

学位论文作者签名：

指导教师签名：

日期：      年    月    日

日期：      年    月    日





## 摘 要

股票市场因其高流动性和高回报潜力，一直是全球投资者关注的焦点。然而，在追求收益的同时，如何在风险可控的前提下准确预测股票价格趋势，始终是投资者和研究者的重要课题。近年来，通过聚合股票间相关性提升深度学习模型预测性能的研究取得了一定进展，但仍面临诸多挑战，包括股票相关性的动态变化尚未充分建模、行业与个股关系的动态捕捉存在难点，以及模型泛化能力和实际应用价值亟待提高。

针对这些问题，本文提出了一种基于 RWKV (Receptance Weighted Key Value) 的股票预测模型，以系统性提升预测性能与实用性。该模型以股票因子特征为输入，通过多层次聚合机制逐步提炼信息。首先，通过编码器模块与 RWKV 模块并行运行，分别对股票间非线性联动关系和单只股票的内部关系进行建模。编码器模块捕捉不同股票之间的依赖性，而 RWKV 模块利用时间混合与通道混合机制，建模单只股票的趋势及周期性波动等特性。两者的输出结果经过合并处理后，再通过时间注意力机制聚焦历史数据中对股票价格预测具有关键影响的时间点，从而提升时间维度的聚合效率。实验结果表明，该模型在多个数据集中取得最优效果，IC 值和 RankIC 值分别平均高出次优模型约 9.93% 和 10.16%。

其次，本文提出了一种动态估值方法，通过动态评估股票与所属行业的关系，全面分析个股的相对估值水平，从而判断其未来的上涨潜力。该方法评估股票价格与行业平均估值之间的偏离程度，结合行业整体表现以及个股特征，精准识别被低估或高估的股票。通过动态调整估值标准，该方法能灵活响应市场条件变化，显著提升策略的灵活性。实验结果表明，该方法在复杂多变的市场环境中能够捕捉潜在投资机会，并在风险可控的前提下实现稳定收益。

为了验证模型与策略的实际应用价值，本文采用投资策略对股票历史数据进行了回测，模拟真实交易场景，对模型在不同市场条件下的表现进行了全面评估。实验结果显示，模型能够准确捕捉市场变化趋势，展现出较高的稳定性与鲁棒性。结合实际应用场景，模型不仅在量化投资中具有可靠性，还为投资者提供了一种科学实用的工具，对未来研究与实践具有重要参考价值。

**关键词：**股票预测，机器学习，RWKV，动态估值

## ABSTRACT

The stock market, with its high liquidity and potential for high returns, has consistently attracted significant attention from global investors. However, alongside the pursuit of returns, accurately predicting future trends in stock prices under controllable risk conditions has become one of the core challenges for both investors and researchers. In recent years, some researchers have attempted to enhance the predictive performance of deep learning models for stock prices by aggregating inter-stock correlations, achieving certain progress. Nevertheless, this field still faces numerous challenges: first, stock correlations are not only instantaneous but may also undergo dynamic changes over time, and efficiently and accurately capturing such complex correlations remains an unresolved issue. Second, stock price movements are often closely related to the industries they belong to, yet effectively capturing the dynamic changes in the relationships between industries and individual stocks remains a challenge. Lastly, most current research validates models only on specific datasets, lacking long-term testing and validation in real-world application scenarios, thereby leaving the models' generalization capabilities and practical applicability in need of further exploration. To address the above issues, this paper proposes a novel method based on stock correlation aggregation, aiming to systematically enhance the performance and practicality of stock price prediction models. Specifically, the main contributions of this paper include the following:

To address these issues, this paper proposes a stock prediction model based on RWKV (Receptance Weighted Key Value) to systematically improve predictive performance and practicality. The model takes stock factor features as input and refines information through a multi-level aggregation mechanism. Firstly, the encoder module and the RWKV module run in parallel to model the non-linear interconnections between stocks and the internal relationships within individual stocks, respectively. The encoder module captures dependencies among different stocks, while the RWKV module uses time-mixing and channel-mixing mechanisms to model trends and cyclical fluctuations of individual stocks. After the outputs of the two modules are merged, they are processed through a temporal attention mechanism, which focuses on key historical data points critical for stock price prediction, thereby enhancing the efficiency of temporal

aggregation. The experimental results show that the proposed model achieves the best performance across multiple datasets, with the IC and RankIC values being on average approximately 9.93% and 10.16% higher than those of the second-best model, respectively.

Secondly, this paper proposes a dynamic valuation method that comprehensively analyzes the relative valuation level of individual stocks by dynamically evaluating their relationship with the corresponding industry, thereby assessing their potential for future growth. The method monitors the deviation between stock prices and the industry's average valuation, combining industry-wide performance with stock-specific characteristics to accurately identify stocks that may be undervalued or overvalued. By dynamically adjusting valuation benchmarks, the method can promptly respond to changes in market conditions, uncover investment opportunities embedded in market fluctuations, and enhance the flexibility and accuracy of investment strategies. Experimental results demonstrate that this integrated approach not only captures potential investment opportunities in complex and volatile market environments but also achieves stable investment returns under controllable risk conditions.

To validate the practical application value of the model and strategy, this paper conducted backtesting on historical stock data using investment strategies, simulated real-world trading scenarios, and thoroughly evaluated the model's performance under different market conditions. Experimental results demonstrate that the model can accurately capture market trends while exhibiting strong stability and robustness. In practical application contexts, the model not only proves reliable in quantitative investment but also provides investors with a scientific and practical tool, offering significant reference value for future research and real-world implementation.

**Keywords:** stock prediction, machine learning, RWKV, dynamic valuation



## 目 录

学位论文原创性声明 .....	1
学位论文版权使用授权书 .....	1
摘 要 .....	1
<b>ABSTRACT</b> .....	2
第 1 章 绪论 .....	1
1.1 研究背景和意义 .....	1
1.1.1 研究背景 .....	1
1.1.2 研究意义 .....	3
1.2 国内外文献综述 .....	3
1.2.1 股票时间序列预测研究现状 .....	3
1.2.2 时间序列预测相关技术的演进 .....	6
1.2.3 文献评述 .....	9
1.3 研究内容与方法 .....	9
1.3.1 研究内容 .....	9
1.3.2 研究方法 .....	10
1.4 主要工作和创新 .....	11
1.5 论文的基本结构 .....	12
第 2 章 相关理论与方法 .....	13
2.1 时序预测经典模型 .....	13
2.1.1 一维卷积神经网络 .....	13
2.1.2 循环神经网络 .....	13
2.1.3 Transformer 神经网络 .....	16
2.2 RWKV 模型 .....	18
2.3 本章小结 .....	20
第 3 章 基于 RWKV 模型的股票回报率预测方法 .....	21
3.1 股票回报率预测问题及相关定义 .....	21
3.1.1 符号说明 .....	21

3.1.2 问题定义.....	21
3.2 基于 RWKV 模型的设计与构建.....	21
3.2.1 股票间关系聚合.....	22
3.2.2 股票内关系聚合.....	24
3.2.3 跨时间相关性聚合.....	27
3.2.4 预测层及损失函数.....	28
3.3 实验与结果分析.....	29
3.3.1 数据集.....	29
3.3.2 实验超参数与环境设置.....	30
3.3.3 评估指标.....	30
3.3.4 基准模型介绍和对比实验分析.....	31
3.3.5 消融实验分析.....	34
3.3.6 超参数实验分析.....	36
3.3.7 鲁棒性实验分析.....	37
3.4 本章小结.....	39
第 4 章 基于动态估值的 RWKV 投资组合方法.....	40
4.1 股票量化投资任务介绍.....	41
4.2 量化投资策略模型设计.....	41
4.2.1 模型网络整体设计与构建.....	41
4.2.2 基于深度学习的股票动态估值方法.....	42
4.2.3 策略模块.....	43
4.3 实验与分析.....	44
4.3.1 评估指标.....	44
4.3.2 基准模型介绍.....	45
4.3.3 实验及结果分析.....	46
4.4 本章小结.....	48
第 5 章 结论与展望.....	50
5.1 结论.....	50
5.2 展望.....	51
附录.....	52

附录 1 57 个技术因子 .....	52
参考文献 .....	54





# 第 1 章 绪论

## 1.1 研究背景和意义

### 1.1.1 研究背景

全球金融市场自最初的简单交易形式发展至今，已逐步演变为一个复杂多样且高度全球化的体系，经历了无数次的变革与挑战。在这一过程中，股票市场作为金融市场的重要组成部分，始终发挥着至关重要的作用。股票市场规模庞大，股票投资已成为人们实现资产增值和保值的重要手段。然而，随着数据量的急剧增加，人工分析股票时间序列的局限性愈发明显<sup>[1]</sup>，尤其是在数据处理能力、分析速度、情绪影响、信息获取全面性以及认知偏差等方面。这些局限性导致传统分析方法的准确性和一致性较低，且难以快速响应市场波动。因此，如何通过先进的计算机程序有效地分析股票价格变动的内在逻辑，并据此做出理性、及时的投资决策，已经成为金融领域和投资者广泛关注的研究热点。

基于上述背景，许多学者尝试利用传统的计量经济学模型对股票市场进行研究。传统计量经济模型在金融研究中确实起到了推动作用，帮助学者从定量角度分析市场动态。然而，这些传统方法存在一些局限性，它们无法全面捕捉市场中各种因素的复杂互动，尤其是在处理股票市场中多样化和非线性关系时存在明显不足<sup>[2][3]</sup>，导致其在实际应用中的预测能力较为有限<sup>[4]</sup>。随着人工智能技术的飞速发展，尤其是深度学习和神经网络的兴起，学者们逐渐转向更为灵活且具有强大学习能力的计算模型来分析股票市场。相比于传统的计量经济学模型，深度学习方法能够更好地捕捉股票市场中的非线性关系、时序特性以及多维度数据的复杂性<sup>[5]</sup>。例如，循环神经网络（RNN）被广泛应用于捕捉股票价格的长期依赖性，能够有效处理股票市场中存在的时间序列数据，并捕获过去的市场信息对未来股价的影响。此外，Transformer 模型因其自注意力机制，能够同时捕捉全局的股票间关系，强化模型对整个市场数据的理解，特别是在处理多维度、长时间跨度的数据时，展现出了其强大的优势。尽管深度学习在股票预测领域取得了显著进展<sup>[6][7]</sup>，其仍然面临一些挑战，限制了其在实际应用中的广泛推广和落地。例如，RNN

虽然能够捕捉股票价格的长期依赖性，但仍然存在“梯度消失”问题，尤其是在处理长时间序列时，远程信息对当前预测的影响会被逐渐遗忘，从而导致模型在长期依赖性捕捉上的能力受限。Transformer 模型凭借其自注意力机制，虽然能够很好地捕捉全局关系，但其计算复杂度较高，尤其是在处理长序列数据时，模型的参数量会呈指数级增长。这导致了在训练和推理过程中，计算资源需求大幅增加，模型的训练时间和推理时间也会显著延长。

近年来，RWKV 模型<sup>[8]</sup>的出现，巧妙地结合了 RNN 和 Transformer 的优点，成为一种新兴的深度学习网络。RWKV 的设计灵感源于对传统神经网络模型的优化。RWKV 模型的结构与 RNN 类似，能够处理长时间序列数据，捕捉股票价格的长期依赖性。但与传统 RNN 模型不同的是，RWKV 引入了键值对机制和时间混合机制，使得模型能够更加高效地处理长时间跨度的数据，并有效克服了传统 RNN 中常见的“梯度消失”问题，从而保持了对长时间序列的记忆能力。另一方面，RWKV 融合了 Transformer 的自注意力机制，使得模型能够捕捉股票间的全局关系，并高效地识别关键时间点的数据。Transformer 模型在处理长序列时常面临计算复杂度过高的问题，特别是在捕捉长时间依赖和跨时间段的关系时，计算开销和训练难度较大。而 RWKV 通过优化注意力机制，减少了计算量的同时，保留了 Transformer 在处理多维度关系时的优势，使得模型不仅能够高效地捕捉股票价格的非线性波动，还能在较低的计算成本下处理大规模的市场数据。RWKV 通过通道混合和时间混合的设计将 RNN 在处理时间序列数据中的优势与 Transformer 在捕捉全局关系方面的强大能力相结合，旨在弥补这两者在股票市场预测中的局限性。

尽管深度学习模型在股票预测中展现了强大的关系捕捉能力，但这些模型通常侧重于短期内的价格波动和相对时间序列关系，缺乏对股票在整体市场环境中的全面估值判断。这使得模型在面对市场整体波动、行业变化或股票内在价值的波动时，无法提供足够的稳健性和准确性。针对这一问题，本文提出了一种动态估值方法，旨在评估股票与所属行业的关系，判断股票是否被高估或低估，从而为预测提供更加全面的背景信息。通过结合动态估值与深度学习模型的预测能力，能在基于股票动态估值的情况下做出买卖决策，提升模型在复杂市场条件下的稳健性。这一方法使得股票预测不仅更加精准，还能有效地管理风险，优化投资决策。

### 1.1.2 研究意义

将股票动态估值方法融入基于 RWKV 模型的股票预测算法，对于稳健投资具有重要的理论意义和现实意义。

在理论意义上，这一方法创新性地结合了深度学习模型在捕捉股票价格波动中的强大能力与动态估值在市场整体趋势分析中的优势，提供了一种更为全面的股票预测框架。通过引入动态估值，模型不仅能捕捉股票的短期波动，还能在长期趋势上进行理性判断，从而提高了预测的准确性和可靠性。这种综合方法有助于深化对股票市场内在规律的理解，丰富金融领域的理论研究。

在现实意义上，将动态估值与 RWKV 模型结合，有助于提升股票预测的稳健性，特别是在面对金融市场波动和复杂风险环境时。动态估值能够实时评估股票是否被高估或低估，为投资者提供更加科学的决策依据，有助于优化投资组合、降低风险并提高回报。对于金融机构和投资者而言，这一方法为风险管理提供了更加精准的工具，使其在动态变化的市场环境中能够做出更加理性和有效的决策，从而增强了金融市场中的投资稳定性和市场效率。

## 1.2 国内外文献综述

股票时间序列预测是金融研究中的关键领域，近年来取得了显著进展。本节将重点讨论股票相关性建模、资产投资组合管理以及深度学习技术的演变，系统梳理当前的研究进展，分析不同方法的应用特点、优势与挑战。结合最新研究成果，探索市场预测技术的创新应用及未来发展方向，为后续研究提供有价值的参考和启示。

### 1.2.1 股票时间序列预测研究现状

时间序列分析是一种用于研究和预测基于时间变化的数据序列的方法。股票时间序列作为时间序列分析的重要应用领域，具有高度的随机性、非线性和噪声特征。通过对股票时间序列的分析与建模，可以捕捉股票价格、交易量等金融指标的历史规律和潜在模式，从而实现对未来走势的预测与估值。本节将从股票相关性建模方法现状和股票资产投资组合管理研究现状两个方向对股票时间序列预测的研究现状进行综述。

### （1）股票相关性建模方法现状

股票相关性建模方法通过利用神经网络，深入挖掘股票与行业、股票与股票之间以及单个股票内部的多层次相关性。这种方法使用大规模的市场数据，包括价格、交易量、财务指标和新闻等信息，构建复杂的特征向量，进行高维度数据的处理和分析。神经网络能够捕捉非线性关系和复杂的模式，揭示出传统统计模型难以发现的潜在关联。通过这种方法，投资者可以更好地理解股票之间的互动关系、行业的影响力以及个股的内部结构变化，从而提升投资组合的优化、风险管理和市场预测的准确性和动态性。

在股票相关性建模领域，Li 等人<sup>[9]</sup>提出了一个基于张量的事件驱动长短期记忆(LSTM)模型，以解决金融市场中多模式数据的交互和采样时间异质性的问题，实验证明其在中国证券市场上的表现优于现有的先进算法。Yoo 等人<sup>[10]</sup>提出了一种名为多层次上下文的数据轴转换器的新方法，通过端到端学习股票之间的相关性，特别关注处理股票间的非对称和动态性质，以提高股票走势预测的准确性和投资回报率。Li 等人<sup>[11]</sup>提出了一种基于 LSTM 关系图卷积网络模型，通过建模股票之间的关联矩阵来预测隔夜股票在前一个收盘价和开盘价之间的运动，充分利用股票间的联系以优化市场走势预测的准确性。Agrawal 等人<sup>[12]</sup>提出了一个进化深度学习模型，结合股票技术指标和 LSTM，以改进股票市场趋势预测，显著提高了预测准确率，特别是通过分析股票相关性，帮助投资者做出更有利可图的投资决策。Ye 等人<sup>[13]</sup>提出了一个名为 Multi-GCGRU 的深度学习框架，通过结合图卷积网络（GCN）和门控循环单元（GRU）来考虑股票之间的交叉效应，从而改进股票价格走势预测，并在中国股市的实验中优于其他基准模型。宁苡鹤<sup>[14]</sup>提出了一个结合股票相关性的混合神经网络预测模型，在考虑股票之间价格相互影响的基础上，通过实验验证其在股票走势预测中的准确性优于单一的 ANN 模型或 SVM 模型。Han 等人<sup>[15]</sup>提出了一个行业关系驱动的超图关注网络，通过高阶关系和动态更新机制来改进股票价格预测，在标准普尔 500 和 CSI500 数据集上表现出色，显著提高了预测性能和盈利能力。Ma 等人<sup>[16]</sup>提出了多图卷积对抗框架 VGC-GAN，通过生成多个相关图和结合多图卷积网络与门控循环单元，全面描述股票间的相关性并增强预测性能，在多个实际数据集上有效提升了股票价格预测的准确性。Qian 等人<sup>[17]</sup>提出了多关系动态图神经网络框架，通过离散动态图全面捕捉股票市场中多方面的关系及其随时间的演变，利用 Transformer 结构编码多重

关系的时间演化，从而在股票价格预测中展现了优越性。

现有的股票相关性建模方法存在一些缺点。首先，这些方法往往单一地关注股票间的相关性或是股票内的相关性，而忽略了两者的复杂交互作用。其次，许多方法依赖于预定义的静态图神经网络来捕捉股票的相关性，但这种方法在应对动态市场环境时显得力不从心，无法及时反映市场中不断变化的关系和趋势。需要引入更加灵活的股票关系捕捉方法，并根据数据进行实时自适应调整，才能有效应对股票市场的复杂性和多变性。

## （2）股票资产投资组合管理研究现状

股票资产投资组合管理研究现状主要集中在优化投资组合以实现风险与收益的平衡。现有研究应用了多种方法，包括均值-方差模型<sup>[18]</sup>、资本资产定价模型<sup>[19]</sup>、风险价值模型<sup>[20]</sup>以及各类机器学习和深度学习算法。这些方法致力于提高投资决策的准确性，减少市场波动带来的风险。然而，传统模型在处理高维数据和动态市场变化方面存在局限，近年来越来越多的研究转向利用神经网络或图卷积网络等技术，以更好地捕捉股票间的复杂关系和时序特征，从而优化投资组合管理。

