

基于多步分解的股价预测模型

李 秀

兰州交通大学数理学院, 甘肃 兰州

收稿日期: 2024年6月23日; 录用日期: 2024年7月16日; 发布日期: 2024年7月24日

摘 要

金融领域一直备受关注, 且股价受多种因素影响, 其预测也存在一定的挑战性。为准确预测股价, 为投资者和交易者提供有益的决策支持, 本研究旨在提出一种基于多步分解的股价预测模型。首先通过变分模态分解(VMD)分解原始序列, 重构高复杂成分, 然后应用鲁棒局部均值分解(RLMD)进行二次分解, 最后利用PSO-LSTM模型进行预测。为验证所提模型的有效性, 将股票数据经VMD-MFE-RLMD分解与没有分解、只有VMD分解以及传统模型CNN、SVR、GRU进行对比, 在沪深300指数数据集上的结果显示: 股票数据经VMD-MFE-RLMD分解的预测误差MAE、MSE、RMSE、MAPE均小于没有分解以及只经VMD分解的预测误差, 且低于传统预测模型的预测误差, 提高了预测精度。最后, 将此模型应用在上证50指数数据集上, 同样取得了较好的预测结果, 再次证明了所提模型在股价预测上具有更高的预测精度。

关键词

股价预测, 二次分解, 变分模态分解, 鲁棒局部均值分解, 长短期记忆网络

A Stock Price Prediction Model Based on Multi-Step Decomposition

Xiu Li

School of Mathematics and Physics, Lanzhou Jiaotong University, Lanzhou Gansu

Received: Jun. 23rd, 2024; accepted: Jul. 16th, 2024; published: Jul. 24th, 2024

Abstract

The financial sector has always been highly regarded, and stock prices are influenced by various factors, making their predictions somewhat challenging. To accurately predict stock prices and provide useful decision support for investors and traders, this study aims to propose a stock price prediction model based on multi-step decomposition. Firstly, the original sequence is decomposed using Variational Mode Decomposition (VMD) and the high complexity components are recon-

structured. Then, robust Local Mean Decomposition (RLMD) is applied for secondary decomposition, and finally, the PSO-LSTM model is used for prediction. To verify the effectiveness of the proposed model, the stock data was decomposed by VMD-MFE-RLMD and compared with non-decomposition, only VMD decomposition, and traditional models such as CNN, SVR, and GRU. The results on the Shanghai and Shenzhen 300 Index dataset showed that the prediction errors MAE, MSE, RMSE, and MAPE of stock data decomposed by VMD-MFE-RLMD were smaller than those without decomposition and only VMD decomposition, and lower than those of traditional prediction models, improving prediction accuracy. Finally, applying this model to the Shanghai Stock Exchange 50 Index dataset also achieved good prediction results, once again proving that the proposed model has higher prediction accuracy in stock price prediction.

Keywords

Stock Price Prediction, Quadratic Decomposition, Variational Mode Decomposition, Robust Local Mean Decomposition, Long Short-Term Memory Network

Copyright © 2024 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

股价预测一直备受金融领域的关注，准确预测股价走势，对投资者和交易者来说至关重要。然而，股价受到多种因素的影响，包括市场供需关系、宏观经济指标、公司业绩等，其预测具有一定的挑战性。因此，寻找一种准确且可靠的股价预测方法一直是学术界和实践中的关注焦点。

由于近年来数据科学和机器学习技术发展迅速，研究人员开始尝试利用这些技术进行股价预测。然而，传统的机器学习方法往往只能捕捉到股价数据的表面特征，无法充分考虑其内在的趋势、随机波动等特征。因此，需要提出更加准确和全面的股价预测模型。鉴于股票等金融价格是由多种因素引起的，如趋势、季节和噪音，研究人员越来越多的重点研究了时间序列的分解，并将时间序列的分解和机器学习相结合，提出了综合预测模型，具有良好的预测性能和泛化能力。Song & Huang 等[1]提出 VMD-EEMD 技术，证明多重价格分解作为输入型工程技术，是提高金融股票数据深度学习模型预测准确率的有效途径。Wang & Cui 等[2]提出了一种基于二次分解、多因素分解和基于注意力的 LSTM 混合模型，证明了二次分解方法在提取原始数据中隐藏的非线性关系方面优于单一分解方法。

对此本文提出了 VMD-MFE-RLMD 技术，我们将 VMD-MFE-RLMD 二次分解技术与 PSO-LSTM 模型相结合，构建混合模型。该模型首先使用 VMD 将数据进行分解，对于高复杂性数据则通过 RLMD 进行二次分解，以充分分解原始股票价格时间序列，从而提高对复杂信号的预测能力。

2. 模型原理

2.1. VMD

VMD 是将复杂信号分解为具有不同频率特性的成分，是一种信号处理方法。其基本思想是通过迭代优化过程将信号分解成多个具有不同频率的本征模态函数(IMFs)，并通过优化问题的求解获得最佳的 IMFs 分量[3]，从而更好地理解和分析原始信号，并能充分考虑信号的时频特征，捕捉到信号中的局部和全局特征，具有较好的自适应性。VMD 算法的操作步骤如下：

步骤 1: 利用希伯特变换求出每个模态的解析信号, 得到单向频谱。

步骤 2: 将解析信号与对应的中心频率 $e^{-jw_k t}$ 混合, 并向基带移动各模态频谱。

步骤 3: 通过解调信号的高斯平滑度, 即梯度的 L2 范数来估计带宽。

$$\begin{cases} \min \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{k=1}^K u_k = f \end{cases} \quad (1)$$

