

朱峰, 郭文静, 阎希平. 基于深度学习的高频交易金融数据的波动率预测[J]. 智能计算机与应用, 2024, 14(9): 82-87. DOI: 10.20169/j.issn.2095-2163.240912

基于深度学习的高频交易金融数据的波动率预测

朱峰¹, 郭文静¹, 阎希平²

(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: 随着信息化技术的发展,许多在线交易平台都可以提供高频的实时交易数据,为基于大数据的高频交易数据的波动率研究提供了基础。使用机器学习和深度学习算法分析大量的交易数据,建立波动率预测模型,可以帮助投资者更好地把握市场风险和机会,但金融高频交易数据存在大量噪声和非平稳性,导致模型的预测效果不佳。针对以上问题,本文构建了基于降噪自动编码器和不稳定注意力机制的深度学习模型,并利用该模型对高频交易数据波动率预测。实验结果表明该模型相较于常用的机器学习和深度学习方法拥有更准确的预测效果。

关键词: 降噪自动编码; 不稳定注意; 波动率预测

中图分类号: TP301

文献标志码: A

文章编号: 2095-2163(2024)09-0082-06

Volatility prediction of high-frequency trading financial data based on deep learning

ZHU Feng¹, GUO Wenjing¹, YAN Xiping²

(1 College of Computer Science and Technology, Donghua University, Shanghai 201620, China;

2 Shanghai Zhaoqian Investment Co., Ltd, Shanghai 201107, China)

Abstract: With the development of information technology, many online trading platforms can provide high-frequency real-time trading data, which provides a basis for the study of volatility based on high-frequency trading data of big data. Many scholars use machine learning and deep learning algorithms to analyze a large amount of trading data to establish volatility forecasting models, helping investors to better grasp market risks and opportunities. However, financial high-frequency trading data has a lot of noise and non-stationarity, which leads to poor predictive performance of models. To address these issues, This paper constructs a deep learning model based on noise reduction autoencoder and unstable attention mechanism, and uses this model to predict the volatility of high-frequency trading data. The experimental results show that this model has more accurate predictive performance than commonly used machine learning and deep learning methods.

Key words: noise reduction automatic coding; Instability attention; Volatility forecasting.

0 引言

金融市场的波动率指的是资产价格的波动程度,衡量的是金融产品价格围绕其平均值波动的程度,是金融市场的基本属性之一。波动作为金融市场与生俱来的基本属性,不仅影响投资者的风险偏好,也直接关系到金融产品的价格,自2008年全球金融危机后再次成为学术界关注的焦点^[1]。在金融市场中,较高的波动率通常意味着高风险,因此波动率的测量和预测对于金融机构和投资者来说是非常重要的。较高的波动率也会对金融市场的稳定性

造成威胁,可能会导致大规模的投资组合损失和市场崩盘,因此正确地测量和预测波动率对于金融稳定性和风险管理至关重要^[2-4]。

高频交易数据相较于低频数据来说,其数据量非常大,具有非常高的频率和密度,而且还有微结构噪音等因素的干扰,数据的方差在不同的时间段内会发生很大变化,导致其中包含了大量的噪声和跳跃^[5-6]。在对高频交易数据进行波动率预测的过程中,对金融高频交易数据进行去噪是一个重要步骤;其次,高频交易市场是一个高度动态的市场,交易者的行为和策略在短时间内不断发生变化,导致数据

作者简介: 朱峰(1998-),男,硕士研究生,主要研究方向:金融股票基金波动率预测; 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: 郭文静(1986-),女,博士,讲师,主要研究方向:传感器网络,体域网,物联网等。Email: wjguo@dhu.edu.cn

收稿日期: 2023-05-03

的波动性非常大,使得交易数据呈现出非平稳的状态,而模型很难从非平稳数据获取有利信息,从而无法准确的进行预测分析,因此在建模预测的时候需要进行平稳化^[7]。但是,传统的平稳化方法类似于移动平均法、k 阶差分法等用在订单簿和交易数据上,很可能会有过度平稳化的问题,可能会导致模型不能学习到数据的区分度,从而降低模型的预测能力,导致模型的预测效果不佳。