在股票资产投资组合管理领域，Ta 等人<sup>[21]</sup>提出了一种基于历史数据的股票走势预测方法，利用 LSTM 网络结合多种投资组合优化技术，包括等权建模、仿真建模和均值变量优化，以提高投资组合的绩效。研究结果显示，该方法不仅在股票预测精度上表现突出，而且在构建的投资组合的收益率和夏普比率方面明显优于传统模型，且超过了标准普尔 500 指数。Sen 等人<sup>[22]</sup>通过分析印度股市九个不同行业的前五大股票的历史价格时间序列，利用 LSTM 模型进行未来股票价格预测，并构建了各行业的最优投资组合，结果显示 LSTM 模型在预测和实际收益上表现出较高的精度。Gondkar 等人<sup>[23]</sup>探讨了利用混合网络和堆叠 LSTM 网络进行股票价格预测的方法，特别关注了基于行业分类的投资组合优化。研究结果显示，混合神经网络在不同情况下都表现出更高的预测能力，对 SBI、印度银行和印度银行的股价预测效果显著。Ma 等人<sup>[24]</sup>通过将收益预测与随机森林、支持向量回归等机器学习模型以及 LSTM 神经网络、深度多层感知器和卷积神经网络等深度学习模型相结合，改进了均值-方差和 omega 投资组合优化模型，并基于中国证券 100 指数的历史数据验证了这些组合模型在投资组合形成中的优越性能。Ma 等人<sup>[25]</sup>提出了一种基于异构图注意网络的多任务学习模型，通过同时预测股票收益和波动风

险来优化投资组合，并在实验证明中显示该方法在股票排序和回溯测试交易评估任务中优于现有最先进的方法。

现存的股票资产投资组合管理方法在捕捉股票之间的多重关系和相互影响方面存在不足，通常仅依赖于单一的关系建模，忽略了多重溢出效应和不同关系类型的权重差异。此外，现有方法主要关注收益预测，而没有综合考虑股票价格的波动风险和估值预估，这导致了投资决策的不全面性和潜在的收益不稳定性。为了克服这些缺陷，需要引入更为全面的关系建模和同时考虑风险的学习方法，以实现更精确和稳健的投资组合优化。

### 1.2.2 时间序列预测相关技术的演进

时序预测模型的演进历程深刻反映了技术架构在金融领域的适应性创新。早期研究中，RNN 及其衍生模型（如 LSTM、GRU）凭借循环结构在捕捉序列时序依赖性上的优势，成为股票预测的主流方法。例如，Shi 等人<sup>[26]</sup>通过集成 GCN 与 LSTM 构建多关系资本流动模型，显著提升了股价预测精度。Zhang 等人<sup>[27]</sup>设计的注意力增强型 Bi-GRU 模型，在比亚迪股价指数预测中表现出超越传统 RNN 变体的稳定性；而黄丽明等人<sup>[28]</sup>提出的多路循环神经网络，则通过融合新闻文本特征与交易数据实现了高精度涨跌分类。然而，RNN 系列模型受限于梯度消失、爆炸问题，在长序列建模中难以平衡计算效率与长期依赖捕捉能力。

为突破 RNN 的瓶颈，Transformer 模型<sup>[29]</sup>凭借自注意力机制开启了时序预测的新范式。其多头注意力架构可并行处理全局时间步关联，有效捕捉市场动态中的非线性特征与复杂模式。研究实践中，Tao 等人<sup>[30]</sup>提出的序列分解 Transformer 通过周期相关机制增强历史模式挖掘，显著提升了股价预测泛化性；Zhang 等人<sup>[31]</sup>构建的 TEANet 框架融合社交媒体文本与股价数据，验证了多模态注意力机制对交易策略收益的优化作用；陈诗乐等人<sup>[32]</sup>提出了一种基于遗传算法进行特征选择和 Transformer 模型进行时序预测的 GA-Transformer 混合模型，用于多因子股票价格预测，实验结果表明其在六支股票数据集上的预测精度优于主流模型。

近年来，RWKV 模型<sup>[8]</sup>作为融合 RNN 与 Transformer 优势的新架构，为时序预测提供了更高效的解决方案。该模型通过线性注意力机制替代传统键值乘积，结合 RNN 式递归解码实现长序列的高效处理。其版本迭代持续优化性能：RWKV-4<sup>[8]</sup>引入通道定向乘积与时间依赖 Softmax，通过残差块结构平衡信息传递稳定性；

RWKV-5<sup>[33]</sup>新增矩阵值注意力状态与 SiLU 门控机制, 结合 LayerNorm 与参数初始化优化, 强化了复杂特征表达能力; RWKV-6<sup>[33]</sup>则通过数据依赖性衰减机制与令牌动态转移策略, 进一步提升对市场时序变化的适应性。Li 等人<sup>[34]</sup>首次系统综述了 RWKV 模型, 全面梳理了其架构原理、版本演进、多领域应用, 并对比分析了其与传统 Transformer 及改进模型的优劣, 同时探讨了当前挑战与未来研究方向。

RWKV 模型凭借其高效的计算特性和强大的序列建模能力, 近年来在多个方向得到了广泛应用。研究人员不断优化其架构和训练方法, 以提升模型性能、计算效率和资源利用率, 推动其在更多复杂场景中的实践与发展。

在通用领域, Dumitru 等人<sup>[35]</sup>提出了一种在 RWKV 架构中引入多时间视角的方法, 通过维持多个时间视图增强语言模型对复杂序列数据的处理和解释能力, 在仅增加少量参数的情况下显著提升模型性能。Kang<sup>[36]</sup>提出 DiSHA 框架, 利用维度分片适应技术改进大型语言模型的参数高效微调 (PEFT), 实现更快的收敛速度和更高的计算效率, 并通过 Bone 和 Bat 方法进一步优化性能与资源利用率。Knunyants 等人<sup>[37]</sup>提出了一种无需训练的激活稀疏化方法, 在循环大语言模型 (R-LLMs) 的线性层前动态设置阈值以消除低幅值激活, 从而提升神经形态处理器的能效, 同时保持模型性能, 并验证了其在自注意力模型上的泛化能力。Nardone 等人<sup>[38]</sup>提出的 SNAP 混合神经网络架构, 通过在芯片间通信的带宽受限区域部署脉冲神经网络, 减少数据传输开销, 同时保留芯片内人工神经网络的高效计算能力, 在语言处理和视觉任务中实现最高 5.3 倍的能效提升和 15.2 倍的延迟降低。Yueyu 等人<sup>[39]</sup>提出基于 RWKV-7 注意力的 ARWKV 模型, 采用三阶段知识蒸馏方法将 Transformer 架构的 Qwen 模型转化为具有更强状态跟踪能力的 RNN 架构。Choe 等人<sup>[40]</sup>针对 RNN-based 大型语言模型 RWKV 提出一套压缩技术, 结合低秩近似、稀疏预测等架构优化, 以及嵌入缓存、分层权重分解等训练后压缩技术, 在保证模型准确性的同时, 将内存占用减少 3.4-5 倍, 相比同精度 Transformer 模型降低 4 倍, 使其能够高效部署于可穿戴设备等边缘计算平台。

在序列预测和强化学习领域, Xie 等人<sup>[41]</sup>提出基于 RWKV 线性注意力机制的方法, 结合三次径向基函数插值与 BYOL 对比学习, 实现 PANI-CeO<sub>2</sub>氨气传感器在复杂温湿度环境下的非线性补偿, 降低预测误差。Hou 等人<sup>[42]</sup>提出的 RWKV-TS 模型采用线性时间复杂度架构设计, 并增强长时序信息捕捉能力, 在时间序列预

测、分类、异常检测等六大任务中取得与 Transformer 和 CNN 等先进模型相当的性能，同时显著降低计算延迟和内存消耗，重新验证了 RNN 在时间序列任务中的潜力。Dong 等人<sup>[43]</sup>提出 Decision-RWKV 模型，结合决策变换器与经验回放机制，在机器人操作任务中实现高效序列决策和持续学习能力，并通过 D4RL 数据库与 D'Claw 平台验证了其在单任务测试与终身学习场景下的优势。Hao 等人<sup>[44]</sup>提出结合多尺度 RWKV 模型和二维时间卷积网络的混合方法，利用快速傅里叶变换识别光伏数据的周期性特征，并通过改进的时间混合块与多尺度依赖关系建模，实现短期光伏发电功率的高精度预测。Cao 等人<sup>[45]</sup>提出的 MATCC 股票预测方法，通过显式提取市场趋势指导个股分析，并分解股票数据为趋势与波动成分，结合跨时间相关性建模，显著提升预测稳健性与准确性。Xie 等人<sup>[46]</sup>提出的 AutoGMM-RWKV 方案，结合自编码器、高斯混合模型和 K-means，利用 RWKV 线性注意力机制分析无线传感器网络的时间序列数据，并通过自适应阈值与聚类方法，有效检测恶劣环境下的选择性转发攻击。Thapa 等人<sup>[47]</sup>提出基于 Mamba 和 RWKV 的高效序列模型 MAM 与 MARWKV，用于解决多智能体强化学习中的奖励分配问题，在保持与 Transformer (MAT) 相当性能的同时，显著提升训练与推理效率，并通过多环境实验验证其有效性。Dong 等人<sup>[48]</sup>提出 DFT 双分支框架，将股票数据分解为趋势与波动成分，并分别建模其时间相关性和股票间动态关联，在多个指标上实现 300%-400% 的性能提升。Fu 等人<sup>[49]</sup>提出结合 RWKV 线性注意力机制和数据-导频辅助的信道估计方法，以改善 V2X 通信中的双重选择性衰落信道估计性能，在降低计算复杂度的同时显著提升准确度。王庆勇等人<sup>[50]</sup>开发 ChemRB 生成模型，引入双向分子环约束机制，并结合环级特征预测和全局跨度闭合预测两项预训练任务，提高 SMILES 表示中的分子环生成有效性，在 EGFR 抑制剂设计中验证了其生成高质量药物分子的能力。Lang 等人<sup>[51]</sup>提出基于 RWKV 序列模型的多模态动态脑图表示学习方法，融合结构与功能连接数据，在 ABIDE II 和 ADNI 数据集上实现高精度脑疾病诊断，并识别出与疾病相关的关键脑区。

纵观模型演进路径，从 RNN 的时序归纳偏置到 Transformer 的全局注意力突破，再到 RWKV 的架构融合创新，技术发展始终围绕“依赖关系建模精度”与“计算效率”的双重目标推进。RWKV 系列通过持续迭代，为高噪声、长序列的金融预测场景提供了兼顾效率与性能的新范式，也为未来智能金融决策系统的构建奠定了技术基石。



### 1.2.3 文献评述

本节对股票时间序列预测研究现状和时间序列预测相关技术的演进两个方面的研究现状进行梳理，并做出以下总结：

第一，通过股票时间序列预测研究现状可知，近年来，股票市场的预测问题越来越受到深度学习领域的关注。现有的研究大多集中在利用时间序列数据通过模型捕捉股价的趋势和波动模式，典型的模型包括深度学习的 LSTM、GRU 和 Transformer 等。尽管这些方法在一定程度上能够揭示股票市场的规律，但它们在面对复杂的股市动态关系、市场噪声和外部因素干扰时仍然表现出较大的局限性。尤其是现有深度学习模型往往假设市场的固定结构，而实际市场关系却是高度非线性和动态变化的。因此，如何动态捕捉股票市场中随时间变化的复杂关系和动态性，仍然是当前研究中亟待解决的关键问题之一，本文在此基础上探讨了如何提升模型动态捕捉股票关系的性能。

第二，通过分析时间序列预测相关技术的演进可知，时序预测模型的架构不断创新和迭代，以解决传统方法无法有效捕捉长期依赖关系的问题。RWKV 模型作为一种新型的序列建模方法，结合了 Transformer 和 RNN 的优势，在处理长序列依赖、建模时间动态性方面展现了巨大的潜力。RWKV 模型利用具有长记忆能力的设计，能够有效捕捉股市中长期的复杂非线性关系，相比传统的 RNN 和 LSTM，具备更高的效率和性能。然而，尽管 RWKV 在一些领域取得了显著进展，其在股票市场数据建模中的应用仍处于探索阶段，特别是如何结合股票市场的多维度数据进行有效建模，以提高预测准确性，仍然是研究中的挑战。因此，如何在股票关系建模中充分利用 RWKV 模型的优势，并探索其在动态市场环境中的适应性，是本文研究的关键方向。

## 1.3 研究内容与方法

### 1.3.1 研究内容

本文探讨了股票时序预测问题，介绍了 RNN 和 Transformer 在处理时间序列中的优缺点，并指出 RWKV 模型旨在克服这两者的局限性，同时保持其优势。通过分析现有股票预测领域的不足，本文提出了一种基于 RWKV 的股票时间序列预测方法，并在此基础上进行了进一步的扩展和完善。

该方法采用了一种多模块协同的深度学习框架，首先通过 Transformer 编码器对股票之间的非线性联动关系进行建模，捕捉市场中不同股票相互影响的复杂特征，提取出全局性的跨股票信息。与此同时，RWKV 模块专注于单只股票内部的时间序列特性，凭借其高效的记忆机制和动态适应能力，捕捉个股在长时间跨度内的价格波动规律及其动态变化模式。接着，将编码器生成的股票间关系特征与 RWKV 提取的股票内动态特征进行融合，形成一种综合性的多维度特征表示，这种合并不仅保留了市场整体的联动效应，还充分考虑了个股的独特时间特性。最后，引入时间注意力机制对融合后的特征进行进一步优化，通过对历史数据中关键时间点的学习，强化模型对短期波动和长期趋势的敏感性，从而显著提升预测精度。整个流程通过股票间关系聚合、股票内关系建模以及时间注意力机制的结合，实现了对金融市场复杂动态特性的全面捕捉与精准分析。

在此基础上，本文进一步提出了一个基于动态估值的 RWKV 投资组合方法。该方法首先使用 RWKV 模型对股票回报率进行预测，与前述的回报率预测方法相同；然后，将每只股票的估值与其所属行业的平均估值进行对比，从而判断股票是否被高估或低估，评估其相对价值；最后，结合回报率预测和动态估值，选择合适的股票进行买卖操作，确保在追求回报最大化的同时，尽量控制风险。

这一方法不仅继承了基于 RWKV 的股票回报率预测的优势，同时通过引入动态估值，进一步提升了投资决策的稳健性和准确性，为股票预测与投资组合管理提供了更加全面和科学的指导。

### 1.3.2 研究方法

本文通过使用 Qstock 等 Python 库获取全面的股票相关数据，这些数据涵盖了不同市场和多个时间段，确保了研究的数据来源广泛且多样。为了构建一个准确且高效的股票回报率预测模型，首先，对现有文献进行了深入分析，汲取了许多有价值的理论成果和方法。在此基础上，本文基于最新的深度学习理论，设计并开发了一个新的股票回报率预测模型。

本文的研究将通过多个不同市场的股票数据集，进行广泛且严格的实验验证。为了全面评估模型的预测能力和稳定性，本文将提出的模型与现有的深度学习基准模型进行对比，确保能够在各种市场环境下表现出较好的适应性和可靠性。实验结果将通过多个评价指标进行量化，以确保模型在实际应用中的高性能和可操

作性。

此外，本文还将股票估值方法与深度学习模型相结合，旨在提升模型的风险管理能力。结合动态估值方法，模型不仅可以提高预测的准确性，还能够增强其在动态市场中的稳健性，确保在面对不同市场条件时具备更高的稳定性。这种结合方法不仅能在理论上完善模型，还能在实际市场中提供更有效的应用。

为了进一步验证模型的实际效果，本文将在 Qlib 等量化回测平台上进行详细的回测分析。通过模拟真实市场环境，本文将全面评估模型的有效性和稳定性，确保其在实际投资决策中能够发挥重要作用。综上所述，本研究的目标是为投资者和金融分析师提供更加科学和可靠的决策支持，提升投资策略的精准性和风险防控能力。

## 1.4 主要工作和创新

在国内外相关研究成果的基础上，本文确定了研究的重点方向：结合动态估值的 RWKV 股票预测方法研究。本文利用 RWKV 模型建模股票之间的非线性关系，并通过其独特的时间序列处理能力捕捉股票的长期依赖性和动态变化。同时，结合动态估值方法对股票进行实时估值判断，从而在合理风险控制的前提下做出更为精准的买卖决策。通过在多个股票市场的数据集上进行广泛实验，本文验证了所提出方法的合理性和有效性，实验结果表明该方法在不同市场环境下具有较强的适应性和稳健性，能够为投资者提供有效的决策支持。

(1) 本文提出了一种基于 RWKV 的股票预测模型，旨在捕捉股票之间的相关性。本文通过 Transformer 编码器中的注意力机制，有效聚合股票间的关系，捕捉不同股票之间的依赖性和相互影响。其次，利用改进后的 RWKV 模型中的时间混合和通道混合技术，对股票内部的关系进行细致的聚合。本文提出的方法旨在全面整合股票间和股票内的关系，从而提升股票预测的准确性和可靠性。

(2) 提出了一种动态估值方法，专注于通过评估股票与行业之间的关系，深入挖掘个股的估值水平，进而预测其未来的上涨空间。该方法通过分析股票价格与行业平均估值的差异，结合行业整体走势及个股特性，精确定位可能被市场低估或高估的目标。通过该模块的应用，投资者可以更加科学地判断个股的投资价值，优化投资组合的配置，减少风险，提高整体投资收益率。

(3) 将动态估值方法与 RWKV 股票预测模型相组合, 并结合改进后的投资策略应用于投资组合优化。该方法通过结合股票回报率预测和实时估值判断, 优化了股票选择的过程, 并能动态调整投资组合, 以适应不断变化的市场条件。通过这种灵活的策略, 模型能够在追求最大化回报的同时, 有效控制风险, 从而增强投资决策的稳健性和有效性。

## 1.5 论文的基本结构

本文共由五个章节构成, 包括引言、相关理论与方法、基于 RWKV 的股票回报率预测方法、基于动态估值的 RWKV 投资组合方法, 结论与展望。具体如下:

第一章引言部分介绍了股票市场的重要性, 阐述了股票预测问题的背景与意义, 并深入分析了当前股票预测研究的重点领域。通过回顾国内外相关文献, 本章详细探讨了股票市场预测研究的现状与发展趋势, 为后续章节奠定了理论基础。

第二章探讨了股票预测方法所采用的深度学习技术。包括: 时序预测经典模型和 RWKV 模型。

第三章提出了一种基于 RWKV 模型的股票回报率预测方法, 其包含股票间关系聚合模块、股票内关系聚合模块、跨时间关系聚合模块。此外, 本章还详细介绍了对比实验部分, 实验中使用了多个数据集和基准模型。实验结果表明, 基于 RWKV 的股票回报率预测方法在多个数据集上均表现出优越的预测性能。

第四章介绍了基于动态估值的 RWKV 投资组合方法, 提出了一种结合动态估值的投资策略, 通过实时评估股票相对于行业的估值状态, 进一步优化投资决策。该方法通过结合 RWKV 模型的预测能力与动态估值机制, 旨在最大化回报的同时有效控制风险, 提供更加科学的投资组合管理方案。

第五章对本文提出的投资组合方法进行了总结与展望, 本章回顾了模型的优点和实际效果, 并讨论了未来研究的方向和潜在的改进空间。展望了该方法在实际金融市场中的应用前景, 以及如何进一步提升模型的稳定性、准确性和实时适应能力。

## 第 2 章 相关理论与方法

本章主要介绍时间序列预测的深度学习方法相关理论。具体内容包括：一维卷积神经网络、循环神经网络、Transformer 神经网络、以及 RWKV 模型。通过对这些方法的详细介绍，为后续章节中的模型设计和实验提供理论支持。

### 2.1 时序预测经典模型

#### 2.1.1 一维卷积神经网络

一维卷积神经网络（1D CNN）是一种专门用于处理一维数据序列的深度学习模型，广泛应用于时间序列分析、语音处理、自然语言处理等领域。1D CNN 的核心机制是通过卷积层利用滤波器（或卷积核）对输入序列进行局部感知，从而提取数据中的局部特征。具体而言，卷积操作通过滑动卷积核沿输入数据进行计算，生成特征图，并捕捉输入数据中的局部模式和时间依赖性。