步骤 4: 引入二次惩罚函数 ε 和拉格朗日乘子 ξ , 将有约束问题转化为无约束问题。

$$\begin{aligned} L(u_k, w_k, \xi_k) = & \varepsilon \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-jw_k t} \right\|_2^2 + \|f(t) - u_k(t)\|_2^2 \\ & + \left[\xi(t), f(t) - \sum_{k=1}^K u_k(t) \right] \end{aligned} \quad (2)$$

步骤 5: 利用交替方向乘法求最优解。求得等式(1)的最优解的方法是交替更新。

u_k^{s+1} 、 w_k^{s+1} 、 ξ_k^{s+1} (s 为迭代次数)。其模态分解及中心频率更新方向如下:

$$u_k^{s+1}(w) = \frac{f(w) - \sum_{i \neq k} u_i(w) + \frac{u(w)}{2}}{1 + 2\varepsilon(w - w_k)^2} \quad (3)$$

$$w_k^{s+1} = \frac{\int_0^\infty w |u_k(w)|^2 dw}{\int_0^\infty |u_k(w)|^2 dw} \quad (4)$$

在上述各式中, u_k 为模态分量, w_k 为模态对应的中心频率, $\delta(t)$ 为单位脉冲函数。

2.2. RLMD

RLMD 的原始版本, 也被称为局部平均分解(LMD) [4]算法, 依赖于将一个整体信号分解成一个较少通信的子信号子集的思想, 这是一种分析时频数据集的自适应方法。RLMD 算法将不稳定的信号分解为其乘积函数[5], 该算法可以从股票收盘价序列 $x(t)$ 中抽离纯调频信号和包络信号, 两者相乘得到 PF 分量, 并进一步得到原序列中没有 PF 分量的剩余信号, 这一过程重复至剩余信号变为常数或单调。最终得到 k 个分量和一个残差分量。

$$x(t) = \sum_{i=1}^k PF_i(t) + u(t) \quad (5)$$

然而, LMD 也会受到末端效应和模态混合问题的影响, 这可以通过集成局部平均分解(ELMD)来改善[6]。类似于 EEMD, ELMD 在滑动平滑过程中同样也存在局部均值发散问题。为了解决这些问题, RLMD 在 LMD 的基础上, 对移动平均算法及其筛选过程进行了优化。RLMD 算法是通过设定初始条件、估计信号包络和筛选终止准则来解决问题的一种具有广阔前景的序列信号处理工具。在第一步中, RLMD 算法使用图像扩展算法确定序列信号左右两端的对称点, 然后计算最佳子集 λ^* 来估计包络。最后, 根据连续三次迭代的关系, 最小化积函数 f 来筛选停止准则。

$$\lambda^* = \text{odd}(u_s + 3 * \delta_s) \quad (6)$$

$$u_s = \sum_{i=1}^{K-1} s(i) f(i) \quad (7)$$

$$\delta_s = \sqrt{\sum_{i=1}^{K-1} [s(i) - u_s]^2 \cdot f(i)} \quad (8)$$

$$f = RES[z(t)] + EK[z(t)] \quad (9)$$

$$RMS[z(t)] = \sqrt{\frac{1}{n} \sum_{t=1}^n [z(t)]^2} \quad (10)$$

$$EK[z(t)] = \frac{\frac{1}{n} \sum_{t=1}^n [z(t) - \bar{z}]^4}{\frac{1}{n} \sum_{t=1}^n [z(t) - \bar{z}]^2} - 3 \quad (11)$$

$$\bar{z} = \frac{1}{n} \sum_{t=1}^n z(t) \quad (12)$$

其中, $s(i)$ 为局部均值信号或局部幅度信号之间的步长, u_s 为 $s(i)$ 的均值, δ_s 为 $s(i)$ 的标准差, $odd(\cdot)$ 提供大于或等于输入的最小奇数, $z(t)$ 为零基线包络信号。

2.3. LSTM

LSTM 是一种专门针对序列数据处理的递归神经网络模型, 是循环神经网络(RNN)的改进, 消除了梯度消失和梯度爆发问题, 由于能更好地发现长期依赖关系而被广泛用于处理序列信息[7]。它引入了控制信息流动的门控制机制, 由两部分组成: Sigmoid 神经网络层和点乘法运算。LSTM 包含决定遗忘信息、导入新信息、输出信息的三个门即遗忘门、输入门和输出门, 主要作用是对细胞的状态进行维护和调节。其基本结构如图 1 所示。

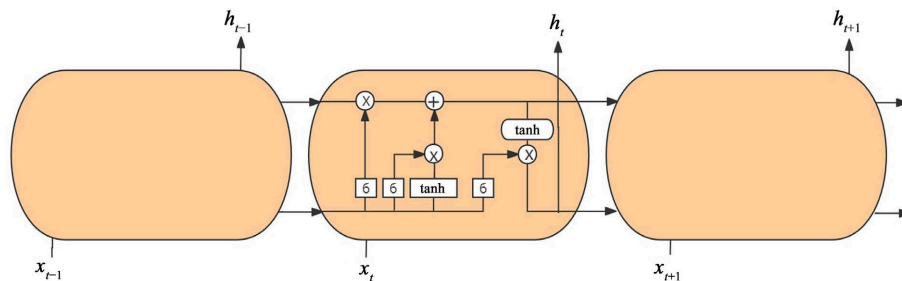


Figure 1. LSTM neuron structure diagram

图 1. LSTM 神经元结构图

1) 遗忘门, 首先需要在遗忘门中通过 Sigmoid 层确定要丢弃的信息并加以实现。通过 h_{t-1} 和 x_t 来确定单元格状态中数字的保留情况。

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (13)$$

2) 输入门, 接着需要在细胞状态下确定所保留的数据。Sigmoid 层决定了需要更新的值, Tanh 层创建一个候选向量加入到细胞状态中, 结合这两个向量来创建更新值, 并计算遗忘的部分来更新候选值。

