

基于 VMD-DBO-LSTM 的空气质量管理预测*

张诗云¹ 朱菊香² 张涛¹ 孙君峰¹ 张赵良²

(1. 南京信息工程大学自动化学院 南京 210000; 2. 无锡学院轨道交通学院 无锡 214105)

摘要:针对传统空气质量预测模型收敛速度慢,精度低的问题,提出一种基于变分模态分解(variational mode decomposition, VMD)和蜣螂优化算法(dung beetle optimizer, DBO)优化长短期记忆网络(long short term memory, LSTM)的预测模型。首先,针对 AQI 原始数据具有大量噪声的问题,使用 VMD 方法对非平稳信号进行模态分解以降低噪声对预测结果的影响从而获得多个不同特征的模态分量;其次,针对 LSTM 靠人工经验调参存在一定局限性,利用 DBO 算法对 LSTM 模型参数进行优化;最后,对分解后的各个子序列使用 LSTM 模型预测,将各个子序列进行叠加得到最后的预测结果。实验结果表明,VMD 对非平稳数据的分解有助于提高预测精度,VMD-DBO-LSTM 模型的性能较其他模型均有不同程度的提高,该模型预测的均方根误差为 $4.73 \mu\text{g}/\text{m}^3$,平均绝对误差为 $3.61 \mu\text{g}/\text{m}^3$,拟合度达到了 97.8%。

关键词:空气质量预测;变分模态分解;蜣螂优化算法;长短期记忆网络

中图分类号: TP393 **文献标识码:** A **国家标准学科分类代码:** 510.4030

Air quality predication based on VMD-DBO-LSTM

Zhang Shiyun¹ Zhu Juxiang² Zhang Tao¹ Sun Junfeng¹ Zhang Zhaoliang²

(1. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210000, China;

2. School of Rail Transportation, Wuxi University, Wuxi 214105, China)

Abstract: Aiming at the slow convergence speed and low accuracy of traditional air quality prediction models, a predication model based on variational mode decomposition (VMD) and dung beetle optimizer (DBO) was proposed to optimize long short term memory (LSTM). First of all, for the problem that the AQI raw data has a large amount of noise, the VMD method was used to decompose the nonstationary data to reduce the influence of noise on the prediction results, so as to obtain multiple modal components with different features. Secondly, there are some limitations in rely on manual parameter tuning based on human experience for LSTM, the DBO algorithm was used to optimize the LSTM model parameters. Finally, the LSTM model was used to predict each subseries after decomposition, and the subseries are superimposed to obtain the final prediction result. The experimental results show that the decomposition of nonstationary data by VMD can help improve the prediction accuracy, and the performance of VMD-DBO-LSTM model is improved to varying degrees compared with other models, the root mean square error of this model prediction is $4.73 \mu\text{g}/\text{m}^3$, the average absolute error is $3.61 \mu\text{g}/\text{m}^3$, the goodness of fit reach 97.8%.

Keywords: air quality predication; variational mode decomposition; dung beetle optimizer; long short-term memory network

0 引言

随着经济、城市化、工业化发展,各种大气污染问题开始出现,空气质量问题一直受全球各国的关注,空气质量

指数(AQI)是公众了解未来空气质量状况最直接的途径^[1]。空气质量的下降会对经济,生态平衡,人类健康产生负面影响,因此,对 AQI 浓度的精准预测具有重要意义。

收稿日期:2023-09-13

* 基金项目:“太湖之光”科技攻关项目(k20221050)资助

大气污染物的预测模型大致可分为物理模型、统计模型和深度学习模型。物理模型是基于已知的物理原理和观测数据,通过建立方程来预测系统的行为或响应。统计模型包括自回归模型^[2](autoregressive model, AR)、自回归综合移动模型^[3](autoregressive integrated moving average, ARIMA)、多元线性回归模型^[4](multiple linear regression, MLR)和灰色模型^[5](grey model, GM)。上述模型原理简单,预测速度快,但不适应于对像大气污染物这种随机不确定性和非平稳的信号进行预测。近年来,随着技术发展,深度学习模型被许多学者广泛应用到空气质量的时间序列预测中。其中,Gennaro等^[6]使用神经网络预测西班牙东北部两个不同环境中的PM₁₀浓度,并取得了令人满意的效果。Himawan等^[7]使用循环神经网络(recurrent neural network, RNN),利用PM₁₀浓度、湿度和温度来预测年度和季节性室内空气质量,实验证明,多级RNN模型在空气质量预测领域表现出了优异的结果。更进一步的,作为RNN的改进,长短期记忆网络模型(long short-term memory, LSTM)克服循环神经网络梯度消失的问题,捕获具有周期性时间序列的不同特征^[8],在序列数据的建模和预测中更常见。Yan等^[9]考虑空气质量的时空分布特征,在整体预测中,LSTM相比BP神经网络(backpropagation neural network, BPNN)、卷积神经网络(convolutional neural network, CNN)和CNN-LSTM模型,是多小时预测中的最优模型。

