

李文颖,潘乔,阎希平. 基于深度学习的金融市场波动率预测模型[J]. 智能计算机与应用,2024,14(7):79-84. DOI:10.20169/j.issn.2095-2163.240711

# 基于深度学习的金融市场波动率预测模型

李文颖<sup>1</sup>, 潘乔<sup>1</sup>, 阎希平<sup>2</sup>

(1 东华大学 计算机科学与技术学院, 上海 201620; 2 上海兆前投资有限公司, 上海 201107)

**摘要:** 波动率在金融投资和风险管理中扮演着至关重要的角色,能够反映金融资产的收益和风险水平,为构建期权量化投资策略和决策以及风险控制提供重要参考指标。然而,波动率具有非线性和长期依赖性问题,如每日变化趋势不稳定,未来变化趋势与历史数据相关等。为解决这些问题,本文基于改进的 Transformer 构建了波动率预测模型 TGC-FinTrans (TCN-BiGRU-CNN Finance Transformer)。实验结果表明,该模型在预测金融数据波动率方面优于其他基线方法,能够更加准确地预测波动率并捕捉金融市场的复杂变化,为投资者提供更为精准的决策参考。

**关键词:** 波动率预测; Transformer; TCN; BiGRU; CNN

中图分类号: TP311 文献标志码: A 文章编号: 2095-2163(2024)07-0079-06

## Research and application of financial market volatility prediction based on deep learning

LI Wenyi<sup>1</sup>, PAN Qiao<sup>1</sup>, YAN Xiping<sup>2</sup>

(1 School of Computer Science and Technology, Donghua University, Shanghai 201620, China;  
2 Shanghai Zhaoqian Investment Co., Ltd, Shanghai 201107, China)

**Abstract:** Volatility plays a crucial role in financial investment and risk management, as it can reflect the return and risk level of financial assets, and provide an important reference indicator for constructing quantitative investment strategies and decisions on options and risk control. However, volatility has the problems of nonlinearity and long-term dependence, such as the unstable trend of daily changes and the correlation of future trends with historical data. To solve these problems, this paper constructs a volatility prediction model TGC-FinTrans (TCN-BiGRU-CNN Finance Transformer) based on an improved Transformer. The experimental results show that the model outperforms other baseline models in predicting the volatility of financial data and has significant advantages in predicting volatility more accurately and capturing the complex changes in the financial market, providing investors with more accurate decision-making references.

**Key words:** volatility forecasting; Transformer; TCN; BiGRU; CNN

## 0 引言

波动率是金融分析的关键指标之一,用于评估未来风险、资产价格波动幅度,反映资产收益率的不确定性<sup>[1]</sup>。波动率在金融投资和风险管理中应用广泛,是投资者决策和风险控制的重要参考指标;同时,波动率对价格具有重要影响,也是构建期权量化投资策略的关键基础<sup>[2-3]</sup>。因此,分析和预测金融数据的波动率对于金融投资者来说具有重要意义<sup>[4]</sup>。

目前金融交易市场主要使用传统的金融波动率

模型,例如 ARMA (Auto-Regressive Moving Average) 模型、GARCH (Generalized Auto-Regressive Conditional Heteroskedasticity) 模型和 SV (Stochastic Volatility) 模型来预测波动率,其基于当前和滞后的波动率之间的关系,利用条件方差公式和历史波动率来推导当前波动率的行为特征,并进行预测。张永东等<sup>[5]</sup>发现常用的 GARCH (1,1) 模型对于上海股市波动性的预测效果并不理想;针对上证指数的研究中,张胜杰<sup>[6]</sup>使用 GARCH 族模型预测沪深 300 指数波动率时,选择样本内最优模型并不能保证获得最佳的预测结果。

**作者简介:** 李文颖(1998-),女,硕士研究生,主要研究方向:智慧金融;阎希平(1978-),女,硕士,工程师,主要研究方向:金融大数据。

**通讯作者:** 潘乔(1977-),男,博士,副教授,主要研究方向:大数据,人工智能。Email: panqiao@dhu.edu.cn

收稿日期: 2023-05-03

结合以上调研发现这些传统金融模型只能处理时序中的线性关系,对于时序中的非线性关系处理能力有限。此外,这些方法对于多变量协同预测需要人工处理,主要适用于小规模单变量的时序预测问题<sup>[7]</sup>。