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (14)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (15)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (16)$$

3) 输出门, 最后确定需要输出的内容。通过 Sigmoid 层进行判断细胞状态应该输出哪些部分, 然后使用 tanh 层规范化单元格状态的值到-1 和 1 之间, 并通过 Sigmoid 门的来确定输出部分。

$$O_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (17)$$

$$h_t = O_t * \tanh(C_t) \quad (18)$$

3. 基于双分解的股价预测

本文提出的基于多步分解的股价预测研究的框架为: 股票收盘价原始序列进行预处理后, 依次执行 VMD 分解、MFE 计算、RLMD 分解, 最后输入到预测模型进行预测。具体流程主要分为五步: 第一步是将经预处理后的股票收盘价原始序列进行 VMD 分解, 生成 n 个分量和一个残余项; 第二步为分别计算各分量及残余项的多尺度模糊熵(MFE), 将 MFE 较高的分量重构为高复杂性数据, 其余分量不做处理; 第三步使用 RLMD 对高复杂性数据进行分解, 将其拆分为复杂度较低子序列; 第四步将一次分解的其余分量和二次分解得到的分量分别输入到 PSO-LSTM 预测模型中, 生成单个成份预测; 第五步加总各分量的预测结果得到股价预测的最终结果。具体模型框架如图 2 所示。

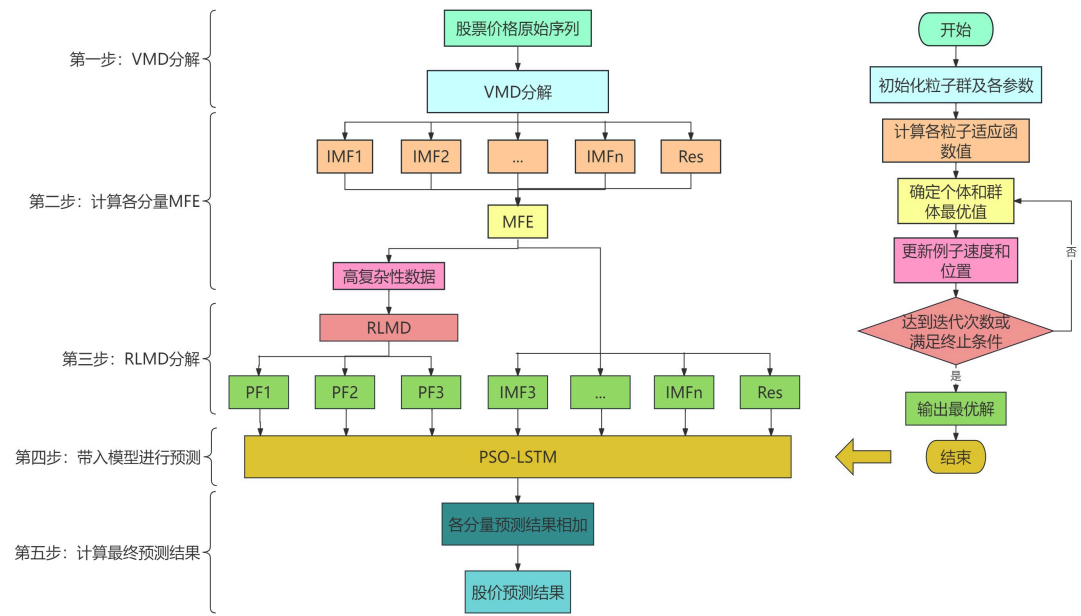


Figure 2. A framework for stock price prediction based on multi-step decomposition
图 2. 基于多步分解的股价预测模型框架

4. 实证研究

4.1. 数据描述

4.1.1. 数据来源

沪深 300 指数是由 300 只股票构成的样本指数, 具有较大的规模和良好的流动性, 被认为可以较好

地代表整个 A 股市场的运行状况。该指数包含了不同行业、不同市值的公司股票，因此对于预测股市整体变化具有一定的泛化能力。这些股票通常有良好的交易流动性，能够反映市场整体的投资情况、交易情况和风险偏好。基于该指数进行预测，可以探讨股票市场的整体走势及行情走势，有助于提高预测的准确性，更好地捕捉市场的整体运行状况，故本文选取沪深 300 指数作为分析预测的标的，以期更全面地了解股票市场行情。本文数据是在东方财富官网上进行爬取的，采用的模型是通过 MATLAB2016 和 Python3.10 实现的。数据起始时间为 2018 年 01 月 12 日，终止时间为 2024 年 01 月 12 日，共 1458 条数据。本文仅预测收盘价这一列时间序列，只保留两列数据，分别为日期及收盘价。沪深 300 指数原始数据见表 1，收盘价的趋势图如图 3 所示。

