

基于多种Transformer架构的股票价格预测模型研究

宁小珊, 杨 燕

南宁师范大学数学与统计学院, 广西 南宁

收稿日期: 2025年4月5日; 录用日期: 2025年5月19日; 发布日期: 2025年5月28日

摘 要

时间序列预测是金融领域的重要研究课题, 对提高决策效率和风险管理具有重要意义。本文提出并构建了多种基于Transformer架构的时间序列预测模型, 包括Vanilla Transformer、Informer、FEDformer、Autoformer、PatchTST和TimesNet, 并以浦发银行股票收盘价作为预测对象, 探讨了单变量和多变量输入下这些模型的预测性能。实验结果表明, 基于Transformer架构的模型在预测精度和稳定性方面显著优于传统模型(ARIMA和VAR)。其中Vanilla Transformer在单变量预测中表现最佳(MSE = 0.0455, MAE = 0.1480), 而TimesNet在多变量预测中性能最优(MSE = 0.0463, MAE = 0.1513)。研究表明, 基于Transformer的模型在处理复杂时间序列问题时具有显著优势, 为金融时间序列预测提供了新的方法和参考。未来可进一步探索Transformer与其他深度学习技术的结合, 以提升模型的泛化能力和实用性。

关键词

时间序列预测, Transformer架构, 自注意力机制, 股票价格预测, TimesNet

Research on Stock Price Prediction Models Based on Various Transformer Architectures

Xiaoshan Ning, Yan Yang

College of Mathematics and Statistics, Nanning Normal University, Nanning Guangxi

Received: Apr. 5th, 2025; accepted: May 19th, 2025; published: May 28th, 2025

Abstract

Time series forecasting is an important research topic in the financial field, with significant implications

文章引用: 宁小珊, 杨燕. 基于多种 Transformer 架构的股票价格预测模型研究[J]. 人工智能与机器人研究, 2025, 14(3): 629-637. DOI: 10.12677/airr.2025.143062

for improving decision-making efficiency and risk management. This paper proposes and constructs several time series forecasting models based on the Transformer architecture, including Vanilla Transformer, Informer, FEDformer, Autoformer, PatchTST, and TimesNet. Using the closing price of Shanghai Pudong Development Bank stock as the forecasting target, the paper examines the predictive performance of these models under univariate and multivariate inputs. Experimental results show that Transformer-based models significantly outperform traditional models (ARIMA and VAR) in terms of prediction accuracy and stability. Among them, the Vanilla Transformer performs the best in univariate forecasting (MSE = 0.0455, MAE = 0.1480), while TimesNet has the best performance in multivariate forecasting (MSE = 0.0463, MAE = 0.1513). The study demonstrates that Transformer-based models have a significant advantage in handling complex time series problems, providing new methods and references for financial time series forecasting. Future research can further explore the combination of Transformer with other deep learning technologies to enhance the model's generalization ability and practical utility.

Keywords

Time Series Forecasting, Transformer Architecture, Self-Attention Mechanism, Stock Price Forecast, TimesNet

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

时间序列预测是一种利用时间顺序排列的数据,通过分析其过去的变化模式和趋势,对未来的数值或状态进行预测的方法。传统的时间序列预测方法主要是在确定时间序列参数模型的基础上求解出模型参数,并利用求解出的模型完成预测工作[1]。在统计模型中,有许多经典的时间序列预测方法来对时间序列问题进行分类和预测,如周静雯[2]、张巨峰[3]、田行健[4]使用 ARIMA 模型对数据进行分析与预测;张一弛[5]构建向量自回归模型(VAR)对上证 50ETF 期权的日收盘价实证研究该价格的影响因素。传统时间序列预测方法适用于短期线性和简单时间序列的预测,但在非线性、大规模、复杂数据上往往表现有限。随着深度学习技术的快速发展,近些年卷积神经网络、循环神经网络、注意力机制和图神经网络等方法[6]在时间预测领域得到广泛应用。同时,各种混合模型的研究也解决了传统的机器学习模型的缺陷,如王嘉增[7]针对个股价格进行预测,提出基于 MDT-Bi LSTM 的股价预测模型,同 BiLSTM、LSTM 和 CNN-LSTM 模型对比,预测结果得到有效提升。