针对高频交易数据存在噪声和非平稳性的特点,本文提出了基于降噪自动编码器和不稳定注意力机制的高频交易数据波动率预测模型。首先,将存在噪声的数据带入降噪自动编码器模块进行训练,使得编码器能够学习到数据中的噪声并输出降噪后的数据,解决了金融高频数据存在大量噪声的问题;其次,引入不稳定注意力机制学习序列信息,通过改变权重来还原数据中的非平稳属性,并在 Transformer 中编码器注意力层之后引入混合卷积层帮助不稳定注意机制捕捉序列中的时空关系,以此来增强非平稳变压器的表现,解决非平稳数据影响模型性能的问题,提高模型的性能。

1 相关研究

近年来,许多针对金融高频交易数据波动率预测的深度学习模型的研究取得了很大的进展。2014 年, Kercheval 等^[8]提出了两种支持向量机模型,一种是基于价格级别信息的模型,另一种是基于时间序列信息的模型,用于建模和预测高频限价订单簿的数据,从而提高了对市场深度和价格波动的预测能力;2016 年, Wang L 等^[9]将 SDA (Stacked Denoising Autoencoders) 应用于高频时间序列的预测,提出了一种自动更新 SDA 模型训练数据集的方法,证实了 DNN (Deep Neural

Network) 神经网络在高频时间序列的预测性;2018 年, Zhou X 等^[10]将生成对抗网络用于股票预测,通过大量的实验表明对抗神经网络进行股票波动率预测,可以取得显著的效果;2021 年, Yang 等^[11]在使用订单簿的数据来预测金融时间序列的工作中,提出了一种基于自适应学习的方法,并通过实验验证了该方法在预测金融高频时间序列方面比传统方法表现更好,能够有效的提高预测精度和计算效率。

随着深度学习时间序列预测模型的不断发展,金融高频交易数据波动率预测任务得到越来越多的关注,但由于金融高频交易数据存在大量的噪声和非平稳性的问题导致数据分布产生偏移,应采取相应的措施进行处理。

2 模型

2.1 模型总体网络结构

针对金融高频交易数据中存在大量噪声和非平稳数据的问题,本文提出了基于降噪自动编码器和不稳定注意力机制的金融高频交易数据波动率预测模型,总体框架如图 1 所示。该模型将 DAE (Denoising Autoencoder) 与非平稳变压器相结合,首先将原数据添加高斯噪声后带入降噪自动编码器中进行训练,使得模型能够识别数据中的噪声,并对数据进行降噪和降维;其次,对降噪后的数据进行平稳化处理,同时计算不同时间窗口数据中的均值和标准差信息,并利用这些信息计算控制变量 τ 和移位向量 Δ ,将 τ 和 Δ 带入不稳定注意力机制来区别不同时间窗口的数据,使得模型能够学习到数据之间的差异性,并在注意力机制后添加混合卷积层,增强模型的表达能力;最后,根据不同时间窗口的均值和标准差对输出的结果进行还原,提高模型的预测精度。

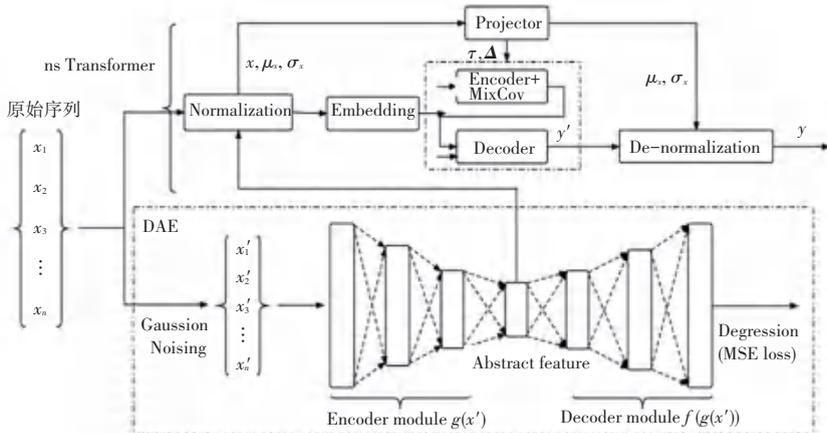


图 1 预测模型总体框架

Fig. 1 Overall framework of the prediction model

2.2 降噪自动编码器模块

在金融高频交易数据的波动率预测问题中,由于数据中存在大量的噪声信息,导致模型的预测误差较大,尤其是在测试集验证模型效果时,预测的结果往往较低。对金融高频数据进行数据降噪并且不丢失数据中的重要信息是一个需要解决的问题。

