

基于二次分解重构的股指预测研究

吉如沁*, 秦江涛#

上海理工大学管理学院, 上海

收稿日期: 2024年6月24日; 录用日期: 2024年7月18日; 发布日期: 2024年7月29日

摘要

针对股票指数复杂难预测的问题, 提出一个基于二次分解重构的混合预测模型(ICEEMDAN-EMD-LSTM)。使用改进自适应噪声互补集成经验模态分解(ICEEMDAN)方法将股指序列分解为一系列子序列, 并根据模糊熵(FE)的值将这些子序列重构为高频、低频和趋势分量。再使用经验模式分解(EMD)方法对高频分量进行分解, 并再次应用FE使高频分量的子序列重构为新的高频、低频和趋势分量。将低频分量、趋势分量和新的高频、低频和趋势分量共五个分量进行线性集成, 得到股票指数的最终预测值。为了验证所提出模型的有效性, 对沪深300指数序列进行一预测。实验结果表明, 与基准模型相比, 本文提出的模型方法提高了预测精度, 并具有良好的稳健性。

关键词

股指预测, 二次分解重构, 长短期记忆网络

Stock Index Prediction Based on Quadratic Decomposition and Reconstruction

Ruqin Ji*, Jiangtao Qin#

College of Management, University of Shanghai for Science & Technology, Shanghai

Received: Jun. 24th, 2024; accepted: Jul. 18th, 2024; published: Jul. 29th, 2024

Abstract

A mixed prediction model based on quadratic decomposition and reconstruction (ICEEMDAN-EMD-LSTM) is proposed to address the problem of complex and difficult prediction of stock indices. The

*第一作者。

#通讯作者。

improved adaptive noise complementary ensemble empirical mode decomposition (ICEEMDAN) method is used to decompose the stock index sequence into a series of subsequences, and these subsequences are reconstructed into high-frequency, low-frequency, and trend components based on the value of fuzzy entropy (FE). The Empirical Mode Decomposition (EMD) method is used to decompose the high-frequency components, and FE is applied again to reconstruct the subsequences of the high-frequency components into new high-frequency, low-frequency, and trend components. The low-frequency component, trend component, and new high-frequency, low-frequency, and trend components are linearly integrated to obtain the final predicted value of the stock index. To verify the effectiveness of the proposed model, a prediction was made on the Shanghai and Shenzhen 300 index series. The experimental results show that compared with the benchmark model, the proposed model method improves prediction accuracy and has good robustness.

Keywords

Stock Index Prediction, Second Decomposition and Reconstruction, 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

近年来,基于深度神经网络的模型在时间序列建模中取得了不错的效果。如长短时记忆(LSTM) [1]、门控递归单元(GRU) [2]等。同时,一些研究将不同的深度学习模型组合成一个更复杂的混合模型,以获得更高的预测精度[3]。然而,对于非线性和高度复杂的时间序列,尽管相关研究的结果显示了人工智能方法的优势,但这些方法仍然存在局限性,包括参数敏感性和过度拟合。它们很容易陷入局部极小值[4]。

因此,为了进一步提高预测性能,有学者开始使用基于分解的组合预测方法来进行预测。如,何凯等[5]利用集成经验模态分解(EEMD)对上证基金指数进行分解,并使用支持向量机(SVM)对各分量分别预测,与直接用 SVM 相比,混合模型得到了更好的拟合效果。刘铭等[6]对金融股值序列的日收盘价通过经验模态分解(EMD)分解,并对每个子序列应用 LSTM 模型分别预测,最后对预测结果求和,有效地提升预测的精确度,证明相对于单个模型,复杂混合模型的优越性。后来有学者发现,对分解后的高频子序列进一步分解能够有效提高预测精度。如殷豪等[7]在对电力价格序列一次分解的情况下,对高频子序列进行二次分解,到了比一次分解更高的预测精度。高凯悦等[8]在 EMD 分解的基础上,运用变分模态分解(VMD)技术将分解后高频的股票数据子序列进一步分解,然后使用神经网络对每个最终的每个子序列进行预测,结果表明基于二次分解的混合预测模型拥有比单分解混合模型更高的预测精度。

尽管组合预测方法已经取得了良好的预测能力,但仍有进一步改进的空间。当原始信号被分解为许多分量时,多个子分量被预测并集成到最终预测结果中,可能会出现误差累积。因此,许多学者采用了重建技术来减少子序列的数量。例如,陈凯杰等[9]通过 EEMD 对股指序列分解后再使用模糊熵(FE)将序列重构为高频,低频和趋势分量并使用不同的预测模型对各分量进行预测,证实了解析重构后进行预测能有效提高预测的性能。高海翔等[10]使用 EEMD 方法对油价时序进行分解,并使用 fine-to-coarse (FTC) 重构技术将子成分重构为三个项目,然后将三个子成分的预测结果进行整合。他们发现,所提出的模型