1D CNN 的工作流程通常包括卷积操作、激活函数、池化层等几个步骤。卷积层通过滤波器提取输入序列中的低级特征，激活函数引入非线性因素，使网络能够捕捉到更复杂的模式。池化层则用于减少特征图的尺寸，降低计算复杂度，同时避免过拟合问题。通过多层卷积层和池化层的堆叠，1D CNN 能够逐层提取更加抽象和高级的特征，提升对输入数据的理解和表示能力。

由于 1D CNN 具备较强的特征提取能力，并且采用参数共享机制，大大减少了需要训练的参数数量，因此其在处理大规模数据时表现出较高的计算效率和较强的泛化能力。1D CNN 尤其适合于需要提取局部特征和捕捉时间依赖关系的任务，如音频信号分析、文本分类和时间序列预测等应用。

#### 2.1.2 循环神经网络

RNN 由 Elman 等人提出<sup>[52]</sup>，是一种适用于处理序列数据的深度学习模型，特别擅长捕捉时间序列中的长期依赖关系，具体结构如图 2.1 所示。与传统的前馈神经网络不同，RNN 通过在网络的隐层中引入循环连接，使得模型能够在处理当前输入时，结合先前的状态信息，从而记忆过去的信息。这种结构使得 RNN 能够有效

地处理时间序列数据。RNN 的公式如下：

$$s_t = Ux_t + Ws_{t-1} + b \quad (2.1)$$

$$o_t = \Theta(s_t) \quad (2.2)$$

其中  $x_t$  是  $t$  时刻的输入矩阵， $U$  是输入层到隐藏层的权重矩阵， $s_t$  是  $t$  时刻隐藏层向量， $W$  是权重矩阵， $b$  是偏置， $\Theta$  是激活函数， $o_t$  是  $t$  时刻的输出向量。

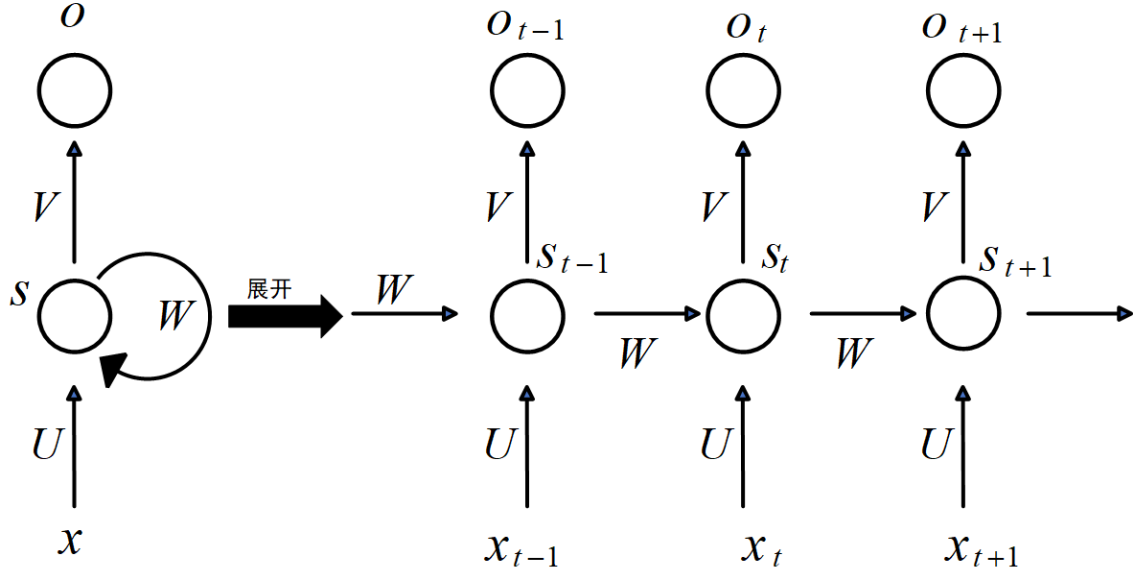


图 2.1 RNN 神经网络结构

传统 RNN 在处理长序列数据时面临梯度消失和爆炸问题，导致难以捕捉长时间依赖关系，因此引出了 LSTM，其通过引入记忆单元和门控机制有效解决了这些问题，能够更好地处理和记忆长期依赖信息。

Hochreiter 等人<sup>[53]</sup>在 RNN 的基础上提出了 LSTM，旨在解决传统 RNN 在处理长时间依赖关系时所面临的梯度消失问题。LSTM 引入了独特的结构设计，特别是其记忆单元和三个门控机制：遗忘门、输入门和输出门。记忆单元用于存储长期依赖信息，而遗忘门则通过 sigmoid 激活函数决定哪些信息应该被遗忘，避免无关信息的积累；输入门控制哪些新的信息可以被写入到记忆单元中，确保新的有用信息能够被保留；输出门则决定记忆单元中哪些信息应该被输出并传递到下一个时刻的隐层状态。通过这三种门的合作，LSTM 能够在每个时间步有选择性地保留、更新或遗忘信息，从而保持长时间序列中的有效记忆，使得网络在学习时能够更加灵活、有效地捕捉复杂的时序模式和长期依赖。

LSTM 的这一结构使其在处理长期依赖关系方面表现出色，广泛应用于诸如自然语言处理、语音识别、机器翻译、时间序列预测等领域。尽管 LSTM 在解决

长期依赖问题上取得了显著成效，但它也有一些不足之处，主要体现在计算复杂度和效率上。由于 LSTM 模型在每个时间步需要进行多个门的计算，其计算开销较大，尤其是在处理长序列数据时，训练和推理过程可能会非常缓慢。此外，LSTM 尽管能够缓解梯度消失问题，但它仍然可能面临梯度爆炸的风险，特别是在训练深层网络时，可能会导致模型的训练不稳定，无法有效地收敛。因此，尽管 LSTM 在许多任务中表现优异，它的高计算开销和潜在的不稳定性仍然限制了其在一些实时应用和大规模数据集上的应用。

Chung 等人<sup>[54]</sup>在 LSTM 的基础上提出了 GRU，旨在简化 LSTM 的结构并提高计算效率，GRU 结构如图 2.2 所示。GRU 通过引入两个门——更新门和重置门来控制信息的流动。更新门决定了当前时刻的隐藏状态应该保留多少来自前一刻的记忆，而重置门则决定了当前输入与过去记忆之间的关联程度。与 LSTM 不同，GRU 将输入门和遗忘门合并为一个更新门，从而减少了模型的参数数量，简化了计算过程。GRU 在许多任务中表现出与 LSTM 相当的效果，尤其在需要较少计算资源或较快训练速度的情况下，GRU 比 LSTM 更具优势。

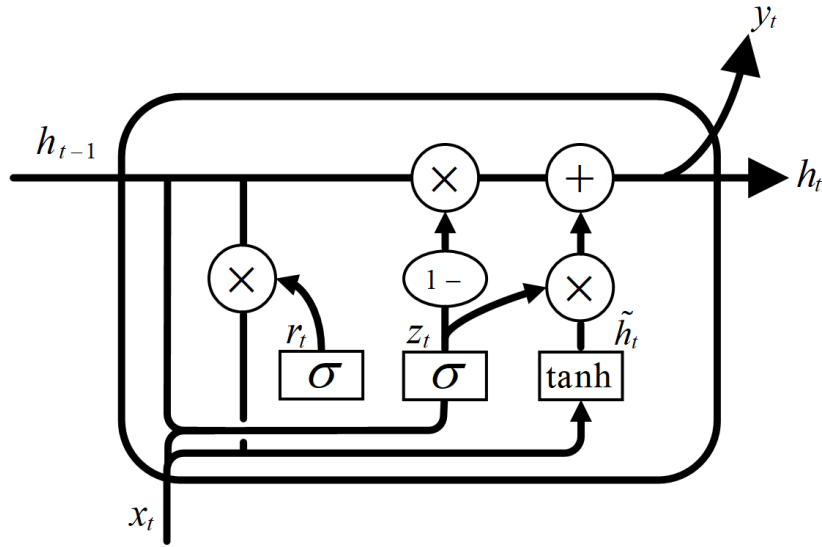


图 2.2 GRU 结构图

由于其结构较为简洁，GRU 不仅能够有效捕捉长时间依赖关系，还能够较短的时间内完成训练，适用于实时任务和大规模数据集。GRU 对于输入序列中的每个元素，每一层都进行如下计算：

$$r_t = \sigma(W_{ir}x_t + b_{ir} + W_{hr}h_{(t-1)} + b_{hr}) \quad (2.3)$$

$$z_t = \sigma(W_{iz}x_t + b_{iz} + W_{hz}h_{(t-1)} + b_{hz}) \quad (2.4)$$

$$\tilde{h}_t = \tanh(W_{in}x_t + b_{in} + r_t \odot (W_{hn}h_{(t-1)} + b_{hn})) \quad (2.5)$$

$$h_t = (1 - z_t) \odot n_t + z_t \odot h_{(t-1)} \quad (2.6)$$

其中  $h_t$  是  $t$  时刻的隐藏状态,  $x_t$  是  $t$  时刻的输入,  $h_{(t-1)}$  是  $t-1$  时刻的隐藏状态或初始隐藏状态。 $r_t$ ,  $z_t$  和  $\tilde{h}_t$  分别是重置门, 更新门和候选隐藏状态。 $\sigma$  是 Sigmoid 激活函数,  $\odot$  是 Hadamard 积。

### 2.1.3 Transformer 神经网络

Transformer 是一种深度学习模型, 由 Vaswani 等人<sup>[29]</sup>提出, 用于解决序列到序列的任务。其最重要的部分是注意力机制 (Attention Mechanism)。该机制是一种模仿人类视觉注意力机制的计算方法, 旨在通过动态聚焦于输入数据中最相关的部分, 提高模型对关键信息的处理能力。其基本思想是, 在处理输入数据时, 模型并不是平等地对待所有信息, 而是根据当前任务的需要, 给予不同部分不同的权重或关注度。该机制的核心结构包括三个主要部分: 查询 (Query)、键 (Key) 和值 (Value)。查询用于确定模型当前关注的信息, 键用于描述输入数据的特征, 而值则包含实际的信息内容。通过计算查询与键之间的相似度, 可以得到每个值的权重, 从而聚焦于最相关的信息。

注意力机制的主要功能是动态选择和加权输入的不同部分, 使模型能够在处理序列数据时更加灵活和高效, 特别是在捕捉长距离依赖关系时。它能够显著提高模型的表现, 尤其在机器翻译、语音识别、图像描述生成等任务中, 已成为不可或缺的技术。与 RNN 不同, Transformer 完全依赖于自注意力机制来捕捉输入序列中各个部分之间的关系, 而不需要像 RNN 那样通过逐步处理序列。它的核心公式为:

$$Attention(Q, K, V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.7)$$

其中  $Q$ ,  $K$ ,  $V$  分别代表查询, 键和值。 $\sqrt{d_k}$  表示缩放系数。

Transformer 的核心结构有两个主要部分, 分别是编码器和解码器, 其简化结构如图 2.3 所示。编码器由多个相同的层堆叠而成, 每个编码器层包含两个主要部分: 自注意力机制和前馈神经网络。自注意力机制的作用是通过计算输入序列中每个元素与其他元素之间的关系, 动态地为每个元素分配一个权重, 使得模型能够关注到与当前输入最相关的部分。编码器的每一层都对输入进行处理, 生成一



个新的表示，该表示保留了输入序列中的关键信息和上下文依赖关系。为了增强模型的顺序感知能力，编码器还引入了位置编码，将位置信息添加到输入中，使得模型能够理解元素在序列中的相对位置。每个编码器层输出的结果将作为下一个编码器层的输入，直到最终的编码器层生成整个序列的隐层表示。

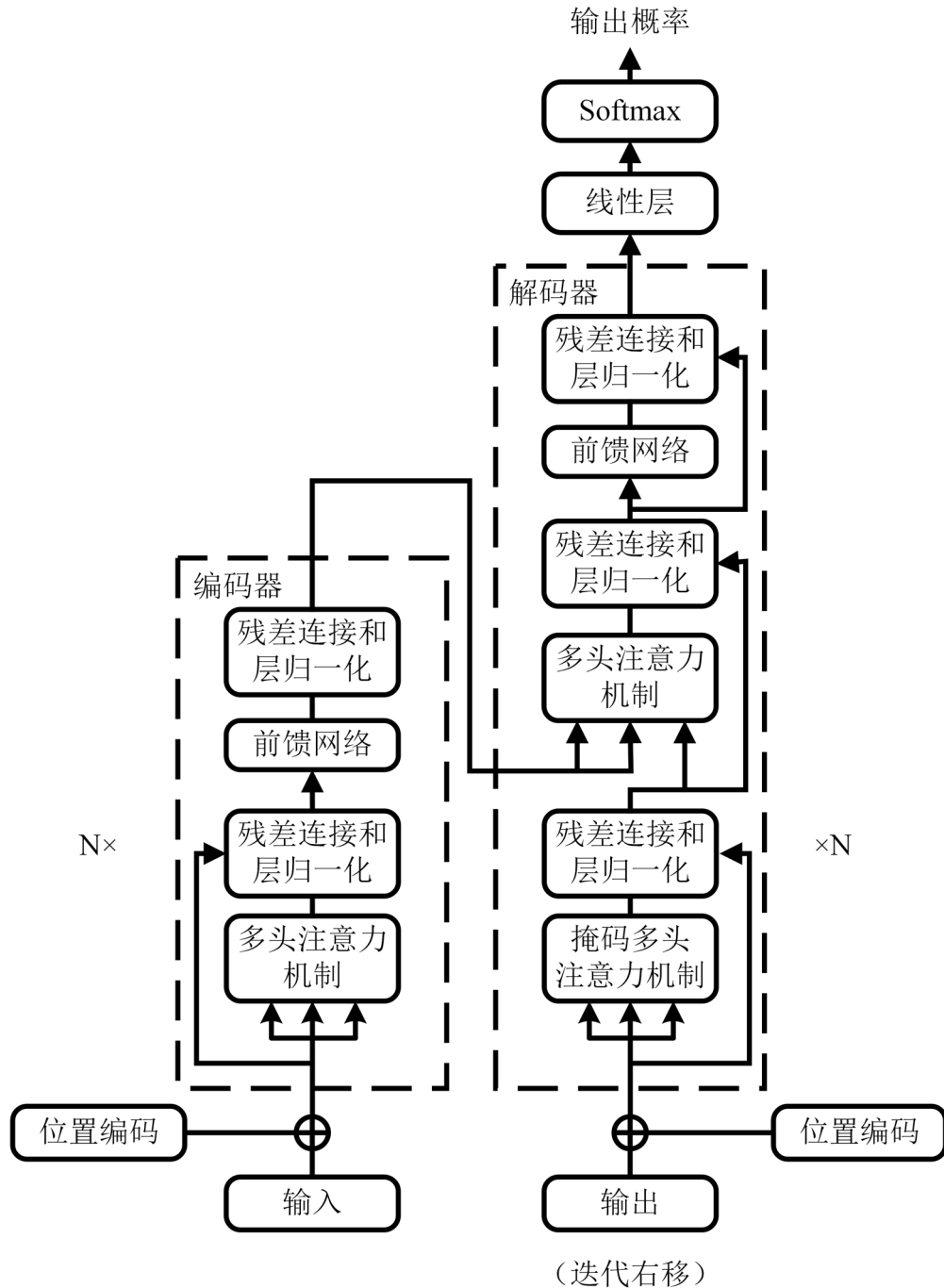


图 2.3 Transformer 结构图

解码器与编码器结构相似，也是由多个相同的层堆叠而成，但解码器在每一层中除了包含自注意力机制和前馈神经网络外，还引入了一个额外的编码器-解码器

注意力层。解码器的自注意力机制确保模型能够根据已生成的部分（即目标序列的前面部分）进行上下文建模，而编码器-解码器注意力层则允许解码器通过与编码器的输出进行交互，聚焦于输入序列中相关的部分。解码器的输出经过最终的线性变换和 Softmax 操作，生成下一个词的预测。通过这种结构，解码器能够根据编码器提供的上下文信息逐步生成目标序列中的每个元素，处理长序列时展现出较强的并行计算能力和信息传递效率。

Transformer 的功能在于通过自注意力机制动态地为输入序列中的每个元素分配权重，关注序列中最相关的部分，从而高效捕捉长距离依赖关系。由于 Transformer 并行计算的特性，相比于 RNN，其在处理长序列时具有更高的计算效率和更好的扩展性。

## 2.2 RWKV 模型

RWKV 模型由 Pneg 等人<sup>[8]</sup>提出,旨在结合传统 RNN 和 Transformer 的优点,并克服它们的局限性,具体结构如图 2.4 所示。

RWKV 模型中, R 是接收向量的缩写, 指充当过去信息的接收者, W 是位置权重衰减向量的缩写, 指模型内的可训练参数, K 是键向量的缩写, 其作用类似于传统注意力机制中的键, V 是价值向量, 其功能与传统注意力过程中的 V 类似。

与 RNN 相比, RWKV 模型能够更好地捕捉长时间跨度的依赖关系, 并通过其独特的时间混合机制有效地解决了 RNN 在处理长序列时面临的梯度消失问题。与 Transformer 相比, RWKV 模型的计算开销显著更低, 因为它避免了复杂的自注意力机制, 而是采用了基于键值对的加权机制来优化信息的传递, 减少了多头注意力和全连接层的计算负担。

RWKV 模型的结构设计包括时间混合机制和通道混合机制, 允许模型在序列的不同时间步中有效地传递信息, 保持对关键模式的关注。通过这种设计, RWKV 模型能够在大规模数据上实现快速训练和推理, 尤其在处理长时间序列数据时, 具有较高的计算效率和较低的内存消耗。该模型在多个任务中表现出了优异的性能, 特别是在自然语言处理、时间序列预测等需要长距离依赖建模的应用中, RWKV 模型展现了与传统模型相比的优势, 并为解决实际问题提供了更加高效和可扩展的解决方案。

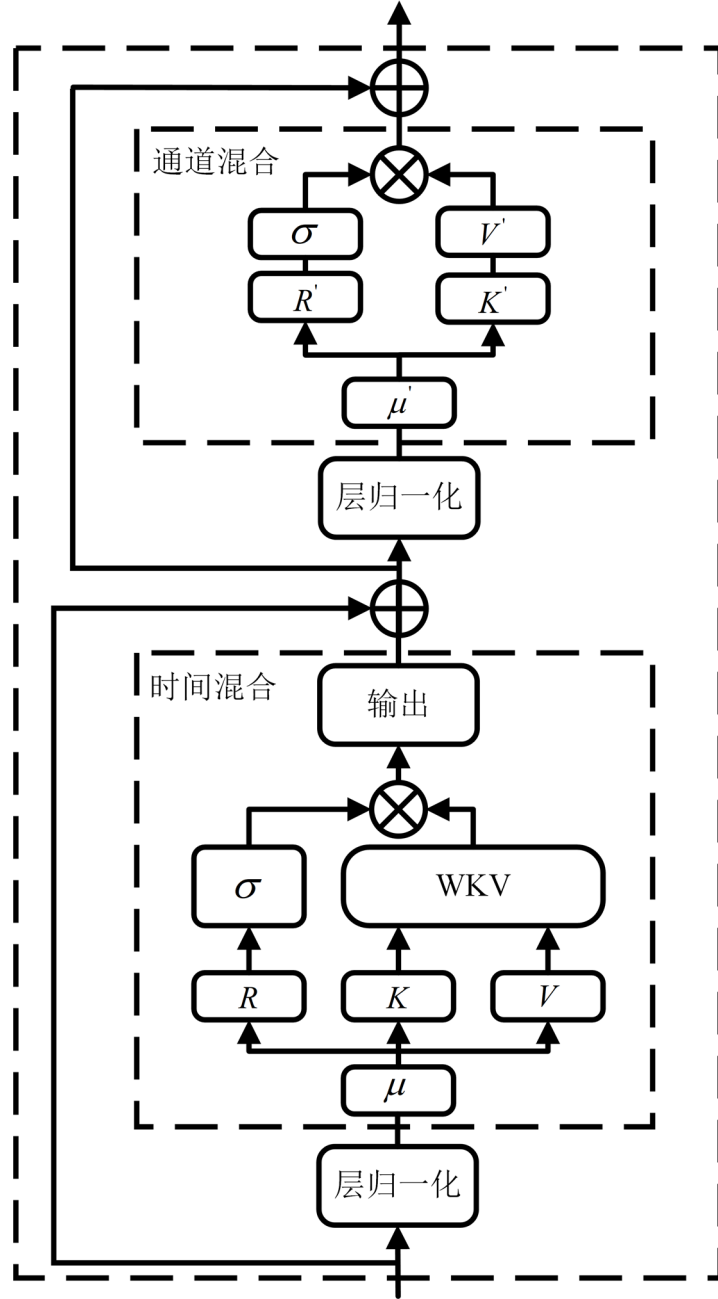


图 2.4 RWKV 块结构图

RWKV 的核心公式如下：

$$r_t = W_r \cdot (\mu_r \odot x_t + (1 - \mu_r) \odot x_{t-1}) \quad (2.8)$$

$$k_t = W_k \cdot (\mu_k \odot x_t + (1 - \mu_k) \odot x_{t-1}) \quad (2.9)$$

$$v_t = W_v \cdot (\mu_v \odot x_t + (1 - \mu_v) \odot x_{t-1}) \quad (2.10)$$

$$wkv_t = \frac{\sum_{i=1}^{t-1} e^{-(t-1-i)w+k_i} \odot v_i + e^{u+k_t} \odot v_t}{\sum_{i=1}^{t-1} e^{-(t-1-i)w+k_i} + e^{u+k_t}} \quad (2.11)$$

