

吉如沁, 秦江涛. 基于 ICEEMDAN-DBO-LSTM 模型的沪深 300 指数预测研究[J]. 智能计算机与应用, 2026, 16(1): 30-36.
DOI: 10.20169/j.issn.2095-2163.24032504

基于 ICEEMDAN-DBO-LSTM 模型的沪深 300 指数预测研究

吉如沁, 秦江涛

(上海理工大学 管理学院, 上海 200093)

摘要: 针对股票指数复杂难预测的问题, 本文采用改进的完全自适应噪声集合经验模态分解(ICEEMDAN)、蜣螂优化算法(DBO)和长短期记忆网络(LSTM)相结合的模型预测沪深 300 股指收盘价。首先, 使用 ICEEMDAN 分解方法将股指序列分解为一系列子序列, 并利用模糊熵(FE)评估序列复杂度将子序列重构为高频、低频和趋势分量。其次, 使用 DBO 优化过的 LSTM 进行分量预测。最后, 将分量预测值线性求和, 得到最终预测值。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, 表现最佳。

关键词: 沪深 300 指数; 改进自适应噪声互补集经验模态分解; 蜣螂优化算法; 长短期记忆网络

中图分类号: F832.51 **文献标志码:** A **文章编号:** 2095-2163(2026)01-0030-07

Research on predicting the CSI 300 index based on the ICEEMDAN-DBO-LSTM model

Ji Ruqin, Qin Jiangtao

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: To solve the problem that stock index is complex and difficult to predict, this paper proposes a model combining Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), Dung Beetle Optimization algorithm (DBO) and Long Short-Term Memory network to predict the closing price of CSI 300 stock index. Firstly, the stock index sequence is decomposed into a series of sub-sequences by using the ICEEMDAN decomposition method, and the sub-sequences are reconstructed into high-frequency, low-frequency and trend components by using Fuzzy Entropy (FE) to evaluate the complexity of the sequence. Secondly, the LSTM optimized by DBO is used for component prediction. Finally, the prediction results of each sub-sequence are superimposed to obtain the final prediction results. The experimental results show that compared with the benchmark model, the proposed model method improves prediction accuracy and performs the best.

Key words: CSI 300 index; improved adaptive noise complementary integration empirical mode decomposition; Dung Beetle Optimization algorithm; Long Short-Term Memory network

0 引言

由于股票数据的高噪声、动态、非线性和混沌特性, 股票市场预测始终是一个挑战。传统的股票走势预测模型主要将其视为一个序列建模问题, 这些模型侧重于技术分析, 通过技术指标, 包括波动率、移动平均线和动量等来预测股票价格走势, 并根据走势确定投资策略。传统的统计方法使用自回归模型, 如 ARMA^[1]、ARIMA^[2]、GARCH^[3]。然而, 此类方法对

高度非线性和不稳定的股票数据的拟合能力较差。

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)^[4]、门控递归单元(GRU)^[5]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^[6]。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。还很容易陷

基金项目: 国家自然科学基金(72174121, 71774111); 上海市 2022 年度“科技创新行动计划”软科学研究项目(22692112600); 上海市自然科学基金(21ZR1444100)。

作者简介: 吉如沁(1998—), 男, 硕士研究生, 主要研究方向: 人工智能理论及应用。

通信作者: 秦江涛(1966—), 男, 副教授, 主要研究方向: 制造系统分析研究。Email: qinjiangtao_usst@126.com。

收稿日期: 2024-03-25

入局部极小值^[7]。

因此, 有很多学者采用启发式算法对网络参数进行优化, 以提高整体模型的预测效能。李婧琦^[8]使用鲸鱼算法 (WOA) 优化 LSTM 的超参数进行股票价格预测模型, 降低人为因素对模型性能的影响, 提高了股票价格预测的准确性。Gülmez^[9]使用人工兔子优化算法 (ARO) 优化 LSTM 网络模型, 使参数与 ARO 算法变量相连。当变量发生变化时, LSTM 架构也会发生变化。结果清楚地表明, LSTM-ARO 模型取得了较好的预测效果。Ji 等学者^[10]提出了一种改进的粒子群优化算法 (IPSO) 和长期-短期记忆 (LSTM) 混合模型用于股价预测。使用 IPSO 模型来优化 LSTM 的超参数。仿真结果证明, 启发式算法优化过的 LSTM 模型的表现优于其他相关基线模型。