LSTM模型的参数对模型的性能和训练效果有着至关重要的影响,而靠人工经验调参存在一定局限性,所以长期以来,优化问题一直是研究的焦点。很多仿生优化算法在空气质量预测领域取得了显著的成功,如根据鸟群和鱼群觅食行为提出的粒子群算法(particle swarm optimization, PSO)^[10]、根据鲸鱼捕食行为提出的鲸鱼搜索算法(whale optimization algorithm, WOA)^[11]和根据非洲秃鹫觅食行为提出的非洲秃鹫优化算法(african vulture optimization algorithm, AVOA)等^[12]为预测参数优化提供了新的途径。但是上述算法随着迭代次数增加均存在一些不足,例如PSO算法收敛速度较慢,WOA算法和AVOA算法会不同程度的陷入局部最优,相比较,蜣螂优化算法(dung beetle optimizer, DBO)在全球搜索能力,收敛速度、稳定性等方面更具有竞争力,能够有效搜索模型的最优超参数。Wu^[13]等使用蜣螂算法优化SVR模型的参数,建立变分模态分解(variational mode decomposition, VMD)-DBO-SVR组合模型预测锂离子电池健康状态。类似地,本文尝试将该算法部署到空气质量的时间序列预测中。

虽然深度学习有望提高空气质量的预测精度,但当遇到更复杂的问题时,单个网络的准确性会受到限制,于是越来越多的混合模型被用来提高预测精度。黄婕等^[14]将记忆能力较强的RNN和特征表达能力较强的CNN结合

起来建立RNN-CNN集成深度学习预测模型,经证明,RNN-CNN模型在PM_{2.5}时间序列预测上的表现明显优于集成之前的RNN和CNN。Zeng等^[15]通过带有自适应噪声的互补集成经验模态分解(complete ensemble empirical mode decomposition with adaptive noise analysis, CEEMDAN)对PM_{2.5}数据进行分解,Huang等^[16]采用经验模态分解(empirical mode decomposition, EMD)处理空气质量数据。虽然EMD和CEEMDAN算法充分发挥了优势,提高了预测精度,但在面对噪声水平很高时仍可能出现问题。VMD相比EMD和CEEMDAN在信号分解方面有一定优势,特别是在稳定性、参数选择和处理噪声信号方面。

针对上述问题,本文旨在建立VMD和DBO算法优化LSTM参数的混合模型在无锡市空气质量预测中的应用。首先使用VMD方法对AQI数据进行分解,其次利用DBO算法优化LSTM模型权重和阈值,最后将各个子序列的预测值叠加得到最终预测值。

1 基本原理

1.1 VMD

VMD模型是Dragomiretskiy等^[17]于2014年提出,是一种非递归变分模式的信号分解方法。假设原始输入信号为 $f(t)$,传统VMD模型的约束变分模型为:

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \cdot u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (1)$$

s. t. $\sum u_k = f(t)$ (2)

式中: $\{u_k\} = \{u_1, u_2, u_3, \dots, u_k\}$ 表示分解后的模态分量; $\{\omega_k\} = \{\omega_1, \omega_2, \omega_3, \dots, \omega_k\}$ 表示模态中心频率; $\delta(t)$ 表示狄拉克分布。约束条件表明,所有模态分量之和应为原始信号。为了求解上述方程的约束最优化问题,在拉格朗日展开表达式中利用二次惩罚因子 α 和拉格朗日乘数算子 $\lambda(t)$,引入增广函数,得到如下:

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \partial \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \cdot u_k(t) \right] \cdot e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \quad (3)$$

式(3)的解在频域中进行,步骤如下。

1) 初始化 $\{\hat{u}_k^1\}$ 、 $\{\hat{\omega}_k^1\}$ 、 $\hat{\lambda}^1 = 0$ 。

2) 对所有的 $\omega \geq 0$,更新 \hat{u}_k :

$$\hat{u}_k^{n+1}(\omega) \leftarrow \frac{\hat{f}(\omega) - \sum_{i < k} \hat{u}_i^{n+1}(\omega) - \sum_{i > k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (4)$$

3) 更新 ω_k :

$$\omega_k^{n+1} \leftarrow \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (5)$$

4)对所有的 $\omega \geq 0$,进行双重提升,更新 λ_k :

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}(\omega) + \tau(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)) \quad (6)$$

其中, τ 表示噪声容限,当信号含有强噪声时,可设定 $\tau=0$ 达到更好的去噪效果。

5)重复步骤 2)~4),直到满足迭代约束条件,则迭代结束。

$$\sum_k \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \epsilon \quad (7)$$

式中: n 表示整个过程的迭代次数; u_k^n 、 ω_k^n 、 λ^n 分别表示模态分量的序列,中心频率和热乘数。

1.2 DBO 算法

DBO 算法是 Shen 团队^[18]最近提出的一种新的基于群体智能优化算法,主要是受到蜣螂的滚球、跳舞、觅食、偷窃和繁殖行为的启发。根据所开发的 DBO 算法,蜣螂群由滚球蜣螂、孵球蜣螂、小蜣螂和偷窃蜣螂组成。

为了模拟滚球行为,蜣螂需要在整个搜索空间中沿着给定的方向移动,假设光源的强度也会影响蜣螂的路径。在滚动过程中,滚球蜣螂的位置被更新:

$$\begin{cases} x_i(t+1) = x_i(t) + \alpha \times k \times x_i(t-1) + b \times \Delta x \\ \Delta x = |x_i(t) - X^w| \end{cases} \quad (8)$$

式中: t 表示当前迭代次数; $x_i(t)$ 表示第 i 只蜣螂在第 t 次迭代时的位置信息; $k \in (0, 0.2]$; b 表示 $(0, 1)$ 的常数; α 是赋值为 -1 或 1 的自然系数; X^w 表示全局最差位置; Δx 模拟光强的变化,值越高,表示光源越弱。