随着金融市场的复杂化,越来越多的研究人员开始探索深度学习在波动率预测中的应用。Lin等<sup>[8]</sup>提出一种基于支持向量机的波动率预测方法,在台湾股市数据集上的表现优于传统方法,具有更好的泛化性能和更高的命中率;Zhao等<sup>[9]</sup>在 LSTM (Long Short-Term Memory) 神经网络中引入时间加权函数,效果比其他模型更优秀;魏何平<sup>[10]</sup>提出 FNN-AE-LSTM-RV-mv 模型,综合考虑上证 50ETF 波动率因素,使用全连接神经网络和长短时记忆神经网络来进行预测研究。

上述深度学习方法对波动率预测由于时间序列数据的非线性和长期依赖性,很难有效捕捉到关键的时间序列特征。此外,由于梯度消失的问题,这些

模型无法充分利用过去的信息进行预测,导致模型的预测能力下降。

基于以上问题,旨在解决传统深度学习模型在捕捉时间序列数据的非线性和长期依赖性方面效果不佳、容易出现梯度消失等问题,本文提出了新的基于深度学习的金融市场波动率预测模型,该模型构建了基于改进 Transformer 的波动率预测模型 TGC FinTrans (TCN-GRU-CNN Finance Transformer),并在上证 50ETF 基金数据上进行实验验证。

## 1 模型

### 1.1 TGC FinTrans 搭建

TGC-FinTrans 模型的整体框架图如图 1 所示,模型在 Transformer 的基础上进行改进,增加了稳定性增强模块 TCN(Temporal Convolutional Network)、长期依赖性捕捉模块 BiGRU(Bidirectional GRU)和非线性表达增强模块 CNN(Convolutional Neural Network)模块,使其更加满足金融时间序列预测要求。