为了对金融高频交易数据进行数据去噪,需要让降噪自动编码器能够识别出数据中的噪声^[12]。因此,本文对交易数据因子特征向量序列 $X = (x_1, x_2, x_3, \dots, x_i)$ 添加高斯噪声,用下式表示:

$$x' = x + \epsilon \quad (1)$$

其中, x 是原始因子特征数据; x' 为添加噪声后的数据; ϵ 是服从均值为 0, 方差为 σ^2 的高斯分布随机噪声, $\epsilon \sim N(0, \sigma^2)$ 。

为了增强金融高频交易数据中因子特征的表达能力,得到数据中的潜在的因子特征信息,需要将添

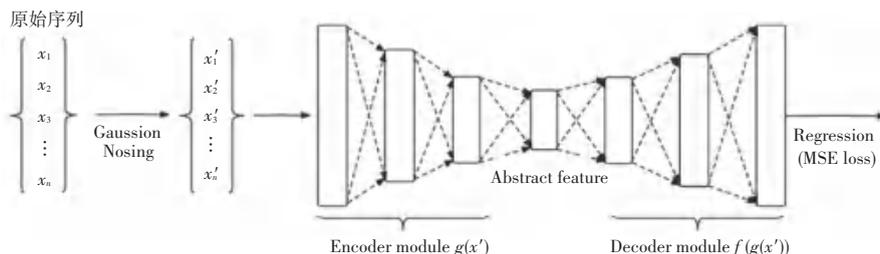


图2 降噪自编码器网络结构

Fig. 2 Denoising-AutoEncoder network structure

DAE 训练完成之后,将输入向量序列 x 输入到编码器中,使用编码器的输出 h 作为时间序列的低维表示,这个低维表示了数据的有用因子特征,同时过滤掉了噪声,提高了模型的预测精度。

2.3 规范化模块

对高频交易数据去噪处理后,需要应用一个可学习的缩放和平移操作模块对低维因子特征向量进行归一化处理。由于高频交易数据存在非平稳时间序列,其均值和标准差可能会随时间变化而变化。为了消除其在不同时间点上的尺度差异,使得数据更易于比较和分析,在规范化的过程中需要考虑时间变化的因素。对于一个非平稳时间序列 x_t , 其 t 时刻的规范化值计算公式如下:

$$z_t = \frac{x_t - \mu_t}{\sigma_t} \quad (4)$$

其中, μ_t 和 σ_t 分别代表时间序列 x_t 在 t 时刻的均值和标准差。

由于均值和标准差会随时间变化而变化,因此需要计算每个时刻的均值和方差。使用滑动窗口的

加高斯噪声后的数据映射到一个低维度的因子特征向量序列 $H = (h_1, h_2, h_3, \dots, h_i)$, 编码器通过下式来实现对数据的映射:

$$h = f(Wx + b) \quad (2)$$

其中, W 是权重矩阵; b 是偏置向量; f 是激活函数 sigma。

最后,根据解码器的参数和权重对因子特征向量进行解码,并对解码后的因子特征向量和原始噪声因子特征进行最小化重构误差,根据重构的结果来调整编码器和解码器的权重和参数,最小化重构误差可以通过下式表示:

$$L(x, x') = \frac{1}{T} \sum_{t=1}^T \|x_t - \hat{x}_t\|^2 \quad (3)$$

其中, $\|$ 为欧几里得范数; T 为每一个因子特征向量; \hat{x}_t 为解码器解码后的因子特征向量。

降噪自编码器网络结构如图 2 所示。

方式来计算均值和标准差,具体流程如下:设滑动窗口大小为 ω , 则在时刻 t 的均值和标准差需要通过下式表示:

$$\mu_t = \frac{1}{\omega} \sum_{i=t-\omega+1}^t x_i \quad (5)$$

$$\sigma_t = \sqrt{\frac{1}{\omega} \sum_{i=t-\omega+1}^t (x_i - \mu_t)^2} \quad (6)$$