比所有基准模型都表现得更好。然而, 单一分解重构的方法可能无法完全提取高频分量的信号。

因此, 为了进一步提高时间序列预测的准确性, 本文对分解后构建的高频分量进行二次分解重构, 进一步降低信号的复杂度, 构造了一个二次分解重构模型(ICEEMDAN-EMD-LSTM)。首先, 采用了一种更高效、更稳健的改进自适应噪声互补集成经验模态分解(ICEEMDAN)分解算法来分解原始数据, 根据模糊熵的值对分量进行重构, 然后对重构后的高频分量通过 EMD 再一次分解重构, 最后通过 LSTM 网络对各重构后的分量预测, 且通过线性集成, 得到股指最终预测值。

2. 理论方法介绍

2.1. EMD

EMD 分解的目的是将信号分解成内在模式函数(intrinsic mode function, IMF)的叠加, 它必须满足以下两个条件: 一是该函数在整个时间范围内必须有相等数量的局部极值和多余的零点, 或最多有一个差值; 二是在任何时间点, 局部极值的包络(上包络)和局部极值的包络(下包络)的平均值必须为零。EMD 的具体分解方法如下:

- 1) 确定所有的局部极值, 包括 $x(t)$ 的局部最大值和局部最小值。
- 2) 生成它们的上、下包络 $u(t)$ 和 $l(t)$ 。
- 3) 从上、下包络计算逐点包络平均数 $m(t)$, 即:

$$m(t) = \frac{u(t) + l(t)}{2} \quad (1)$$

4) 提取细节 $(t) = x(t) - m(t)$, 验证 $h_1(t)$ 是否满足 IMF 的两个条件, 若满足, $h_1(t)$ 则为第一个 IMF, 否则 $h_1(t)$ 做上述同样的处理。以此类推直到 $h_k(t)$ 符合 IMF 的条件或达到停止标准, 停止标准的计算值为:

$$y_{SD} = \sum_{t=0}^T \frac{|h_{k-1}(t) - h_k(t)|^2}{h_{k-1}^2(t)} \quad (2)$$

其中: $h_{k-1}(t)$ 为第 $k-1$ 个 IMF 分量; $h_k(t)$ 为第 k 个 IMF 分量。当 y_{SD} 小于某一特定值即达到停止标准。

2.2. ICEEMDAN

ICEEMDAN 方法是 Colominas 等改进的一种新的分解算法[11], 在 EMD 分解后添加高斯白噪声分量。它计算每个模式分量的噪声和局部均值, 分解分量定义为残差信号和局部均值之间的差。实验表明, ICEEMDAN 方法克服了残余噪声和模式混合的缺点, 有效地提高了降噪效果[12]。其分解的过程如下:

- 1) 向初始信号 $x(t)$ 中添加一组白噪声 $E_k(\omega^i)$ (均值和协方差均为 0), 得到一组新信号:

$$x_i(t) = x(t) + E_k(\omega^i) \quad (i = 1, 2, 3, \dots; k = 0, 1, 2, \dots) \quad (3)$$

- 2) 计算 $x_i(t)$ 的局部的最大值和局部的最小值, 再求出所有上, 下包络线 $u(t)$ 、 $l(t)$ 。得出均值:

$$m_{i,1}(t) = \frac{u(t) + l(t)}{2} \quad (4)$$

- 3) 重复上述两个过程 i 次, 得到第一个剩余函数 $r_1(t)$ 和第一个 IMF₁ 分量:

$$r_1(t) = \frac{1}{i} \sum (x(t) + E_1(\omega^i)) \quad (5)$$

$$\text{IMF}_1 = x(t) - r_1(t) \quad (6)$$

4) 重复上述步骤, 得到第二个剩余函数 $r_2(t)$ 和第二个 IMF 分量 IMF_2 :

$$r_2(t) = \frac{1}{i} \sum (x(t) + E_2(\omega^i)) \quad (7)$$

$$\text{IMF}_2 = r_1(t) - r_2(t) \quad (8)$$

5) 重复迭代, 得到第 k 个剩余函数 $r_k(t)$ 和第 k 个 IMF 分量 IMF_k 为:

$$r_k(t) = \frac{1}{i} \sum (x(t) + E_k(\omega^i)) \quad (9)$$

$$\text{IMF}_k = r_{k-1}(t) - r_k(t) \quad (10)$$

2.3. 模糊熵

模糊熵[13]是计算时间序列的复杂度与随机性的动力学方法, 是对样本熵的改进, 用于衡量时间序列随维数变化产生新模式的概率。其具体步骤如下:

1) 将时间序列 $x(t)$, 构造成 m 维向量 X_i^m , 即

$$X_i^m = \{x(i), x(i+1), \dots, x(i+m-1)\} - x_0(i), i = 1, 2, \dots, N-m+1 \quad (11)$$