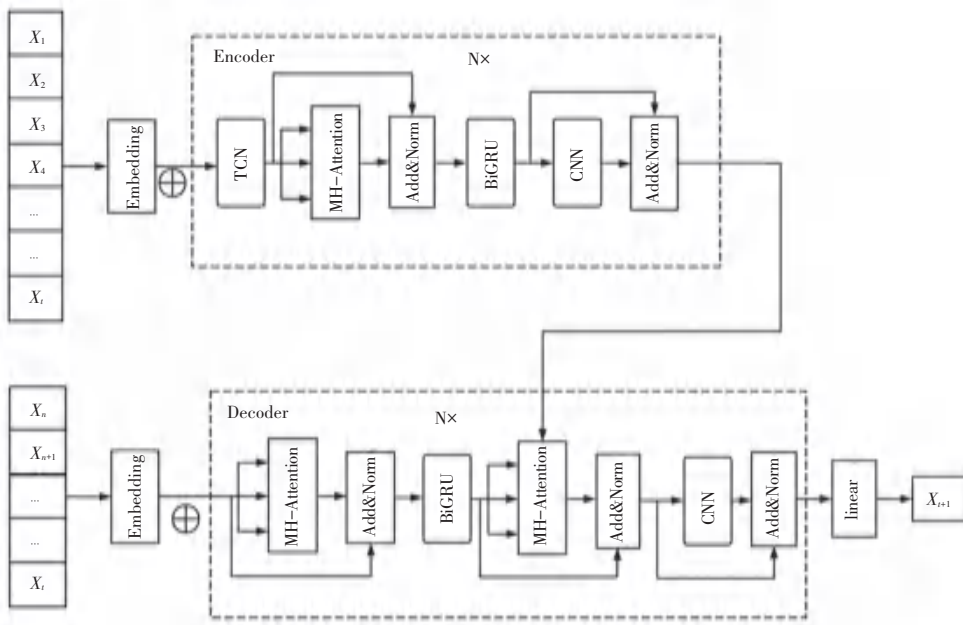


图 1 TGC-FinTrans 整体框架图

Fig. 1 Overall architecture of TGC-FinTrans

TGC-FinTrans 模型包含多个编码器和解码器。编码器中输入的  $X_1, X_2, X_3, \dots, X_i$  表示  $t$  天内的历史金融数据依次经过词嵌入 (Word Embedding)、稳定性增强模块 (TCN)、多头注意力机制 (Mulit Head-Attention Mechanism)、残差连接和层归一化模块 (Add&Norm)、长期依赖性捕捉模块 (BiGRU)、非线性表达增强模块 (CNN) 和残差连接和层归一化模块处理,得到  $X_{encoder}$ ; 解码器中,输入的  $X_1, X_2, X_3,$

$\dots, X_i$  历史金融数据依次经过词嵌入、多头注意力机制、残差连接和层归一化模块、长期依赖性捕捉模块、多头注意力机制、残差连接和层归一化模块、CNN 和残差连接和层归一化模块处理。编码器的输出向量  $X_{encoder}$  通过解码器的第二个多头注意力机制进行交互,最后经过线性层生成预测值。

针对上证 50ETF 基金波动率预测任务,将金融数据特征集统一表示为  $X$ , 目标值为上证 50ETF 基

金波动率  $y$ 。在该任务中需要使用滑动窗口来标记初始数据,即对于每个时间步,从数据集中选择大小为  $w$  的窗口获取金融数据点序列,作为输入金融数据集输出  $X$ ,并从相应时间步后  $w$  处的值来获取波动率值作为输出  $y$ 。使用数据分割部分生成的  $X$  和  $y$  列表来训练 TGC-FinTrans 模型,以预测目标波动率值  $y_{\text{pred}}$ 。该模型将输入序列  $X$  传递给 TGC-FinTrans 模型,以获取隐藏表示  $h(X)$ ,然后将这个隐藏表示传递到全连接层  $fc$  中,以获取预测的目标波动率值  $y_{\text{pred}}$ 。在训练过程中,使用均方误差作为损失函数来计算预测值和实际目标波动率值之间的误差,并使用 Adam 优化器来反向传播误差并更新模型权重,模型输出预测出来的波动率  $y_{\text{pred}}$ 。

TGC-FinTrans 的完整预测算法如算法 1 所示。

**算法 1** TGC-FinTrans 的完整预测算法

#数据集划分

**Input** 金融数据集  $data$ , 波动率数值  $target$

**Output** 具有滑动窗口的金融数据集  $X$ , 具有滑动窗口的波动率数值  $y$

1  $X = [], y = [], n$  是时间序列的长度,  $w$  是滑动窗口的大小

2 for all  $i \in [1, n - w]$  do:

3  $X_{\text{window}} \leftarrow data[i:i + w]$  // 获取数据集中大小为  $w$  的窗口

4  $y_{\text{window}} \leftarrow target[i + w]$  // 获取相应时间步后  $w$  处的波动率值

5  $X \leftarrow X_{\text{window}}$  // 将  $X_{\text{window}}$  添加到  $X$  列表中

6  $y \leftarrow y_{\text{window}}$  // 将  $y_{\text{window}}$  添加到  $y$  列表中

7 end for

8 return  $X, y$

#预测

**Input** 具有滑动窗口的金融数据集  $X$ , 具有滑动窗口的波动率数值  $y$

**Output** 目标值为  $y_{\text{pred}}$

1 Embedding and normalization processing // 对输入序列  $X$  进行嵌入和归一化处理

2 for  $e = 1$  to epoch do:

3  $h(X) \leftarrow \text{TGC-FinTrans}(X)$  // 将  $X$  传递给 TGC-FinTrans 网络, 获取隐藏表示  $h(X)$

4  $y_{\text{pred}} \leftarrow fc(h(X))$  // 将隐藏表示  $h(X)$  传递给全连接层  $fc$ , 获取预测的目标波动率值  $y_{\text{pred}}$