虽然时间序列预测任务在以上预测模型的加持下已取得了卓有成效的性能,但对于更复杂的时间序列任务,以上的模型仍存在一定的局限性。因此,研究人员开始探索基于 Transformer 的模型和强化学习(RL)模型在时间序列预测任务中的可能性。2017 年,谷歌提出 Transformer 模型[8],随后出现基于 Transformer 架构的各种创新变体模型[9]。钟袁丰等[10]利用 Informer 模型对海洋环境要素数据集进行实验,结果 Informer 模型比 LSTM 模型、GRU 模型、GRU-D 模型等深度学习方法的均方根误差降低了 25.42%,平均绝对误差降低了 26.53%。王钰涵等[11]用 iTransformer 模型对工商银行的股票价格进行预测,其精度普遍优于统计模型 ARIMA、神经网络模型 LSTM 和 Transformer 模型。

基于 Transformer 模型及其变体在时间序列预测上初步展现的性能优势,但多种 Transformer 架构变体模型在时间序列预测上的性能是否普遍提升,针对预测问题,哪一类 Transformer 架构变体模型能取得

最优预测性能等仍有进一步探索的空间。因此，本文以浦发银行股票价格作为研究对象(浦发银行作为国内大型上市银行，股票交易活跃度高，且银行股票受宏观经济政策、行业监管等多重因素的影响，时间序列特征更为复杂，更适合验证不同模型的预测能力)，构建传统模型(ARIMA 模型、VAR 模型)以及基于 Transformer 架构的变体模型(Vanilla Transformer 模型、Informer 模型、FEDformer 模型、Autoformer 模型、PatchTST 模型、TimesNet 模型等)，分别利用以上模型预测股票收盘价的未来走势，以期进一步明确 Transformer 架构的变体模型在股票预测研究中的预测能力提高上限，并为寻找最优的预测模型提供实证支撑。

2. 相关原理

2.1. 自回归综合移动平均线

自回归综合移动平均线模型(ARIMA)的核心原理是利用序列自身的历史数据来进行预测，同时通过处理非平稳性和随机波动来提高预测精度。ARIMA 模型通常表示为 $ARIMA(p, d, q)$ ，其中 p 是自回归项的阶数(AR 部分)， d 是差分次数(I 部分)， q 是移动平均项的阶数(MA 部分)，用数学形式表示为：

$$\hat{y}_t = \mu + \phi_1 \cdot y_{t-1} + \cdots + \phi_p \cdot y_{t-p} + \theta_1 \cdot e_{t-1} + \cdots + \theta_q \cdot e_{t-q},$$

其中， ϕ 表示 AR 的系数， θ 表示 MA 的系数。

2.2. 向量自回归

向量自回归(VAR)模型是基于数据的统计性质建立模型，将单变量自回归模型推广到由多元时间序列变量组成的“向量”自回归模型。VAR 模型将系统中每一个内生变量作为系统中所有内生变量的滞后值的函数来构造模型，从而处理多个相关经济指标的分析与预测。

VAR 模型的一般形式为：

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t,$$

其中， Y_t 是一个内生变量列向量， $A_1, A_2, \cdots, A_p$ 是系数矩阵， ε_t 是误差项向量，满足 $\varepsilon_t \sim N(0, \Omega)$ ， Ω 是误差项的协方差矩阵。

2.3. Vanilla Transformer

Vanilla Transformer 是一种基于自注意力机制的序列到序列模型，主要由编码器(Encoder)和解码器(Decoder)两部分组成。编码器包含多个编码层，每层包含多头自注意力机制和前馈神经网络。输入序列经过每个编码层处理后，最终得到编码器的输出。解码器由多个解码层组成，除了包含自注意力机制外，还包含编码器 - 解码器注意力层，允许解码器在生成输出时关注编码器的输出[12]。