当蜣螂遇到障碍物需要通过跳舞来重新定位,为了模拟舞蹈行为使用正切函数获得新的滚动方向。将滚球蜣螂的位置更新定义如下:

$$x_i(t+1) = x_i(t) + \tan\theta |x_i(t) - x_i(t-1)| \quad (9)$$

式中: $|x_i(t) - x_i(t-1)|$ 是第 i 只蜣螂在第 t 次迭代与第 $t-1$ 次迭代时的位置之差; θ 表示倾斜角度,当 θ 为 0 、 $\pi/2$ 、 π 时,位置不更新。

选择合适的产卵地点对蜣螂来说至关重要,于是提出了一种边界选择策略来模拟雌性蜣螂产卵的区域,产卵区域的动态更新可以促进对局部地区的开发,定义如下:

$$\begin{cases} Lb^* = \max(X^* \times (1 - R), Lb) \\ Ub^* = \min(X^* \times (1 + R), Ub) \end{cases} \quad (10)$$

式中: X^* 表示当前局部最优位置; Lb^* 和 Ub^* 分别表示产卵区域的下界和上界; $R = 1 - t/T_{max}$, 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。

确定产卵区域后,雌性蜣螂就会选择此区域的卵球产卵,由式(10)可看出,产卵区的边界范围是动态变化的。因此,孵球蜣螂的位置在迭代过程中也是动态的,定义如下:

$$B_i(t+1) = X^* + b_1 \times (B_i(t) - Lb^*) + b_2 \times (B_i(t) - Ub^*) \quad (11)$$

式中: $B_i(t)$ 表示第 i 个孵球蜣螂在第 t 次迭代时的位置信息; b_1 和 b_2 表示大小为 $1 \times D$ 的两个独立随机向量, D 表示优化问题的维数。

长大的蜣螂会从地下钻出来寻找食物,被称为小蜣螂,因此还需建立最佳觅食区域来指导蜣螂觅食,最优觅食区域边界定义如下:

$$\begin{cases} Lb^b = \max(X^b \times (1 - R), Lb) \\ Ub^b = \min(X^b \times (1 + R), Ub) \end{cases} \quad (12)$$

式中: X^b 表示全局最佳位置; Lb^b 和 Ub^b 分别表示最优觅食区域的下界和上界;其他参数定义如式(10)。小蜣螂的位置更新如下:

$$x_i(t+1) = x_i(t) + C_1 \times (x_i(t) - Lb^b) + C_2 \times (x_i(t) - Ub^b) \quad (13)$$

式中: $X_i(t)$ 表示第 i 只小蜣螂在第 t 次迭代时的位置信息; C_1 表示服从正态分布的随机数; C_2 表示属于 $(0, 1)$ 的随机向量。

一些蜣螂会从其他蜣螂身上偷粪球是自然界很常见的现象,它们被称为偷窃蜣螂,其偷窃行为保证了算法能够跳出局部最优解。在迭代过程中,小偷的位置信息更新如下:

$$x_i(t+1) = X^b + S \times g \times (|x_i(t) - X^*| + |x_i(t) - X^b|) \quad (14)$$

式中: $x_i(t)$ 表示第 i 个小偷在第 t 次迭代时的位置信息; g 是服从正态分布,大小为 $1 \times D$ 的随机向量; S 表示常数。

DBO 算法步骤大致可以概括为:初始化蜣螂群和 DBO 算法参数、根据目标函数计算适应度值 f_b 、更新滚球蜣螂、孵球蜣螂、小蜣螂和偷窃蜣螂的位置、判断各个蜣螂是否在边界外、更新当前最优解及其适应度值、重复上述步骤,直到 t 满足终止条件,输出全局最优解 X^b 及其适应度值 f_b 。

1.3 LSTM 网络

LSTM 是一种用于处理序列数据的 RNN 的变体,有效缓解了 RNN 无法避免的梯度爆炸问题,能更好预测时间序列。LSTM 存储单元的体系结构如图 1 所示。

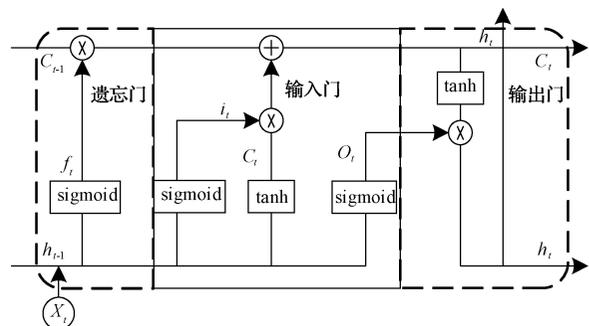


图 1 LSTM 结构

LSTM 在状态路径上利用遗忘门、输入门和输出门调节细胞状态。遗忘门决定丢弃哪些信息,方向为将前一个单元的输出 h_{t-1} 和输入向量 x_t 输入遗忘门,通过 sigmoid 激活函数计算遗忘门的输出值 f_t ;输入门负责决定哪些新信息必须传递到单元状态,方向为将 h_{t-1} 和 x_t 输入输入门,分别通过 sigmoid 函数和 tanh 激活函数得到输入门神经网络层的输出向量 i_t 和更新记忆细胞信息的向量 \tilde{C}_t ;输出门决定当前时间的输出信息,方向为将当前时间步的隐藏状态和输入数据作为输入,利用 sigmoid 函数计算输出门的输出 O_t ,最后将输出门的向量与当前时间步的隐藏状态相乘,得到 LSTM 的输出。

2 基于 VMD-DBO-LSTM 预测模型

2.1 数据来源和处理

本文以江苏省无锡市的空气质量作为研究对象,无锡市是一座工业城市,拥有众多的重工业,重工业不仅对当地经济发展产生不容小觑的作用还会对空气质量产生一定的影响,所以对现代化工业企业在环境保护方面有更高的标准和要求。综上,对无锡市空气质量的研究具有重要意义。