其中, X_i^m 为重构后的时间序列, $x_0(i) = \frac{1}{m} \sum_{k=0}^{m-1} x(i+k)$ 。

2) 计算空间向量 X_i^m 与 X_j^m 之间的切比雪夫距离 d_{ij}^m :

$$d_{ij}^m = d[X_i^m, X_j^m] = \max \{ |[x(i+k) - x_0(i)] - [x(j+k) - x_0(j)]| \} \quad (12)$$

其中 $i, j = 1, 2, \dots, N-m+1; i \neq j$ 。

3) 运用模糊函数 $\mu(d_{ij}^m, n, r)$ 计算向量 X_i^m 与 X_j^m 的相似程度 D_{ij}^m :

$$D_{ij}^m = \mu(d_{ij}^m, n, r) = e^{-\ln 2 \left(\frac{d_{ij}^m}{r} \right)^n} \quad (13)$$

其中 r 表示 $\mu(d_{ij}^m, n, r)$ 的相似容限; n 表示边界梯度。

4) 计算 m 维下的关系维度 $\varphi^m(n, r)$, 并增加到 $m+1$ 维, 重复上述步骤得到 $\varphi^{m+1}(n, r)$:

$$\varphi^m(n, r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \left(\frac{1}{N-m} \sum_{j=1, j \neq i}^{N-m+1} D_{ij}^m \right) \quad (14)$$

$$\varphi^{m+1}(n, r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \left(\frac{1}{N-m} \sum_{j=1, j \neq i}^{N-m+1} D_{ij}^{m+1} \right) \quad (15)$$

5) 计算模糊熵:

$$S_{FE}(m, n, r) = \lim_{N \rightarrow \infty} [\ln \varphi^m(n, r) - \ln \varphi^{m+1}(n, r)] \quad (16)$$

2.4. LSTM

作为一种改进的循环神经网络(RNN), 长短时记忆(LSTM)得益于 RNN 模型的优点, 利用门的独特结构, 有效地解决了 RNN 中梯度爆炸和梯度消失的问题[14]。

LSTM 由多个循环单元组成, 其输入包含当前时刻的输入数据、上一时刻网络单元的状态向量和隐藏层的输出向量。LSTM 首先通过遗忘门计算出单元的废弃信息, 其数学表达式如下:

$$f_t = \sigma[W_f \cdot (h_{t-1}, x_t) + b_f] \tag{17}$$

其中, W_f 是“遗忘门”到较长向量中的权重矩阵, b_f 的偏置值, σ 为 sigmoid 函数。

“遗忘门”读取 x_t 和 h_{t-1} , 其中输出值的大小代表遗忘的程度, 并给细胞状态 C_{t-1} 赋值。 f_t 的大小为 $[0, 1]$, 值越小, 遗忘程度就越高。

输入数据 x_t 和隐层输出 h_{t-1} 经过“输入门”, 得到待输入的细胞状态 \tilde{C}_t 和向量 i_t 表达式为:

$$\tilde{C}_t = \tanh[W_c \cdot (h_{t-1}, x_t) + b_c] \tag{18}$$

$$i_t = \sigma[W_i \cdot (h_{t-1}, x_t) + b_i] \tag{19}$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{20}$$

其中, W_i 、 W_c 为“输入门”的权值矩阵, b_i 、 b_c 为“输入门”的偏差向量, C_t 为当前输入单元状态。

如图 1 所示, 基于 LSTM 的结构, “遗忘门”和“输入门”在单元状态更新后进入“输出门”。由于其优化的结构, LSTM 可以有选择地记住重要的信息并忘记不重要的信息。因此, RNN 的梯度消失问题得到了改善。