Table 1. Raw data of the Shanghai and Shenzhen 300 index
表 1. 沪深 300 指数原始数据

日期	开盘价(元)	收盘价(元)	最高价(元)	最低价(元)	成交量(股)	成交额(元)
2018-01-12	4205.14	4225.00	4227.39	4199.03	111,354,793	157,606,000,000
2018-01-15	4229.84	4225.24	4262.93	4216.36	170,907,896	234,987,000,000
2018-01-16	4215.62	4258.47	4260.21	4213.13	161,494,963	228,271,000,000
...
2024-01-10	3281.67	3277.13	3312.92	3268.25	94,199,366	159,778,000,000
2024-01-11	3275.31	3295.67	3311.11	3266.79	108,789,735	180,979,000,000
2024-01-12	3284.36	3292.89	3310.47	3279.70	68,159,762	110,710,000,000



Figure 3. Trend chart of the closing price of the Shanghai and Shenzhen 300 index
图 3. 沪深 300 指数收盘价的趋势图

4.1.2. 描述性统计

为更加细致、准确地了解数据，本文对沪深 300 指数进行了描述性统计。见表 2。

Table 2. Descriptive statistics on the closing price of the Shanghai and Shenzhen 300 index
表 2. 沪深 300 指数收盘价描述性统计

最大值	最小值	中位数	均值	标准差	偏度	峰度
5807.7200	2964.8400	3983.375	4129.8831	593.8921	0.4878	-0.5331

表2显示,沪深300指数收盘价的均值为4129.8831,中位数为3983.375。收盘价数据的偏度为0.4878,大于0,为右偏分布。峰度为-0.5331,小于0,其分布较为平坦。因此在描述性统计中没有指标表明沪深300指数收盘价数据服从正态分布,为有效分析股票价格数据的相关信息,本文选择了适合处理非平稳性和非线性波动特征的方法对沪深300指数收盘价进行处理。

4.2. 数据处理

4.2.1. 数据预处理

本研究的数据来源于东方财富官网,由于节假日没有数据,传统的缺失值处理方法一般是采用插值法或者用平均值进行代替,考虑到数据的真实性,本文未对缺失数据进行处理,而是采用原始数据进行分析。对于原始时间序列需要进行训练集和测试集的划分,按照7:3的比例进行划分,即前70%的数据作为训练集,后30%的数据作为测试集。训练集、测试集划分前需要对数据进行滑动窗口处理,窗口大小设置为60,以第*i*至*i+59*天的数据作为模型的输入,以第*i+60*天的数据作为模型的输出,即利用最近60日内的数据信息对未来一日收盘价进行预测,以此来构建预测模型的训练和预测。数据集的构建如图4所示。

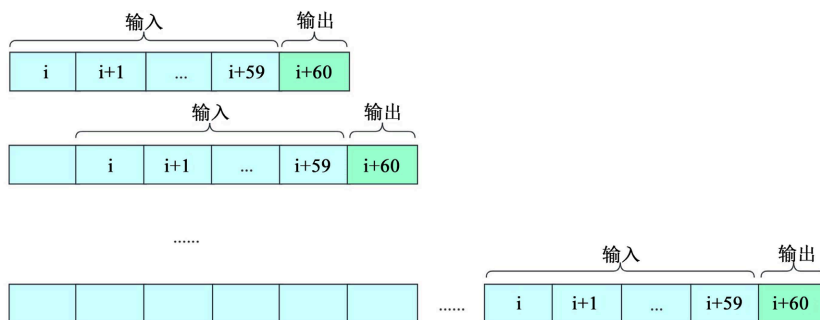


Figure 4. Dataset construction

图4. 数据集构建

4.2.2. 数据归一化

归一化的目的是确保不同数据具有可比性,同时统一它们的量纲和数量级。对模型可靠性的提升有一定的帮助,同时也可以增强算法收敛的效率。通过归一化,我们可以消除特征间由于量纲和数量级不同而产生的偏差,使得模型更准确地学习到数据之间的内在关系,进而提高整体训练效果。

$$X = \frac{x - \max(x)}{\max(x) - \min(x)} \quad (19)$$

4.3. 指标评价

评估函数用于量化预测结果与实际观测值的差异,同时评估模型的预测精度和拟合程度,对预测结果的评价起着重要的作用,在模型评估方面扮演着非常关键的角色。可用于衡量预测模型的整体性能,从多个角度来评估模型的准确性、泛化能力和稳定性。本研究主要通过常用指标对预测结果进行评估,包括平均绝对值误差、均方误差、均方根误差、平均绝对值百分比误差,即MAE、MSE、RMSE、MAPE,其值越小表示预测结果与实际值之间的差异越小,模型的准确性和精度越高。计算公式如下:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (20)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (21)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (22)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (23)$$