本文选取无锡市 3 460 组空气质量数据,其中 80% 的数据用于训练,20% 的数据用于测试,这些数据包括 2014 年~2023 年 6 月 22 日的 AQI 浓度,原始 AQI 浓度序列如图 2 所示,其来自中国空气质量在线监测分析平台(www.aqistudy.cn/historydata/)。由于原始数据存在特征差距较大现象,训练时间过长,为了提高模型的训练速度,通过归一化操作将数据映射到(0,1]。归一化公式如下:

$$x_i = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (15)$$

式中: x 表示原始数据; x_{\max} 和 x_{\min} 分别表示最大数据和最小数据; x_i 是归一化后的数值。

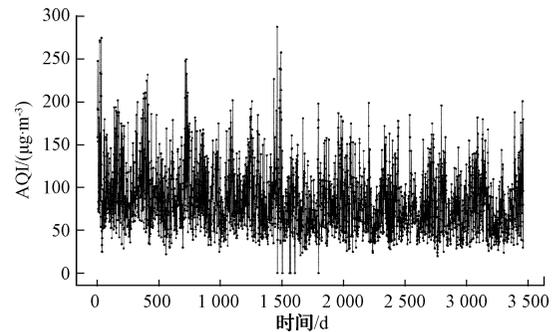


图 2 原始 AQI 浓度序列

本文需要提前确定 VMD 参数 K , 预实验表明 K 取值过大会导致过分解,反之,则会欠分解。通过观察不同 K 值下是否具有相似的中心频率,即当最后一层 IMF 分量的中心频率保持相对稳定时,参数 K 就被确定下来。对应的中心频率如表 1 所示,多次调试最终确定最佳模态 $K=6$ 时的分解效果最好。使用 VMD 方法对原始数据进行分解,结果如图 3 所示。

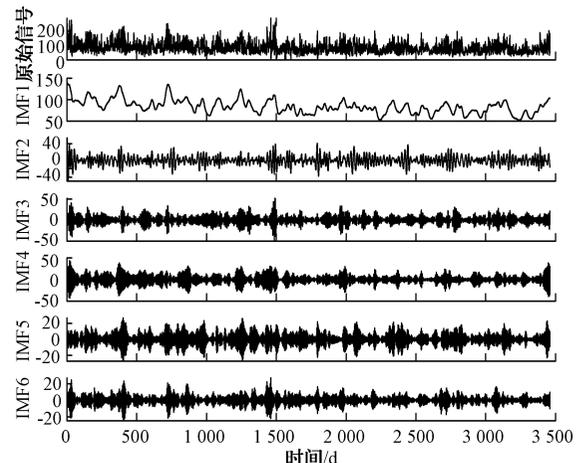


图 3 利用 VMD 分解后的 AQI 浓度序列

表 1 不同 K 值下 IMF 分量对应的中心频率

K	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8
4	0.000 1	0.091 9	0.188 8	0.296 8				
5	0.000 1	0.087 4	0.170 9	0.277 7	0.416 8			
6	0.000 1	0.063 2	0.130 2	0.196 4	0.282 7	0.418 6		
7	0.000 1	0.045 7	0.104 5	0.171 1	0.238 8	0.299 0	0.418 6	
8	0.000 1	0.044 6	0.102 2	0.167 2	0.223 9	0.280 1	0.335 0	0.419 5

由图 3 可知,VMD 方法抑制了 EMD 方法存在模态混叠的现象,降低了复杂度高和非线性强的时间非平稳性,分解得到了 6 个清晰且相对平稳的子序列。VMD 方法分解后的重构如图 4 所示,可以看出,原始信号序列大于重构信号序列之和,这是因为 VMD 分解可以达到去除噪声的效果。

2.2 评估指标

为了评估 VMD-DBO-LSTM 网络模型的预测性能,

本文采用均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)和拟合度 R^2 对算法模型进行评价,公式如下:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (16)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (17)$$

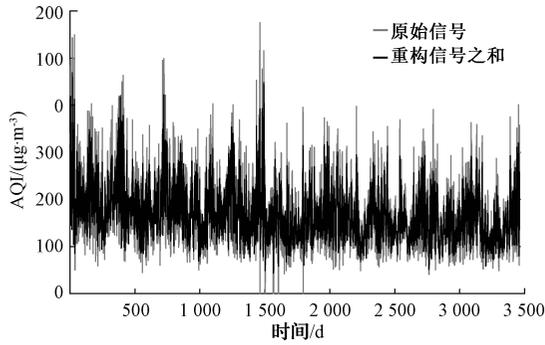


图4 AQI 重构浓度序列

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18)$$

式中： y_i 和 \hat{y}_i 分别是测试集的真实值和算法模型的预测值； n 为测试总次数； \bar{y} 表示测试集的真实值的平均值。其中 RMSE 和 MAE 的值越小，模型预测效果越好， R^2 的值越接近 1，预测模型的拟合效果越好。