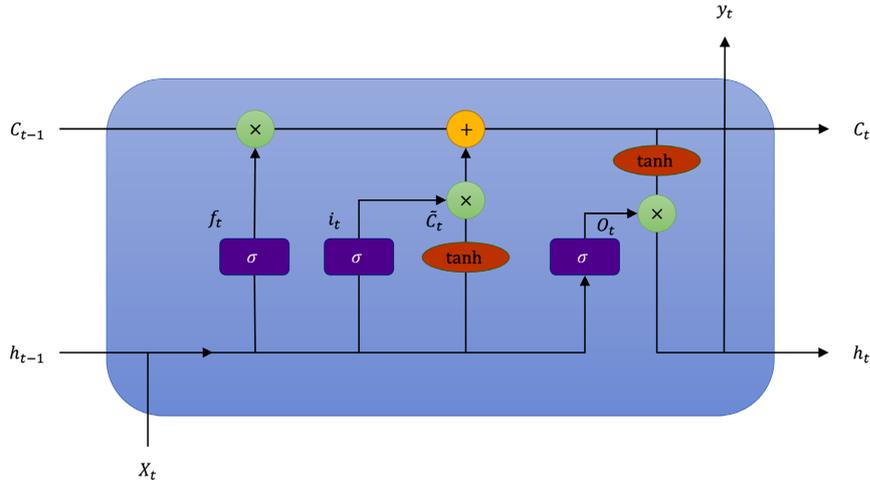


Figure 1. LSTM model structure diagram
图 1. LSTM 模型结构图

2.5. 基于二次分解重构模型的股指预测模型

本文的框架首先利用 ICEEMDAN 分解方法将股票指数数据分解为简单的 IMF 成分, 然后计算每个 IMF 成分的 FE 值, 并根据 FE 值将分解的 IMF 成分重新构建为高频、低频和趋势项。由于现有的分解方法存在缺陷, 如不能完全提取高频信号, 本文提出了一种新的二次分解技术, 作为单一分解重构方法的补充。在这项技术中, EMD 方法被用来进一步分解由单一分解重构产生的高频成分, 以减少高频信号的复杂性。构建了一种新的组合预测方法(ICEEMDAN-EMD-LSTM)。模型构建的工作流程如图 2 所示。

- 1) 分解数据: 使用 ICEEMDAN 分解方法, 将原始股票价格序列分解一系列 IMF 成分。
- 2) 重构分量: 使用 FE 算法计算每个 IMF 分量的熵值, 并将这些分量重建为高频、低频和趋势分量。
- 3) 高频分量分解重构: 因为重构后的高频分量的复杂性且信息难以提取, 所以使用 EMD 方法应用于高频分量, 并再次使用 FE 算法将分解后的成分重建为新的高频、低频和趋势成分。
- 4) 模型预测: 将上述分量划分为训练集和测试集, 并进行归一化处理。使用 LSTM 网络模型预测上

述分解重构的各个分量, 并进行反归一化。

5) 线性集成: 对上述五个分量的预测值进行线性求和, 得到股票指数数据收盘价的预测值。

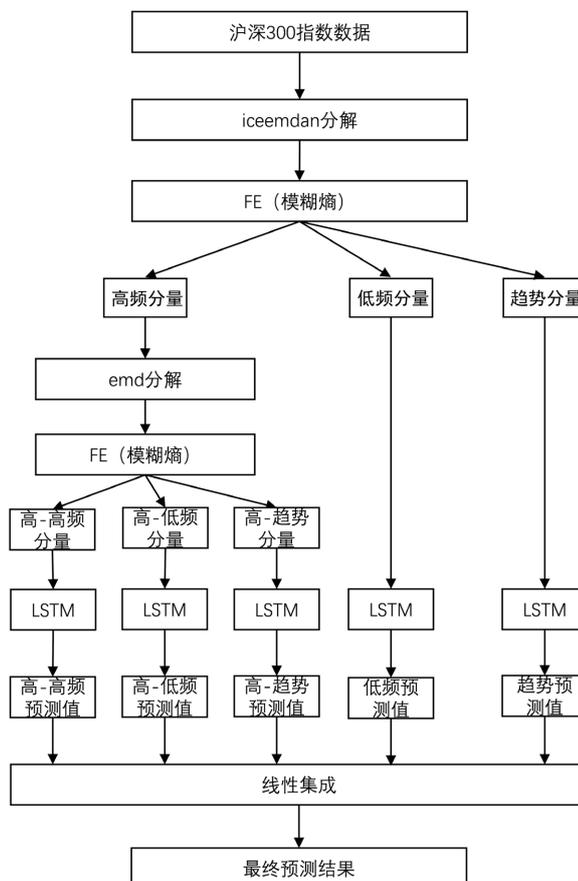


Figure 2. Algorithm framework flowchart

图 2. 算法框架流程图

3. 实证分析

3.1. 数据分析

为了测试本文 ICEEMDAN-EMD-LSTM-组合模型的预测效果, 数据集采用中国沪深 300 指数。数据来源于东方财富旗下 Choice 金融终端。股票指数数据跨度为 2017 年 1 月 3 日至 2023 年 3 月 1 日, 共 1519 个交易日数据。其中训练集占总数据 90%, 测试集占总数据 10%。数据统计信息如表 1 所示。

Table 1. Descriptive statistics of stock index data

表 1. 股指数据描述性统计

序列	样本值	最大值	最小值	均值	标准差	偏度	峰度
收盘价	1519	5807.7	2964.8	4105.8	596.78	0.525	2.4675

3.2. 数据归一化处理及预测结果评价指标

为了避免个别极大值对整个模型的预测效果, 本文使用最大最小值标准化来对各模态分量序列数

据进行归一化处理, 其表达式为:

$$X^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (21)$$

其中, X^* 为标准化后的子序列, x 为初始值, x_{\max}, x_{\min} 分别为各模态分量的最大值和最小值。

本文选取了四种评价指标来检验模型的预测效果, 包括相对平均误差绝对值(MAPE)、均方根误差(RMSE)、绝对平均误差(MAE)以及拟合优度(R-squared), 其计算公式为:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (22)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (23)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (24)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (25)$$