其中, ω 是滑动窗口的大小, i 是窗口中的时间序列的下标。

2.4 注意力模块

1) 不稳定注意力机制

传统的 Transformer 模型中注意力权重是固定不变的,即每个 Query 向量和每个 Key 向量之间的相似度是固定的,不会随着序列位置的变化而改变^[13]。而金融高频交易数据中的非平稳部分的数据通常表现为趋势或者周期变化,这些变化的幅度和频率可能会随着时间的推移而改变,如果使用固定的注意力权重,模型可能会将这些变化视为噪声进行过滤,导致预测的序列失去了数据原有的趋势

或者周期变化,模型的预测效果下降。相比之下,不稳定注意力机制可以动态调整注意力权重,根据序列的时间信息来改变权重大小和分配,更好的还原

数据中的非平稳属性,这种机制可以让模型更好的适应数据中的变化,提高模型的预测性能^[14],具体结构如图 3 所示。

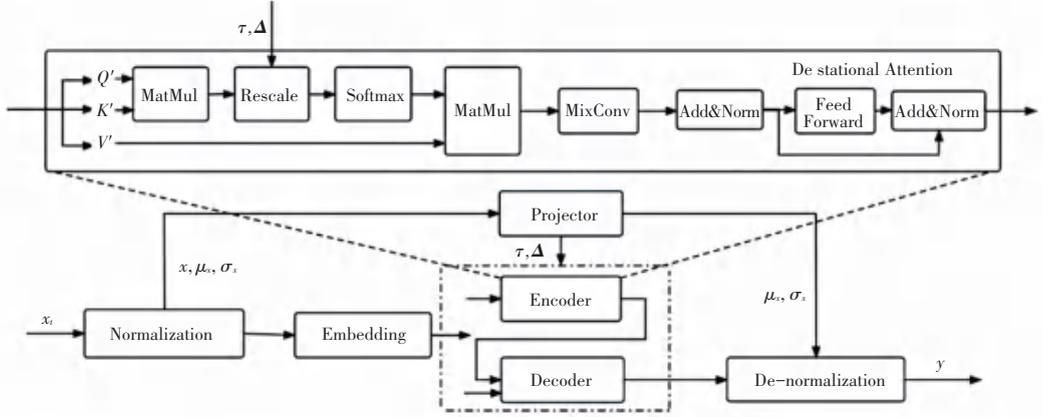


图 3 注意力机制结构图

Fig. 3 Structure diagram of attention mechanism

为了使注意力机制能够根据不同的时间段交易数据的序列信息调整自身的权重,增强模型的适应能力,本文将平稳化模块得到的每个窗口的交易数据的方差 σ_x 和均值 μ_x 输入到多层感知机 MLP (Multilayer Perceptron) 中,计算得到一个控制变量 τ 和移位向量 Δ , 具体的计算公式如下:

$$\log \tau = \text{MLP}(\sigma_x, x) \quad (7)$$

$$\Delta = \text{MLP}(\mu_x, x) \quad (8)$$

其中, x 为输入序列; σ_x 为序列 x 的方差; μ_x 为序列 x 的均值。

在模型学习的过程中,注意力模块的权重会根据交易数据的因子特征向量不同进行动态调整,利用计算得到的 τ 和 Δ 动态调整 Key 向量,并根据序列位置 i 进行调整。在训练的过程中控制变量 τ 和移位向量 Δ 也会根据因子特征向量的不同不断地反向传播来优化其值。