在上述各式中, y_i 为真实值, \hat{y}_i 为预测值。

4.4. 实验结果及分析

4.4.1. VME-MFE-RLMD 分解

预处理后的股票收盘价的原始时间序列第一步要经 VMD 分解, 根据 VMD 算法的分解原理, 其分解结果受分解个数 k 影响, 因此需要事先确定原始信号的分解个数 K 。基于这个问题, 本研究考虑 CEEMDAN 分解不需要事先设定分解个数, 可以自适应地确定原始信号的分解个数[8]。因此, 将 CEEMDAN 分解的个数作为 VMD 算法的 K 值, 经验证 CEEMDAN 分解的分量个数为 8, 因此本研究将 VMD 算法中的 K 值设置为 8, 每个分量代表原始时间序列信号内部的不同信号特征, 确定分解个数 K 后, 使用 VMD 算法对碳价数据进行分解, 分解结果如图 5 所示。为了更科学、更准确地理解初级分解产生的 IMF 的复杂程度, 用 MFE 来形容其复杂程度。MFE 利用了模糊神经网络以及多尺度神经网络的优势, 并且具有非线性特点[9]。它减少了对初始参数的依赖, 可以描述不同阈值下的数据, 具有较强的抗干扰能力。各分量模糊熵值见表 3, 模糊熵值对应的折线图如图 6 所示, 可以看出 IMF1 与 IMF2 的复杂程度相对较高, 因此将 IMF1 与 IMF2 合并成高复杂性分量, 使用 RLMD 对其进行分解, 其余成分保持不变。分解结果如图 7 所示。

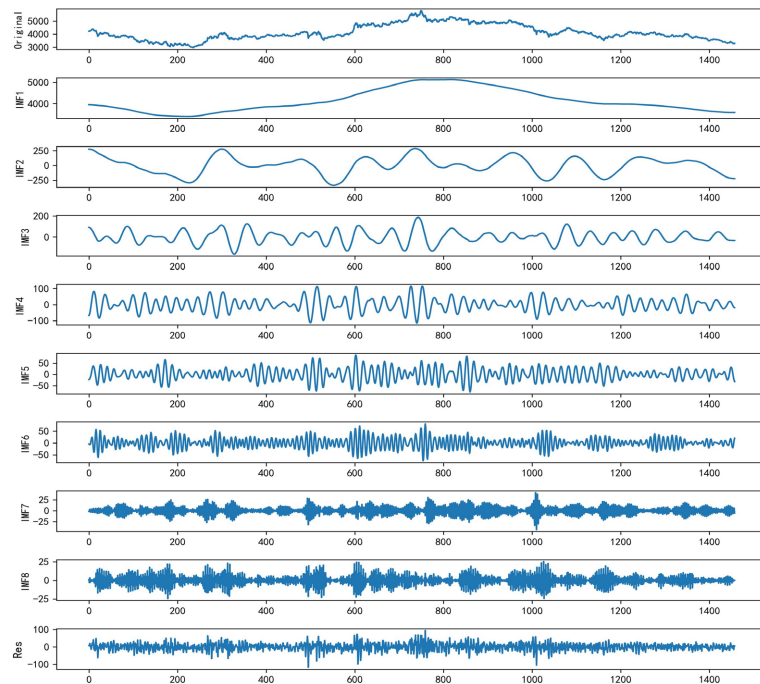


Figure 5. VMD decomposition chart of the closing price of the CSI 300 index

图 5. 沪深 300 指数收盘价 VMD 分解图

Table 3. Fuzzy entropy values of each component
表 3. 各分量模糊熵值

分量	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	Res
模糊熵值	3.0036	2.9757	2.6617	2.6482	2.5812	2.3503	2.1019	2.5311	1.9408