5  $Loss = \text{loss}(y, y_{\text{pred}})$  // 使用均方误差

作为损失函数, 计算预测值  $y_{\text{pred}}$  和实际目标波动率值  $y$  之间的误差

6 feed Loss to train the model by processing the Adam // 使用 Adam 优化器反向传播误差

7 backpropagate and update // 更新模型权重

8 end for

9 return  $y_{\text{pred}}$  // 返回预测的目标波动率值  $y_{\text{pred}}$

## 1.2 稳定性增强模块

本文中采用时间卷积层(Temporal Convolutional Network, TCN)对时间序列进行整体处理。TCN 由因果扩张卷积和残差块组成, 具有稳定的梯度, 扩张卷积具有增大感受野的作用, 残差连接有效避免特征提取过程中的梯度消失等问题, 与循环架构不同, TCN 的反向传播路径与序列的时间方向不同。因此, TCN 避免了梯度爆炸和消失的问题<sup>[11-12]</sup>。

金融数据集  $X$  是形状为  $(batch\_size, window\_size, input\_dim)$  的三维张量, 其中  $batch\_size$  表示输入金融数据的批次大小,  $window\_size$  表示输入的时间步,  $input\_dim$  表示输入数据中每个时间步的特征数量。首先对原始特征进行嵌入操作得到  $X_{\text{embedding}}$ , 公式(1); 其次, 通过位置编码器对  $X_{\text{embedding}}$  进行标记, 得到编码序列  $X_{\text{pos}}$ , 公式(2); 最后, 通过把其叠加到原来的特征嵌入上得到新的嵌入, 公式(3):

$$X_{\text{embedding}} = \text{Embedding}(X) \quad (1)$$

$$X_{\text{pos}} = PE(X_{\text{embedding}}) \quad (2)$$

$$X_{\text{embedding}} = X_{\text{embedding}} + X_{\text{pos}} \quad (3)$$

$X_{\text{embedding}}$  再传输到 TCN 网络层, TCN 使用因果卷积处理序列数据, 因果卷积操作如式(4)所示:

$$X_{\text{causal-conv}} = \sum_{k=1}^K f_k x_{t-K+k} \quad (4)$$

其中,  $f_k$  是一个由  $K$  个滤波器组成的集合, 序列  $x_t$  是一个长度为  $t$  的金融时间序列, 每个时间步的维度是输入数据的维度 ( $input\_dim$ ),  $x_{t-K+k}$  表示输入序列  $X$  在时间步  $t - K + k$  的值,  $K$  的取值范围是  $[1, t]$ , 卷积核在时间上不包含未来信息, 只能使用当前及过去的信息进行卷积操作, 保证了模型的稳定性。

传统卷积增加池化层扩大感受野会丢失信息, TCN 采用空洞卷积, 超参数指定间隔数量, 空洞卷积可以增加感受野并保留更多信息。空洞卷积操作如式(5)所示:

$$X_{\text{dilated-conv}} = \sum_{k=1}^K f_k x_{t-(K+k)d} \quad (5)$$

其中,卷积核长度为  $k$ ,  $d$  为使用固定大小的卷积核。

进行卷积计算,空洞卷积的感受野大小为  $(K-1)d+1$ ,所以增大  $K$  或  $d$  都可以增加感受野。通过增大感受野可以更大范围的金融特征提取,并且可以在不增加计算量的情况下,提高卷积层的感受野,进而提升模型的性能。

经过空洞卷积之后,将  $X_{\text{dilated-conv}}$  进行残差连接和权重归一化处理如图 2 所示,同时在卷积层添加  $\text{ReLU}$  激活函数引入非线性;每个卷积层都应用权重归一化 (WeightNorm),并在残差块中引入正则化 Dropout,以避免过拟合。具体操作如式(6)所示:

$$X_{\text{res}} = X_{\text{dilated-conv}} + F(X_{\text{dilated-conv}}) \quad (6)$$