其中， $r_t$ ， $k_t$  和  $v_t$  分别代表  $t$  时刻的门控因子，键和值。 $\mu_o$  表示更新因子，用于控

制信息在当前时间步的更新或混合程度。 $wkv_i$ 用于模拟注意力机制的效果。具体来说， $wkv_i$ 通过将过去的键值对按照相对位置调整的权重 $w$ 进行指数衰减加权，并结合当前时刻的键 $k_i$ 和值 $v_i$ 来计算一个新的加权平均值。这一机制使得 RWKV 能够在保持线性时间和内存复杂度的同时，有效地处理长序列依赖。

## 2.3 本章小结

本章详细介绍了研究所需的相关理论和方法，为后续的研究工作奠定了理论基础和技术框架。首先，在时序预测经典模型部分，探讨了几种广泛应用于时间序列预测的深度学习模型，包括 1D CNN、RNN 以及 Transformer 神经网络。这些模型分别代表了从局部特征提取到全局依赖捕捉的不同方式，为时序数据的处理提供了多样化的技术路径。具体而言，1D CNN 通过卷积核在数据中的滑动窗口机制，有效捕捉短期依赖关系；循环神经网络利用其递归结构，能够处理具有时序性质的序列数据；Transformer 模型则通过自注意力机制，能够在不依赖递归结构的情况下捕捉序列中的长距离依赖，显著提高了训练效率，并在多个领域中取得了突破性进展。

接下来，本章重点介绍了本研究的核心模型——RWKV 模型。RWKV 模型的创新之处在于其结合了 RNN 在时序处理中的优势与 Transformer 模型在并行计算和长依赖捕捉方面的强大能力。该模型通过引入窗口化递归计算的机制，能够高效地处理大规模时序数据，并提取深层次的时间依赖特征。此外，RWKV 还具有较低的计算开销，能够在处理大规模数据集时保持高效性能。其独特的网络结构使得模型在处理时序数据时，能够同时保持较强的表达能力和计算效率，为实际应用中的实时预测和动态分析提供了有力支持。RWKV 模型的这种设计理念，特别适用于需要捕捉复杂动态模式和依赖关系的任务，如金融市场预测和资源调度等领域。

通过以上内容，本章为论文提供了股票时序预测的理论支持，同时也为后续的研究设计和实验分析指明了方向。

## 第3章 基于 RWKV 模型的股票回报率预测方法

### 3.1 股票回报率预测问题及相关定义

#### 3.1.1 符号说明

为便于后续描述，本章及以下章节定义了相关符号。本文将斜体英文字母表示为标量，如  $x$  和  $D$ ；将加粗斜体小写英文字母表示为向量，如  $\mathbf{x}$ 。将加粗斜体大写英文字母表示为矩阵，如  $\mathbf{X}$ 。将希腊字母表示为超参数，如  $\lambda$ 。特定的，我们将  $\mathcal{S}$  表示为股票的集合。

#### 3.1.2 问题定义

对于股票集中的每一支股票  $u \in \mathcal{S}$ ，收集其在每一个时间步  $\tau \in [1, L]$  上的指标，形成特征向量  $\mathbf{x}_{u,\tau} \in \mathbb{R}^F$ ， $F$  是特征维度。本文重点关注于预测股票收益率，因为它能同时体现股票价格变化的幅度和趋势。我们将股票收益率定义为  $d$  天内的股票收盘价变动率，即  $y_u = (c_{u,\tau+d} - c_{u,\tau+1}) / c_{u,\tau+1}$ ，其中  $c_{u,\tau+1}$  表示股票  $u$  在  $\tau+1$  日的收盘价。本文将股票预测任务正式描述为：给定股票特征  $\{\mathbf{x}_{u,\tau}\}_{u \in \mathcal{S}, \tau \in [1, L]}$ ，函数  $f(\mathbf{X}; \Theta)$  将股票特征映射为预测收益率  $\{\hat{y}_u\}_{u \in \mathcal{S}}$ ，其中  $\mathbf{X}$  是股票特征矩阵， $\Theta$  是模型参数。

### 3.2 基于 RWKV 模型的设计与构建

本章提出了一种基于 RWKV 的预测模型，简称为 Stock-RWKV。该模型通过三个主要的聚合过程来处理股票数据并进行预测。首先，Stock-RWKV 通过股票间聚合模块来捕捉不同股票间的相关性，以便于识别多个股票之间的潜在依赖关系。同时，模型通过股票内聚合模块对单只股票的内部特征进行处理，聚焦于该股票的历史价格等信息，捕捉其自身的趋势和波动规律。最后，Stock-RWKV 通过跨时间相关性聚合模块处理时间序列中的长期依赖性和动态变化，确保模型能够充分考虑股票价格随时间变化的复杂模式。通过这三个层次的聚合，Stock-RWKV 能够综合考虑股票间、股票内及时间维度的信息，提升预测的准确性和鲁棒性，

其整体框架如图 3.1 所示。

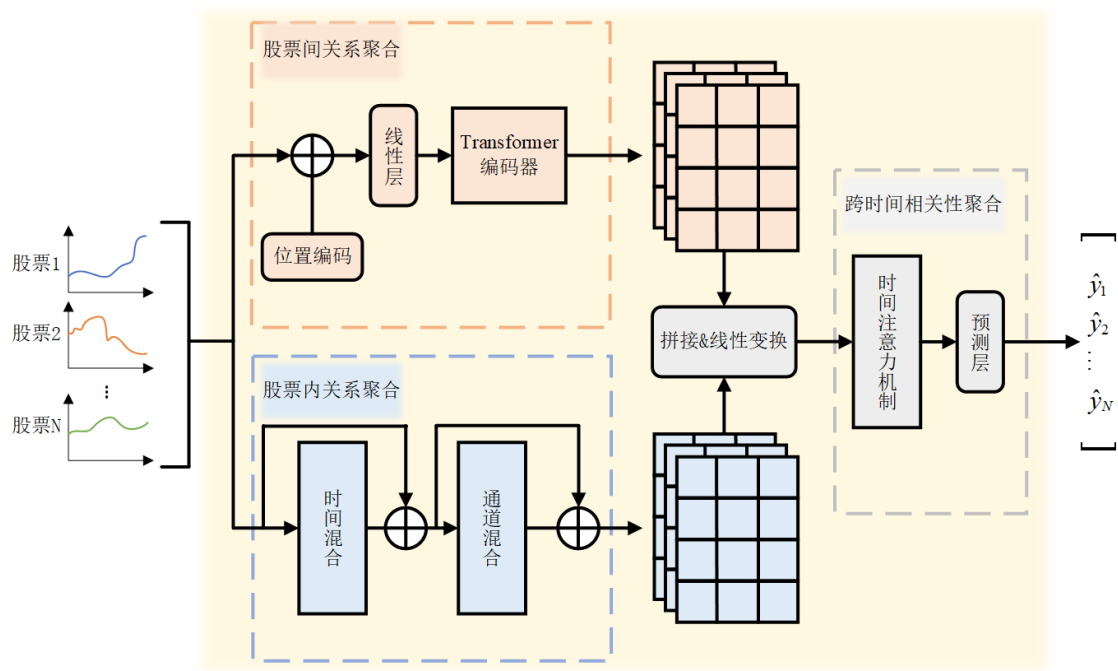


图 3.1 Stock-RWKV 的整体框架图

3.2.1 股票间关系聚合



图 3.2 2024 年 1-6 月古井贡酒与泸州老窖收盘价走势图

股票间相关性指的是不同股票之间的关系，它衡量的是两只或多只股票的价格变动是否倾向于一起上升或下降。这种相关性反映了不同公司股票价格在一定时期内可能会受到相似的经济因素或市场情绪的影响，例如行业政策变化、宏观

经济环境波动或投资者对特定行业的整体预期。

如图 3.2 所示，图中展示了古井贡酒（股票代码：000586.SZ）和泸州老窖（股票代码：000568.SZ）在 2024 年 1 月 1 日至 2024 年 6 月 30 日期间的股票价格走势。可以明显看出，这两只股票的价格走势高度相关，呈现出相似的波动模式，说明它们可能受到相似的市场因素或行业动态的共同影响。

在股票价格预测中，建模股票间的动态相关性至关重要，但现有方法多依赖于预定义的静态关系等，难以捕捉实时和跨时间的复杂相关性。这些方法往往忽略了股票相关性的瞬时性和非对齐特性，导致模型无法充分反映市场的真实动态。

为此，我们引入了基于注意力机制的动态相关性建模方法，能够在每个时间步上计算股票之间的瞬时相关性，并有效捕捉跨时间的复杂相关性。具体来说，我们通过使用 Transformer 的编码器来挖掘不对称和动态的股票间相关性，并通过多头注意力机制进行信息聚合，生成包含股票间瞬时相关性的时空嵌入。多头注意力机制的引入允许模型在不同的特征空间中并行捕捉股票间的多样化关系，从而提升了对复杂市场结构的建模能力。与传统的静态关系建模方法相比，该方法能够自适应地调整股票间的权重分配，从而更准确地刻画出动态市场环境下的潜在模式与变化趋势。这一设计不仅提高了模型的表达能力，还能够更好地适应市场的实时变化。

为便于描述，本文首先以单个股票信息的视角进行描述。我们将有  $F$  个特征，时间窗口为  $L$  的股票  $\mathbf{X}_u \in \mathbb{R}^{L \times F}$  加上  $F$  维度的位置编码  $\mathbf{p}$ ，再通过线性层将股票信息的维度转换为模型的维度  $D$  得到  $\tilde{\mathbf{X}}_u \in \mathbb{R}^{L \times D}$ 。

$$\tilde{\mathbf{X}}_u = \mathbf{W}_x(\mathbf{X}_u + \mathbf{p}) + \mathbf{b}_x \quad (3.1)$$

其中  $\mathbf{W}_x$  是可学习的矩阵， $\mathbf{b}_x$  是可学习的偏置项。

我们将  $\tilde{\mathbf{X}}_u$  进行层归一化操作  $LN(\cdot)$  得到  $\mathbf{H}_u$ 。然后，本文考虑汇总相关股票的信息。具体而言，在每一个时间步中，我们收集所有股票的局部嵌入，并应用带有  $N_1$  个头的多头注意力机制来处理这些数据。

$$\mathbf{H} = \parallel_{u \in \mathcal{S}, \tau \in L} LN(\tilde{\mathbf{X}}_u) \quad (3.2)$$

$$\mathbf{Q}^1 = \mathbf{W}_Q^1 \mathbf{H}, \quad \mathbf{K}^1 = \mathbf{W}_K^1 \mathbf{H}, \quad \mathbf{V}^1 = \mathbf{W}_V^1 \mathbf{H}, \quad (3.3)$$

$$\mathbf{Y} = FFN(MHA(\mathbf{Q}^1, \mathbf{K}^1, \mathbf{V}^1) + \mathbf{H}) \quad (3.4)$$

其中， $\parallel$  代表连接， $\mathbf{H} \in \mathbb{R}^{N \times L \times D}$  指  $N$  支股票在所有时间窗口  $L$  内的信息， $\mathbf{W}_Q^1$ ， $\mathbf{W}_K^1$  和

$W_l^i$  是可学习的参数矩阵， $MHA$  指多头注意力机制， $FFN$  是具有 ReLU 激活函数和残差连接的两层多层感知机（MLP）。 $Y$  是股票间关系聚合模块的输出。

### 3.2.2 股票内关系聚合

股票内相关性是指同一支股票在不同特征之间的相互关系，这些特征可以包括价格、成交量、财务指标、技术指标等。股票内相关性体现为多个维度中不同特征的关联性。例如，价格与成交量的关系可以反映市场对特定股票的供需变化，财务指标与技术指标的关系可以揭示股票基本面与市场表现之间的相互影响。此外，不同特征之间的相互作用也可能反映出股票在特定市场环境下的行为模式与风险特征。对股票内相关性的深入研究，有助于更全面地理解股票的内在规律，并为股票预测、风险控制和投资策略优化提供重要支持。

以中兴通讯（000063.SZ）为例，如图 3.3 所示，在 2024 年 10 月 20 日至 2024 年 12 月 20 日这段时间内，股票的收盘价与成交量之间表现出高度的相关性。这意味着，股价的波动与成交量的变化密切相关，通常在价格上涨时，成交量也会出现相应的增加，反之亦然。通过捕捉这些内在的相关性，投资者可以更好地理解市场行为，并在股票预测和投资决策中提高准确性。了解股票内相关性的变化趋势，有助于发现潜在的交易信号，优化交易策略。

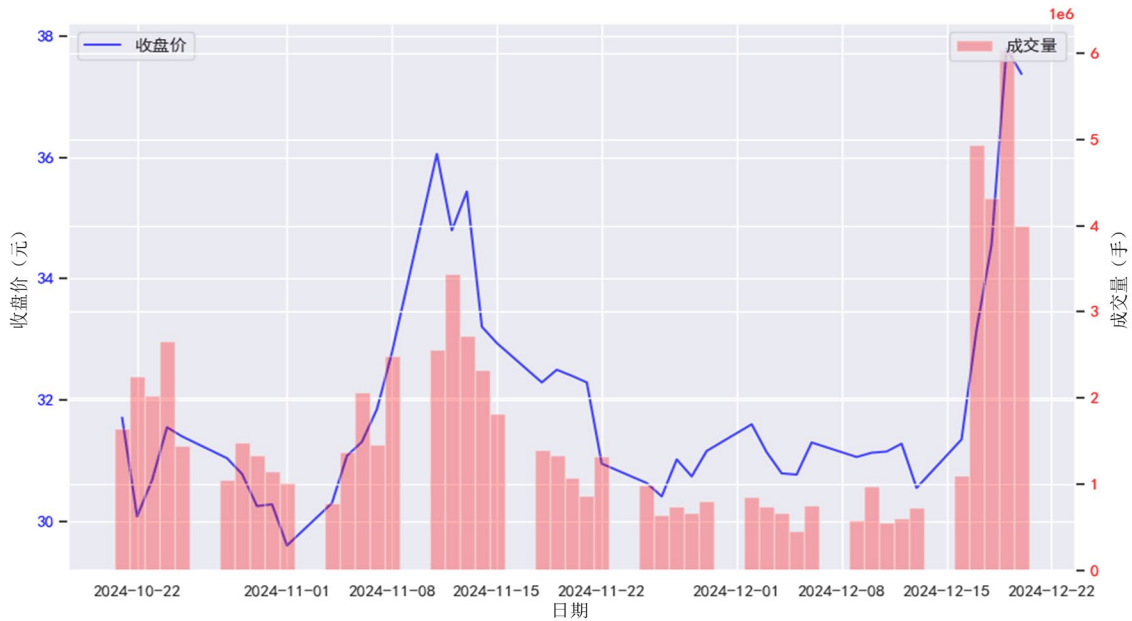


图 3.3 2024 年 10-12 月中兴通讯收盘价和成交量走势图

股票内相关性建模在股票预测中具有重要意义，因为它涉及到单只股票的历史数据、趋势变化及其内部波动的深度理解。然而，目前的相关建模方法仍面临一



些挑战。首先，传统方法通常采用静态特征提取，无法捕捉股票价格波动中的非线性关系和复杂的长期依赖；其次，大多数现有模型在平衡短期和长期依赖时存在困难，难以有效应对快速变化的市场条件。

为了解决这些问题，本文提出一种改进的 RWKV 来进行股票内相关性建模。RWKV 结合了 RNN 和 Transformer 的优势，通过时间混合和通道混合机制，有效捕捉股票内部的复杂依赖关系。该模块能够自适应地加权和聚合股票的内部特征，解决了传统方法在处理非线性关系和长期依赖时的局限性，显著提升了预测的准确性和效率。具体而言，本文将原先的时间混合模块的 WKV 计算部分设计成多头 WKV 计算，以增强时间混合模块的建模能力。同时，在通道混合模块增加了 ReLU 激活函数，使通道混合模块更好捕捉股票内复杂模式。通过这一方法，不仅增强了模型对股票内在动态变化的理解，还提高了整体预测性能，使得在多变的市场环境中，模型能够更加灵活和稳健地进行预测，具体结构如图 3.4 所示。

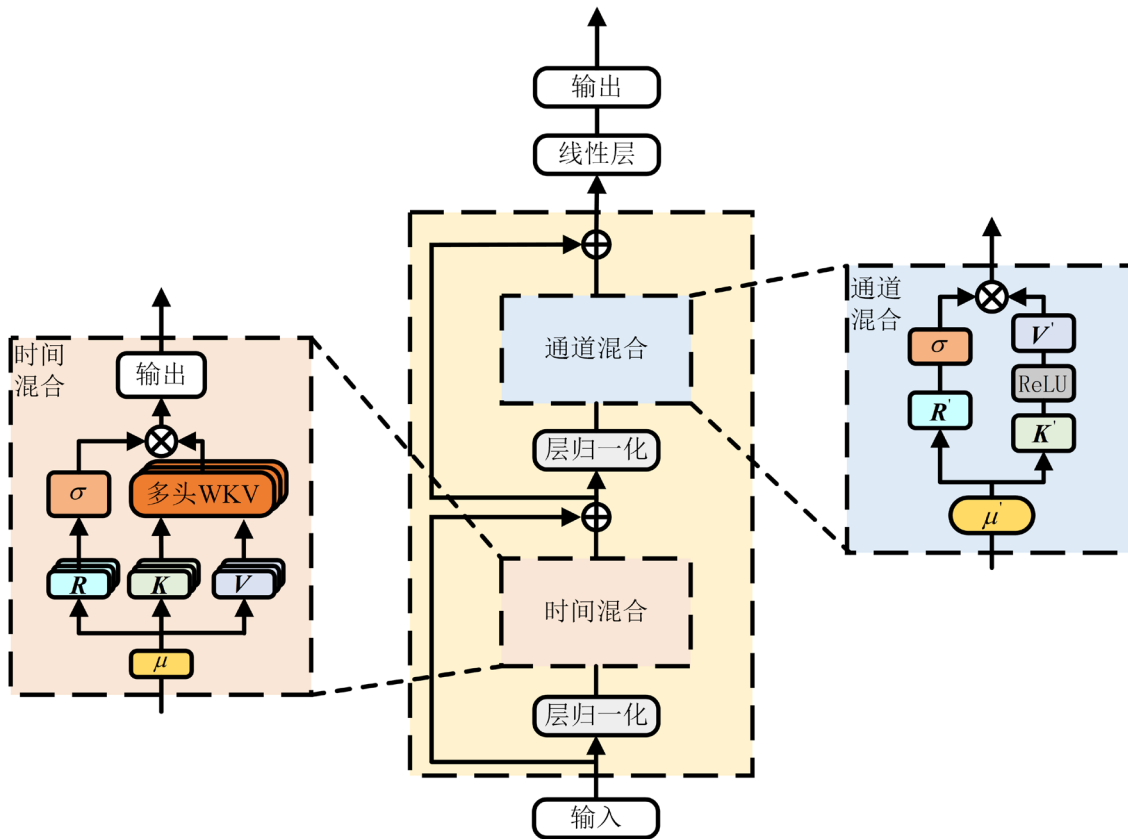


图 3.4 改进后的 RWKV 模型图

在这个模块中，计算中涉及的所有线性投影向量（时间混合中的  $R$ 、 $K$ 、 $V$ ，以及通道混合中的  $R'$ 、 $K'$ ）都是通过当前和先前时间步输入之间的线性插值生成的，从而实现信息的逐步传递与融合。用于时间混合计算的向量是 RWKV 块的当