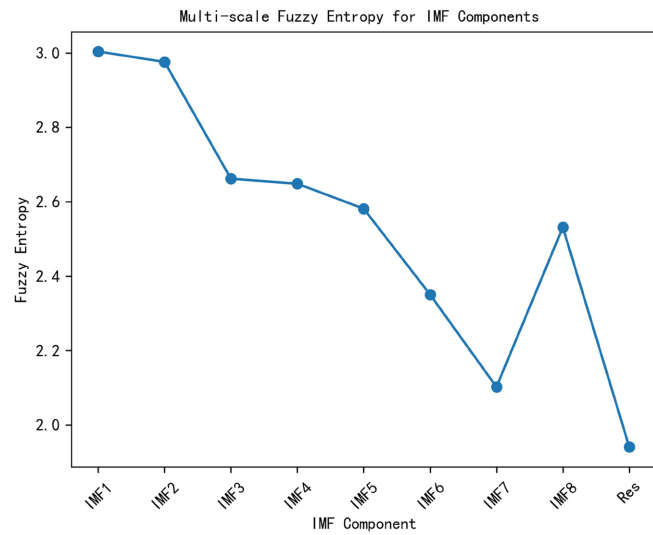


Figure 6. Line chart of fuzzy entropy values for each component
图 6. 各分量模糊熵值折线图

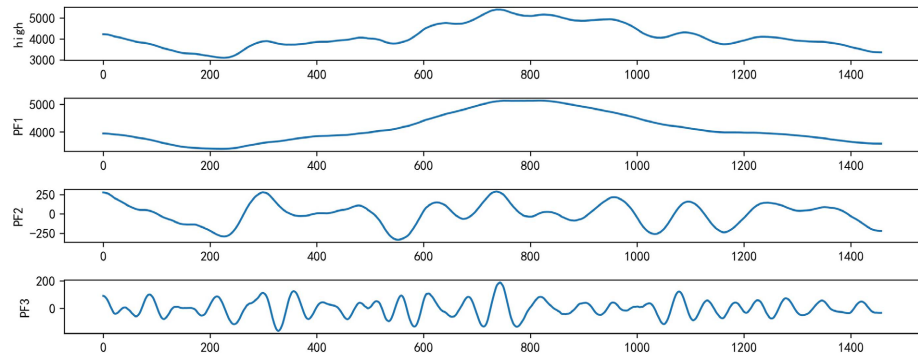


Figure 7. RLMD decomposition diagram of high complexity data
图 7. 高复杂数据 RLMD 分解图

4.4.2. PSO-LSTM 模型预测

为了使预测结果更加接近真实值，LSTM 部分参数的优化采用了粒子群优化算法(PSO)。在 PSO 中，将待优化问题看作是一个多维空间中的搜索问题，通过模拟鸟群中个体和群体的行为来不断调整解空间中的粒子位置，以寻找最优解[10]。经 VMD-MFE-LSTM 分解后的分量分别输入到 PSO-LSTM 模型中，对每个分量的 LSTM 模型均进行 PSO 优化，生成最优参数。本实验的 LSTM 网络包含了两层 LSTM 层，然后是两层 LSTM 层后的全连接层，并采用 ReLU 激活函数，最后是一个 Dense 层，Dense 层中只有一个神经元作为输出层。且模型的损失函数采用均方误差，优化器采用 Adam 优化器。本文选取了六个参数作为 PSO 算法的寻优对象，包括迭代次数、Batchsize、第一层 lstm 的输出神经元数、第二层 lstm 的输出神经元数、全连接层的神经元数、学习率。

为了验证本文提到的二次分解模型的有效性，分别对比了不进行分解直接预测、进行一次分解后预测、进行两次分解后预测的效果。

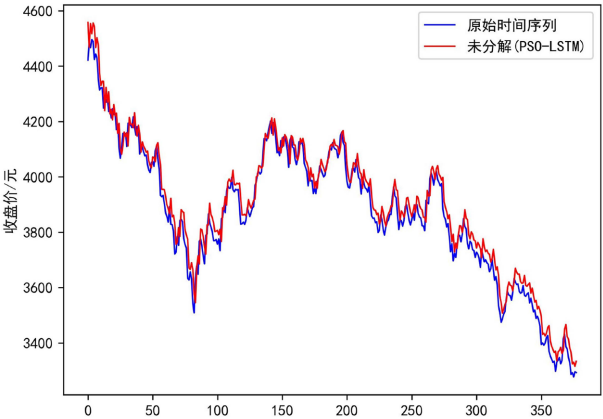


Figure 8. Trend chart of PSO-LSTM prediction without decomposition
图 8. 未分解下 PSO-LSTM 预测趋势图

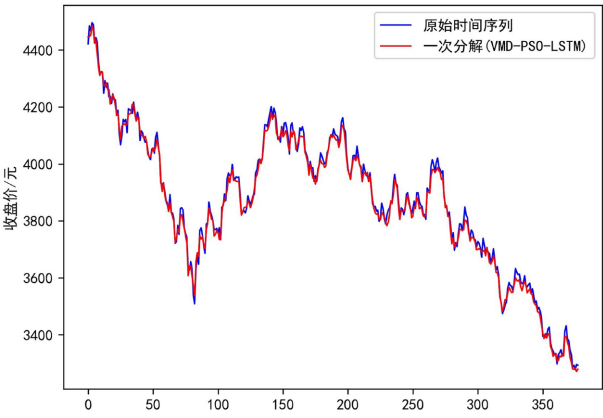


Figure 9. Trend chart of PSO-LSTM prediction under one decomposition
图 9. 一次分解下 PSO-LSTM 预测趋势图

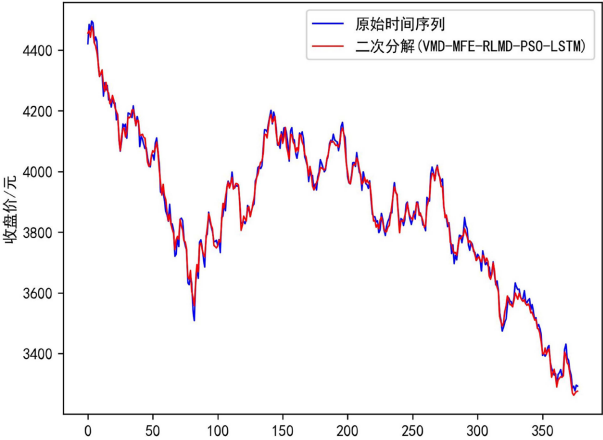


Figure 10. Trend chart of PSO-LSTM prediction under quadratic decomposition
图 10. 二次分解下 PSO-LSTM 预测趋势图

未分解状态、一次分解状态、二次分解状态下 PSO-LSTM 预测结果的趋势图如图 8~10 所示，从图中可以看出经二次分解得到的预测结果的趋势图比一次分解和未分解的趋势图更接近原始时间序列，更能够充分地反映出股票价格的变化趋势和波动状态。

Table 4. Evaluation results of predictive indicators for different decomposition times of the closing price of the Shanghai and Shenzhen 300 index

表 4. 沪深 300 指数收盘价不同分解次数预测指标评价结果

Name of the model	MAE	MSE	RMSE	MAPE (%)
LSTM	73.225	6718.6057	81.9671	1.9098
PSO-LSTM	40.187	2284.4309	47.7957	1.0489
VMD-PSO-LSTM	14.9198	324.7969	18.0221	0.386
VMD-MFE-RLMD-PSO-LSTM	13.8424	301.1527	17.3538	0.3596