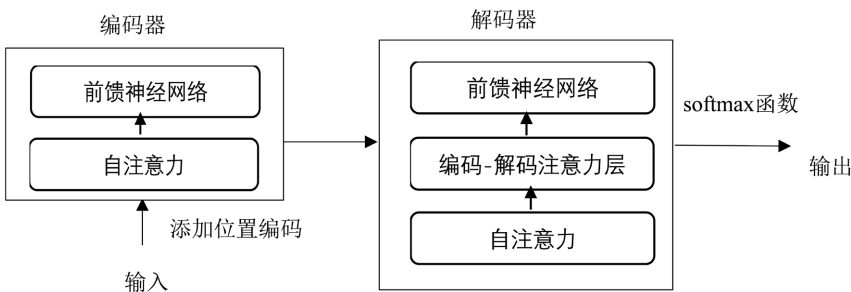


Figure 1. Workflow of Vanilla Transformer
图 1. Vanilla Transformer 的工作流程

Vanilla Transformer 的主要组件包括自注意力机制、多头注意力、位置编码、前馈神经网络、残差连接和层归一化。Vanilla Transformer 的工作流程见图 1。

2.4. Informer

Informer 是一种用于时间序列预测的深度学习模型, 它在 Transformer 的基础上进行了一些改进, 以提高对长序列数据的处理能力和预测精度[13]。Informer 引入了一个高效的自注意力机制, 称为 ProbSparse Self-Attention, 它通过选择性地计算最重要的注意力分数来减少计算复杂度。Informer 主要由编码器和解码器组成, 编码器采用一种基于稀疏自注意力机制的架构, 解码器则采用基于多头注意力机制的架构, 可以更好地捕捉时间序列中的长期依赖关系。

2.5. FEDformer

FEDformer 类似于其他 Transformer 模型, 引入频率增强和分解机制。它采用了基于傅里叶变换的频率增强机制和基于趋势 - 周期 - 残差分解的机制, 同时引入了时间维度上的稀疏自注意力机制, 减少计算复杂度[14]。

FEDformer 的工作流程并不复杂, 首先将输入时间序列转换为向量表示, 并进行频域分解, 得到不同频率的成分。接着输入序列经过多个编码器层, 层内使用频率增强的自注意力机制和前馈神经网络来提取特征。然后解码器根据编码器的输出生成未来的预测结果, 使用多头自注意力机制来关注不同的时间步。最后解码器的输出通过线性变换和 softmax 函数生成最终的预测分布。

2.6. Autoformer

Autoformer [15] 引入深度分解架构和自相关机制, 并结合了自注意力机制和自回归策略。模型引入了季节性分解模块, 该模块将输入的时间序列分解为趋势成分和季节性成分, 将趋势成分和季节性成分分别进行建模, 有效捕捉时间序列中的趋势和季节性变化。

2.7. PatchTST

PatchTST 是一种针对时间序列预测问题的深度学习模型, 旨在通过将时间序列数据转化为补丁 (patches) 的方式, 充分利用 Transformer 模型的自注意力机制, 提高对时间序列的建模和预测能力[16]。PatchTST 首先使用一个卷积层对时间序列数据进行预处理, 将其分割成多个小块; 然后, 每个小块被输入到一个线性层中进行编码, 这些编码被输入到 Transformer 编码器中, 其中包括多头自注意力机制和前馈神经网络层; 最后, 通过一个全连接层进行预测或分类。

2.8. TimesNet

TimesNet 采用 2D 卷积网络来捕捉时间序列的时间变化模式, 通过对时间序列数据进行 2D 化处理, TimesNet 能够处理建模时序中的短期和长期依赖关系[17]。TimesNet 采用了一种自适应的时间序列建模方法, 自动调整模型的结构和参数, 同时通过多个自注意力层逐层提取数据的特征。TimesNet 关注时域和频域特征的提取, 通过对时间序列进行傅里叶变换, 获取频域信息, 模型还采用残差连接和层归一化技术, 以提升训练的稳定性和加速收敛。

3. 模型构建

3.1. 模型评估指标

在时间序列预测问题中, 模型评估指标对于判断模型性能至关重要。均方误差(MSE)和平均绝对误差

(MAE)是常用的模型评估指标, 通常用于衡量预测值与实际值之间的偏差。一般来说, MSE 和 MAE 的值越小, 代表模型的预测误差越小, 因此在模型评估中, 这两个指标通常是越小越好。本文采用均方误差和平均绝对误差作为模型评估的指标。它们的计算公式为:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