根据学习到的 Q, K 和 V 向量,计算注意力分数 $e_{i,j}$:

$$e_{i,j} = \tau Q_i K_j^T + \Delta \quad (9)$$

其中, Q_i 表示第 i 个位置的 Query 向量; K_j 表示第 j 个位置的 Key 向量; T 表示向量的转置操作。

使用 Softmax 函数将注意力分数归一化,得到注意力权重 $a_{i,j}$:

$$a_{i,j} = \frac{\exp(e_{i,j})}{\sum_{k=1}^n \exp(e_{i,k})} \quad (10)$$

最后,使用注意力权重对 Value 向量进行加权求和,得到最终的输出向量:

$$y_i = \sum_{j=1}^n a_{i,j} V_j \quad (11)$$

其中, V_j 表示第 j 个位置的 Value 向量。

2) 混合卷积层

在金融高频交易数据预测时,添加混合卷积层可以帮助不稳定注意力机制捕捉序列中的时空关系,增强非平稳变压器的表现^[15-16]。混合卷积层结合了 1D 卷积和逐点卷积,同时还引入了一个可学习的权重向量,可以在时间和因子特征维度上对输入进行有效的滤波和变换^[17]。对于金融时间序列预测问题,时间维度能够提供序列中不同时间点之间的关系。混合卷积层可以通过对时间维度的滤波来捕捉时间序列中的时空关系,在因子特征维度上也可以进行变换以捕捉更高级别的因子特征,使模型能够更好地学习到非平稳序列中的信息,这种方法还可以减少模型的参数数量,提高模型的计算效率。

经过注意力机制得到一个加权平均后的因子特征向量为 y_i , 混合卷积通过使用 n 个不同大小的卷积核得到的因子特征向量进行混合卷积操作,计算如下式所示:

$$y = \sum_{i=1}^n \omega_i \cdot \text{Conv}(X, k_i) \quad (12)$$

其中, $\text{Conv}(X, k_i)$ 表示对输入序列 X 应用大小为 k_i 的卷积核得到的输出, ω_i 表示第 i 个卷积核的权重。

为了保证权重 ω_i 的和为 1,需要使用 Softmax 函数对每个输出的低维交易数据的向量权重进行归一化处理,即:

$$\omega_i = \frac{\exp(a_i)}{\sum_{j=1}^n \exp(a_j)} \quad (13)$$

其中, a_i 表示第 i 个通过训练得到的卷积核的权重参数。

最后,为了保持序列长度不变,在卷积操作后添加了残差连接和归一化层:

$$y = \text{LayerNorm}(X + y) \quad (14)$$

其中,LayerNorm 代表归一化操作, $X + y$ 表示将上一层的输出结果和当前层的输入相加得到的结果。

通过这种方法,将因子特征维度变换以捕捉更高级别的因子特征,使得模型能够更好地学习时间序列的非平稳性,从而提高时间序列预测的性能。

2.5 去规范化模块

在注意力子层输出交易数据特征张量后,需要将张量数据转换回原始尺度,以便能够理解和解释分析结果,因此需要进行去规范化处理,将规范化后的数据转换为原始数据的形式。去规范化过程可以看作是规范化过程的逆操作,对于一个交易数据特征张量 z_i , 其在时刻 t 的去规范化值:

$$x_t = \sigma_t z_t + \mu_t \quad (15)$$

其中, μ_t 和 σ_t 分别代表 x_t 在 t 时刻的窗口数据的均值和标准差。

在去规范化时同样需要考虑到时间变化的因素,因此需要计算每个时刻的均值和标准差。同样地,使用滑动窗口的方式来计算均值和标准差。

3 实验结果分析

3.1 实验数据

本文从两个方面验证模型的性能,一方面使用2021年9月至2022年4月上证50ETF基金真实的交易数据集,将模型和多个主流的基准模型进行对比实验,验证本文提出的模型对金融高频交易数据的波动率预测的性能;另一方面使用国外市场126只股票验证模型对不同类别的股票的波动率预测效果,检验模型的稳定性和鲁棒性。

3.2 模型预测误差对比实验

1) 本文预测模型和基准模型预测误差对比实验

为了对比和分析本文所提出的基于降噪编码器和改进的不稳定注意力机制模型的预测效果,将该模型和其他的基准模型进行对比预测对比实验。本文用于对比的模型包括传统的机器学习模型,常用的时间序列预测模型和未改动的 ns Transformer,机器学习模型包括 KFord_LGBM、KFord_LGBM_NN、