PSO-LSTM 模型在二次分解、一次分解、未分解条件下的预测效果见表 4，股票收盘价原始序列未经分解直接用 PSO 优化后的 LSTM 进行预测，其 MAE、MSE、RMSE、MAPE 分别为 40.1870%、2284.4309%、47.7957%、1.0489%；原序列经 VMD 一次分解后输入 PSO 优化后的 LSTM 模型，其 MAE、MSE、RMSE、MAPE 为 14.9198%、324.7969%、18.0221%、0.3860%，均低于原始序列直接进行 PSO-LSTM 预测的结果；原序列经 VMD-MFE-RLMD 二次分解后，其 MAE、MSE、RMSE、MAPE 为 13.8424%、301.1527%、17.3538%、0.3596%，预测效果的指标评估均比直接进行 PSO-LSTM 预测、进行 VMD 分解后预测的结果低。一次分解相较于未分解，MAE、MSE、RMSE、MAPE 下降了 0.687%、0.8578%、0.629%、0.6320%；二次分解相较一次分解，MAE、MSE、RMSE、MAPE 下降了 0.0722%，0.0728%，0.0371%，0.0684%，即本文提出的二次分解下的股票价格预测相较于一次分解和未分解更精确，预测准确度更高。

4.5. 预测模型对比

除上述不同分解次数的对比，本文还将提到的模型与传统预测模型 CNN、SVR、GRU 进行了对比。最终预测结果如图 11 所示，各项指标的结果见表 5。

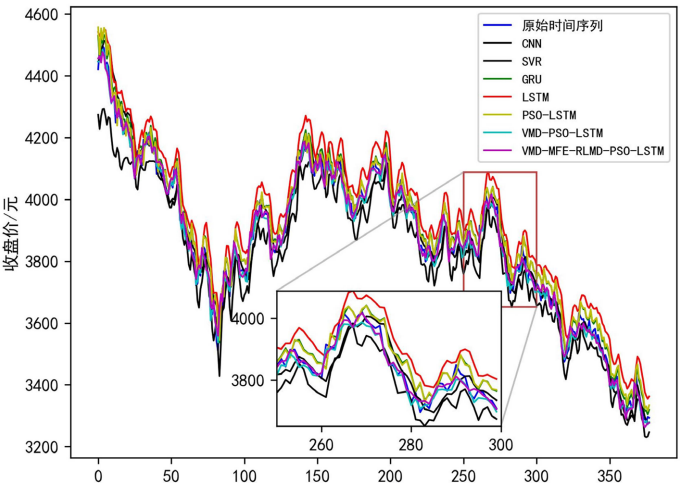


Figure 11. Comparison of Shanghai and Shenzhen 300 index prediction models
图 11. 沪深 300 指数预测模型对比

Table 5. Evaluation results of closing price prediction indicators for the Shanghai and Shenzhen 300 index
表 5. 沪深 300 指数收盘价预测指标评价结果

Name of the model	MAE	MSE	RMSE	MAPE (%)
CNN	61.8241	5645.5208	75.1367	1.5762
SVR	36.4257	2215.2301	47.0662	0.9365
GRU	39.8883	2272.3678	47.6694	1.0367
LSTM	73.225	6718.6057	81.9671	1.9098
PSO-LSTM	40.187	2284.4309	47.7957	1.0489
VMD-PSO-LSTM	14.9198	324.7969	18.0221	0.386
VMD-MFE-RLMD-PSO-LSTM	13.8424	301.1527	17.3538	0.3596

本文提出的模型 VMD-MFE-RLMD-PSO-LSTM 与传统模型相比，预测误差 MAE、MSE、RMSE、MAPE 均较传统预测模型有所降低。具体来说，相较于 CNN、SVR 和 GRU 模型，MAE、MSE、RMSE、MAPE 分别降低了 0.7761、0.9467、0.7690、0.7719%和 0.6200、0.8641、0.6313、0.6161%以及 0.6530、0.8675、0.6360、0.6532%，降低幅度较大。由此可得出，本文所提出的基于二次分解条件下 PSO-LSTM 模型的预测结果最好，预测精度较一次分解、未分解更高，比传统模型的预测更准确可靠，对股价的预测精确度有明显的改善作用。

除了本文使用的沪深 300 指数对所提模型进行了验证，本文还利用上证 50 数据来验证模型的有效性。最终预测结果如图 12 所示，各项指标的结果见表 6。

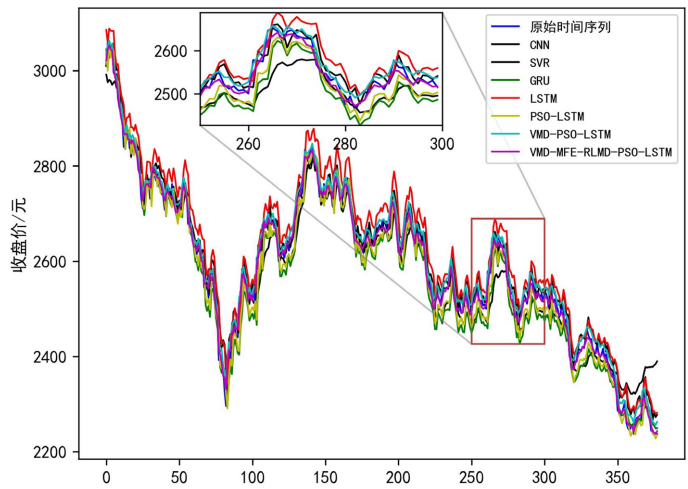


Figure 12. Comparison of Shanghai stock exchange 50 index prediction models
图 12. 上证 50 指数预测模型对比

