

控制与决策

Control and Decision

基于EMD-PSO-LSSVM的碳排分解集成预测方法

张雯, 吴志彬, 徐玖平

引用本文:

张雯, 吴志彬, 徐玖平. 基于EMD-PSO-LSSVM的碳排分解集成预测方法[J]. *控制与决策*, 2022, 37(7): 1837-1846.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.1787>

您可能感兴趣的其他文章

Articles you may be interested in

基于小波变换与差分变异BSO-BP算法的大坝变形预测

Dam deformation prediction based on wavelet transform and differential mutation BSO-BP algorithm

控制与决策. 2021, 36(7): 1611-1618 <https://doi.org/10.13195/j.kzyjc.2019.1431>

基于FWADE-ELM的短时交通流预测方法

Short-term traffic flow forecasting based on hybrid FWADE-ELM

控制与决策. 2021, 36(4): 925-932 <https://doi.org/10.13195/j.kzyjc.2019.1103>

基于FWADE-ELM的短时交通流预测方法

Short-term traffic flow forecasting based on hybrid FWADE-ELM

控制与决策. 2021, 36(4): 925-932 <https://doi.org/10.13195/j.kzyjc.2019.1103>

基于多种群分解预测的动态多目标引力搜索算法

Dynamic multi-objective gravitational searching algorithm based on multi-population decomposition prediction

控制与决策. 2021, 36(12): 2910-2918 <https://doi.org/10.13195/j.kzyjc.2020.1002>

基于R2指标和目标空间分解的高维多目标粒子群优化算法

R2 indicator and objective space partition based many-objective particle swarm optimizer

控制与决策. 2021, 36(9): 2085-2094 <https://doi.org/10.13195/j.kzyjc.2020.0113>

基于EMD-PSO-LSSVM的碳排分解集成预测方法

张雯, 吴志彬[†], 徐玖平

(四川大学 商学院, 成都 610065)

摘要: 二氧化碳排放量的发展趋势作为能够反映各国减排措施的指标之一, 近些年来受到广泛关注. 为了缓解碳排放数据的非线性和波动性对预测精度造成的影响, 提出一种高效的分解集成预测方法用于预测二氧化碳的年排放量. 碳排的原始序列数据被经验模态分解 (empirical mode decomposition, EMD) 方法分解为不同频率的振动模块和残余项, 粒子群优化算法 (particle swarm optimization, PSO) 优化后的最小二乘支持向量机 (least squares support vector machine, LSSVM) 用于预测每个分解模块. 选取世界上 12 个国家的真实碳排数据进行实例验证, 预测结果表明: EMD 方法能够有效提高碳排预测的精准度; 与其他预测模型相比, 分解集成预测方法能够将平均绝对误差 (mean absolute error, MAE) 的均值最少提高 46.46%, 最多提高 90.09%, 将平均 Pearson 相关系数 (Pearson correlation coefficient, PCC) 值最少提高 10.45%, 最多提高 45.10%.

关键词: 碳排预测; 时间序列; 数据分解; 最小二乘支持向量机; 粒子群优化算法; 分解集成

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.1787

开放科学(资源服务)标识码(OSID):



引用格式: 张雯, 吴志彬, 徐玖平. 基于 EMD-PSO-LSSVM 的碳排分解集成预测方法[J]. 控制与决策, 2022, 37(7): 1837-1846.

A decomposition-integration forecasting method of carbon emission based on EMD-PSO-LSSVM

ZHANG Wen, WU Zhi-bin[†], XU Jiu-ping

(Business school, Sichuan University, Chengdu 610065, China)

Abstract: The emission of carbon dioxide has received much attention in recent years, as it can reflect the effectiveness of those low-carbon measures. To alleviate the nonlinearity and volatility of the annual carbon dioxide emissions, which may affect the forecast accuracy, this paper proposes an efficient decomposition-integration forecasting method to forecast the annual carbon emissions. The empirical mode decomposition (EMD) is used to decompose the original emission series into intrinsic oscillatory modes and a residual with different frequencies. The least squares support vector machine (LSSVM) is optimized using the particle swarm optimization (PSO) algorithm to predict each decomposed part. This paper chooses the real annual carbon emission of 12 countries all over the world to do the case study. The forecasting results indicate the validity of the EMD on improving the accuracy of carbon emission prediction. Furthermore, the comparison results between the EMD-PSO-LSSVM method and other forecasting models show the EMD-PSO-LSSVM can improve the average accuracy of the mean absolute error (MAE) at least 46.46% and at most 90.09%, and can improve the average Pearson correlation coefficient (PCC) at least 10.45% and at most 45.10%.

Keywords: carbon emission forecasting; time series; data decomposition; LSSVM; PSO; decomposition-integration

0 引言

能源的快速消耗所带来的全球温室效应已被认为是 21 世纪最主要的全球性环境问题之一^[1]. 研究表明, 全球气温升高带来的环境退化、人类健康威胁、自然灾害等副作用越来越难以忽视^[2]. 世界各国相继开启了节能减排、绿色发展的道路, 而碳排放量的变

化则是衡量各国减排成果的重要指标之一.

由于环境问题的严峻性和控制碳排的重要性, 碳排的相关研究不断发展, 其中碳排的预测问题受到了广泛关注. 已有的碳排预测文献可以从 3 个方面进行总结: 碳排的单维预测、碳排的多维预测以及不同行业的碳排预测. 碳排的单维预测主要集中于研究方

收稿日期: 2020-12-21; 录用日期: 2021-04-21.