2.3 预测模型叙述

本文基于 VMD-DBO-LSTM 预测模型的流程如图 5 所示。

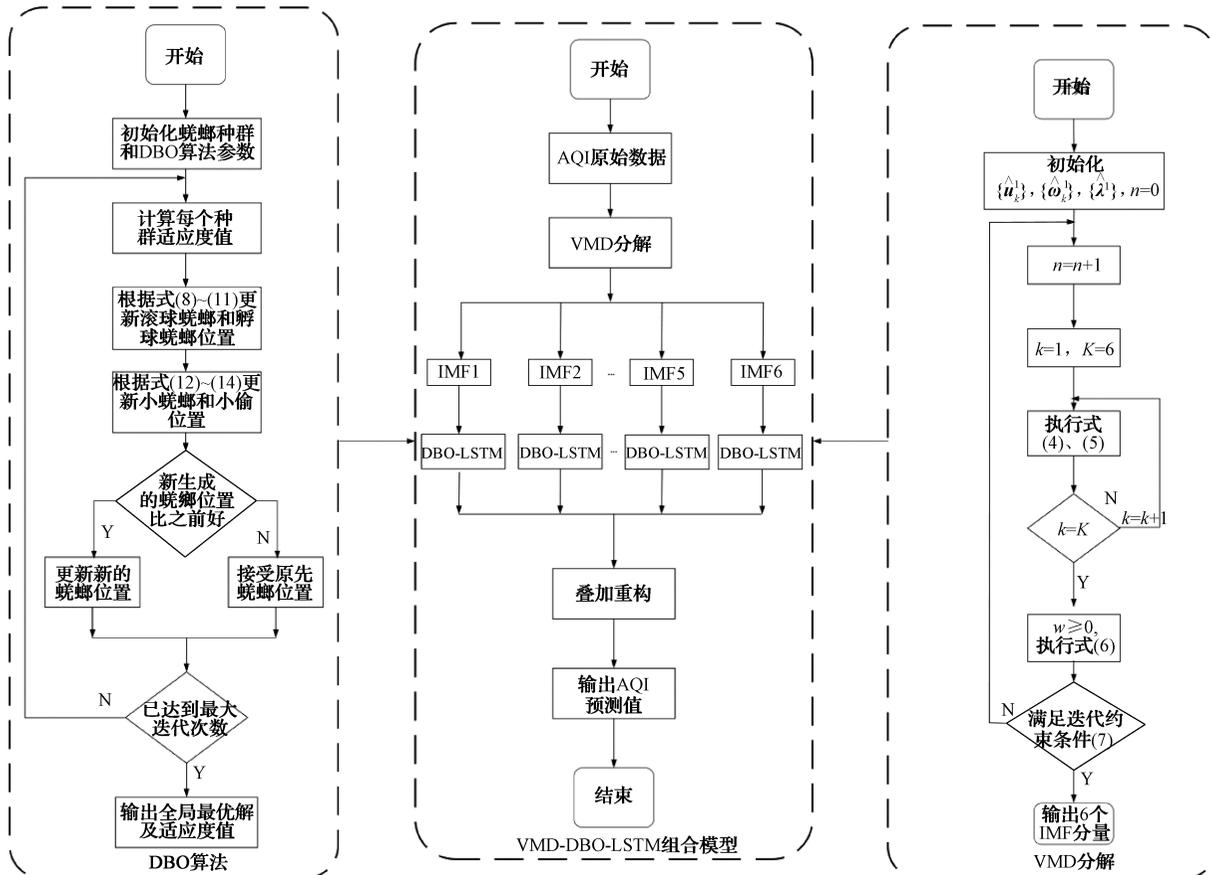


图5 VMD-DBO-LSTM 模型流程

1)使用中心频率法,计算不同 K 值下的中心频率来确定各模态分量。

2)利用 VMD 算法将原始数据分解为多组模态分量 IMF,以抑制和去除测试集数据的噪声,分解非平稳数据从而提高预测精度。

3)将 AQI 数据分为两部分,一部分为训练样本,占比为 80%,另一部分为测试样本,占比为 20%。

4)待数据归一化处理后将训练集数据输入到 DBO-LSTM 模型中,对 AQI 时间序列进行时间特征挖掘。

5)将重构后的 AQI 测试集输入到优化的 LSTM 模型中来,并将各个子序列的预测结果进行叠加求和得到最终的 AQI 的预测结果。

3 实验结果与讨论

3.1 DBO 算法性能分析

为了验证 DBO 技术的搜索性能,本文使用 6 种测试函数对 PSO 算法、WOA 算法、SSA 算法、GWO 算法进行了比较,测试函数如表 2 所示。为了公平比较不同优化方法的综合搜索能力,实验应在相同的环境中进行,因此设置测试函数种群数量为 30,最大迭代次数为 500,所有测试函数独立运行 30 次进行寻优,单峰函数和多峰函数维度设为 30,固定维多峰函数维度设为 2,各算法参数设置

如表 3 所示。以下测试函数仿真实验均采用 MATLAB (位),处理器为 Intel(R) Core(TM) i5-10210U CPU@ R2019b 作为仿真软件,操作系统为 Microsoft Win11(64 1.60 GHz 2.11 GHz,内存为 16 G。

表 2 测试函数

函数名	测试函数	n	测试范围
Sphere	$F_1(x) = \sum_{i=1}^n x_i^2$	30	[-100,100]
Schwefel's 1.2	$F_3(x) = \sum_{i=1}^n (\sum_{j=1}^j x_j)^2$	30	[-100,100]
Ackley	$F_{10}(x) = -20 \exp\left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + e$	30	[-32,32]
Generalized penalized Function1	$F_{12}(x) = \frac{\pi}{n} \left\{ 10 \sin(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 [1 + 10 \sin^2(\pi y_{i+1})] + (y_n - 1)^2 \right\} + \sum_{i=1}^n u(x_i, 10, 100, 4) y_i = 1 + \frac{x_i + 1}{4} u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m x_i > a \\ 0 - a < x_i < a \\ k(-x_i - a)^m x_i < -a \end{cases}$	30	[-50,50]
Six-hump camel back	$F_{16}(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$	2	[-5,5]
Goldstein-Price function	$F_{18}(x) = [1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \times [30 + (2x_1 - 3x_2)^2 \times (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$	2	[-2,2]