另外, 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。何凯等学者^[11]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) 对各分量分别进行预测, 与直接使用 SVM 相比, 混合模型得到了更好的拟合效果。刘铭等学者^[12]对金融股值序列的日收盘价通过 EMD 分解, 并对每个子序列应用 LSTM 模型分别进行预测, 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。陈凯杰等学者^[13]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, 并进一步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。

因此, 在上述研究的基础上, 本文构建了 ICEEMDAN-DBO-LSTM 模型进行沪深 300 指数收盘价预测。利用 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, 以此降低原始数据的复杂度。利用 DBO 算法优化 LSTM 超参数, 减少网络参数对模型的影响。

1 模型方法研究

1.1 蜚螂优化算法

蜚螂优化算法 (DBO)^[14] 主要基于蜚螂的滚球、跳舞、觅食、偷窃和繁殖行为而提出的一种元启发式算法。研究分析后可知, 算法主要包括蜚螂滚球、蜚螂觅食、蜚螂繁衍、蜚螂偷窃四个优化过程。每个过程具体描述如下。

1.1.1 蜚螂滚球

(1) 无障碍模式。当蜚螂前行无障碍时, 蜚螂

在滚动过程中通过太阳来导航, 以保持粪球在直线上滚动。滚球蜚螂的位置更新公式如下:

$$\begin{cases} x_i^{t+1} = x_i^t + \alpha \times k \times x_i^{t-1} + b \times \Delta x \\ \Delta x = |x_i^t - x_{\text{worst}}^t| \end{cases} \quad (1)$$

其中, t 表示当前迭代次数; x_i^t 表示蜚螂的位置信息; $k \in (0, 0.2]$ 表示偏转系数; $b \in (0, 1)$ 表示常数; α 的具体取值根据概率方法来确定, 用以模拟复杂的环境, 当 $\alpha = 1$ 时表示无偏差, 当 $\alpha = -1$ 时表示偏离原来方向; x_{worst}^t 表示全局最差位置; Δx 模拟了光强度变化, 越大表示光源越弱, 该参数可以提升算法搜索的性能, 尽量使算法遍历整个搜索空间。

(2) 跳舞。当蜚螂遇到障碍物而无法前进时, 会通过跳舞来重新定向自己, 以获得新的路线。因此, 蜚螂的位置更新公式如下:

$$x_i^{t+1} = x_i^t + \tan(\theta) |x_i^t - x_{\text{worst}}^t| \quad (2)$$

其中, $\theta \in [0, \pi]$ 。当 $\theta = 0, \frac{\pi}{2}$ 或 π 时, $\tan(\theta)$ 没有意义, 所以此时不会更新蜚螂的位置。

1.1.2 繁殖

为了给后代提供一个安全的环境, 选择一个边界选择策略来模拟雌性蜚螂产卵的区域, 其定义公式为:

$$\begin{cases} \text{Lb}^* = \max(x_{\text{gbest}}^t \times (1 - R), \text{Lb}) \\ \text{Ub}^* = \min(x_{\text{gbest}}^t \times (1 + R), \text{Ub}) \end{cases} \quad (3)$$

其中, x_{gbest}^t 表示当前局部最优位置; $R = 1 - t/T_{\text{max}}$; T_{max} 表示最大迭代次数; Lb 和 Ub 分别表示优化问题的下界和上界; Lb^* 和 Ub^* 分别表示产卵区域的下界和上界。蜚螂产卵区域是随迭代次数调整的。产卵球的位置在迭代过程中也是动态的, 推得的公式如下:

$$B_i^{t+1} = x_{\text{gbest}}^t + b_1 \times (B_i^t - \text{Lb}^*) + b_2 \times (B_i^t - \text{Ub}^*) \quad (4)$$

其中, B_i^t 表示位置信息; b_1 和 b_2 表示 2 个独立的大小为 $1 \times D$ 的随机向量; D 表示优化问题的维度。

1.1.3 觅食

一些成熟的小蜚螂会从地下出来寻找食物, 小蜚螂的觅食区域是动态更新的, 表示如下:

$$\begin{cases} \text{Lb}^l = \max(x_{\text{lbest}}^t \times (1 - R), \text{Lb}) \\ \text{Ub}^l = \min(x_{\text{lbest}}^t \times (1 + R), \text{Ub}) \end{cases} \quad (5)$$