其中, n 为样本数量, y_i 为实际值, $\hat{y}_i$ 为预测值。

3.2. 数据选取

Transformer 架构适用于各种时间序列预测任务, 包括但不限于经济和金融数据预测、交通流量预测、能源需求预测、销售和库存管理。在分析国内外对时间序列预测方法的基础上, 为了能充分验证基于原始 Transformer 进行改进和优化后模型的性能, 本文选取浦发银行(600000.SH)为实证对象开展实证研究, 我们选择 2023.01.03~2024.07.01 内每天 K 线的数据, 以开盘价、收盘价、最低价、最高价、前收盘价、成交量和总金额这七个指标作为交易数据, 构建传统时间序列预测模型 ARIMA 和 VAR, 以及 Vanilla Transformer、Informer、FEDformer、Autoformer、PatchTST、TimesNet 的多种 Transformer 架构的时间序列预测模型。

3.3. 构建传统时间序列预测模型

传统的时间序列预测方法主要包括线性模型和统计模型, 我们选择 ARIMA 模型来进行单变量预测, 即用股票历史的收盘价数据来预测测试集的收盘价, 对于多变量预测选择 VAR 模型, 采用以上七个指标的交易数据来预测股票收盘价。检查数据的缺失值和异常值, 选择合适的数据特征构建模型, 考虑到股票的时序性, 实验数据集划分保持连贯性, 前 75% 为训练集, 后 25% 为测试集[18]。

3.4. 构建多种 Transformer 模型

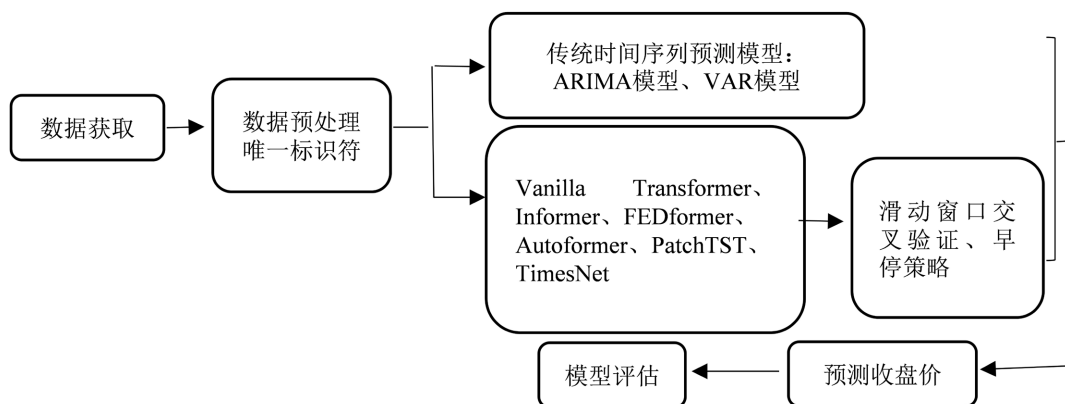


Figure 2. Modeling flowchart

图 2. 建模流程图

基于上述原理, 我们使用多种基于 Transformer 的深度学习模型来预测时间序列数据, 构建 Vanilla Transformer、Informer、FEDformer、Autoformer、PatchTST 和 TimesNet 模型, 并评估其在浦发银行股票收盘价数据上的表现。模型参数参考了 Informer (Zhou 等人, 2021 年) [13]、Autoformer (Wu 等人, 2021 年) [15]、FEDformer (Zhou 等人, 2022 年) [14] 等模型原始论文的默认设置, 为保证公平性与可比性, 所有模型统一设定预测范围在未来 10 步, 最大训练迭代次数 100 次, 每 10 步验证一次, 学习率为 0.0001,

采用 Adam 优化器，批次大小设置为 32，注意力头数为 4，隐藏层维度为 128。该设置在多个组合的巡参巡优的对比实验中表现最佳，兼顾预测精度与计算效率。部分模型如 Informer 采用 256 维隐藏层，但在此次研究的 Google Colab 环境(GPU 显存 15 GB，系统内存为 12.7 GB)下，128 维更稳定高效。参考 Autoformer 与 Informer 等模型的默认参数，并结合消融实验结果，4 个注意力头已具备良好的表达能力，进一步增加并未显著提升性能，反而增加资源消耗，故最终选择当前配置。模型还使用滑动窗口交叉验证策略，用 cross_validation 函数进行交叉验证，这样模型能在滑动窗口下适应不同的数据子集，保证预测的稳健性；模型在训练时使用了早停策略(early_stop_patience_steps)以防止过拟合。