其中, y_i 是真实值, \hat{y}_i 是预测值, n 是样本数量。 \bar{y}_i 是真实值的平均值。

3.3. 股指序列分解

首先, 本文使用改进自适应噪声互补集成经验模态分解(ICEEMDAN)方法对初始序列进行分解, 相较于经验模态分解(EMD), ICEEMDAN 分解能够有效较少模态中的残余噪声, 解决模态混叠问题, 避免伪模态的产生, 从而可以有效提高预测的准确性。图 3 展示了沪深 300 分解结果。

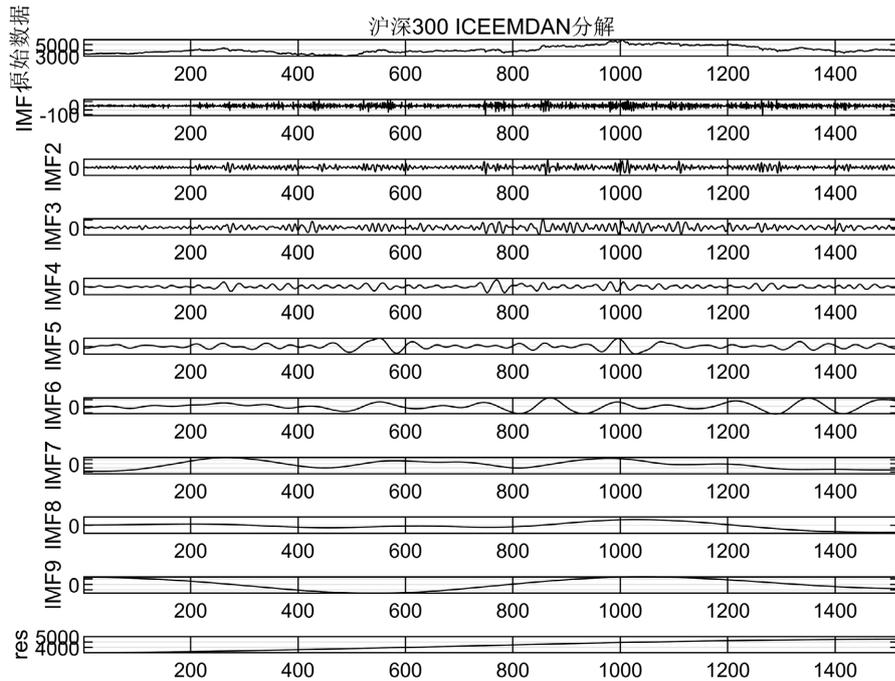


Figure 3. ICEEMDAN decomposition results

图 3. ICEEMDAN 分解结果

本文通过模糊熵算法(FE)计算各子序列的频率特征,其参数设定采用嵌入维度 $m=2, r=0.15, n=2$ 。本文设置了两个阈值, $\lambda_1=0.1, \lambda_2=0.001$ 。当分解后的序列模糊熵的值大于 λ_1 时, 被判断为高频分量, 当序列模糊熵的值处于 λ_1 和 λ_2 之间时, 被判断为低频分量, 其余的被判断为趋势分量。表 2 展现了各子序列的 FE 值和分类状况。其中, 模态分量 1~4 为高频分量, 5~7 为低频分量, 8~10 为趋势分量。

Table 2. Fuzzy entropy values and classification results of each subsequence
表 2. 各子序列模糊熵值及分类结果

	imf_1	imf_2	imf_3	imf_4	imf_5
CSI300	2.926	2.298	1.479	0.879	0.649
	imf_6	imf_7	imf_8	imf_9	res
CSI300	0.498	0.227	0.088	0.061	0.002

因为一次分解重构后的股指高频分量的信息可能很难被 LSTM 模型有效提取, 所以本文使用 EMD 方法对重构后的高频分量进一步分解, 以此降低数据的复杂度, 沪深 300 重构后高频分量的分解结果由图 4 展示。