表 3 算法参数设置

算法模型	参数设置
SSA	$V_0 = 0$
PSO	$C_1 = C_2 = 2, W = (0.9, 0.1)$
WOA	$a = (0, 2)$
GWO	$a_{\min} = 0, a_{\max} = 2$
DBO	$k = \lambda = 0.1, b = 0.3, s = 0.5$

测试函数的三维模型及收敛曲线如图 6 所示,表 4 是各种算法对基本函数的测试结果,为了评估算法稳定性,引入函数均值(the mean value, Mean)和标准差(standard deviation, Std),公式如式(19)、(20)。由图 6 可看出,对于单峰函数而言,DBO 算法在迭代前期收敛速度更快且在后期尽可能收敛到全局最优位置,其他算法收敛速度远不及于 DBO 算法。对于多峰函数而言,DBO 算法收敛速度仍然最快,且能够避免陷入局部最优,而其他算法均会不同程度陷入局部最优值。对于固定维数的多模态函数而言,各算法的搜索能力相似,综上 DBO 算法得到的优化结果具有竞争性。由表 4 可看出,DBO 在绝大多数情形下适应度均值及标准差都更小,即有明显的优越性。总之,DBO 相比其他算法表现的更稳定,收敛速度更快,也不容易陷入局部最优陷阱。

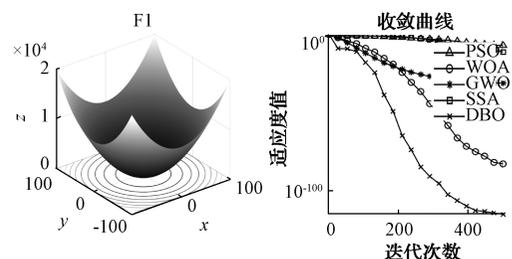
$$Mean = \frac{1}{P} \sum_{i=1}^P f_i \quad (19)$$

$$Std = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (f_i - Mean)^2} \quad (20)$$

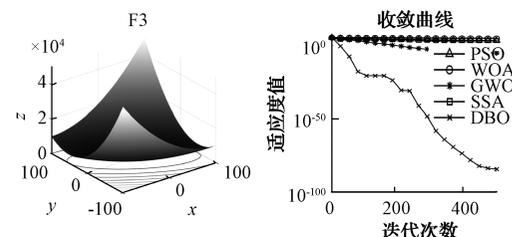
式中:P 表示优化实验次数; f_i 表示每次独立运行的最优值。

3.2 预测模型性能对比分析

为了验证引入 VMD 算法的优越性,本文将 EMD-LSTM、CEEMDAN-LSTM、VMD-LSTM 和 VMD-DBO-LSTM 4 个预测模型结果进行对比,得到的预测结果曲线如图 7 所示,不同模型得出的空气质量预测值与真实值之间的误差评价如表 5 所示。从图 7 可以看出,VMD-



(a) Sphere函数三维模型和收敛曲线



(b) Schwefel's 1.2函数三维模型和收敛曲线

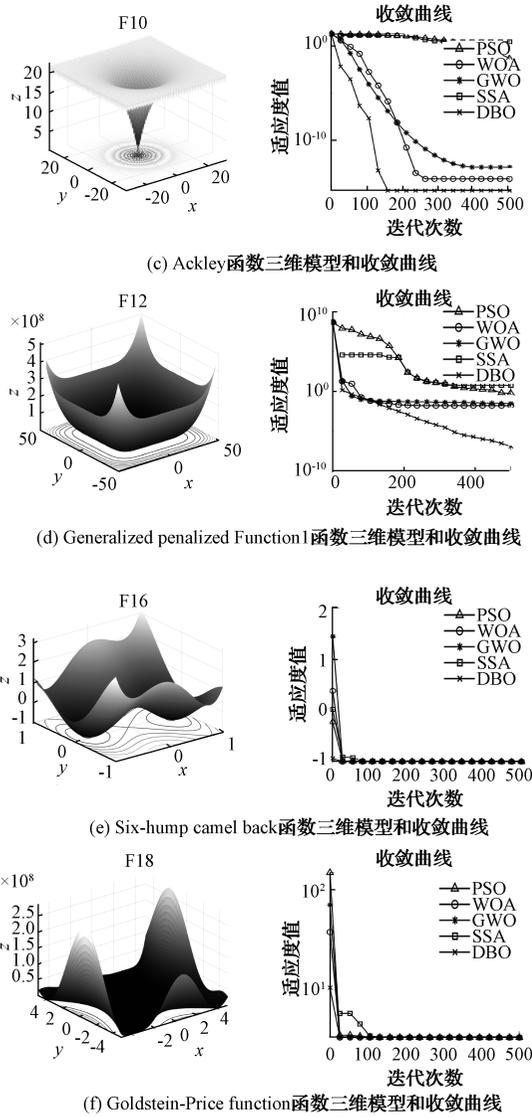


图6 函数三维模型及收敛曲线

DBO-LSTM模型的预测值与原始的真实数据拟合程度最高,从表5可知,基于VMD-DBO-LSTM模型的均方根误差为 $4.73 \mu\text{g}/\text{m}^3$,平均绝对误差为 $3.61 \mu\text{g}/\text{m}^3$,拟合度达到了97.8%。对比拟合度可得,VMD-DBO-LSTM模型与EMD-LSTM相比,预测精度提高了35%;与CEEMDAN-LSTM相比,预测精度提高了29.7%;与VMD-LSTM相比,预测精度提高了5.1%。结果证明VMD方法去噪效果显著,提高了DBO-LSTM模型的预测精度。