综上，构建模型来预测股票收盘价的总体流程图如图 2 所示。

4. 实验结果与分析

针对上述的交易数据提取及建模步骤，我们对收盘价进行单变量预测和多变量预测，本文研究的问题属于预测问题，因此，我们采用时间序列预测模型常用性能评价指标来评价模型的预测性能，包括均方误差(MSE)和平均绝对误差(MAE)。

4.1. 单变量预测结果

为了验证基于多种 Transformer 架构的时间序列预测模型的性能，我们以传统时间序列预测方法的 ARIMA 模型作为基线模型，对照 Vanilla Transformer 模型、Informer 模型、FEDformer 模型、Autoformer 模型、PatchTST 模型和 TimesNet 模型来比较。通过股票历史收盘价的数据来预测未来收盘价的走势，在测试集上模型的预测性能见表 1，结果见图 3。

Table 1. Comparison of univariate prediction results
表 1. 单变量预测结果对比

模型	MSE	MAE
ARIMA	0.7617	0.6987
Vanilla Transformer	0.0455	0.1480
Informer	0.0457	0.1481
FEDformer	0.0506	0.1572
Autoformer	0.0495	0.1558
PatchTST	0.0459	0.1478
TimesNet	0.0518	0.1543

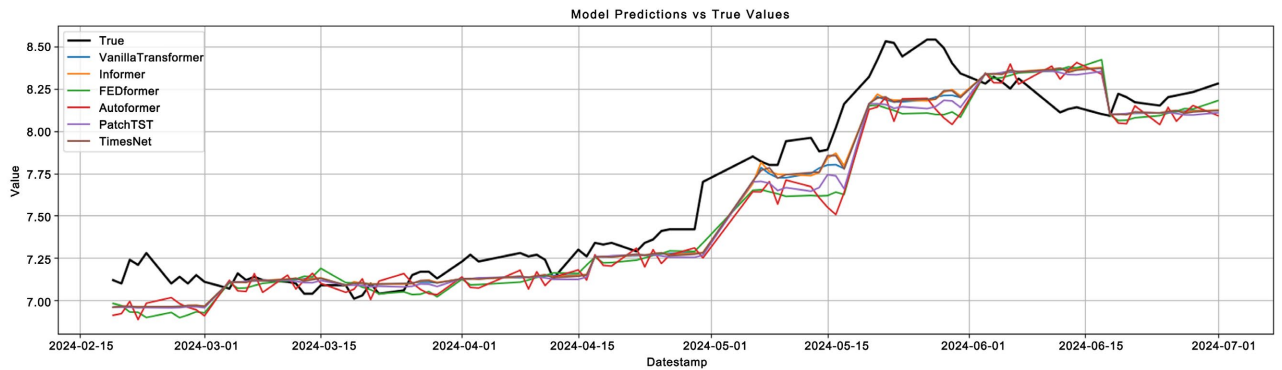


Figure 3. Univariate prediction result diagram
图 3. 单变量预测结果图

从实验结果来看，基于多种 Transformer 架构的时间序列预测模型得到的预测结果与真实值较为接近。由表中数据可见，传统的时间序列预测 ARIMA 模型的预测效果最差，MSE 为 0.7617，MAE 为 0.6987。基于多种 Transformer 架构的时间序列预测模型的 MSE 和 MAE 均较低，比传统的 ARIMA 模型具有更好的预测性能。其中，Vanilla Transformer 模型的 MSE 是最低，为 0.0455，其次是 Informer 模型；PatchTST 模型的 MAE 是最低，为 0.1478，其次是 Vanilla Transformer 模型。Vanilla Transformer 模型在单变量预测中表现最优，因为其简单而有效的结构更适合处理单一变量的时间依赖关系，而不会受到过度复杂架构带来的噪声影响。相比之下，TimesNet 等复杂模型在处理多变量预测时优势更明显，这体现了模型复杂度与任务复杂度的匹配原则。总体来说，基于多种 Transformer 架构的时间序列预测模型具有更好的预测性能和稳定性。