前和先前输入的线性组合的线性投影:

$$\mathbf{r}_t = \mathbf{W}_r \cdot (\mu_r \odot \mathbf{z}_\tau + (1 - \mu_r) \odot \mathbf{z}_{\tau-1}), \quad (3.5)$$

$$\mathbf{k}_t = \mathbf{W}_k \cdot (\mu_k \odot \mathbf{z}_\tau + (1 - \mu_k) \odot \mathbf{z}_{\tau-1}), \quad (3.6)$$

$$\mathbf{v}_t = \mathbf{W}_v \cdot (\mu_v \odot \mathbf{z}_\tau + (1 - \mu_v) \odot \mathbf{z}_{\tau-1}), \quad (3.7)$$

其中,  $\mu_r$ ,  $\mu_k$  和  $\mu_v$  都是可训练变量,  $\mathbf{W}_r$ ,  $\mathbf{W}_k$  和  $\mathbf{W}_v$  为可训练矩阵,  $\mathbf{z}_\tau \in \mathbf{Z}$  表示时间步  $\tau$  上的股票隐藏特征。

在 RWKV 中, WKV 计算是自注意力机制的近似, 本文单头的 WKV 公式如下<sup>[33]</sup>:

$$wkv_\tau = \text{diag}(u) \cdot \mathbf{k}_\tau^L \cdot \mathbf{v}_\tau + \sum_{i=1}^{\tau-1} \text{diag}(w)^{\tau-1-i} \cdot \mathbf{k}_i^L \cdot \mathbf{v}_i \quad (3.8)$$

其中,  $\text{diag}(\cdot)$  表示创建一个对角矩阵,  $u$  和  $w$  是两个可训练的参数, 参数  $u$  是奖励模型第一次遇到令牌 (特别是当前令牌) 的奖励。这有助于模型更多地关注当前令牌并避免  $w$  的任何潜在退化。  $w$  是每个头的通道时间衰减向量。此外, 我们将参数  $w$  变换如下:

$$w = \exp(-\exp(w)) \quad (3.9)$$

此变换确保  $w$  的所有值都在 (0,1) 范围内, 从而确保  $\text{diag}(w)$  表示收缩矩阵。

本文将多头 WKV 正式定义为:

$$\text{multihead } wkv_\tau = \text{Concat}(wkv_\tau^1, \dots, wkv_\tau^{N_2}) \quad (3.10)$$

其中  $N_2$  是多头 WKV 操作头的数量。

多头 WKV 计算后的输出由下式给出:

$$\mathbf{o}_\tau = \mathbf{W}_o \cdot (\sigma(\mathbf{r}_t) \odot wkv_\tau) \quad (3.11)$$

其中,  $\mathbf{W}_o$  为可学习的参数矩阵。

在股票信息通过时间混合模块后进行相加并通过层归一化处理:

$$\mathbf{z}'_\tau = \text{LN}(\mathbf{o}_\tau + \mathbf{z}_\tau) \quad (3.12)$$

$\mathbf{z}'_\tau$  成为通道混合模块的输入, 和时间混合模块一样, 在通道混合模块中首先进行令牌移位:

$$\mathbf{r}'_\tau = \mathbf{W}'_r \cdot (\mu'_r \odot \mathbf{z}'_\tau + (1 - \mu'_r) \odot \mathbf{z}'_{\tau-1}), \quad (3.13)$$

$$\mathbf{k}'_\tau = \mathbf{W}'_k \cdot (\mu'_k \odot \mathbf{z}'_\tau + (1 - \mu'_k) \odot \mathbf{z}'_{\tau-1}), \quad (3.14)$$

其中,  $\mathbf{W}'_r$  和  $\mathbf{W}'_k$  是可学习的参数矩阵,  $\mu'_r$  和  $\mu'_k$  都是可训练变量。在求得  $\mathbf{v}'_\tau$  的

过程中，与 RWKV 原始过程有所不同：

$$\mathbf{v}'_t = \mathbf{W}'_v \cdot (\text{ReLU}(\mathbf{k}'_t) + b'_v) \quad (3.15)$$

其中， $\mathbf{W}'_v$  是可学习的参数矩阵， $b'_v$  是可学习的偏置项。

最后，通道混合和改进后整体 RWKV 的输出如下：

$$\mathbf{o}'_t = \sigma(r'_t) \odot (\mathbf{W}'_o \cdot \max(\mathbf{k}'_t, 0)^2), \quad (3.16)$$

$$\mathbf{j}_t = \mathbf{W}_j \mathbf{o}'_t + b_j \quad (3.17)$$

$$\mathbf{J} = \parallel_{u \in \mathcal{S}, \tau \in L} \mathbf{j}_{u, \tau} \quad (3.18)$$

其中， $\max(\mathbf{k}'_t, 0)^2$  表示的是 ReLU 激活函数，它将所有负值设为零，保留正值不变。之后，这个结果再被平方，即每个元素都被其自身相乘<sup>[56]</sup>。最后通过一个线性变化得到 RWKV 的输出，其中  $\mathbf{W}_j$  是可学习的参数矩阵， $b_j$  是可学习的偏置项， $\mathbf{J}$  是股票内相关性聚合模块的输出。

### 3.2.3 跨时间相关性聚合

股票的跨时间相关性指的是同一支股票或相同市场指数在不同时期之间的关系。这种相关性可以体现为同一股票在不同天、周、月或年间的价格或收益率的关联性，也可以表现为不同时间段内市场整体趋势的一致性与延续性。捕捉跨时间相关性至关重要，因为它揭示了市场或个股在不同时间段内的行为模式和趋势延续性，帮助投资者识别周期性和季节性特征，优化交易策略，提高预测准确性，并有效管理风险。



图 3.5 2024 年 9-12 月百邦科技蜡烛图走势

如图 3.5 所示，百邦科技（300736.SZ）在 2024 年 9 月 12 日至 2024 年 12 月 20 日期间，经历了三次相似的市场模式波动，这些波动具有相似的趋势和特征，值得进一步分析和研究。

为了解决跨时间步长之间的关联问题，我们采用了时间注意力机制。这种机制能够有效地捕捉不同时间点之间的依赖关系，从而增强模型在处理时间序列数据时的表现。在此之前，我们将股票间聚合模块的输出与股票内聚合模块的输出拼接在一起，并通过线性变换对拼接后的特征进行融合，以生成一个综合性的表示。过程如下：

$$\tilde{\mathbf{M}} = \text{concat}(\mathbf{Y}, \mathbf{J}) \quad (3.19)$$

$$\mathbf{M} = \mathbf{W}_m \tilde{\mathbf{M}} + b_m \quad (3.20)$$

其中， $\text{concat}(\cdot)$  函数将  $\mathbf{Y}$  和  $\mathbf{J}$  按最后一个维度拼接在一起。 $\mathbf{W}_m$  是可训练的参数矩阵， $b_m$  是可学习的偏置项。这一过程确保了股票间的相互影响和个股内部动态特性得以充分整合。随后，通过引入时间注意力机制，模型可以基于这一综合表示自适应地聚焦于不同时间段的关键特征，进一步提升对时序模式的理解和预测能力。因此，我们使用注意机制将隐藏表示聚合为：

$$\mathbf{a}_{u,\tau} = \frac{\exp(\mathbf{m}_{u,\tau}^\top \mathbf{W}_a \mathbf{m}_{u,\tau})}{\sum_{\tau \in [1,L]} \exp(\mathbf{m}_{u,\tau}^\top \mathbf{W}_a \mathbf{m}_{u,\tau})} \quad (3.21)$$

$$\mathbf{e}_u = \sum_{\tau \in [1,L]} \mathbf{a}_{u,\tau} \mathbf{m}_{u,\tau} \quad (3.22)$$

其中， $\mathbf{m}_{u,\tau} \in \mathbf{M}^{N \times L \times D}$ ， $\mathbf{W}_a$  是可训练的参数矩阵， $\mathbf{a}_{u,\tau}$  是隐藏空间中的注意力分数， $\mathbf{e}_u$  是股票的综合嵌入。

### 3.2.4 预测层及损失函数

我们用一个全连接层作为预测层：

$$\tilde{y}_u = \mathbf{W}_p \mathbf{e}_u + b_p \quad (3.23)$$

其中， $\mathbf{W}_p$  是可学习的参数矩阵， $b_p$  是可学习的偏置项， $\tilde{y}_u$  是模型预测出的股票回报率。

最后，我们用均方误差作为损失函数：

$$MSE(y, \tilde{y}) = \frac{1}{N} \sum_{i=1}^N (y_i - \tilde{y}_i)^2 \quad (3.24)$$

### 3.3 实验与结果分析

本节将详细介绍实验的各项细节，包括所使用的数据集、模型的设定以及评估指标等。此外，本节还将引入多种基准模型进行对比分析，以更全面地评估所提出模型的性能。在实验设计方面，除了基本的对比实验外，本节还将详细介绍消融实验，以验证各个模型组件的作用和重要性。同时，超参数实验将进一步探讨模型在不同超参数配置下的表现，并通过鲁棒性实验评估模型在不同数据环境和条件下的稳定性与适应性。通过一系列严谨的实验验证，旨在全面展示模型的有效性和优越性。

#### 3.3.1 数据集

##### （1）CSI300 数据集

CSI300 数据集包含了中国 A 股市场中 300 只市值最大、流动性最好的股票，覆盖了广泛的行业和公司类型。该指数由中证指数公司编制，代表了中国股市的整体表现，因此常被用作衡量中国股市行情和投资表现的重要基准。

##### （2）HSI 数据集

恒生指数(HSI)数据集包含了香港股市中市值最大、流动性最强的股票，涵盖了多个行业，如金融、地产、能源、消费品、科技等，全面反映了香港市场的整体表现。

##### （3）NDX100 数据集

纳斯达克 100 指数（NDX100）由纳斯达克交易所上市的 100 家市值最大、创新领先的非金融公司组成，涵盖信息技术、消费品、医疗健康、通信等多个行业。该指数以其科技驱动的成长性和创新性为特征，代表了全球科技行业的核心力量。作为全球最具影响力的股指之一，纳斯达克 100 指数广泛用于衡量科技领域的整体表现和市场趋势，并在全球金融市场中占据重要地位。

本文收集了 2018 年 1 月 1 日至 2022 年 12 月 31 日期间的 CSI300 指数、恒生指数和纳斯达克 100 指数的成分股作为数据集。根据数据样本比例为 0.85: 0.05: 0.05: 0.05，将数据按顺序划分为训练集、验证集、测试集和回测集。数据集包含 57 个技术因子，这些因子涵盖了常用的技术分析指标（具体因子信息见附录 1）。为了保证数据的分布特征不受影响，采用十进制缩放标准化方法进行预处理，使

得不同量纲的数据可以在相同的尺度下进行比较与分析。标签为未来 5 天的收益率，旨在预测短期内市场的表现。

### 3.3.2 实验超参数与环境设置

本文通过网格搜索法优化超参数。优化器 Adam 的学习率在  $\{0.001, 0.0001, 0.00001\}$  范围内选择，时间窗口大小  $L$  在  $\{24, 48, 64, 86, 98\}$  之间确定，模型维度大小  $D$  在  $\{64, 128, 256\}$  范围内调优。Transformer 编码器多头注意力机制中头的数量  $N_1$  和 RWKV 多头 WKV 中头的数量  $N_2$  都在  $\{1, 2, 3, 4, 5\}$  范围内选择，所有超参数通过验证集进行优化。训练过程中，epoch 设置为 150，并采用早停机制以减少过拟合的风险。

实验在 Windows 11 操作系统上进行，使用 NVIDIA GeForce RTX 4070 GPU，所有模型均在 PyTorch 框架下实现。为了减少随机初始化参数的影响，每个实验独立重复五次，并报告结果的平均值。

### 3.3.3 评估指标

本文使用信息系数（Information Coefficient, IC）、RankIC、收益率（Rate of Return, RoR）和年化夏普率（Annualized Sharpe Ratio, ASR）来评估模型的预测效果。IC 和 RankIC，其取值范围均为  $[-1, 1]$ ，值越接近 1 表示预测与实际值之间的相关性越强；RoR 取值范围为  $[-\infty, +\infty]$ ，值越大表示策略的盈利能力越强；ASR 取值范围为  $[-\infty, +\infty]$ ，值越大表示在单位风险下获得超额回报越高。这些指标综合衡量了模型的预测准确性、盈利能力和风险调整后的回报。

（1）IC 计算公式如下：

$$\text{corr}(\mathbf{X}, \mathbf{Y}) = \frac{\sum (X_i - \bar{X}) - (Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \cdot \sum (Y_i - \bar{Y})^2}}, \quad (3.25)$$

$$\text{IC}(\mathbf{y}^t, \hat{\mathbf{y}}^t) = \text{corr}(\mathbf{y}^t, \hat{\mathbf{y}}^t). \quad (3.26)$$

其中， $X_i$  和  $Y_i$  分别是样本数据集中第  $i$  个观测值。 $\bar{X}$  和  $\bar{Y}$  分别是  $X$  和  $Y$  的样本均值。

（2）RankIC 计算公式如下：

$$\text{RankIC}(\mathbf{y}^t, \hat{\mathbf{y}}^t) = \text{corr}(\text{rank}_{\mathbf{y}^t}, \text{rank}_{\hat{\mathbf{y}}^t}) \quad (3.27)$$

其中,  $rank_{y'}$  和  $rank_{\hat{y}'}$  分别是标签目标值和模型预测值的降序排名。

(3) RoR 计算公式如下:

$$RoR = \frac{P_t - P_0}{P_0} \quad (3.28)$$

其中,  $P_t$  和  $P_0$  分别是投资结束时的总金额和初始投资金额。

(4) ASR 计算公式如下:

$$ASR = \frac{\mathbb{E}[R_p - R_f]}{\sigma_p} \times \sqrt{T} \quad (3.29)$$

其中,  $R_p$  是投资组合的每日收益率或其他频率的收益率,  $R_f$  是无风险利率, 通常使用国债收益率,  $\mathbb{E}[R_p - R_f]$  指平均超额收益率,  $\sigma_p$  是投资组合的年化波动率 (标准差),  $T$  是一年中的交易天数, 通常设为 252。

特别的, 年化夏普率使用 252 个交易日来代表一年, 因为它主要衡量市场交易相关投资的风险调整后收益, 且只在工作日交易。这样的计算方式确保了每个指标能准确反映实际情况。

此外, 为了衡量模型的盈利能力, 在回测阶段所有模型统一采用 TopkDropout 交易策略, 该策略通过选择模型预测得分排名前 K 的股票进行投资, 并排除预测得分较低的股票, 从而构建高效的投资组合。具体而言, 该策略每日根据模型预测收益率对股票进行排序, 选择预测得分排名前 K 的股票进行投资, 并替换持仓内得分较低的 K 只股票, 从而构建高效的投资组合。

### 3.3.4 基准模型介绍和对比实验分析

本文提出的模型将在实验中与以下基准模型进行对比分析, 以评估其性能和有效性:

(1) MLP: 由多层全连接层构成的神经网络, 每层之间使用非线性激活函数。

(2) LSTM<sup>[53]</sup>: 一种特殊的 RNN, 通过引入记忆单元和门控机制来有效处理和预测时间序列数据中的长期依赖关系。

(3) GRU<sup>[54]</sup>: 一种简化版的 RNN, 通过合并遗忘门和输入门来有效捕捉序列数据中的长期依赖关系, 同时减少模型复杂度。

(4) ALSTM<sup>[57]</sup>: 一种 LSTM 变体, 通过引入注意力机制来自适应地聚合时间序列信息。

(5) Transformer<sup>[29]</sup>: 一种基于自注意力机制的神经网络架构, 能够并行处理序列数据中的所有位置, 有效捕捉长距离依赖关系。

(6) MASTER<sup>[55]</sup>: 通过建模瞬时和跨时间的股票相关性, 并利用市场信息进行自动特征选择, 有效解决了股票价格预测中的复杂相关性问题的。

(7) StockMixer<sup>[58]</sup>: 一种基于 MLP 的简单架构, 通过指标混合、时间混合和股票混合的方式, 显著提升了股票价格预测的性能。

(8) MATCC<sup>[45]</sup>: 一种基于 RWKV 的股票预测模型, 通过显式提取市场趋势、将股票数据分解为趋势和波动成分, 并挖掘跨时间的相关性, 解决了以往模型忽视市场趋势和跨时间相关性的问题。

本节将从四个关键指标出发, 对本章所提出的 Stock-RWKV 模型与基线模型的性能进行全面评估。通过这些指标, 我们旨在深入分析和比较不同模型在不同场景下的表现, 以验证 Stock-RWKV 模型在股市预测任务中的有效性与优势。这些评估指标包括 IC、RankIC、RoR 和 ASR。在对比试验中, 将加粗最优模型指标, 并为次优模型指标加下划线。

表 3.1 基线模型和 Stock-RWCV 在 CSI300 数据集上的实验结果

模型	CSI300			
	IC(↑)	RankIC(↑)	RoR(↑)	ASR(↑)
MLP	0.021	0.019	0.026	0.261
LSTM	0.032	0.031	0.061	0.562
GRU	0.034	0.032	0.072	0.587
ALSTM	0.037	0.039	0.106	0.624
Transformer	0.042	0.041	0.141	0.769
MASTER	0.044	0.043	0.165	0.972
StockMixer	0.043	0.042	0.159	0.951
MATCC	<u>0.046</u>	<u>0.044</u>	<u>0.192</u>	<u>1.016</u>
Stock-RWKV	<b>0.052</b>	<b>0.051</b>	<b>0.215</b>	<b>1.124</b>

从表 3.1 中的实验结果可以得出以下几点观察: 在所有基线模型中, MLP 表现较为薄弱, 这可能是由于其在处理时序数据时缺乏有效的时间序列建模能力, 而这一能力在股市预测任务中显得尤为重要。相比之下, RNN 类模型 (如 LSTM 和 GRU) 和注意力机制类模型 (如 Transformer) 各有其独特优势, 可能与它们在捕捉时间依赖性和建模长短期记忆方面的不同机制有关。具体而言, RNN 类模型通过门控机制有效学习时序数据的长期依赖, 而注意力机制则能够灵活地捕捉不