其中, R 与前文表示相同; x_{lbest}^t 表示当前种群的全局最优位置; Lb^l 和 Ub^l 分别表示小蜚螂觅食区域的上、下界。小蜚螂的位置更新公式具体如下:

$$x_i^{t+1} = x_i^t + C_1 \times (x_i^t - \text{Lb}^t) + C_2 \times (x_i^t - \text{Ub}^t) \quad (6)$$

其中, C_1 表示服从正态分布的随机数, 即 $C_1 \sim N(0,1)$, C_2 表示 $1 \times D$ 的属于 $(0,1)$ 之间的随机向量。

1.1.4 偷窃

在种群中, 有一些蜣螂从其他同伴那里偷食物, 偷窃蜣螂的位置更新公式为:

$$x_i^{t+1} = x_{\text{lbest}}^t + S \times g \times (|x_i^t - x_{\text{gbest}}^t| + |x_i^t - x_{\text{lbest}}^t|) \quad (7)$$

其中, x_{lbest}^t 表示最优食物源, 即假设该位置附近是竞争食物的最优位置; g 表示大小为 $1 \times D$ 服从正态分布的随机向量; S 表示一个常数。

1.2 ICEEMDAN

改进自适应噪声互补集成经验模态分解 (ICEEMDAN) 方法是 Colominas 等学者^[15] 改进的一种新的分解算法, 在 EMD 分解后添加高斯白噪声分量。通过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^[16]。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:

$$x_i(t) = x(t) + E_k(\omega^i) \quad (i = 1, 2, 3, \dots; k = 0, 1, 2, \dots) \quad (8)$$

(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, 可由下式计算求得:

$$m_{i,1}(t) = \frac{u(t) + l(t)}{2} \quad (9)$$

(3) 重复上述 2 个过程 i 次, 得到第 1 个剩余函数 $r_1(t)$ 和第 1 个 IMF₁ 分量。即:

$$r_1(t) = \frac{1}{i} \sum (x(t) + E_1(\omega^i)) \quad (10)$$

$$\text{IMF}_1 = x(t) - r_1(t) \quad (11)$$

(4) 重复上述步骤, 得到第 2 个剩余函数 $r_2(t)$ 和第 2 个 IMF 分量 IMF₂, 即:

$$r_2(t) = \frac{1}{i} \sum (x(t) + E_2(\omega^i)) \quad (12)$$

$$\text{IMF}_2 = r_1(t) - r_2(t) \quad (13)$$

(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:

$$r_k(t) = \frac{1}{i} \sum (x(t) + E_k(\omega^i)) \quad (14)$$

$$\text{IMF}_k = r_{k-1}(t) - r_k(t) \quad (15)$$

1.3 模糊熵

模糊熵 (FE)^[17] 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。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:

$$X_i^m = \{x(i), x(i+1), \dots, x(i+m-1)\} - x_0(i), \quad i = 1, 2, \dots, N-m+1 \quad (16)$$

其中, X_i^m 表示重构后的时间序列, $x_0(i) =$

$$\frac{1}{m} \sum_{k=0}^{m-1} x(i+k).$$

(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 , 定义公式如下:

$$d_{ij}^m = d[X_i^m, X_j^m] = \max\{|[x(i+k) - x_0(i)] - [x(j+k) - x_0(j)]|\} \quad (17)$$

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 , 定义公式如下:

$$D_{ij}^m = \mu(d_{ij}^m, n, r) = e^{-\ln 2 \left(\frac{d_{ij}^m}{r}\right)^n} \quad (18)$$

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 $\varphi^{m+1}(n, r)$ 。这一过程可以表示为:

$$\varphi^m(n, r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \left(\frac{1}{N-m_j=1, i \neq j} \sum_{j=1}^{N-m+1} D_{ij}^m \right) \quad (19)$$

$$\varphi^{m+1}(n, r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \left(\frac{1}{N-m_j=1, i \neq j} \sum_{j=1}^{N-m+1} D_{ij}^{m+1} \right) \quad (20)$$