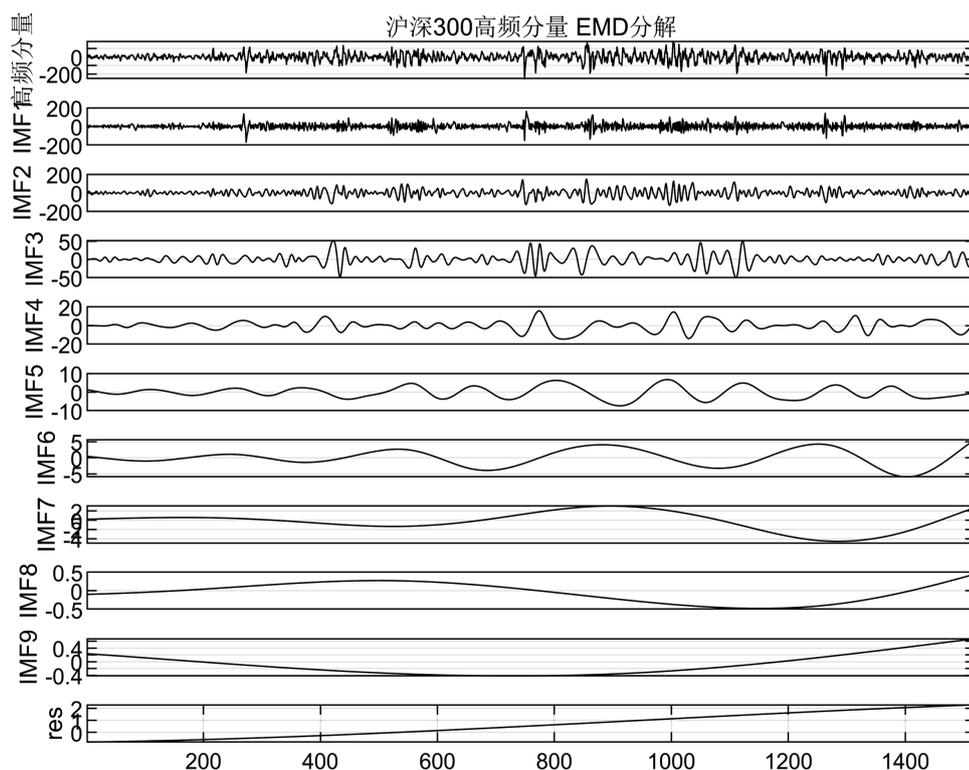


Figure 4. High frequency component EMD decomposition results
图 4. 高频分量 EMD 分解结果

与第一分解后一样, 对 EMD 分解后的各模态分量计算 FE 值, 并根据阈值 λ_1, λ_2 再次对 10 各分量进行分类重构。表 3 展现了高频分量分解后各子序列的模糊熵值和分类结果。其中, 模态分量 1~2 为高 - 高频分量, 3~5 为高 - 低频分量, 6~10 为高 - 趋势分量。

Table 3. Fuzzy entropy values and classification results of each subsequence of high-frequency components
表 3. 高频分量各子序列模糊熵值及分类结果

	imf_1	imf_2	imf_3	imf_4	imf_5
CSI300	3.029	1.826	0.6899	0.2456	0.051
	imf_6	imf_7	imf_8	imf_9	res
CSI300	0.011	1.7e-3	1.0e-4	5.4e-5	1.6e-6

3.4. 模型比较

为了验证本文提出的模型在股票指数预测上的有效性, 本文在沪深 300 股票指数预测上通过多个模型进行实验分析。首先, 通过对比单一模型, 实验分析发现 LSTM 模型相对其他单一模型更加适合复杂的股票指数预测。常用于时间序列预测的机器学习方法大多有 LSTM、SVM、ELM 和 BP 神经网络, 因此本文选择这四种作为单一模型对比, 比较出 LSTM 模型相对其他的单一预测模型在股票指数预测上的效果更好。

不同单一模型的预测曲线图如下:

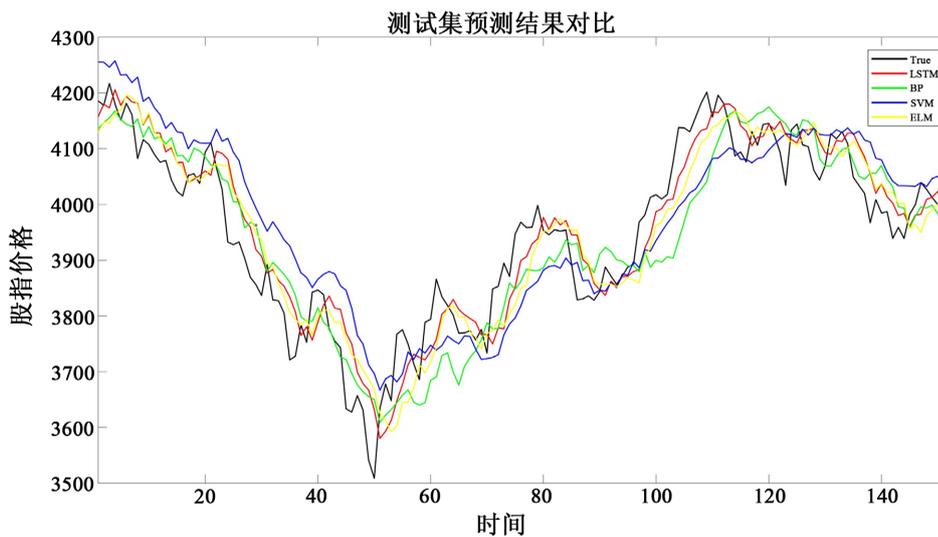


Figure 5. Curve chart of CSI 300 index forecast by different single models

图 5. 不同单一模型沪深 300 指数预测曲线图

从图 5 的预测图形中可以看出单一模型在沪深 300 股票指数序列预测上的效果并不佳, 只能预测出大概的趋势, 对于一些极端点的预测偏离很明显。同时随着时间的不断推移, 整体预测效果也变得更差, 预测误差变得愈发明显。