其中,  $F(\cdot)$  表示通过卷积层和激活函数的深度神经网络。

由于  $X_{\text{dilated-conv}}$  和  $F(X_{\text{dilated-conv}})$  之间的通道数不一样,需要一个  $1 \times 1$  的 Conv 来对  $X_{\text{dilated-conv}}$  做一个简单的变换,使得变换后  $X_{\text{dilated-conv}}$  和  $F(X_{\text{dilated-conv}})$  可以相加。残差连接能够让网络更好地传递梯度和避免梯度消失的问题,从而使得网络更容易训练和收敛。

### 1.3 长期依赖性捕捉模块

LSTM 参数多,结构复杂,容易出现过拟合问题,GRU 结构简单,参数少,更适合放到 Transformer 的后续层中捕捉序列中的时间信息<sup>[13]</sup>。双向 GRU 可以分别从当前时间步之前的时间步和之后的时间步获取信息,从而更好地捕捉序列中的长期依赖关系,避免时间信息的丢失。

金融特征经过 TCN 层得到的表达  $X_{\text{res}} = (x_0, x_1, \dots, x_t)$ , 经过 Transformer 的多头注意力机制后得到经过注意力机制处理后的金融时序数据的区域信息  $X_{\text{Atten}}$ 。Transformer 的注意力机制将有限的注意力集中在重要的局部区域,从而获得有用金融因子的信息。Vaswani 等<sup>[14]</sup>提出的 Transformer 的注意力机制定义如式(7)所示:

$$X_{\text{Atten}} = \text{softmax} \left( \frac{\mathbf{Q} \mathbf{K}^T}{\sqrt{d_k}} \right) \cdot \mathbf{V} \quad (7)$$

其中,  $d_k$  为维度,长度为  $n$ ;  $\mathbf{Q} \in \mathbb{R}^{n \times k}$ ,  $\mathbf{K} \in \mathbb{R}^{n \times k}$  和  $\mathbf{V} \in \mathbb{R}^{n \times v}$  分别是每个嵌入的查询矩阵、键矩阵和值矩阵,其是具有相同输入的 3 个不同线性层的输出。

在自我注意力机制中每个输入向量都会与其他输入向量进行比较,以确定各向量之间的关系。这种比较是基于向量之间的点积计算得出的,由于点积计算是排列不变的,因此自我注意力机制也是排列不变的,即无论输入向量的顺序如何,都会得到相同的输出,虽然 Transformer 使用位置编码来保留时间信息,但也只是在输入层添加位置信息。因此为了更好地提取有序连续时间点之间的时间关系,通过添加双向 GRU 来更好地捕捉序列中的长期依赖关系。

输出  $X_{\text{Atten}}$  的形状为  $[\text{batch\_size}, \text{window\_size}, \text{hidden\_dim}]$ , 将输入张量沿着时间步的维度 (即  $\text{window\_size}$  维度) 进行切片,得到一个长度为  $\text{window\_size}$  的列表,列表中的每个元素形状为  $[\text{window\_size}, \text{hidden\_dim}]$ 。将每个元素输入双向 GRU 中进行处理,双向 GRU 由两个单向 GRU 层组成,如图 3 所示。一个按正向顺序处理输入序列,另一个按反向顺序处理序列,如公式(8)和公式(9);两个单向 GRU 层的输出会被拼接起来作为双向 GRU 层的最终输出,公式(10):

$$\vec{H}_t = \overrightarrow{\text{GRU}}(x_t, \vec{H}_{t+1}) \quad (8)$$

$$\overleftarrow{H}_t = \overleftarrow{\text{GRU}}(x_t, \overleftarrow{H}_{t-1}) \quad (9)$$