(5) 计算模糊熵值。计算公式为:

$$S_{FE}(m, n, r) = \lim_{N \rightarrow \infty} [\ln \varphi^m(n, r) - \ln \varphi^{m+1}(n, r)] \quad (21)$$

1.4 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^[18]。

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:

$$f_t = \sigma[\mathbf{W}_f \cdot (h_{t-1}, x_t) + \mathbf{b}_f] \quad (22)$$

其中, \mathbf{W}_f 表示遗忘门到较长向量中的权重矩阵; \mathbf{b}_f 表示偏置值; σ 表示 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:

$$\tilde{C}_t = \tanh[\mathbf{W}_c \cdot (h_{t-1}, x_t) + \mathbf{b}_c] \quad (23)$$

$$i_t = \sigma[\mathbf{W}_i \cdot (h_{t-1}, x_t) + \mathbf{b}_i] \quad (24)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (25)$$

其中, \mathbf{W}_i 、 \mathbf{W}_c 表示输入门的权值矩阵; \mathbf{b}_i 、 \mathbf{b}_c 表示输入门的偏差向量; C_t 表示当前输入单元状态。

LSTM 模型结构如图 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。

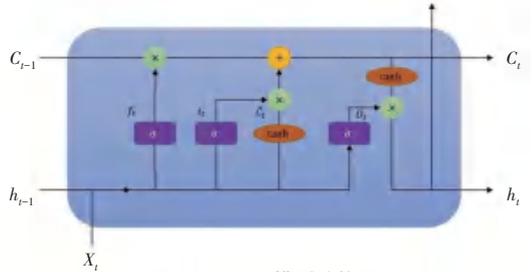


图 1 LSTM 模型结构图

Fig. 1 LSTM model structure diagram

1.5 基于 ICEEMDAN-DBO-LSTM 的股票指数预测模型

本文构建了一种新的组合预测方法 ICEEMDAN-DBO-LSTM 来提高股指的预测精度。模型构建的工作流程如图 2 所示。

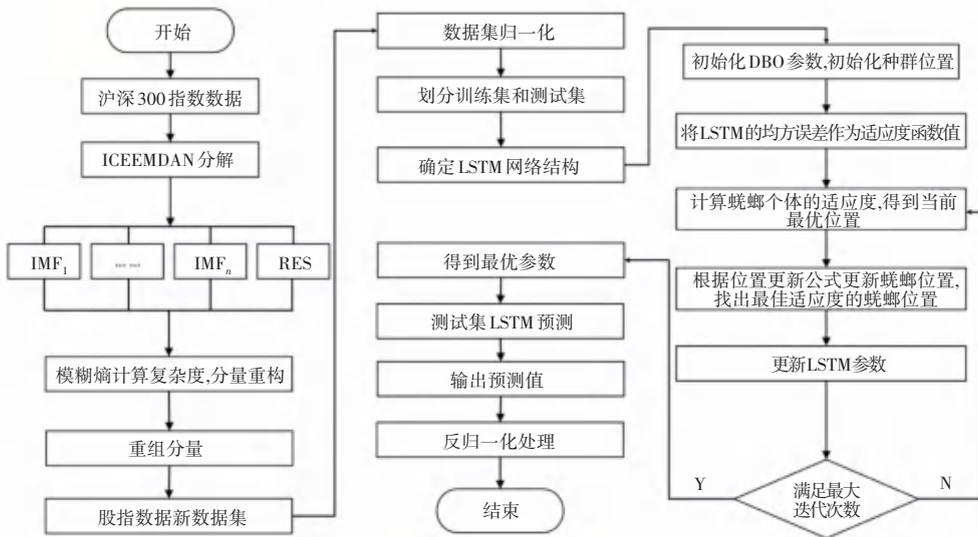


图 2 模型框架流程图

Fig. 2 Flowchart of the model framework

基于图 2, 现对模型框架流程给出阐释分述如下。

(1)原始序列分解:使用 ICEEMDAN 分解方法,将原始股票价格序列分解为有限的相对简单和有意义的 IMF 成分。

(2)模型分量重建: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)分量预测:使用 DBO 优化过的 LSTM 模型来预测这 3 个不同的分量,得到各自的预测值。