表 4 详细列出了总体输出值与期望值之间的误差情况, 通过多个评估指标的综合考量, 能够更加客观、全面地评价模型的预测结果。从表中数据可以看出, 本节选择的 LSTM 模型 MAPE 值为 1.57%, 虽然这个数值相对较低, 但仍表明模型在预测精度方面存在一定的问题。同时, RMSE 值为 58.5921, MAE 值为 49.4722, 这两个指标都较高, 进一步反映了模型预测误差较大的情况。此外, R^2 值为 86.19%, 拟合效果欠缺。但是相对其他三个单一预测模型, LSTM 的性能最优, 说明在股票指数数据的时间序列预测方面, LSTM 模型更佳适合。

Table 4. Evaluation indicators of single model prediction results
表 4. 单一模型预测结果评价指标

预测模型	MAPE	RMSE	MAE	R ²
LSTM	0.0157	58.5921	49.4722	0.8601
BP	0.0176	73.6874	60.7766	0.7817
SVM	0.0192	88.8510	73.1969	0.7126
ELM	0.0163	64.3251	52.4690	0.8336

为了验证分解重构方法在股票指数预测中的有效性以及本文提出的二次分解重构模型(ICEEMDAN-EMD-LSTM)的优越性。选择了一系列的模型作为对比模型。对于分解重构的所有模型,均包含了数据的分解和重构,为了突出二次分解重构比一次分解重构更有效,选择了 EMD-LSTM、ICEEMDAN-LSTM 一次分解重构模型,EMD-EMD-LSTM、ICEEMDAN-ICEEMDAN-LSTM 和 IE-LSTM 二次分解重构模型来对沪深 300 指数序列进行预测并作对比分析。各分解重构股指预测模型如下图:

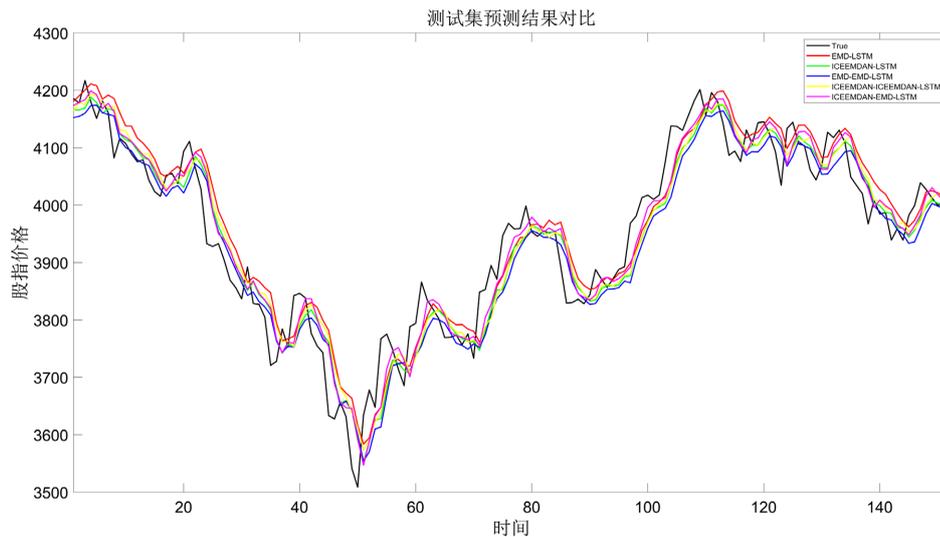


Figure 6. Prediction curves of the Shanghai and Shenzhen 300 Index using different decomposition and reconstruction models
图 6. 不同分解重构模型沪深 300 指数预测曲线图

从图 5 和图 6 可以清晰看出相对于单一股票指数预测模型,引入分解技术的混合预测模型的预测效果得到了很大的提升,预测曲线的偏离程度得到改善。说明将股票指数序列通过分解能有效降低其复杂度,提高模型的预测精度。从图 6 还可以看出经过二次分解的模型预测曲线相对于一次分解也更加贴近真实情况,再次验证对于噪音多,非平稳的股票指数序列进行二次分解是有必要的。

从表 5 的评价指标可以看出,经过一次分解重构处理过的 EMD-LSTM 和 ICEEMDAN-LSTM 模型,其预测精度相对于表 4 中单一预测模型的精度都是得到提升的。从表 5 可以看出,一次分解 ICEEMDAN-LSTM 模型的预测效果较 EMD-LSTM 模型有所提升,因为 ICEEMDAN 算法是在 EMD 算法的基础上添加了高斯白噪音,能够极大降低原始序列的噪音,从而降低模型预测的难度,提升预测效果。同时,所有经过对高频分量再次分解的二次分解重构预测模型的预测效果都是优于一次分解重构模型的,说明高频分量对于 LSTM 模型来说,是难以提取的,需要进一步处理。