同时间步之间的关系，提高了对动态市场数据的响应能力。

此外，结合 RNN 和注意力机制的 ALSTM 模型在各项指标上表现出色，优于单纯的 RNN 和注意力机制模型。这可能源于两者的互补特性：RNN 擅长处理时序数据中的长期依赖，而注意力机制则增强了模型对关键信息的关注，从而提高了预测精度和鲁棒性。

在实验中，混合模型 MASTER 和 StockMixer 展现了卓越的性能，其出色的表现可能来源于模型在不同层次上融合了多种特征学习机制，使其能够更好地适应复杂市场波动的特性。值得注意的是，基于 RWKV 的 MATCC 模型在实验中表现出色，优于其他基线模型，这可能得益于其中 RWKV 模型的优势。RWKV 模型通过结合 RNN 和 Transformer 的特点，能够更好地处理长短期依赖关系，并有效捕捉时间序列中的复杂模式，从而提升了对股市数据的建模能力和预测精度。此外，RWKV 的灵活性和高效性使得 MATCC 在动态变化的市场中表现出较强的适应性。

相比之下，本文提出的 Stock-RWKV 模型在各项指标上达到了综合最优，在显示出其在股市预测任务中的卓越性能。相较于次优模型 MATCC，Stock-RWKV 在 IC 和 RankIC 上分别提升了约 13% 和 16%，在 RoR 和 ASR 上分别提升了约 12% 和 11%。这一结果表明，Stock-RWKV 模型能够有效捕捉股票间、股票内以及跨时间的复杂关系。具体而言，模型通过在时间序列和跨股票的特征交互中进行深度学习，能够挖掘出不同股票之间的潜在关联性以及同一股票在不同时间点的波动规律，从而提高了对市场走势的预测准确性。

表 3.2 基线模型和 Stock-RWCV 在 HSI 数据集上的实验结果

模型	HSI			
	IC(↑)	RankIC(↑)	RoR(↑)	ASR(↑)
MLP	0.034	0.031	0.069	0.602
LSTM	0.042	0.043	0.091	0.725
GRU	0.045	0.043	0.115	0.844
ALSTM	0.049	0.046	0.168	0.994
Transformer	0.051	0.050	0.179	1.125
MASTER	0.053	0.054	0.188	1.136
StockMixer	0.052	0.051	0.184	1.132
MATCC	<u>0.057</u>	<u>0.055</u>	<u>0.221</u>	<u>1.316</u>
Stock-RWKV	<b>0.061</b>	<b>0.060</b>	<b>0.256</b>	<b>1.519</b>

为了验证本章所提出的模型在不同股票市场上的适用性，本节在 HSI 和 NDX100 两个数据集上进行了实验。表 3.2 和表 3.3 展示了实验结果。从实验结果来看，模型在 HSI 市场上的表现最为优异，这可能是由于 HSI 成分股主要涵盖香港市场的大型蓝筹股，其价格波动相对稳定，且受区域性经济因素影响较大。而 NDX100 成分股则以美国科技股为主，具有更高的波动性和不确定性，受全球科技行业动态、政策变化以及投资者情绪的影响更为显著。这种市场特性的差异可能导致模型在 HSI 数据集上的表现优于 NDX100。

表 3.3 基线模型和 Stock-RWCV 在 NDX100 数据集上的实验结果

模型	NDX100			
	IC(↑)	RankIC(↑)	RoR(↑)	ASR(↑)
MLP	0.028	0.026	0.035	0.412
LSTM	0.035	0.034	0.071	0.611
GRU	0.038	0.037	0.082	0.625
ALSTM	0.042	0.039	0.118	0.722
Transformer	0.044	0.042	0.142	0.821
MASTER	0.047	0.048	0.192	1.102
StockMixer	0.046	0.045	0.187	1.068
MATCC	<u>0.051</u>	<u>0.054</u>	<u>0.212</u>	<u>1.128</u>
Stock-RWKV	<b>0.056</b>	<b>0.057</b>	<b>0.248</b>	<b>1.314</b>

在 NDX100 市场上，模型的表现虽然稍逊一筹，但仍达到了次优水平。此现象可能与 NDX100 成分股涉及更多高科技企业和新兴产业有关，其价格走势可能更加复杂和非线性，增加了模型学习的难度。尽管如此，Stock-RWKV 模型在 NDX100 数据集上仍表现出了较强的预测能力，证明了其在不同规模和不同结构市场中的良好适应性。

值得注意的是，在三个数据集上，本章所提出的 Stock-RWKV 模型均实现了综合最优成绩。这进一步证明了该模型在跨市场应用中的有效性和鲁棒性，能够在多样化的市场环境下保持优异的性能，验证了其在股市预测任务中的广泛适用性和优势。

### 3.3.5 消融实验分析

本章所提出的 Stock-RWKV 模型主要由三个关键模块组成：股票间关系聚合

模块、股票内关系聚合模块和跨时间相关性聚合模块。为了全面评估每个模块对模型性能的贡献，本节将针对这三个模块进行消融实验。通过逐一移除或替换各个模块，分析模型在不同配置下的表现变化，旨在揭示各个模块在提升模型效果中的具体作用和重要性。

(1) **Inter**: 仅保留了 Stock-RWKV 中的股票间聚合模块，用以验证该模块对模型性能的影响。

(2) **Intra**: 仅保留了 Stock-RWKV 中的股票内聚合模块，用以验证该模块对模型性能的影响。

(3) **Inter+CTC**: 仅保留 Stock-RWKV 中股票间聚合模块和跨时间相关性聚合模块，以验证两者组合对模型性能的影响。

(4) **Intra+CTC**: 仅保留 Stock-RWKV 中股票内聚合模块和跨时间相关性聚合模块，以验证两者组合对模型性能的影响。

(5) **Stock-RWKV**: 本章提出的完整模型，组合了三个模块，用于与以上分块模型进行对比。

表 3.4 Stock-RWCV 在 CSI300 数据集上的消融实验结果

模型	CSI300			
	IC(↑)	RankIC(↑)	RoR(↑)	ASR(↑)
Inter	0.029	0.027	0.055	0.546
Intra	0.031	0.030	0.062	0.558
Inter+CTC	0.041	0.042	0.136	0.761
Intra+CTC	0.044	0.045	0.167	0.976
Stock-RWKV	0.052	0.051	0.215	1.124

表 3.4 展示了 Stock-RWKV 模型不同变体的实验性能。从中可以看出，单独使用股票间关系聚合模块（Inter）或股票内关系聚合模块（Intra）的效果最差，这表明这两个模块各自对市场预测能力的提升是有限的，可能是因为单一的关系聚合方式无法充分捕捉到市场的复杂动态特征。具体而言，股票间关系聚合模块主要捕捉股票间的相互关系，而股票内关系聚合模块则专注于单只股票内部的变化，单独使用这些模块时，模型未能充分利用股市中多层次的时序信息。

另一方面，股票间关系聚合加跨时间相关性聚合（Inter+CTC）和股票内关系聚合加跨时间相关性聚合组合模型（Intra+CTC）在实验中取得了更好的成绩。这

表明，跨时间相关性聚合模块能够有效地增强模型在不同时间尺度下对市场波动的理解，弥补了单一关系聚合模块在动态市场环境下的不足。

Stock-RWKV 模型通过三者模块的有效结合达到了最优效果。这表明，股票间关系聚合、股票内关系聚合与跨时间相关性聚合三者的结合，能够充分利用股市数据中的多维信息和时间序列特征，从而使模型在捕捉市场复杂关系和长期趋势方面表现得更加精确，进一步验证了三者组合对提升预测性能的关键作用。

### 3.3.6 超参数实验分析

本节通过调整超参数，探讨 Transformer 编码器中多头注意力的数量  $N_1$  和 RWKV 模型中的时间混合模块中 WKV 的数量  $N_2$  对模型性能的影响。实验在 CSI300 数据集上进行。在改变这些超参数时，其他超参数均保持为先前所提到的默认数值。通过这一实验设计，本节旨在深入分析这些关键超参数对模型效果的作用，以便为 Stock-RWKV 模型的优化提供理论依据和实践指导。

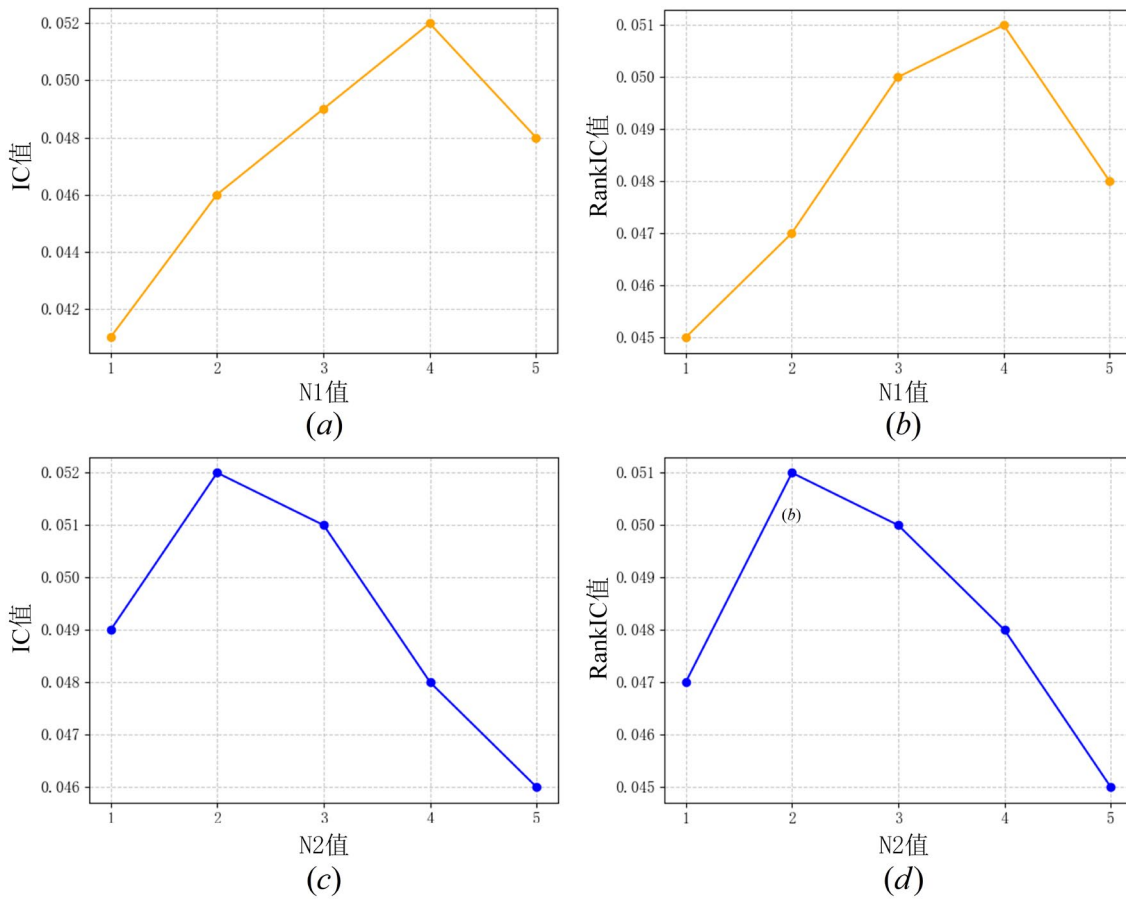


图 3.6 超参数敏感性分析。(a)图为  $N_1$  值和 IC 值的关系；(b)图为  $N_1$  值和 RankIC 值的关系；(c) 图为  $N_2$  值和 IC 值的关系；(d) 值为  $N_2$  值和 RankIC 值的关系。

图 3.6 (a) 和图 3.6 (b) 展示了在调整  $N_1$  时, 模型性能的变化。结果表明, 当  $N_1 = 4$  时, 模型性能达到最优。这表明, 适中的多头注意力数量有助于模型更好地捕捉股票市场中复杂的长期依赖关系, 并有效平衡计算效率与表现。当  $N_1$  过小时, 模型的多头注意力数量不足, 无法充分建模股票间的复杂关系, 导致模型的表达能力受限, 影响了预测精度。反之, 当  $N_1$  过大时, 虽然模型能够捕捉更多的信息, 但计算复杂度随之增加, 可能导致模型的训练时间显著延长, 且容易出现过拟合问题, 从而影响其在未见数据上的泛化能力。因此,  $N_1 = 4$  达到的最优性能表明, 适中的注意力头数有助于在准确性和效率之间找到最佳平衡点。

图 3.6 (c) 和图 3.6 (d) 展示了在调整  $N_2$  时, 模型性能的变化。实验结果表明, 当  $N_2 = 2$  时, 模型性能最为优越。较少的 WKV 头数能够较好地捕捉到股票数据中的时序依赖关系, 并有效融合多尺度的信息。若  $N_2$  过小, 模型的长期依赖性建模能力受限, 无法充分捕捉复杂的时间动态, 导致预测精度降低。相反, 当  $N_2$  过大时, 模型可能会引入过多的冗余信息, 导致过多的计算开销, 同时增加模型的复杂度, 可能使其难以有效训练, 甚至发生过拟合, 从而降低了泛化性能。因此, 适中的  $N_2$  数值表现出最佳的平衡, 既能保持模型对时序数据的敏感性, 又能有效避免不必要的计算复杂性和过拟合问题。

这两个超参数的调优实验表明, Transformer 中的多头注意力数量和 RWKV 中的 WKV 头数量都对模型性能有着重要影响。适当的参数选择能够显著提升模型的预测能力, 同时保证计算效率和训练的稳定性。

### 3.3.7 鲁棒性实验分析

鲁棒性实验旨在验证模型在数据扰动情况下的稳定性和适应性。通过对原始数据引入随机噪声、异常值或其他扰动因素, 对模型性能进行测试, 观察其预测结果的变化幅度。该实验能够评估模型在应对真实市场数据中的噪声和异常情况时的表现, 确保其在复杂数据环境下依然具有良好的预测能力, 从而验证模型的可靠性与实用性。由于股票数据信噪比较低, 天然包含大量的随机波动和噪音特征, 同时股票数据量较为丰富, 因此无需额外生成人工干扰数据即可模拟真实市场环境中的复杂性和不确定性。基于这一特点, 本文设计了基于 CSI300 数据集的鲁棒性实验。具体而言, 实验首先选定 CSI300 数据集中 200 只股票的历史数据进

行模型训练，随后在测试阶段逐步引入剩余的部分股票数据，分别按照 5%、10%、15%的比例逐步增加测试集规模。通过观察模型在不同规模和噪声水平下的表现变化，能够全面评估其在应对复杂股票市场环境时的稳定性与适应性。

(1) Stock-RWKV-5%: 选定 CSI300 数据集中 200 支股票进行训练，并在测试时额外增加 10 支股票的信息。

(2) Stock-RWKV-10%: 选定 CSI300 数据集中 200 支股票进行训练，并在测试时额外增加 20 支股票的信息。

(3) Stock-RWKV-15%: 选定 CSI300 数据集中 200 支股票进行训练，并在测试时额外增加 30 支股票的信息。

(4) Stock-RWKV-all: 选定 CSI300 数据集中所有股票进行训练和测试，作为实验参照。

其中，(1)、(2)、(3) 所选 200 支股票相同，(1) 中额外增加的 10 支股票是 (2) 中额外增加 20 支股票的一部分，同样的，(2) 中额外增加的 20 支股票是 (3) 中额外增加 30 支股票的一部分。

表 3.5 Stock-RWCV 在 CSI300 数据集上的鲁棒性实验结果

模型-数据量	CSI300			
	IC(↑)	RankIC(↑)	RoR(↑)	ASR(↑)
Stock-RWKV-5%	0.051	0.050	0.206	1.116
Stock-RWKV-10%	0.048	0.049	0.192	1.102
Stock-RWKV-15%	0.046	0.045	0.187	1.008
Stock-RWKV-all	0.052	0.051	0.215	1.124

表 3.5 展示了 Stock-RWKV 模型在 CSI300 数据集上的鲁棒性实验结果。从中可以看出，随着测试阶段逐步增加额外股票数量，模型的性能有所下降。这可能是因为新增股票带来的数据分布变化和噪声增加，对模型的预测精度造成了一定影响。然而，这种性能下降幅度较为有限，这可能得益于 Stock-RWKV 模型在捕捉股票间关系、股票内动态特征以及跨时间相关性方面的强大能力，使其能够在复杂环境下保持较高的鲁棒性和稳定性。实验结果与本文的预期一致，进一步证明了 Stock-RWKV 模型在应对高噪声、多样化市场环境时的适应能力与可靠性。

### 3.4 本章小结

本章提出了一种基于改进 RWKV 模型的股票预测方法 Stock-RWKV。该模型首先通过 Transformer 编码器有效捕捉股票间的相关性，同时，利用改进的 RWKV 结构捕捉股票内部的动态特征，最后引入时间注意力机制，以捕捉跨时间的相关性和趋势变化。通过这种多层次、多维度的信息融合，Stock-RWKV 能够从多个角度对股票价格的变动进行建模，从而更准确地预测股票的未来趋势。在多个真实世界数据集上的回测实验中，Stock-RWKV 展现了显著优于其他传统模型的预测性能。尤其是在面对股票市场动态变化和噪声较大的情况下，Stock-RWKV 表现出了更高的稳定性和鲁棒性，能够在复杂的市场环境中持续提供可靠的预测结果。这些实验结果不仅验证了 Stock-RWKV 在股票预测任务中的有效性，还证明了其在实际应用中的强大潜力和优势。

## 第 4 章 基于动态估值的 RWKV 投资组合方法

第三章提出的基于 RWKV 的股票回报率预测方法在回测中展现了一定的优势，但仍存在两个主要问题。首先，模型未考虑行业因素的影响，忽略了不同股票所在行业的特性，这可能导致的预测效果不及预期。其次，与其他主流深度学习预测模型一样，该方法也存在一个共性问题，即未充分考虑相对估值因素。这一问题本质是对风险的把控不足，因为相对估值有助于识别被高估或低估的股票，从而在市场波动中降低潜在风险。



图 4.1 2024 年 3-8 月上港集团蜡烛图走势

如图 4.1 所示，在 2024 年 3 月 28 日至 2024 年 4 月 3 日以及 2024 年 8 月 13 日至 2024 年 8 月 19 日，均出现了五连阳现象。这两个五连阳在某些因子上可能表现得相似，模型可能会给出相似的预测判断。然而，第一个五连阳发生在股价的相对低位，而第二个五连阳则处于相对高位，风险明显较大。因此，尽管这两个五连阳在因子层面可能类似，但应根据股票的相对估值判断股价是否高估，再进行买卖。针对以上两个问题，本章提出了一种基于动态估值的 RWKV 投资组合方法。该方法基于第三章提出的模型融合动态估值因素，该模块能够结合个股的



股价估值与行业整体的股价估值来判断股票是否相对高估或低估。通过对比个股与其所在行业的估值水平，模型可以识别出股价偏离行业均值的情况，从而帮助投资者更好地把握买入或卖出的时机。该方法能够有效规避单纯依赖股价走势的局限，提供更为全面的估值视角，有助于在市场波动中实现更为精确的风险控制和投资决策。

#### 4.1 股票量化投资任务介绍

实战中的量化投资注重在控制风险的前提下实现可观的收益。第三章提出的 Stock-RWKV 模型虽然在回测中表现出优势，但未充分考虑风险因素。为了解决这一问题，本章提出了 RPStock-RWKV 模型（RP 指 Risk Prevention），该模型引入了风险管理机制，进一步提升了模型的实用性和稳定性。本章还引入了最大回撤和盈亏比两个重要的风险衡量指标。通过这些改进，RPStock-RWKV 模型不仅能够在回报预测上保持优势，同时在实际投资过程中也能够更有效地控制潜在风险，确保在追求收益的同时实现更为稳健的风险管理。