Table 6. Evaluation results of Shanghai Stock Exchange 50 index closing price prediction indicators
表 6. 上证 50 指数收盘价预测指标评价结果

Name of the model	MAE	MSE	RMSE	MAPE (%)
CNN	24.7766	964.8563	31.0621	0.9686
SVR	33.9531	1925.6568	43.8823	1.3363

续表

GRU	31.5788	1533.4463	39.1592	1.223
LSTM	39.3423	2017.5263	44.9169	1.5231
PSO-LSTM	25.2109	1063.7934	32.6158	0.9723
VMD-PSO-LSTM	14.2038	283.4887	16.8371	0.5578
VMD-MFE-RLMD-PSO-LSTM	9.1034	130.4607	11.4219	0.3505

上文提到的上证 50 指数的预测结果显示，二次分解的 MAE、MSE、RMSE、MAPE 均低于一次分解和未分解下的 MAE、MSE、RMSE、MAPE，这也说明模型的预测效果在二次分解下较突出，预测准确度较高。且本文所提出模型的预测值趋势图与传统模型 CNN、SVR、GRU 的预测值趋势图相比，更加接近原始数据的趋势图，误差评价指标也比传统模型低，因此该模型应用于不同数据仍可以取的较高的预测精度，具有更好的性能。

5. 结论

本文介绍了基于二次分解的 PSO-LSTM 预测模型。该模型首先使用 VMD 模型将原始股票收盘价时间序列分解成多个子序列，然后利用 RLMD 模型对高复杂性数据进行二次分解。最后，对于两次分解得到的各子序列，分别将其输入 PSO-LSTM 模型进行预测。为验证二次分解模型的有效性，本文将二次分解模型与一次分解、未分解及传统模型的预测效果进行对比，其预测效果均优于其他模型。同时，本文还将上证 50 指数的股票数据代入所提的二次分解的 PSO-LSTM 及其余模型，结果显示二次分解的 LSTM-PSO 模型的预测效果同样优于一次分解、未分解及传统模型。证明了经二次分解的数据运用 PSO-LSTM 进行预测，能够更好的捕捉各子序列所隐藏的信息，得到更精准的预测效果。未来工作可尝试以下两个方面：一是选取其他分解方法来探究多次分解模型的预测性能，并挖掘相关指标；二是将所提模型运用到其他行业的数据。

参考文献

[1] Song, Y.P., et al. (2024) Multi-Decomposition in Deep Learning Models for Futures Price Prediction. *Expert Systems with Applications*, **246**, Article ID: 123171. <https://doi.org/10.1016/j.eswa.2024.123171>

[2] Wang, J., Cui, Q., Sun, X. and He, M. (2022) Asian Stock Markets Closing Index Forecast Based on Secondary Decomposition, Multi-Factor Analysis and Attention-Based LSTM Model. *Engineering Applications of Artificial Intelligence*, **113**, Article ID: 104908. <https://doi.org/10.1016/j.engappai.2022.104908>

[3] Mao, Y. and Yu, X. (2024) A Hybrid Forecasting Approach for China's National Carbon Emission Allowance Prices with Balanced Accuracy and Interpretability. *Journal of Environmental Management*, **351**, Article ID: 119873. <https://doi.org/10.1016/j.jenvman.2023.119873>

[4] Liu, Z., Jin, Y., Zuo, M.J. and Feng, Z. (2017) Time-Frequency Representation Based on Robust Local Mean Decomposition for Multicomponent AM-FM Signal Analysis. *Mechanical Systems and Signal Processing*, **95**, 468-487. <https://doi.org/10.1016/j.ymssp.2017.03.035>

[5] Wang, H., Xiong, M., Chen, H. and Liu, S. (2022) Multi-Step Ahead Wind Speed Prediction Based on a Two-Step Decomposition Technique and Prediction Model Parameter Optimization. *Energy Reports*, **8**, 6086-6100. <https://doi.org/10.1016/j.egyr.2022.04.045>

[6] Ngoc-Lan Huynh, A., Deo, R.C., Ali, M., Abdulla, S. and Raj, N. (2021) Novel Short-Term Solar Radiation Hybrid Model: Long Short-Term Memory Network Integrated with Robust Local Mean Decomposition. *Applied Energy*, **298**, Article ID: 117193. <https://doi.org/10.1016/j.apenergy.2021.117193>

[7] 杨青, 王晨蔚. 基于深度学习 LSTM 神经网络的全球股票指数预测研究[J]. 统计研究, 2019, 36(3): 65-77.

[8] 赵鑫, 陈臣鹏, 毕贵红, 等. 基于多模式分解的碳交易价格组合预测模型[J]. 电力科学与工程, 2022, 38(2): 52-59.

-
- [9] Yang, H., Yang, X. and Li, G. (2023) Forecasting Carbon Price in China Using a Novel Hybrid Model Based on Secondary Decomposition, Multi-Complexity and Error Correction. *Journal of Cleaner Production*, **401**, Article ID: 136701. <https://doi.org/10.1016/j.jclepro.2023.136701>
- [10] 郑雪文. 基于 VMD-SE-PSO-LSTM 的碳交易价格预测研究[D]: [硕士学位论文]. 郑州: 河南大学, 2023.