(4)线性集成:对上述 3 个分量的预测值进行线性集成,得到最终的预测结果。

2 实证分析

2.1 数据描述

为了测试本文 ICEEMDAN-DBO-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。股票指数的数据跨度为 2017 年 1 月 3 日至 2023 年 3 月 1 日,共 1 519 个交易日数据。其中,训练集占总数据 90%,测试集占总数据 10%。ICEEMDAN-DBO-LSTM 模型的参数设置如下:DBO 算法种群规模为 30,最大迭代次数为 30;LSTM 网络结构为 3 层网络结构,隐层神经

元个数为 50, 时间步长为 30。数据统计信息见表 1。

由表 1 可知, 日收盘价序列的偏度为正但有轻微的偏斜, 峰度值为正但没有超过 3, 没有出现明显

的尖峰现象。意味着收盘价数据中的异常值或极端值相对较多, 预测较为困难。

表 1 沪深 300 指数描述性统计

Table 1 Descriptive statistics of CSI 300 index

序列	样本值	最大值	最小值	均值	标准差	偏度	峰度
CSI 300	1 519	5 807.7	2 964.8	4 105.8	596.78	0.525	2.467 5

2.2 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预测效果的影响, 本文使用最大最小值标准化来对各模态分子序列数据进行归一化处理, 其表达式为:

$$X^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (26)$$

其中, X^* 表示标准化后的子序列; x 表示初始值; x_{\max}, x_{\min} 分别表示各模态分量的最大值和最小值。

本文选取了 5 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)、平均绝对误差 (MAE) 及拟合优度 (R^2), 其计算公式为:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| \quad (27)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_t - \hat{x}_t)^2} \quad (28)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_t - \hat{x}_t| \quad (29)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{x}_t - x_t)^2}{\sum_{i=1}^n (x_t - \bar{x}_t)^2} \quad (30)$$

其中, x_t 和 \hat{x}_t 分别表示 t 时刻收益率的真实值和预测值; \bar{x}_t 表示真实值的平均值; n 表示样本量。 R^2 取值范围为 $[0, 1]$, R^2 取值越接近 1 表示拟合效果越好, MAPE、RMSE 以及 MAE 的数值越小, 则越好。

2.3 股指收益率序列分解重构

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 指数收盘价分解图。

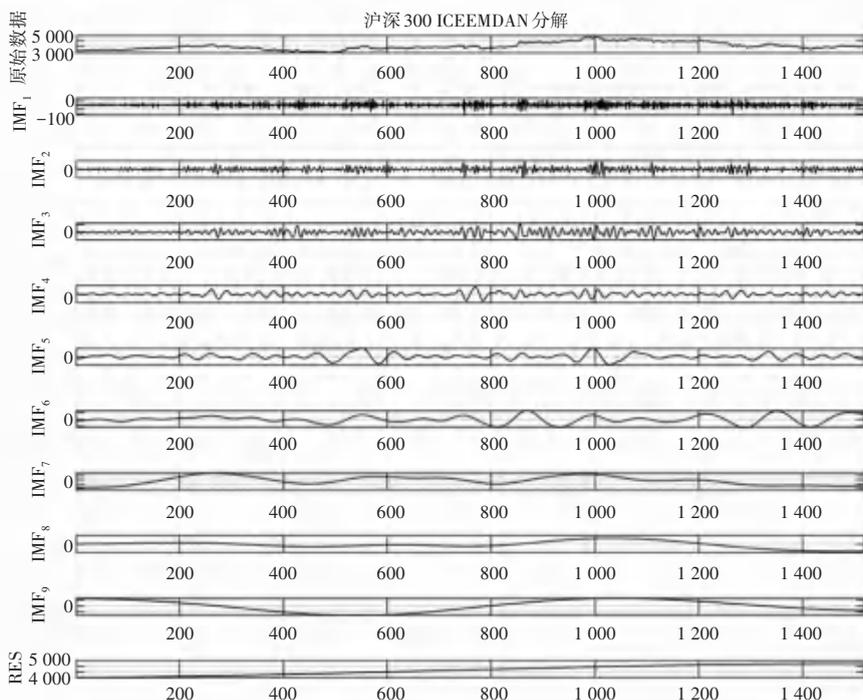


图 3 沪深 300 指数 ICEEMDAN 分解结果

Fig. 3 ICEEMDAN decomposition results of the CSI 300 index

如果对 ICEEMDAN 分解后的每一个分量直接使用 DBO-LSTM 模型进行预测, 会使计算的复杂性增加, 因此可对分量重组以提高预测效率。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 2 个阈值, $\lambda_1 = 1, \lambda_2 = 0.01$ 。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。表 2 中, 模态分量 1~3 被重构为高频分量, 4~9 被重构为低频分量, RES 为趋势分量。分解重构后的高频分量、低频分量和趋势分量的具体情况见表 3。