Wavelet_LGBM_NN 模型;常用的时间序列预测的深度学习方法 (Long Short - Term Memory)、CNN (Convolutional Neural Network)、DAE_LSTM、DAE_CNN 模型。使用 50ETF 高频交易数据集,采用 R^2 (R -Square)、RMSPE (Root Mean Square Percentage Error) 和 MAE (Mean Absolute Error) 作为模型的评价指标,结果见表 1。

表 1 模型对比实验

Table 1 Comparison experiments of different models

模型	R^2	RMSPE	MAE
KFord_LGBM	0.840 1	0.236 1	5.720e ⁻⁴
KFord_LGBM_NN	0.847 1	0.233 5	5.689e ⁻⁴
Wavelet_LGBM_NN	0.761 2	0.310 8	7.035e ⁻⁴
LSTM	0.862 8	0.232 0	5.665e ⁻⁴
CNN	0.850 2	0.219 6	5.622e ⁻⁴
DAE_LSTM	0.889 1	0.196 2	4.109e ⁻⁴
DAE_CNN	0.882 4	0.191 9	4.151e ⁻⁴
ns_Transformer	0.859 0	0.225 5	4.701e ⁻⁴
本文	0.914 7	0.157 1	3.302e⁻⁴

由表 1 可知,传统的机器学习模型在测试集上的效果较差,在对数据使用 wavelet 平稳化之后,效果反而降低,深度学习模型结合 DAE 效果有明显提升,而深度学习模型的预测效果不如本文提出的改进后的深度学习模型。本文提出的模型在所有的模型中具有最好的预测效果和最好的拟合效果,和 LSTM 模型相比, R^2 由 0.862 8 提升到 0.914 7, RMSPE 由 0.202 0 下降到 0.157 1, MAE 由 5.665 e⁻⁴ 下降到 3.302e⁻⁴。实验结果证明本文提出的模型预测的准确度相比于其他的模型有很大的提升。

2) 本文预测模型和基准模型鲁棒性对比实验

为了验证本文模型对不同股票市场的金融交易数据的预测性能,本文设置了模型预测鲁棒性对比实验,对国外金融市场的 126 只股票交易数据进行聚类,将聚类后的股票数据作为数据集验证模型是否具有很好的普适性和鲁棒性。通过利用 Kmeans 对 126 只股票数据进行聚类的结果如图 4 所示。

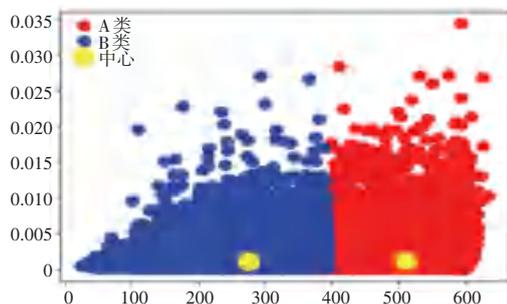


图 4 126 只股票的聚类结果图

Fig. 4 Cluster of 126 stocks

基于波动率对股票进行聚类是风险管理和投资组合优化中常用的技术。通过对数据波动模式相似的股票进行分组,根据每只股票的占比将126只股票交易数据分为A类和B类两类,接着将常用深度学习模型 DAE_LSTM、DAE_CNN 和未改进的 ns_Transformer 与本文提出的模型进行对比实验,实验结果见表2、表3。

表2 A类股票数据模型预测实验结果

Table 2 Experimental results of Class A stock data model prediction

模型	R^2	$RMSPE$	MAE
DAE_LSTM	0.921 5	0.152 7	$3.259 e^{-4}$
DAE_CNN	0.900 2	0.181 9	$3.551 e^{-4}$
ns_Transformer	0.858 1	0.225 5	$4.621 e^{-4}$
本文	0.923 5	0.147 8	$3.122 e^{-4}$