#### 4.2 量化投资策略模型设计

##### 4.2.1 模型网络整体设计与构建

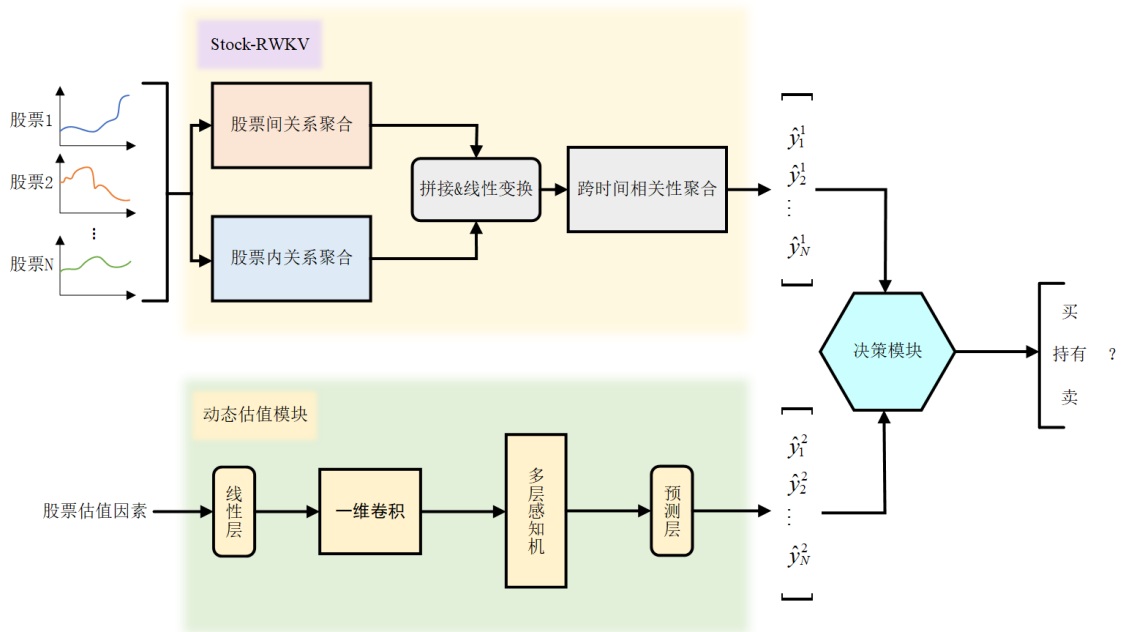


图 4.2 基于动态估值的 RPStock-RWKV 投资组合整体框架图

本章提出的 RPStock-RWKV 整体框架如图 4.2 所示。Stock-RWKV 模型接收股票的因子特征，关注股票的相对模式变化；而动态估值模型则接收股票的绝对价格特征和相对价格位置。通过这两种方式，模型从相对模式和相对价格位置两个角度对股票进行分析，以实现更全面的预测和决策支持。动态估值模块接收股票的开盘价、收盘价、最高价、最低价和成交量，以及每日波动率、当日收盘价与四十五日平均收盘价之比和所属行业平均当日收盘价与四十五日平均收盘价之比这八个特征。而标签  $y_u^2$  则表示为：

$$y_u^2 = \begin{cases} g_{u,\tau+d} / g_{u,\tau+1} > 1, & 1 \\ g_{u,\tau+d} / g_{u,\tau+1} \leq 1, & 0 \end{cases} \quad (4.1)$$

其中， $g_{u,\tau+d}$  表示为  $j_{u,\tau+d} / l_{i,\tau+d}$ ， $j_{u,\tau+d}$  指股票  $u$  在  $\tau+d$  天的收盘价与其最近的四十五日平均收盘价之比， $l_{i,\tau+d}$  指股票  $u$  所属的行业  $i$  在  $\tau+d$  天的行业平均  $j_{u,\tau+d}$  的值。故标签的含义为当未来股票  $u$  当日收盘价与四十五日平均收盘价之比  $j_{u,\tau+d}$  相较于个股所属行业的四十五日平均  $j_{u,\tau+d}$  之比更大时，则该股票价格未来有上升空间，表示相对低估，反之表示相对高估。在本章，我们将 Stock-RWKV 的输入设为  $\mathbf{X}_u^1$ ，其经过 Stock-RWKV 后的输出为  $\hat{y}_u^1$ ，动态估值模块的输入设为  $\mathbf{X}_u^2$ ，其经过动态估值模块后的输出为  $\hat{y}_u^2$ 。将  $\hat{y}_u^1$  和  $\hat{y}_u^2$  同时输入决策模块，最后由决策模块来决定是买入，持有还是卖出。

#### 4.2.2 基于深度学习的股票动态估值方法

动态估值模块结合了一维卷积和 MLP 两种技术。一维卷积能够提取股价时间序列中的短期波动和趋势特征，为动态估值提供局部信息；而 MLP 则通过建模复杂的非线性关系，对股票的绝对价格和其他特征进行全局分析，进一步优化估值。两者结合，使得动态估值模块能够实时调整股票估值，兼顾局部模式和全局趋势，从而实现更精准的动态估值。

$\mathbf{X}_u^2$  首先经过线性变换，将维度转换为模型的维度：

$$\mathbf{Z}_u^2 = \mathbf{W}_Z^2 \mathbf{X}_u^2 + \mathbf{b}_Z^2 \quad (4.2)$$

其中， $\mathbf{W}_Z^2$  是可学习的矩阵， $\mathbf{b}_Z^2$  是可学习的偏置项。

接着，再进行一维卷积操作：

$$\mathbf{V}_u^2 = 1d\_conv(\mathbf{Z}_u^2) \quad (4.3)$$

其中， $1d\_conv(\cdot)$ 表示一维卷积操作。然后，再将输入进 MLP，该 MLP 由两层线性层组成，且每个线性转换后都会通过 ReLU 函数进行激活：

$$\mathbf{H}_u^2 = \text{ReLU}(\mathbf{W}_H^2 \mathbf{V}_u^2 + b_H^2) \quad (4.4)$$

$$\mathbf{M}_u^2 = \text{ReLU}(\mathbf{W}_M^2 \mathbf{H}_u^2 + b_M^2) \quad (4.5)$$

其中， $\mathbf{W}_H^2$  和  $\mathbf{W}_M^2$  是可学习的参数矩阵， $b_H^2$  和  $b_M^2$  是可学习的偏置项。

最后通过预测层，预测层由一个线性变换和一个 softmax 函数构成：

$$y_u^2 = \text{softmax}(\mathbf{W}_P^2 \mathbf{M}_u^2 + b_P^2) \quad (4.6)$$

其中  $\mathbf{W}_P^2$  是可学习的参数矩阵， $b_P^2$  是可学习的偏置项。在二分类任务中，Softmax 函数将两个类的得分转换为概率值，使得输出值表示样本属于每个类别的概率，并且两个概率值的和为 1。

### 4.2.3 策略模块

本章使用改进后的 TopkDropout 策略，具体如图 4.3 所示。

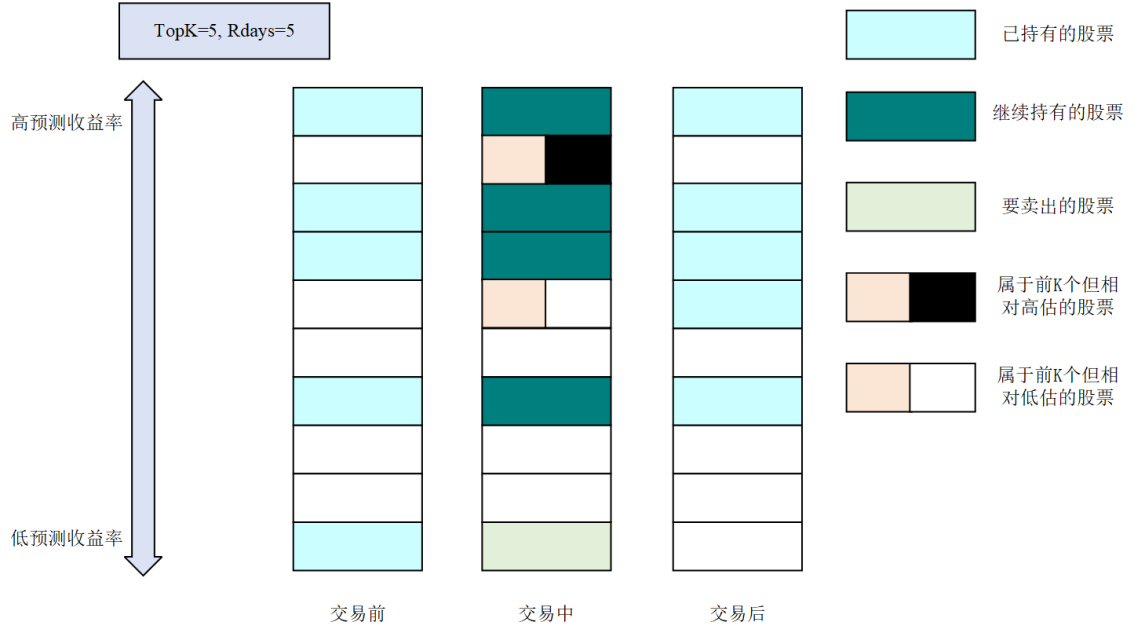


图 4.3 改进后的 TopkDropout 策略示意图

在本策略中，TopK 表示持仓股票的数量，Rdays 表示每只股票的预计持仓交易日数。具体操作如下：

1. 如果股票在持仓中且持仓交易日数未超过 Rdays，则该股票留在持仓中。
2. 如果股票在持仓中且持仓交易日数超过 Rdays，但不在最新的 TopK 列表的前 K 个中，则该股票将被卖出。

3. 如果股票在持仓中且持仓交易日数超过  $Rdays$ ，并且仍然位于最新的 TopK 列表的前 K 个，且相对低估，则该股票将继续持有。

4. 如果股票不在持仓中，但位于最新的 TopK 列表的前 K 个，则需要根据情况决定如何操作。此时，卖出的股票数量等于即将购买的 TopK 列表中前几个股票的数量。

5. 如果 TopK 列表前 K 个相对高估数量太多，不足以补充持仓中被卖出的股票数，则按顺序往后推，购买相对低估的股票。

改进后的 TopkDropout 策略通过  $Rdays$  与动态 TopK 排名的双重约束机制，构建了一个兼顾投资纪律性和灵活性的调仓框架。该策略在保留原 TopK 选股逻辑的基础上，创新性地引入持仓时间阈值和估值动态评估机制：一方面通过  $Rdays$  限制单只股票最长持有周期以避免过度依赖历史表现，另一方面结合实时 TopK 排名与估值状态进行动态筛选，既防止过早卖出潜力股，又能及时剔除超期且表现不佳的标的。同时通过强制轮动机制和资金平均分配原则，在控制风险敞口的同时实现持仓结构的持续优化，叠加复利效应驱动组合收益稳健增长，体现了策略在收益捕捉、风险控制与资金效率间的平衡性设计。

## 4.3 实验与分析

### 4.3.1 评估指标

本章侧重于量化回测的效果评估，因此在采用第三章中提到的收益率和年化夏普率作为评价指标的基础上，还引入了最大回撤率和盈亏比指标。最大回撤率 (Max Drawdown, MDD) 用于衡量投资组合在特定时间段内的最大亏损幅度，从而帮助评估风险控制能力；盈亏比 (Profit/Loss Ratio, P/L) 则用于衡量每一单位风险所获得的回报，进一步揭示策略在风险调整后的收益水平。这些新增指标有助于全面评估模型在实际应用中的表现，尤其在控制风险的前提下的收益能力。

(1) MDD 是指在一段时间内，投资组合从历史最高点到最低点的最大跌幅，通常用于衡量策略的风险。其计算公式为：

$$MDD = \frac{P_{peak} - P_{trough}}{P_{peak}} \quad (4.7)$$

其中， $P_{peak}$  是峰值，指的是历史最高点的投资组合净值。 $P_{trough}$  是谷值，指的是峰

值后出现的最小净值。

(2) P/L 表示盈利交易的平均收益与亏损交易的平均损失之比。其计算公式为:

$$P/L \text{ Ratio} = \frac{\bar{P}}{\bar{L}} \quad (4.8)$$

其中,  $\bar{P}$  指所有盈利交易的总收益除以盈利交易的次数。 $\bar{L}$  指所有亏损交易的总损失除以亏损交易的次数。

#### 4.3.2 基准模型介绍

本节所提方法与传统方法及当前优秀的深度学习方法进行比较, 这些方法包括以下几类: (1) 买入并持有策略; (2) 持续再平衡策略; (3) 基于深度学习的方法 (MASTER+ITD、StockMixer+ITD、MATCC+ITD)。其中, ITD 指改进后的 TopkDropout 策略。

(1) 买入并持有策略 (Buy And Hold, BAH): 在回测开始时, 将资金均匀分配到每项资产中, 一直持有, 直到回测结束。

(2) 持续再平衡策略 (Constant Rebalanced Portfolio, CRP) [59]: CRP 在每个时间段都进行投资组合的再平衡 (调仓)。由于投资组合中的每项资产的价格都会随着时间变化而变化, CRP 的思想是通过不断调仓使得投资组合始终保持在初始的权重配比。

(3) MASTER+ITD: MASTER 模型加上改进过的 TopkDropt 策略。

(4) StockMixer+ITD: StockMixer 模型加上改进过的 TopkDropt 策略。

(5) MATCC+ITD: MATCC 模型加上改进过的 TopkDropt 策略。

(6) Stock-RWKV+ITD: Stock-RWKV 模型加上改进过的 TopkDropt 策略。

实验参数设置: 对于本节提出的 RPStock-RWKV 模型, 使用 Adam 优化器, 学习率在 {0.001, 0.0001, 0.00001} 范围内选择。时间窗口大小设置为 64, 模型维度设置为 64, Transformer 编码器中的多头注意力数量设置为 4, WKV 数量设置为 2。一维卷积核大小设置为 4, 模型训练周期为 150 个 epoch, 并采用早停策略以避免过拟合。此外, 实验中设置了 0.1% 的交易手续费, 以模拟真实交易环境。

4.3.3 实验及结果分析

本实验基于 HSI 数据集，旨在评估所提出的模型在不同关键指标上的表现。通过在该数据集上进行回测分析，我们能够较为全面的衡量模型的收益能力、风险控制水平及整体稳健性。实验过程中，我们对比了多个基准策略和不同模型，以验证所提出方法的有效性和优势，实验结果如表 4.1 所示。

表 4.1 不同基线模型和方法在 HSI 数据集上的实验结果

模型	HSI			
	RoR(↑)	ASR(↑)	MDD(↓)	P/L(↑)
BAH	0.061	1.612	0.172	1.164
CRP	0.072	1.637	0.167	1.215
MASTER+ITD	0.172	1.986	0.142	1.812
StockMixer+ITD	0.166	1.967	0.147	1.708
MATCC+ITD	0.202	2.035	0.131	2.123
Stock-RWKV+ITD	<u>0.266</u>	<u>3.269</u>	<u>0.115</u>	<u>2.611</u>
RPStock-RWKV+ITD	<b>0.298</b>	<b>3.841</b>	<b>0.081</b>	<b>3.264</b>

从表 4.1 的实验结果可以看出，传统的交易策略 BAH 和 CRP 效果较差。这可能是因为这些策略在设计时往往基于固定的规则，缺乏对市场复杂动态特征的深度建模能力，因此无法有效应对股票市场中快速变化的价格波动和多样化的风险因素。在实际应用中，这种单一策略容易受到市场噪音的影响，导致预测不准确和收益不稳定。

从实验结果来看，MATCC+ITD、Stock-RWKV+ITD 和 RPStock-RWKV+ITD 均显著优于基准策略以及其他传统模型。其中，Stock-RWKV+ITD 的 RoR 达到 0.266，ASR 为 3.269，MDD 仅为 0.115，P/L 为 2.611，表明该模型在提升收益率的同时，也能有效控制风险，取得了较为稳定的收益表现。而 RPStock-RWKV+ITD 在此基础上进一步优化，引入了股票估值模块，使得 RoR 提升至 0.298，ASR 提高至 3.841，同时 MDD 下降至 0.081，P/L 提升至 3.264。这表明，RPStock-RWKV+ITD 不仅在收益率方面超越了其他模型，同时在回撤控制和收益稳定性上也表现更为优异。

对比 Stock-RWKV+ITD 和 RPStock-RWKV+ITD 的结果可以看出，引入股票估值模块能够有效提升模型的稳定性，使得回测过程中收益曲线更为平稳，同时减少极端市场环境下的回撤幅度。RPStock-RWKV+ITD 通过综合考虑股票收益率预

测和估值水平，不仅能够识别收益潜力较高的股票，还能进一步筛选被低估的股票，从而优化交易决策，提高整体投资组合的收益风险比。这一结果表明，在量化交易策略中，结合股票估值信息可以显著增强收益率预测模型的稳健性和盈利能力与灵活性。

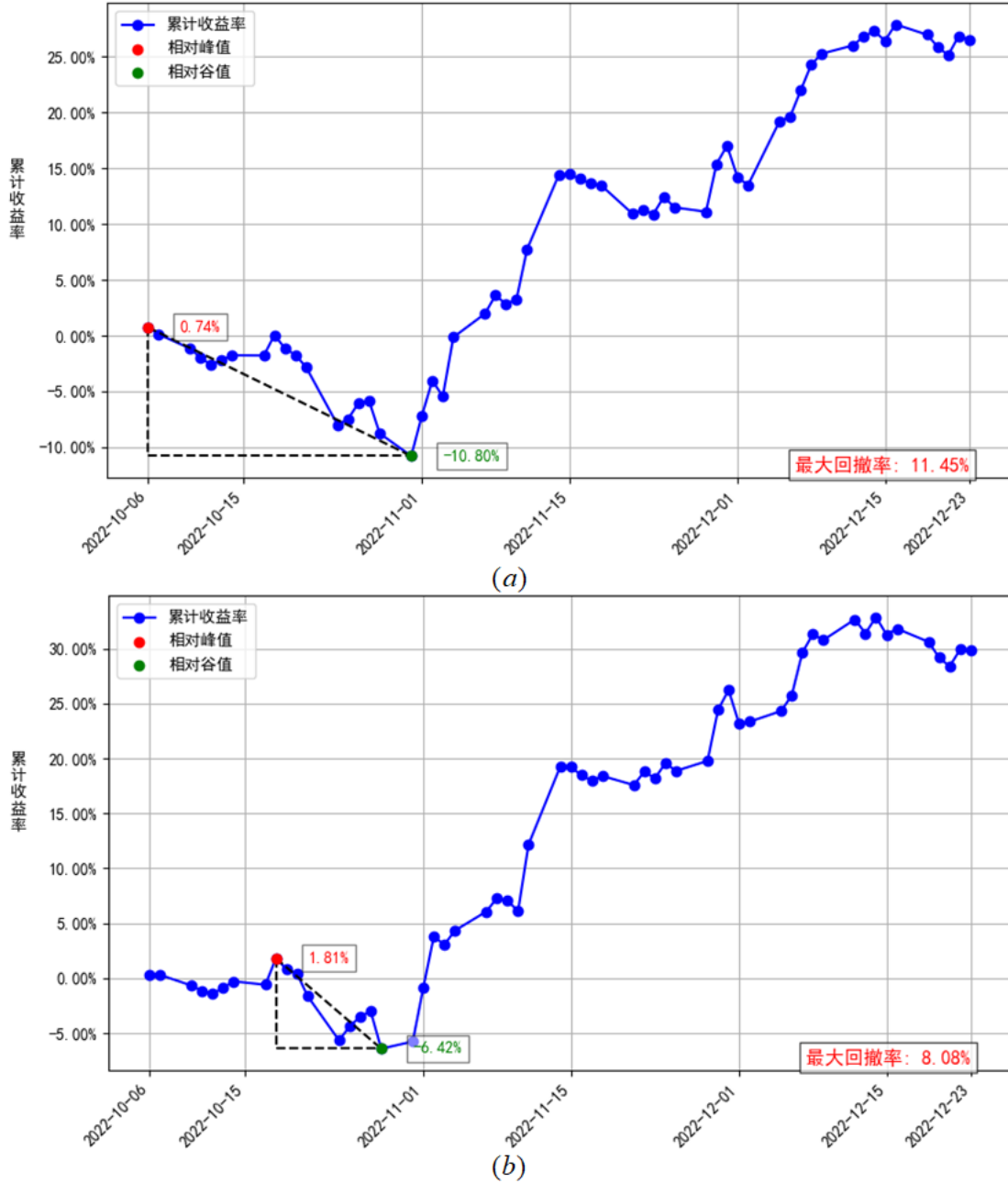


图 4.4 Stock-RWKV+ITD 和 RPSStock-RWKV+ITD 回测对比图。(a) 图为 Stock-RWKV+ITD 的累计收益与最大回撤示意图；(b) 图为 RPSStock-RWKV+ITD 的累计收益与最大回撤示意图。