分解重构后子序列如图 4 所示。其中, IMF_1 为重构后的高频分量, IMF_2 为重构后的低频分量, IMF_3 为重构后的趋势分量。

表 2 各子序列模糊熵值及分类结果

Table 2 Fuzzy entropy values and classification results of each sub-sequence

序列	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅	IMF ₆	IMF ₇	IMF ₈	IMF ₉	RES
CSI300	2.926	2.298	1.479	0.879	0.649	0.498	0.227	0.088	0.061	0.002

表 3 各序列重构结果

Table 3 Reconstruction results of each sequence

序列	高频分量	低频分量	趋势分量
分量	IMF ₁ +IMF ₂ +IMF ₃	IMF ₄ +IMF ₅ +IMF ₆ +IMF ₇ +IMF ₈ +IMF ₉	RES

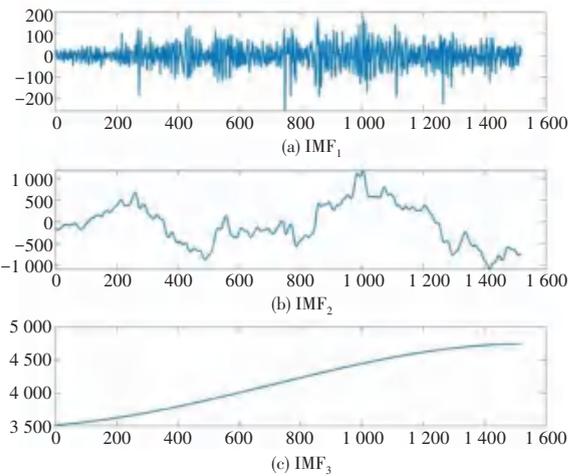


图 4 分解重构后子序列

Fig. 4 Decomposed and reconstructed sub-sequences

2.4 预测结果分析

为验证所提出的 ICEEMDAN-DBO-LSTM 模型的有效性, 建立了混合模型以及单一模型作为对照, 进行实验验证, 得到的预测结果对比如图 5 所示, 评价指标结果见表 4。

通过图 5 看出, 在单个预测模型中, LSTM 模型相对于 BP 模型, 其整体趋势和拟合度更好, 但随着时间的推移, LSTM 模型预测误差也变得愈发明显。ICEEMDAN-LSTM 和 DBO-LSTM 混合模型的预测曲线明显优于单一模型, 而 ICEEMDAN-DBO-LSTM 混合模型的曲线和股指序列的真实值十分接近, 拟合效果最好。因此, ICEEMDAN-DBO-LSTM 混合模型相对其他预测模型拥有更高的预测精度。

表 4 预测结果评价

Table 4 Evaluation of prediction results

模型	MAPE/%	RMSE	MAE	R ²
BP	1.33	63.642 0	52.052 6	0.847 1
LSTM	1.11	53.296 4	43.521 3	0.875 8
DBO-LSTM	0.92	45.569 6	36.044 7	0.916 5
ICEEMDAN-LSTM	0.84	36.399 5	29.290 4	0.929 4
ICEEMDAN-DBO-LSTM	0.69	31.284 3	24.930 9	0.950 1

通过表 4 预测结果分析可以看出, LSTM 预测模型的评价指标 MAPE 为 1.11%, RMSE 为 53.296 4, MAE 为 43.521 3, R² 为 0.875 8, 和 BP 模型相比, 均优于 BP 预测模型。另外, 通过 DBO 算法参数优化过的 DBO-LSTM 混合模型的 MAPE、

RMSE、MAE 和 R² 分别为 0.92%、45.569 6、36.044 7 和 0.916 5, 和基础的 LSTM 模型相比, 预测精度得到很好的改善。数据表明 DBO 算法对于 LSTM 中的参数寻优问题有很好的帮助。最后, ICEEMDAN-DBO-LSTM 预测模型的 MAPE、RMSE、