Table 5. Evaluation indicators for decomposition and reconstruction model prediction results
表 5. 分解重构模型预测结果评价指标

预测模型	MAPE	RMSE	MAE	R ²
EMD-LSTM	0.0112	53.3180	43.6543	0.8857
ICEEMDAN-LSTM	0.0105	51.1412	40.5800	0.8948
EMD-EMD-LSTM	0.0099	47.5388	38.6867	0.9091
ICEEMDAN-ICEEMDAN-LSTM	0.0096	46.1093	37.1870	0.9102
ICEEMDAN-EMD-LSTM	0.0086	41.8349	33.3676	0.9296

4. 结论

本文以“分而治之”思想为指导,构建了一个基于二次分解重构的混合预测模型(ICEEMDAN-EMD-LSTM)。该模型考虑了股指数据的内在发展规律,采用二次分解重构策略处理非线性非平稳的高频分量数据,并对各分量的预测结果线形集成得到最终预测结果。通过沪深 300 股指序列的实验结果表明,在处理股指序列时, LSTM 网络模型相对其他单一模型预测的效果更好;通过分解重构预处理后再进行预测,预测的精度明显提升;通过对一次分解重构后的高频分量再次分解能有效降低数据复杂度,提升模型整体预测效果。因此,本文提出的二次分解重构模型预测效果显著,相对于其他对比模型,取得了较优的预测结果。在后续的工作,将考虑其他因素的输入,来进一步提升预测的精准度。

基金项目

国家自然科学基金项目(72174121, 71774111)。

参考文献

- [1] Chen, S. and Ge, L. (2019) Exploring the Attention Mechanism in LSTM-Based Hong Kong Stock Price Movement Prediction. *Quantitative Finance*, **19**, 1507-1515. <https://doi.org/10.1080/14697688.2019.1622287>
- [2] Niu, H. and Xu, K. (2020) A Hybrid Model Combining Variational Mode Decomposition and an Attention-GRU Network for Stock Price Index Forecasting. *Mathematical Biosciences and Engineering*, **17**, 7151-7166. <https://doi.org/10.3934/mbe.2020367>
- [3] 赵红蕊, 薛雷. 基于 LSTM-CNN-CBAM 模型的股票预测研究[J]. 计算机工程与应用, 2021, 57(3): 203-207.
- [4] Tang, L., Yu, L., Wang, S., Li, J. and Wang, S. (2012) A Novel Hybrid Ensemble Learning Paradigm for Nuclear Energy Consumption Forecasting. *Applied Energy*, **93**, 432-443. <https://doi.org/10.1016/j.apenergy.2011.12.030>
- [5] 何凯, 苏栾芳, 何卫平. 上证基金指数波动结构分解与短期预测: 基于 EEMD 模型[J]. 金融理论与实践, 2014(1): 80-85.
- [6] 刘铭, 单玉莹. 基于 EMD-LSTM 模型的股指收盘价预测[J]. 重庆理工大学学报(自然科学), 2021, 35(12): 269-276.
- [7] 殷豪, 等. 二层分解技术在电价预测中的应用研究[J]. 计算机应用研究, 2019, 36(4): 1101-1105.
- [8] 高凯悦, 牟莉. 基于二次分解和 GRU-attention 的时间序列预测研究[J]. 国外电子测量技术, 2023, 42(2): 80-87.
- [9] 陈凯杰, 唐振鹏, 吴俊传, 杜晓旭, 蔡毅. 基于分解-集成和混频数据采样的中国股票市场预测研究[J]. 系统工程理论与实践, 2022, 42(11): 3105-3120.
- [10] 高海翔, 胡瑜, 余乐安. 基于分解集成的 LSTM 神经网络模型的油价预测[J]. 计算机应用与软件, 2021, 38(10): 78-83.
- [11] Colominas, M.A., Schlotthauer, G. and Torres, M.E. (2014) Improved Complete Ensemble EMD: A Suitable Tool for Biomedical Signal Processing. *Biomedical Signal Processing and Control*, **14**, 19-29. <https://doi.org/10.1016/j.bspc.2014.06.009>

- [12] 李昌兴, 钟清华, 廖金湘. 基于 ICEEMD 的人脸视频心率检测[J]. 激光杂志, 2019, 40(1): 33-36.
- [13] Zhu, K., Chen, L. and Hu, X. (2019) Rolling Element Bearing Fault Diagnosis Based on Multi-Scale Global Fuzzy Entropy, Multiple Class Feature Selection and Support Vector Machine. *Transactions of the Institute of Measurement and Control*, **41**, 4013-4022. <https://doi.org/10.1177/0142331219844555>
- [14] 李洁, 林永峰. 基于多时间尺度 RNN 的时序数据预测[J]. 计算机应用与软件, 2018, 35(7): 33-37, 62.