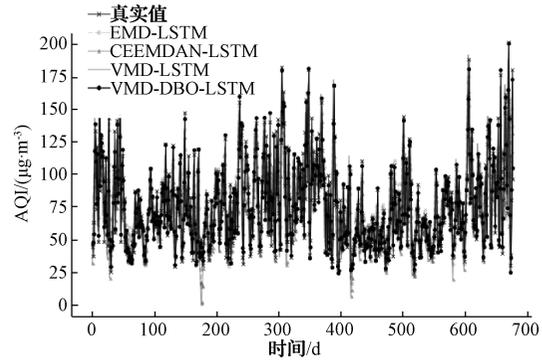


图7 各预测模型结果对比

为了全面评估VMD-DBO-LSTM模型的预测性能,本文将单一模型LSTM、混合模型CEEMDAN-CNN-LSTM、VMD-AVOA-LSTM和VMD-DBO-LSTM模型进行对比,得到的预测结果曲线和相对误差分别如图8和9所示,各预测模型的评估指标如表6所示。由图8和9可知,单一模型LSTM预测精度最差,VMD-DBO-LSTM模型误差最小,预测得到的曲线波动幅度与真实值最接近。由表6可知,VMD-DBO-LSTM模型的均方根误差、平均绝对误差与单一模型LSTM相比,降低了21.67、16.79 $\mu\text{g}/\text{m}^3$,拟合度提高了65.3%;同样对比模型CEEMDAN-CNN-LSTM和VMD-AVOA-LSTM,本文所提模型的均方根误差和平均绝对误差最低,拟合度最

表4 各种算法对基本函数的测试结果

算法	$F_1(x)$		$F_3(x)$		$F_{10}(x)$		$F_{12}(x)$		$F_{16}(x)$		$F_{18}(x)$	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
SSA	6.85×10^{-7}	5.65×10^{-7}	1.39×10^3	2.01×10^2	2.579	8.04×10^{-1}	7.696	1.992	-1.036	5.00×10^{-3}	3.00	2.51×10^{-13}
PSO	3.60×10^{-3}	3.01×10^{-5}	1.63×10^3	3.75×10^1	5.09×10^{-2}	2.15×10^{-4}	1.032	3.90×10^{-2}	-1.036	2.82×10^{-4}	3.00	2.57×10^{-15}
WOA	1.94×10^{-82}	2.26×10^{-68}	4.22×10^3	1.67×10^4	7.99×10^{-15}	3.01×10^{-15}	5.00×10^{-3}	1.80×10^{-2}	-1.036	2.00×10^{-3}	3.00	3.55×10^{-4}
GWO	2.32×10^{-28}	4.55×10^{-27}	8.39×10^{-6}	2.82×10^{-5}	1.04×10^{-13}	1.50×10^{-14}	6.00×10^{-2}	2.90×10^{-2}	-1.036	8.31×10^{-3}	3.00	3.44×10^{-5}
DBO	2.33×10^{-116}	9.7×10^{-104}	1.42×10^{-84}	8.88×10^{-79}	8.68×10^{-16}	0.00	5.47×10^{-6}	3.55×10^{-6}	-1.036	4.02×10^{-4}	3.00	5.00×10^{-15}

表5 模型评价指标

算法模型	RMSE	MAE	$R^2/\%$
	$/(\mu\text{g}\cdot\text{m}^{-3})$	$/(\mu\text{g}\cdot\text{m}^{-3})$	
EMD-LSTM	19.60	14.64	62.8
CEEMDAN-LSTM	18.17	13.24	68.1
VMD-LSTM	8.66	6.37	92.7
VMD-DBO-LSTM	4.73	3.61	97.8

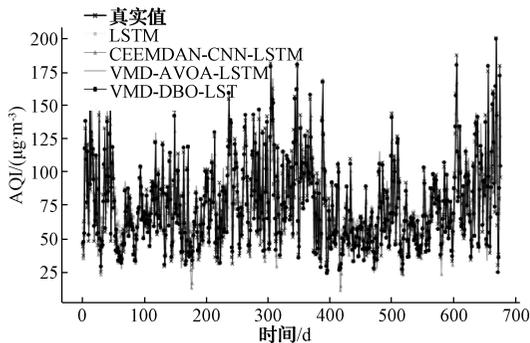


图8 各预测模型结果对比

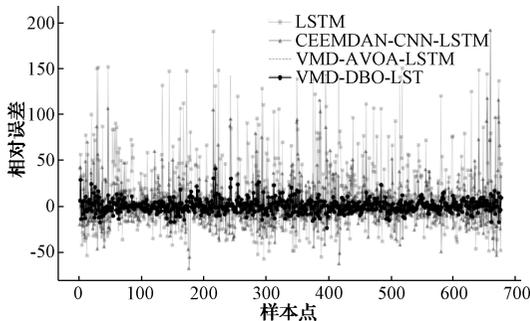


图9 各预测模型相对误差对比

表6 本文模型与单一模型、混合模型对比结果

算法模型	RMSE	MAE	$R^2/\%$
	$/(\mu\text{g}\cdot\text{m}^{-3})$	$/(\mu\text{g}\cdot\text{m}^{-3})$	
LSTM	26.4	20.4	32.5
CEEMDAN-CNN-LSTM	17.52	12.71	70.3
VMD-AVOA-LSTM	7.50	5.54	94.6
VMD-DBO-LSTM	4.73	3.61	97.8