表3 B类股票数据模型预测实验结果

Table 3 Experimental results of Class B stock data model prediction

模型	R^2	$RMSPE$	MAE
DAE_LSTM	0.772 1	0.336 3	$6.537 e^{-4}$
DAE_CNN	0.790 5	0.301 5	$6.201 e^{-4}$
ns_Transformer	0.748 7	0.440 9	$7.105 e^{-4}$
本文	0.874 7	0.207 1	$4.202 e^{-4}$

从表2和表3的结果能够看出,对比模型对A类股票的预测效果有了明显提升,但是对B类股票的预测效果明显下降。本文提出的模型对两类股票高频交易数据的波动率预测都有较好的预测效果,对A类股票的预测的拟合程度能够达到0.9235,平均平方百分比误差和绝对误差降低到了0.1478和 $3.122e^{-4}$;对于B类股票数据的预测精度也是所有对比中最高的, R^2 达到了0.8747, $RMSPE$ 和 MAE 降低到了0.2071和 $4.202e^{-4}$,相比对比模型都有很好的预测效果,实验说明模型具有更好的鲁棒性和普适性。

4 结束语

本文提出基于降噪自动编码器和不稳定注意力的深度学习模型,在上交所的真实交易数据集和国外金融市场股票数据验证了本文提出模型改进后的预测有效性和鲁棒性。实验的结果表明,本文提出的模型的预测精度优于对比模型,能够达到最好的预测效果。

参考文献

- [1] 汪训孝. 基于高频数据金融市场波动率研究[D]. 上海: 上海交通大学, 2016.
- [2] MENKVELD A J. High frequency trading and the new market makers[J]. Journal of Financial Markets, 2013, 16(4): 712-740.
- [3] POON S H, GRANGER C W J. Forecasting volatility in financial markets: A review[J]. Journal of Economic Literature, 2003, 41(2): 478-539.
- [4] VALENTI D, FAZIO G, SPAGNOLO B. Stabilizing effect of volatility in financial markets[J]. Physical Review E, 2018, 97(6): 062307.
- [5] AIT-SAHALIA Y, FAN J, XIU D. High-frequency covariance estimates with noisy and asynchronous financial data[J]. Journal of the American Statistical Association, 2010, 105(492): 1504-1517.
- [6] LUX T. Time variation of second moments from a noise trader/infection model[J]. Journal of Economic Dynamics and Control, 1997, 22(1): 1-38.
- [7] CHEN R Y, MYKLAND P A. Model-free approaches to discern non-stationary microstructure noise and time-varying liquidity in high-frequency data[J]. Journal of Econometrics, 2017, 200(1): 79-103.
- [8] KERCHEVAL A N, ZHANG Y. Modelling high-frequency limit order book dynamics with support vector machines [J]. Quantitative Finance, 2015, 15(8): 1315-1329.
- [9] WANG L, ZHANG Z, CHEN J. Short-term electricity price forecasting with stacked denoising autoencoders [J]. IEEE Transactions on Power Systems, 2016, 32(4): 2673-2681.
- [10] ZHOU X, PAN Z, HU G, et al. Stock market prediction on high-frequency data using generative adversarial nets[J]. Mathematical Problems in Engineering, 2018, 2018(1): 4907423.
- [11] YANG P R. Forecasting high-frequency financial time series: an adaptive learning approach with the order book data [J]. arXiv preprint arXiv:2103.00264, 2021.
- [12] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders [C]// Proceedings of the 25th International Conference on Machine Learning. IEEE, 2008: 1096-1103.
- [13] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer [J]. Journal of Machine Learning Research, 2020, 21(1): 5485-5551.
- [14] LIU Y, WU H, WANG J, et al. Non-stationary transformers: Rethinking the stationarity in time series forecasting [J]. arXiv preprint arXiv:2205.14415, 2022.
- [15] WU B, WAN A, YUE X, et al. Shift: A zero flop, zero parameter alternative to spatial convolutions [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 9127-9135.
- [16] DING M, LIAN X, YANG L, et al. Hr-nas: Searching efficient high-resolution neural architectures with lightweight transformers [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 2982-2992.
- [17] TAN M, LE Q V. Mixconv: Mixed depthwise convolutional kernels [J]. arXiv preprint arXiv:1907.09595, 2019.