4.2. 多变量预测结果

针对开盘价、收盘价、最低价、最高价、前收盘价、成交量和总金额这七个指标，我们通过多个变量的历史数据来预测未来收盘价的走势，以传统时间序列预测方法的 VAR 模型作为基线模型，对照基于多种 Transformer 架构的时间序列预测模型来比较。在测试集上模型的预测性能见表 2，结果见图 4。

Table 2. Comparison of multivariate prediction results
表 2. 多变量预测结果对比

模型	MSE	MAE
VAR	0.5939	0.5862
Vanilla Transformer	0.0523	0.1615
Informer	0.0496	0.1595
FEDformer	0.0641	0.1759
Autoformer	0.0663	0.1816
PatchTST	0.0519	0.1558
TimesNet	0.0463	0.1513

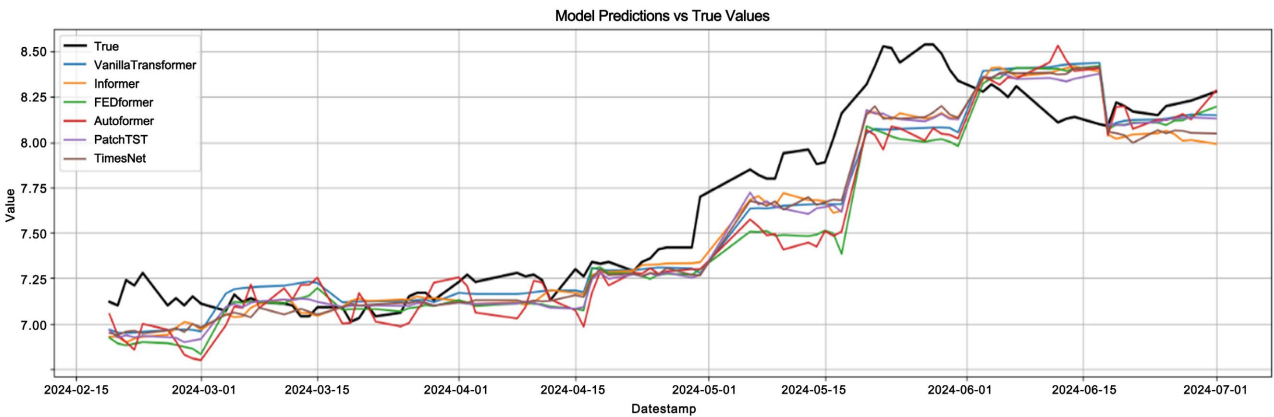


Figure 4. Multivariate prediction result diagram
图 4. 多变量预测结果图

由表 2 可以看出，传统的时间序列预测 VAR 模型得到的 MSE 为 0.5939，MAE 为 0.5862，在所有模型中性能最差。构建的 TimesNet 模型的评估指标最低，MSE、MAE 分别为 0.0463 和 0.1513，其次是

Informer 模型。相比于其他模型, TimesNet 有强大的多尺度建模能力、灵活的输入处理和高效的 Transformer 架构, 它能捕捉时间序列中的长期和短期依赖关系, 并自适应地融合不同特征的影响, 提升预测精度, 同时它能够高效处理长序列数据, 适应复杂的多变量任务, 所以 TimesNet 模型的性能为最好。在基于 Transformer 架构的所有模型中 Autoformer 模型是较差的, 它的 MSE 为 0.663, MAE 为 0.1816, 相比于传统的 VAR 模型, 评估指标明显较低。

5. 总结

本文选取浦发银行(600000.SH)作为对象来预测股票收盘价, 在原始 Transformer 模型的基础上进行改进和扩展, 构建基于多种 Transformer 架构的时间序列预测模型。这些模型(Vanilla Transformer、Informer、FEDformer、Autoformer、PatchTST 和 TimesNet)都基于原始 Transformer 的自注意力机制, 以适应时间序列预测的特殊需求。同时, 与传统的 ARIMA 模型和 VAR 模型进行性能对比分析, 得出以下结论:

1) 在对股票进行单变量预测中, Vanilla Transformer 模型的 MSE 是最低, 为 0.0455, 其次是 Informer 模型; PatchTST 模型的 MAE 是最低, 为 0.1478, 其次是 Vanilla Transformer 模型。说明基于多种 Transformer 架构的时间序列预测模型比传统的 ARIMA 模型具有更好的预测性能。

2) 在对股票进行多变量预测中, 最终构建 TimesNet 模型的 MSE 为 0.0463, MAE 为 0.1513, 其他基于 Transformer 架构的模型性能均明显优于传统的 VAR 模型, 说明优化的 Transformer 架构模型能更好地处理复杂的时间序列预测问题, 能更好地进行未来预测。

以上结论表明, 基于多种 Transformer 架构的时间序列预测模型优于传统的时间序列预测方法。本研究仍存在一些局限性, 如样本期间较短, 未能充分验证模型在不同市场周期的表现; 未考虑市场情绪、新闻等非量化因素的影响; 计算成本较高, 实际应用中需要权衡预测精度和运算效率。随着深度学习模型的发展, 用于预测时间序列任务的方法已经从传统技术转变为先进的深度学习技术, 未来可以考虑更多的预测方法, 例如 Transformer 和 RL 等技术, 这些改进的模型在时间序列预测任务中表现出更好的性能和适应性。

参考文献

- [1] 杨海民, 潘志松, 白玮. 时间序列预测方法综述[J]. 计算机科学, 2019, 46(1): 21-28.
- [2] 周静雯. 基于 ARIMA 模型的股票预测——以中国银行为例[J]. 现代计算机, 2024, 30(14): 89-92.
- [3] 张巨峰. 基于 ARIMA 模型的甘肃省碳排放预测分析[J]. 山西大同大学学报(自然科学版), 2024, 40(4): 31-36.
- [4] 田行健. 基于 ARIMA 模型的碳期货产品设计[D]: [硕士学位论文]. 蚌埠: 安徽财经大学, 2023.
- [5] 张一弛. 基于向量自回归模型的上证 50ETF 期权价格影响因素的实证研究[D]: [硕士学位论文]. 杭州: 浙江财经大学, 2018.
- [6] 毛远宏, 孙琛琛, 徐鲁豫, 等. 基于深度学习的时间序列预测方法综述[J]. 微电子学与计算机, 2023, 40(4): 8-17.
- [7] 王嘉增. 基于改进 LSTM 神经网络的股价预测与股票投资组合研究[D]: [硕士学位论文]. 西安: 西安建筑科技大学, 2023.
- [8] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, Long Beach, 4-9 December 2017, 5998-6008.
- [9] 夏雪, 闫恩来, 李喜武. Transformer 在时间序列预测中的应用综述[J]. 信息技术与信息化, 2024(3): 124-128.
- [10] 钟袁丰, 马丽文. 基于 Informer 模型的海洋环境要素缺失数据时间序列预测[J]. 青岛大学学报(工程技术版), 2024, 39(2): 11-16.
- [11] 王钰涵, 梁志勇. 基于 iTransformer 模型的金融时间序列预测[J]. 产业创新研究, 2024(15): 122-124.
- [12] Lin, T., Wang, Y., Liu, X. and Qiu, X. (2022) A Survey of Transformers. *AI Open*, 3, 111-132. <https://doi.org/10.1016/j.aiopen.2022.10.001>

-
- [13] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., *et al.* (2021) Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 11106-11115. <https://doi.org/10.1609/aaai.v35i12.17325>
 - [14] Zhou, T., Ma, Z., Wen, Q., *et al.* (2022) FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting. arXiv: 2201.12740.
 - [15] Wu, H., Xu, J., Wang, J., *et al.* (2021) Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *Advances in Neural Information Processing Systems*, **34**, 22419-22430.
 - [16] Nie, Y., Nguyen, N.H., Sinthong, P., *et al.* (2022) A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. arXiv: 2211.14730.
 - [17] Wu, H., Hu, T., Liu, Y., *et al.* (2022) TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. arXiv: 2210.02186.
 - [18] 黄柱兴, 杨燕, 刘宇婷. 基于财经新闻的多维情感特征融合交易特征的股票预测模型研究[J]. 南宁师范大学学报(自然科学版), 2023, 40(1): 64-71.