高,预测精度大大提升。以上实验证明了本文所提 VMD-DBO-LSTM 组合预测模型的优越性和适用性。

4 结论

本文以无锡市空气质量数据为基础,提出了一种基于 VMD-DBO-LSTM 的空气质量预测模型,采用 VMD 方法将原始数据分解成多个 IMF 分量,将各个子序列分放入到 DBO 算法优化 LSTM 的模型中预测,建立各分量的 LSTM 子模型,对子模型预测的结果进行叠加得到最终预

测结果。通过实验与分析,得到如下结论。1)变分模态分解能将非平稳信号分解成多个不同频率尺度且相对平稳清晰的子序列,以此降低序列的复杂度和非线性,同时达到克服模态混叠现象的效果。2)DBO 算法在收敛速度、预测精度及稳定性方面与先进的优化方法相比有更好的效果,引入 DBO 算法优化 LSTM 模型参数使 VMD-DBO-LSTM 模型预测精度大幅度提升。3)本文所提模型可对 AQI 浓度进行有效预测,是一个可应用的模型,有助于治理空气污染问题,为无锡市大气污染治理和相关政策制定提供依据支撑。

参考文献

- [1] 王晓彦,陈佳,朱莉莉,等. 城市环境空气质量指数范围预报方法初探[J]. 中国环境监测, 2015, 31(6): 139-142.
- [2] SYAFIQ M S, NORLIN K, MOHD H A A, et al. Impact analysis of the external shocks on the prices of malaysian crude palm oil: Evidence from a structural vector autoregressive model[J]. Mathematics, 2022, 10(23): 4599.
- [3] HERNANDEZ-MATAMOROS A, FUJITA H, HAYASHI T, et al. Forecasting of COVID19 per regions using ARIMA models and polynomial functions[J]. Applied Soft Computing Journal, 2020, 96: 106610.
- [4] CHUNHUA P. Analysing the characteristics of precipitation variation in Hefei in recent years based on multiple linear regression[J]. Academic Journal of Environment & Earth Science, 2022, 4(8): 35-40.
- [5] 黄海宏,颜碧琛,王海欣. 改进灰色 GM(1,1)模型的 EAST 快控电源输出电流预测[J]. 电子测量与仪器学报, 2021, 35(5): 128-136.
- [6] GENNARO D G, TRIZIO L, GILIO D A, et al. Neural network model for the prediction of PM 10 daily concentrations in two sites in the Western Mediterranean[J]. Science of the Total Environment, 2013, 463-464: 875-883.
- [7] HIMAWAN N, TEGUH A P, MORSHED M A, et al. Multilevel RNN-based PM10 air quality prediction for industrial internet of things applications in cleanroom environment[J]. Wireless Communications and Mobile Computing, 2022, DOI: 10.1155/2022/1874237.
- [8] 陈思佳,罗志增. 基于长短时记忆和卷积神经网络的手势肌电识别研究[J]. 仪器仪表学报, 2021, 42(2): 162-170.
- [9] YAN R, LIAO J Q, YANG J, et al. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal

- clustering[J]. Expert Systems with Applications, 2021, 169: 114513.
- [10] 王强,秦华旺,齐春帅,等. 基于自适应 SA-PSO 改进的 XGBoost 气温预测方法[J]. 电子测量技术, 2023, 46(7):67-72.
- [11] 周冬阳,王彦海,刘晓亮,等. 基于 IWOA-BP 神经网络的滑坡区塔线体系应力预测模型[J]. 国外电子测量技术, 2023, 42(7): 121-131.
- [12] 张先起,赵玥,郑志文,等. 基于 VMD-AVOA-LSSVM 模型的月降水量预测研究[J]. 水电能源科学, 2022, 40(12): 1-5.
- [13] WU C, FU J, HUANG X, et al. Lithium-ion battery health state prediction based on VMD and DBO-SVR[J]. Energies, 2023, 16(10): 3993.
- [14] 黄婕,张丰,杜震洪,等. 基于 RNN-CNN 集成深度学习模型的 PM_{2.5} 小时浓度预测[J]. 浙江大学学报(理学版), 2019, 46(3): 370-379.
- [15] ZENG Q L, WANG L H, ZHU S Y, et al. Long-term PM_{2.5} concentrations forecasting using CEEMDAN and deep Transformer neural network [J]. Atmospheric Pollution Research, 2023, 14(9): 101839.
- [16] HUANG Y, YU J, DAI X, et al. Air-quality prediction based on the EMD-IPSO-LSTM combination model[J]. Sustainability, 2022, 14(9): 4889.
- [17] DRAGOMIRETSKIY K,ZOSSO D. Variational mode decomposition [J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
- [18] XUE J, SHEN B. Dung beetle optimizer: A new meta-heuristic algorithm for global optimization[J]. The Journal of Supercomputing, 2023, 79 (7): 7305-7336.

作者简介

张诗云,硕士研究生,主要研究方向为智能控制,检测技术。

E-mail:3109412978@qq.com

朱菊香,副教授,硕士生导师,主要研究方向为自动化及控制技术、检测技术。

E-mail:zjx@cw Xu. edu. cn

张涛,硕士研究生,主要研究方向为传感器应用,检测技术。

E-mail:1849173864@qq.com

孙君峰,硕士研究生,主要研究方向为环境感知、检测技术。

E-mail:850007512@qq.com

张赵良(通信作者),硕士,高级工程师,主要研究方向为控制工程、仪器仪表。

E-mail:zhangzl@cw Xu. edu. cn