为了进一步验证 RPSStock-RWKV 结合改进交易策略的风险控制能力，本节分

析了交易过程中两种模型搭配策略的每日累计收益率（即从初始交易日至每个交易日的累计收益率变化情况）。

图 4.4 的（a）和（b）分别展示了 Stock-RWKV+ITD 和 RPStock-RWKV+ITD 的累计收益率与最大回撤情况。其中，累计收益率曲线反映了策略的收益表现，而最大回撤则衡量了投资过程中可能遭遇的最大损失幅度。

从累计收益率来看，RPStock-RWKV+ITD（图 b）在大部分时间段内的收益均优于 Stock-RWKV+ITD（图 a）。其收益曲线呈现较为平稳的上升趋势，且在市场波动较大的阶段仍能保持较好的增长态势，说明该策略在不同市场环境下具有较强的适应性和稳健性。相较之下，Stock-RWKV+ITD 的收益曲线虽然也呈现增长趋势，但在部分时间段内增长较为缓慢，且波动幅度较大，说明该策略可能对市场变化较为敏感，收益稳定性略逊一筹。

在最大回撤方面，RPStock-RWKV+ITD 的最大回撤在多数时间内低于 Stock-RWKV+ITD，表明其在控制投资风险方面具有更好的表现。具体来看，Stock-RWKV+ITD 在部分时期经历了较深的回撤，且回撤后的恢复速度相对较慢，而 RPStock-RWKV+ITD 的回撤幅度较小，并能较快恢复至较高收益水平。这表明 RPStock-RWKV+ITD 在市场下行期间能够更有效地控制损失，降低投资组合的回撤风险。

综合来看，RPStock-RWKV+ITD 在提高收益率的同时，展现出更优的风险管理能力，使其整体投资表现优于 Stock-RWKV+ITD。其较强的抗回撤能力使得该策略在不同市场环境下更具稳定性，为投资者提供了更好的风险收益比。这表明 RPStock-RWKV+ITD 在实际应用中可能更适用于市场波动较大的场景，能够在提升收益的同时有效减少风险暴露，从而优化投资组合的长期回报。

#### 4.4 本章小结

本章在 Stock-RWKV 模型的基础上，新增了动态估值模块，提升了模型在风险防范方面的能力，并将其整体命名为 RPStock-RWKV。该动态估值模块由一维卷积网络和多层感知机构成，能够有效地评估股票的相对估值，进而为决策提供风险控制依据。通过引入动态估值，模型能够判断股票在当前市场环境中的高估或低估情况，从而避免因市场波动导致的潜在风险。



此外，本章还对 TopkDropout 策略进行了优化，结合动态估值模块，RPStock-RWKV 在实验中的表现优于传统方法。实验结果验证了该模型在一定风险下能够稳定实现收益，且具备较强的适应性和鲁棒性，能够有效减少市场波动带来的回撤并提高盈亏比。总体来看，这些创新提升了 RPStock-RWKV 的实用性，使其不仅具备强大的预测能力，还能够更好地管控风险，在量化投资中实现更稳定的收益。

## 第 5 章 结论与展望

### 5.1 结论

本文介绍了股票时间序列预测的研究背景和意义，综述了国内外学者在该领域的相关研究成果，并总结了几种经典的股票预测模型。通过分析现有研究，发现现有模型存在一些不足。它们在捕捉股票之间复杂关系时能力有限，无法全面反映股票市场的动态特征；同时，模型未考虑相对估值因素，忽视了不同市场环境下股票的相对价格水平。针对这些问题，本文提出了基于 RWKV 模型的股票预测方法，旨在更好地捕捉股票间的相互关系，并结合相对估值机制，提升预测的准确性和鲁棒性。最后，本文将该模型应用于投资组合管理，并通过多个数据集的实验验证，证明了所提方法在提升预测精度、降低风险以及优化投资回报方面的有效性。本文主要工作总结如下：

为了有效捕捉股票之间的相关性，本文提出了一种基于 RWKV 的股票预测模型，命名为 Stock-RWKV。该模型利用 Transformer 的编码器捕捉股票间的相关性，同时，通过改进后的 RWKV 网络捕捉股票的内在相关性。具体来说，RWKV 被改进为多头 WKV 操作，以增强其捕捉能力，并通过优化通道混合模块，提升了模型对股票模式的学习能力。最后，模型通过引入时间注意力机制来捕捉股票的跨时间相关性，从而增强对股票长期趋势和短期波动的预测能力。本文对 Stock-RWKV 模型进行了深入研究，并与其他方法进行了对比，结果表明，Stock-RWKV 在多个数据集上表现出了显著的优势，验证了其在股票预测任务中的有效性。

为解决动态估值问题，本文在 Stock-RWKV 模型的基础上进行了改进，提出了一个集成风险防范机制的动态估值模块，将整体命名为 RPStock-RWKV。该模块结合了一维卷积和多层感知机，能够有效地利用股票及其所属行业的相关信息，进行相对估值判断。通过这种方式，RPStock-RWKV 不仅提升了对股票内在价值的判断能力，还能更好地应对市场波动。此外，本文还提出了改进的 TopkDropout 策略，在结合 RPStock-RWKV 后，显著增强了模型的风险控制能力。实验结果表明，RPStock-RWKV 能够在一定风险条件下实现稳定的收益，验证了其在实际量化投资中的有效性和实用性。

## 5.2 展望

该小节探索了股票预测方法未来的研究方向，主要包括以下几个方面：

### （1）探索多模态股票预测模型

股票预测模型的准确性和鲁棒性可以通过引入多模态数据融合方法显著提升。现有研究主要依赖于量价信息及其衍生因子进行建模，但这类方法仅能从历史交易数据中提取有限的模式和特征。未来的研究可以构建多模态股票预测框架，通过设计高效的数据融合机制与深度学习模型的集成优化，以增强特征表达的丰富性与泛化能力。

### （2）探索强化学习模型

股票交易策略如果固定不变，随着市场环境的变化，其效果会逐渐下降并失效。强化学习模型能够自动制定交易策略，生成内在的交易逻辑，并根据不同的市场行情进行动态调整。通过不断训练和优化，强化学习可以提高策略的适应性和效率，从而提升交易效果和投资回报。

## 附录

### 附录 1 57 个技术因子

因子	因子（英文缩写命名）
布林带上轨	bbands_upperband
布林带中轨	bbands_middleband
布林带下轨	bbands_lowerband
指数移动平均线	ema
希尔伯特瞬时趋势线	ht_trendline
考夫曼自适应移动平均线	kama
移动平均线	ma
梅斯自适应移动平均线	mama
跟随自适应移动平均线	fama
抛物线停损指标	sar
简单移动平均线	sma
三重指数移动平均线(T3)	t3
三重指数移动平均线	tema
三角移动平均线	trima
加权移动平均线	wma
平均趋向指数	adx
平均趋向指数评级	adxr
绝对价格振荡器	apo
阿隆下行线	arrondown
阿隆上行线	arronup
均势指标	bop
商品通道指数	cci
钱德动量振荡器	cmo
方向运动指数	dx
MACD 线	macd
MACD 信号线	macdsignal
MACD 柱状图	macdhist

资金流量指数	mfi
负向动向指标	minus_di
负向运动指标	minus_dm
动量指标	mom
正向动向指标	plus_di
正向运动指标	plus_dm
价格百分比振荡器	ppo
变动率指标	roc
相对强弱指数	rsi
随机指标慢 K 线	stoch_slowk
随机指标慢 D 线	stoch_slowd
终极振荡器	trix
威廉姆斯指标	willr
累积/分布线	ad
加速振荡指标	adosc
平衡交易量	obv
平均真实波幅	atr
归一化平均真实波幅	natr
真实波动范围	trange
希尔伯特周期	ht_dcperiod
希尔伯特相位	ht_dcphase
希尔伯特相量-相位内	phrasor_inphase
希尔伯特相量-正交	phrasor_quadrature
希尔伯特正弦	sine_sine
希尔伯特领先正弦	sine_leadsine
开盘价	open
最高价	high
最低价	low
收盘价	close
成交量	volume

## 参考文献

- [1] Kraft J, Kraft A. Determinants of common stock prices: A time series analysis[J]. The journal of finance, 1977, 32(2): 417-425.
- [2] Cao C Q, Tsay R S. Nonlinear time - series analysis of stock volatilities[J]. Journal of applied econometrics, 1992, 7(S1): S165-S185.
- [3] Meng J, Zhu T Y, Chen X, et al. The nonlinear dynamics characteristics of stock market and its variation[J]. Applied Mechanics and Materials, 2014, 513: 3188-3193.
- [4] 李金轩, 杜军平, 薛哲. 基于多视角股票特征的股票预测研究[J]. 南京大学学报 (自然科学版), 2021, 57(1): 68-74.
- [5] Ting J, Fu T, Chung F. Mining of stock data: intra-and inter-stock pattern associative classification[J]. Threshold, 2006, 5(100): 5-99.
- [6] Ni J, Xu Y. Forecasting the dynamic correlation of stock indices based on deep learning method[J]. Computational Economics, 2023, 61(1): 35-55.
- [7] 张倩玉, 严冬梅, 韩佳彤. 结合深度学习和分解算法的股票价格预测研究[J]. 计算机工程与应用, 2021, 57(05): 56-64.
- [8] Peng B, Goldstein D, Anthony Q, et al. Eagle and finch: RwkV with matrix-valued states and dynamic recurrence[J]. arXiv preprint arXiv:2404.05892, 2024.
- [9] Li Q, Tan J, Wang J, et al. A multimodal event-driven LSTM model for stock prediction using online news[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 33(10): 3323-3337.
- [10] Yoo J, Soun Y, Park Y, et al. Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 2037-2045.
- [11] Li W, Bao R, Harimoto K, et al. Modeling the stock relation with graph network for overnight stock movement prediction[C]//Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence. 2021: 4541-4547.
- [12] Agrawal M, Shukla P K, Nair R, et al. Stock prediction based on technical indicators using deep learning model[J]. Computers, Materials & Continua, 2022, 70(1).
- [13] Ye J, Zhao J, Ye K, et al. Multi-graph convolutional network for

- relationship-driven stock movement prediction[C]//2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021: 6702-6709.
- [14] 宁苡鹤. 基于相关性的股票价格预测模型研究[D]. 北京: 北京邮电大学, 2018.
- [15] Han H, Xie L, Chen S, et al. Stock trend prediction based on industry relationships driven hypergraph attention networks[J]. *Applied Intelligence*, 2023, 53(23): 29448-29464.
- [16] Ma D, Yuan D, Huang M, et al. VGC-GAN: A multi-graph convolution adversarial network for stock price prediction[J]. *Expert Systems with Applications*, 2024, 236: 121204.
- [17] Qian H, Zhou H, Zhao Q, et al. MDGNN: Multi-Relational Dynamic Graph Neural Network for Comprehensive and Dynamic Stock Investment Prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(13): 14642-14650.
- [18] Bauder D, Bodnar T, Parolya N, et al. Bayesian mean–variance analysis: optimal portfolio selection under parameter uncertainty[J]. *Quantitative Finance*, 2021, 21(2): 221-242.
- [19] 程兰芳. 基于 CAPM 框架的组合投资线性规划模型研究[J]. *洛阳理工学院学报 (社会科学版)*, 2011, 26(6): 37-40.
- [20] Bodnar T, Lindholm M, Niklasson V, et al. Bayesian portfolio selection using VaR and CVaR[J]. *Applied Mathematics and Computation*, 2022, 427: 127120.
- [21] Ta V D, Liu C M, Tadesse D A. Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading[J]. *Applied Sciences*, 2020, 10(2): 437.
- [22] Sen J, Dutta A, Mehtab S. Stock portfolio optimization using a deep learning LSTM model[C]//2021 IEEE Mysore sub section international conference (MysuruCon). IEEE, 2021: 263-271.
- [23] Gondkar A, Thukrul J, Bang R, et al. Stock market prediction and portfolio optimization[C]//2021 2nd global conference for advancement in technology (GCAT). IEEE, 2021: 1-10.
- [24] Ma Y, Han R, Wang W. Portfolio optimization with return prediction using deep learning and machine learning[J]. *Expert Systems with Applications*, 2021, 165: 113973.

- [25] Ma Y, Mao R, Lin Q, et al. Quantitative stock portfolio optimization by multi-task learning risk and return[J]. Information Fusion, 2024, 104: 102165.
- [26] Shi Y, Wang Y, Qu Y, et al. Integrated GCN-LSTM stock prices movement prediction based on knowledge-incorporated graphs construction[J]. International Journal of Machine Learning and Cybernetics, 2024, 15(1): 161-176.
- [27] Zhang Y, Tumibay G M. Stock Price Prediction Based on the Bi-GRU-Attention Model[J]. Journal of Computer and Communications, 2024, 12(4): 72-85. Zhang Q, Qin C, Zhang Y, et al. Transformer-based attention network for stock movement prediction[J]. Expert Systems with Applications, 2022, 202: 117239.
- [28] 黄丽明, 陈维政, 闫宏飞, 等. 基于循环神经网络和深度学习的股票预测方法[J]. 《广西师范大学学报》(自然科学版), 2019, 37(1): 13-22.
- [29] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [30] Tao Z, Wu W, Wang J. Series decomposition Transformer with period-correlation for stock market index prediction[J]. Expert Systems with Applications, 2024, 237: 121424.
- [31] Zhang Q, Qin C, Zhang Y, et al. Transformer-based attention network for stock movement prediction[J]. Expert Systems with Applications, 2022, 202: 117239.
- [32] 陈诗乐, 王笑, 周昌军. 基于 GA-Transformer 模型的多因子股票预测[J]. 广州大学学报(自然科学版), 2021, 20(01): 44-55.
- [33] Peng B, Goldstein D, Anthony Q, et al. Eagle and finch: RwkV with matrix-valued states and dynamic recurrence[J]. arXiv preprint arXiv:2404.05892, 2024.
- [34] Li Z, Xia T, Chang Y, et al. A Survey of RwkV[J]. arXiv preprint arXiv:2412.14847, 2024.
- [35] Dumitru R G, Peteleaza D, Surdeanu M. Enhancing transformer rnns with multiple temporal perspectives[J]. arXiv preprint arXiv:2402.02625, 2024.
- [36] Kang J. Bone: Block Affine Transformation as Parameter Efficient Fine-tuning Methods for Large Language Models[J]. arXiv preprint arXiv:2409.15371, 2024.
- [37] Knunyants I, Tavakol M, Sifalakis M, et al. Explore Activation Sparsity in Recurrent LLMs for Energy-Efficient Neuromorphic Computing[J]. arXiv preprint arXiv:2501.16337, 2025.
- [38] Nardone J, Zhu R, Callenes J, et al. Learnable Sparsification of Die-to-Die Communication via Spike-Based Encoding[J]. arXiv preprint arXiv:2501.08645,



2025.

- [39] Yueyu L, Zhiyuan L, Yue P, et al. ARWKV: Pretrain is not what we need, an RNN-Attention-Based Language Model Born from Transformer[J]. arXiv preprint arXiv:2501.15570, 2025.
- [40] Choe W, Ji Y, Lin F. RWKV-Lite: Deeply Compressed RWKV for Resource-Constrained Devices[J]. arXiv preprint arXiv:2412.10856, 2024.
- [41] Xie S, Zhang B, Tang Y, et al. Rwkv: A Linear Attention Mechanism for Temperature and Humidity Compensation for Gas Sensors[J]. Available at SSRN 4612708.
- [42] Hou H, Yu F R. Rwkv-ts: Beyond traditional recurrent neural network for time series tasks[J]. arXiv preprint arXiv:2401.09093, 2024.
- [43] Dong Y, Wu T, Song C. Optimizing Robotic Manipulation with Decision-RWKV: A Recurrent Sequence Modeling Approach for Lifelong Learning[J]. Journal of Computing and Information Science in Engineering, 2025, 25(3).
- [44] Hao J, Liu F, Zhang W. Multi-scale RWKV with 2-dimensional temporal convolutional network for short-term photovoltaic power forecasting[J]. Energy, 2024, 309: 133068.
- [45] Cao Z, Xu J, Dong C, et al. MATCC: A Novel Approach for Robust Stock Price Prediction Incorporating Market Trends and Cross-time Correlations[C]//Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024: 187-196.
- [46] Xie S, Li Y, Ma Y, et al. AutoGMM-RWKV: A Detecting Scheme Based on Attention Mechanisms Against Selective Forwarding Attacks in Wireless Sensor Networks[J]. IEEE Internet of Things Journal, 2024.
- [47] Thapa J. Modern Sequence Models in Context of Multi-Agent Reinforcement Learning/Author Jenish Thapa[J]. 2024.
- [48] Dong C, Cao Z, Zhou S K, et al. DFT: A Dual-branch Framework of Fluctuation and Trend for Stock Price Prediction[J]. arXiv preprint arXiv:2411.06065, 2024.
- [49] Fu Q, Yuan H, Hou Y, et al. Linear Attention Based Channel Estimation Scheme for V2X Communications[C]//2024 8th International Conference on Communication and Information Systems (ICCIS). IEEE, 2024: 96-101.
- [50] 王庆勇, 唐良贵, 王振宇, 辜丽川. ChemRB:基于双向分子环约束的新型生成模型. 陕西师范大学学报(自然科学版). 2025, 53(1): 45-59

<https://doi.org/10.15983/j.cnki.jsnu.2025005>

- [51] Lang J, Yang L Z, Li H. Multi-Modal Dynamic Brain Graph Representation Learning for Brain Disorder Diagnosis Via Temporal Sequence Model[J]. Available at SSRN 5114041.
- [52] Elman J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179-211.
- [53] Schmidhuber J, Hochreiter S. Long short-term memory[J]. Neural Comput, 1997, 9(8): 1735-1780.
- [54] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.
- [55] Li T, Liu Z, Shen Y, et al. MASTER: Market-Guided Stock Transformer for Stock Price Forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(1): 162-170.
- [56] So D, Mañke W, Liu H, et al. Searching for efficient transformers for language modeling[J]. Advances in neural information processing systems, 2021, 34: 6010-6022.
- [57] Qin Y, Song D, Chen H, et al. A dual-stage attention-based recurrent neural network for time series prediction[J]. arXiv preprint arXiv:1704.02971, 2017.
- [58] Fan J, Shen Y. StockMixer: A Simple Yet Strong MLP-Based Architecture for Stock Price Forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(8): 8389-8397.
- [59] Cover T M. Universal portfolios[J]. Mathematical finance, 1991, 1(1): 1-29.