$$X_{\text{biGRU}} = \vec{H}_t \oplus \overleftarrow{H}_t \quad (10)$$

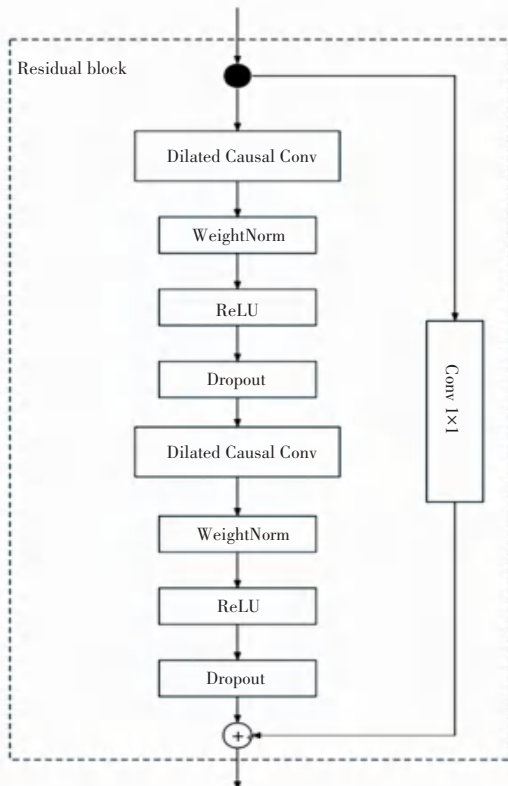


图 2 TCN 中的残差连接部分

Fig. 2 Residual connection part in TCN

其中,  $\vec{H}_t$  表示前向隐藏层状态;  $\overleftarrow{H}_t$  表示向后隐藏层状态;  $\oplus$  表示拼接操作。

得到的最终数据  $X_{\text{biGRU}}$  的输出形状为 [batch\_size, window\_size, hidden\_dim], 作为前馈神经网络的输入, 由于数据  $X_{\text{Atten}}$  通过双向 GRU 层, 双向 GRU 允许模型从正向和反向两个方向同时观察数据, 能够更好地捕捉数据中的上下文信息, 因此可以更好地捕捉序列数据中的长期依赖性。

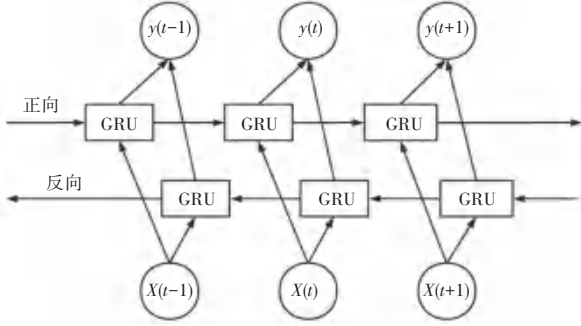


图 3 双向 GRU 模型

Fig. 3 Bidirectional GRU model

### 1.4 非线性表达增强模块

前馈神经网络 FFN 层 (Feed forward 层) 通过线性变换和激活函数将输入向量映射到高维空间, 再通过线性层映射到低维空间得到输出向量。卷积神经网络能够共享卷积核, 对于输入的不同位置, 使用相同的卷积核进行特征提取, 从而减少模型的复杂度, 并提高模型的泛化能力。在金融时间序列处理中, 卷积核可以捕捉时间序列数据中的局部模式, 因此可以考虑将 FFN 层替换为卷积层来处理金融时间序列, 如式 (11) 所示:

$$X_{\text{FFN}} = \text{RELU}(\text{conv}(\text{conv}(X_{\text{biGRU}}))) \quad (11)$$

其中, conv 表示卷积计算, RELU 为激活函数。

## 2 实验

### 2.1 数据集介绍

本文使用 WIND 数据库中的上证 50 指数、上证 50ETF 基金和上证 50ETF 期权 3 个市场的日度交易数据, 时间跨度为 2015 年 2 月 9 日至 2021 年 12 月 22 日期间, 共 1 653 个交易日。预测目标是上证 50ETF 基金波动率, 选取 11 个金融特征数据作为实验指标, 包括期权隐含波动率、基金 AD、股票收盘价、期权最高价、基金成交量加权平均价格、基金收盘价、期权收盘价、期权在值程度、基金均价、基金最高价和期权无风险利率<sup>[15-16]</sup>。