MAE 和 R^2 的值都是最优,分别为 0.69%、31.284 3、24.930 9 和 0.950 1,和 DBO-LSTM 混合模型相比,预测精度有明显提升,且使用 ICEEMDAN 算法分解的 ICEEMDAN-LSTM 混合模型的预测误差比 DBO-LSTM 模型的预测误差也有所降低,说明在股指预测中,使用 ICEEMDAN 算法分解能够降低序列的复杂度,提高预测精度。

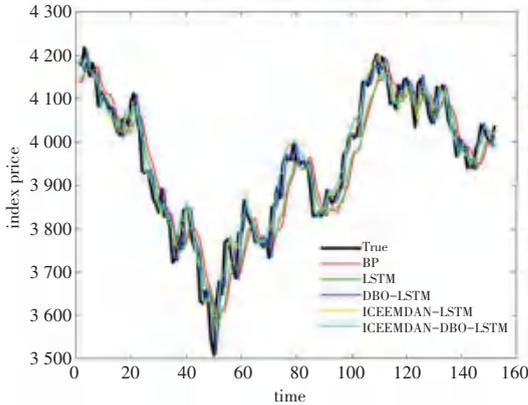


图5 预测曲线对比

Fig. 5 Comparison of prediction curves

3 结束语

针对股指序列数据非线性及噪音多导致难以预测的问题,提出基于 ICEEMDAN-DBO-LSTM 预测模型,以此提高预测准确性。通过实验结果可知,ICEEMDAN-DBO-LSTM 模型的效果最好,预测误差最小;其中,ICEEMDAN 分解方法能够将股指数数据分解为更简单的分量,以便更精确地进行预测;DBO 算法能够有效地优化 LSTM 总的超参数,减少人为因素影响,提高模型的预测精度。

参考文献

[1] BURLANDO P, ROSSO R, CADAVID L G, et al. Forecasting of short-term rainfall using ARMA models [J]. Journal of Hydrology, 1993, 144(1-4): 193-211.

[2] DAREKAR A, REDDY A. Predicting market price of soybean in major India studies through ARIMA model [J]. Journal of Food Legumes, 2017, 30(2): 73-76.

[3] 杨继平,冯毅俊,王辉. 基于结构转换 PTT-GARCH 模型沪深股市波动率的估计 [J]. 系统工程理论与实践, 2016, 36(9): 2205-2215.

[4] CHEN Shun, GE Lei. Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction [J]. Quantitative Finance, 2019, 19(9): 1507-1515.

[5] NIU Hongli, XU Kunliang. A hybrid model combining variational mode decomposition and an attention-GRU network for stock price index forecasting [J]. Mathematical Biosciences and Engineering, 2020, 17(6): 7151-7166.

[6] 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.

[7] TANG Ling, YU Le'an, WANG Shuai, et al. A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting [J]. Applied Energy, 2012, 93: 432-443.

[8] 李婧琦. 基于鲸鱼算法优化 LSTM 的股票价格预测模型 [J]. 智能计算机与应用, 2023, 13(2): 35-40.

[9] GÜLMEZ B. Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm [J]. Expert Systems with Applications, 2023, 227(3): 231-247.

[10] JI Yi, LIEW A W C, YANG Lixia. A novel improved particle swarm optimization with long-short term memory hybrid model for stock indices forecast [J]. IEEE Access, 2021, 9(7): 23660-23671.

[11] 何凯,苏栳芳,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.

[12] 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.

[13] 陈凯杰,唐振鹏,吴俊传,等. 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.

[14] XUE Jiankai, SHEN Bo. Dung beetle optimizer: A new meta-heuristic algorithm for global optimization [J]. Journal of Supercomputing, 2022, 79(7): 7305-7336.

[15] COLOMINAS M A, SCHLOTTHAUER G, TORRES M E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing [J]. Biomedical Signal Processing and Control, 2014, 14(1): 19-29.

[16] 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.

[17] ZHU Keheng, CHEN Liang, HU Xiong. Rolling element bearing fault diagnosis based on multi-scale global fuzzy entropy, multiple class feature selection and support vector machine [J]. Transactions of the Institute of Measurement and Control, 2019, 41(14): 4013-4022.

[18] 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.