### 2.2 评价指标

本文研究的上证 50ETF 基金波动率预测任务,

本质是一个回归任务。因此使用回归任务中的 4 种常用评价指标, 即均方误差 (MSE), 均方根误差 (RMSE), 平均绝对误差 (MAE) 以及平均绝对百分比误差 (MAPE), 如式 (12) ~ (15) 所示:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^{\text{pred}})^2 \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^{\text{pred}})^2} \quad (13)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i^{\text{pred}} - y_i| \quad (14)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i^{\text{pred}} - y_i}{y} \right| \quad (15)$$

### 2.3 实验结果

为了验证本文所提模型 (TGC-FinTrans) 对上证 50ETF 基金波动率预测任务的有效性, 本文通过模型消融实验和综合实验进行对比分析, 从而更好地评估 TGC-FinTrans 在波动率预测任务中的性能表现。

#### 1) 消融实验

为了验证增加了 TCN、BiGRU 和 CNN 模型后对于 Transformer 模型性能的影响, 本文进行了消融实验, 并与本文提出的模型 (TGC-FinTrans) 进行对比, 旨在评估不同模块对模型预测的影响程度和重要性。通过逐步去除或替换这些模块, 分析其在模型性能中的贡献, 并进一步了解这些模块对于模型整体性能的关键作用, 有助于深入理解模型的结构和设计, 以优化和改进波动率预测任务的性能。实验结果见表 1。

由表 1 可见, 加入 TCN、BiGRU 和 CNN 替换 Transformer 的前馈神经网络都能适度降低模型的误差指标; 将其结合起来的 TGC-FinTrans 模型在所有模型中表现最佳, 误差最小, 其 MSE、RMSE、MAE 和 MAPE 分别降低了 0.000 5、0.010 1、0.010 1 和 2.01%。

表 1 消融实验结果

Table 1 Results of ablation experiment

模型	MSE	RMSE	MAE	MAPE / %
Transformer	0.001 0	0.033 5	0.027 5	22.31
Transformer +TCN	0.000 7	0.027 0	0.023 3	21.43
Transformer +BiGRU	0.000 6	0.025 5	0.022 8	21.31
Transformer +CNN	0.000 7	0.026 2	0.023 6	21.58
Transformer+TCN+BiGRU+ CNN(TGC-FinTrans)	<b>0.000 5</b>	<b>0.023 4</b>	<b>0.017 4</b>	<b>20.30</b>

## 2) 综合实验

为了验证本文所提模型对于上证 50ETF 基金波动率预测任务的有效性,选取了常见的金融预测模型:TCN(时间卷积神经网络<sup>[11]</sup>),ARIMA(自回归移动平均模型)<sup>[17]</sup>,SVR(支持向量回归模型)<sup>[18]</sup>,RNN(循环神经网络模型)<sup>[19]</sup>,LSTM-attention<sup>[20]</sup>,LSTNet<sup>[21]</sup>,TCN,Transformer,与本文所提出的 TGC-FinTrans 模型实现上证 50ETF 基金波动率预测任务,实验结果见表 2。

表 2 综合实验结果

Table 2 Results of comprehensive experiment

模型	MSE	RMSE	MAE	MAPE/%
ARIMA	0.003 6	0.059 6	0.050 1	30.96
SVR	0.005 8	0.076 1	0.069 2	37.76
RNN	0.001 9	0.043 1	0.037 2	27.00
LSTM-attention	0.001 1	0.035 4	0.028 8	22.78
LSTNet	0.001 3	0.035 6	0.028 6	22.81
TCN	0.001 0	0.034 0	0.027 7	22.76
Transformer	0.001 0	0.033 5	0.027 5	22.31
TGC-FinTrans	<b>0.000 5</b>	<b>0.023 4</b>	<b>0.017 4</b>	<b>20.30</b>

实验结果表明,TGC-FinTrans 模型在波动率预测任务中表现优于其他模型,证明其在金融数据中应用的可靠性和有效性,该模型在各个评估指标上误差最小,MSE、RMSE、MAE 和 MAPE 分别为 0.000 5、0.023 4、0.017 4 和 20.30%,显示了其优越性能。

## 3 结束语

本文提出了一种波动率预测模型 TGC-FinTrans。基于改进的 Transformer 结构,通过引入 TCN 解决梯度消失的问题,引入双向 GRU 更好地捕捉金融时间序列中的长期依赖性;使用卷积神经网络替代 Transformer 的前馈神经网络,增强了模型的非线性能力。TGC-FinTrans 模型在波动率预测任务上表现出色。

## 参考文献

[1] WILMS I, ROMBOUITS J, CROUX C. Multivariate volatility forecasts for stock market indices [J]. International Journal of Forecasting, 2021, 37(2): 484-499.

[2] 孔华强. 金融市场波动率模型及实证研究 [D]. 北京: 首都经济贸易大学, 2006.

[3] 范天腾, 吴佩玉. 基于 GARCH(1,1) 模型的上证 50ETF 波动率研究 [J]. 时代金融, 2017, 660(14): 160.

[4] 陶晓. 基于 MEMD-HAR 波动率预测模型的上证 50ETF 期权量化投资策略研究 [D]. 西安: 西北大学, 2022.

[5] 张永东, 毕秋香. 上海股市波动性预测模型的实证比较 [J]. 管理工程学报, 2003(2): 16-19.

[6] 张胜杰. 基于样本外预测的 GARCH 族模型评价研究 [J]. 经济研究导刊, 2020, 458(36): 58-61, 67.

[7] 徐晓芳, 管瑞. 基于神经网络集成学习算法的金融时间序列预测 [J]. 计算机系统应用, 2022, 31(6): 29-37.

[8] LIN Y, GUO H, HU J. An SVM-based approach for stock market trend prediction [C]// Proceedings of 2013 International Joint Conference on Neural Networks (IJCNN). IEEE, 2013: 1-7.

[9] ZHAO Z, RAO R, TU S, et al. Time-weighted LSTM model with redefined labeling for stock trend prediction [C]// Proceedings of 2017 IEEE 29<sup>th</sup> International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2017: 1210-1217.

[10] 魏何平. 基于多变量-混合神经网络模型的已实现波动率研究 [D]. 济南: 山东大学, 2022.

[11] ZHANG C X, LI J, HUANG X F, et al. Forecasting stock volatility and value-at-risk based on temporal convolutional networks [J]. Expert Systems with Applications, 2022, 207: 117951.

[12] LI W, JIANG X. Prediction of air pollutant concentrations based on TCN-BiLSTM-DMAAttention with STL decomposition [J]. Scientific Reports, 2023, 13(1): 4665.

[13] LIANG R, CHANG X, JIA P, et al. Mine gas concentration forecasting model based on an optimized BiGRU network [J]. ACS Omega, 2020, 5(44): 28579-28586.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31<sup>st</sup> International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017: 6000-6010.

[15] 陶隼. 投资指数基金的时机选择及策略研究 [D]. 上海: 同济大学, 2008.

[16] 王庆宗. 移动平均线建构动量策略研究 [J]. 河南科技, 2010, 448(14): 35-36.

[17] ARIYO A A, ADEWUMI A O, AYO C K. Stock price prediction using the ARIMA model [C]// Proceedings of 2014 UKSim-AMSS 16<sup>th</sup> International Conference on Computer Modelling and Simulation. IEEE, 2014: 106-112.

[18] ZHANG X, SHEN F, ZHAO J, et al. Time series forecasting using GRU neural network with multi-lag after decomposition [C]// Proceedings of Neural Information Processing: 24<sup>th</sup> International Conference. Cham: Springer, 2017: 523-532.

[19] ZHU Y. Stock price prediction using the RNN model [J]. Journal of Physics, 2020, 1650(3): 032103.

[20] HOLLIS T, VISCARDI A, YI S E. A comparison of LSTMs and attention mechanisms for forecasting financial time series [J]. arXiv preprint arXiv:1812.07699, 2018.

[21] LAI G, CHANG W C, YANG Y, et al. Modeling long- and short-term temporal patterns with deep neural networks [C]// Proceedings of the 41<sup>st</sup> International ACM SIGIR Conference on Research & Development in Information Retrieval. IEEE, 2018: 95-104.