基金项目: 国家自然科学基金面上项目 (71971148, 71671118); 中央高校基本科研业务费专项资金项目 (SXYPY202103).

[†]通讯作者. E-mail: zhibinwu@scu.edu.cn.

法的创新,通过结合碳排数据自身特点,得到更精确的预测结果. Zhao等^[3]利用简化计量经济学的方法扩展了碳排的样本时期,其结果表明到2050年碳排放量将减少25%. Fang等^[4]利用改进的粒子群算法对碳排进行高斯过程回归预测,有效地优化了高斯过程回归中协方差函数的超参数,并将模型运用到美国、中国和日本3个案例上.

考虑到经济增长、能源消耗、人口变化等因素对碳排的影响,碳排的多维预测同样值得关注. Wang等^[5]研究了碳排放与经济增长之间的非线性和不确定性关系,将相关变量作为外生变量引入到多变量灰色模型中,以此实现化石能源消耗产生的碳排预测. Wang等^[6]利用情景模拟方法和马尔可夫链模型对中国2020年的碳强度进行了预测,计算了18种可能情况下能源结构调整对中国实现碳排目标的影响. 赵亚涛等^[7]考虑国内生产总值和人口总数对煤电产业发电的影响,结合多元线性回归模型和情景分析对煤电行业的碳排峰值进行了预测. Li等^[8]选取总人口、人均国内生产总值、城镇化水平、一次能源消费结构、技术水平和产业结构5个因素作为碳排的影响因素,计算了1996年~2016年中国能源消费碳排放量.

由于碳排涉及的范围广泛,已有研究针对某一行业的碳排预测也进行了深入探讨. Ge等^[9]分析了天津市2005年~2012年工业能源消耗的碳排情况,结果表明生产要素规模对天津工业碳排放的增长起主要作用. Wen等^[10]以城市为单位研究了住宅行业能源消耗造成的碳排,利用灰度关系进行特征分类,再用主成分分析和改进的鸡群优化算法优化的SVM进行碳排预测. Ofosu-Adarkwa等^[11]利用灰色预测模型分析了中国水泥行业的碳排. 徐勇戈等^[12]利用模糊布谷鸟搜索算法优化的支持向量机模型对建筑行业的碳排放量进行了预测.

已有文献虽然对碳排预测的各方面进行了深入探讨,但是仍然存在一些不足. 首先大部分碳排预测的研究集中在方法的创新上,对碳排数据进行预处理的研究较少. 无论是时间序列模型还是机器学习方法,在预测碳排放量时都要面临碳排数据的非线性和波动特质^[3-4],单纯依靠预测方法的改进难以完全解决这个问题. 其次通过对已有文献进行总结不难发现,现有的碳排放量核算数据多是以年为单位,存在数据体量小的问题^[12],模型的普适性难以在单一的实例数据中得到验证.

为了解决以上两个问题,本文提出一种分解集

成预测方法,结合数据处理方法——经验模态分解(empirical mode decomposition, EMD)和以智能算法优化的机器学习模型进行碳排放量的预测,缓解碳排数据的非线性和波动性对预测结果的影响. 另一方面,将分解集成预测模型运用到多个国家的碳排数据实例中以验证其普适性和有效性. 本文主要贡献是:1)在碳排预测问题中引入数据分解方法,通过EMD方法降低碳排序列的波动性和非线性,以此实现更精确地碳排预测;2)实验涵盖各大洲12个国家的碳排数据,提出的分解集成预测方法能够较为准确地拟合各国不同情况下的碳排趋势.

1 模型与方法

1.1 经验模态分解EMD

碳排放量受到多方面因素的影响,例如能源消耗、经济发展、人口规模等,表现出了非线性、波动性的特点. 非线性和波动性会影响碳排放量预测的准确性,因此需要对碳排数据进行初步处理. 本文采用Huang等^[13]提出的一种后验数据分解方法——经验模式分解(EMD),该方法的有效性在多个领域都得到了验证,例如碳交易价格^[14]. Liu等^[15]利用EMD方法将各国碳排放的时间序列数据分解为高频模态和趋势模态,分别表示为各国碳排放的短期干扰和长期趋势. EMD的核心思想是将不稳定的数据序列通过经验识别分解为更为稳定的内涵模态分量(intrinsic oscillatory mode, IMF)和一个残余项. 具体的数据分解过程如下.

step 1: 初始提取. 采用三次样条插值函数拟合上下包络线,包络线与局部最大最小值连接并覆盖其间所有数据. 设包络线的首个平均值为 h_{11} ,原始数据信号为 M , m_{11} 分量为原始数据和平均值的差,即

$$M - h_{11} = m_{11}. \quad (1)$$

step 2: 重复筛选. 重复筛选直到分量满足IMF的标准,即极值和过零点的数目必须相等或者最多相差1,其次包络平均值必须为0. $m_{i(k-1)}$ 为 $k-1$ 次筛选的分量, h_{ik} 为第 k 次筛选的均值, m_{ik} 为第 k 次筛选后满足IMF条件的第 i 个分量,将其定义为第 i 个IMF分量 l_i . 重复筛选过程如下:

$$m_{i(k-1)} - h_{ik} = m_{ik}, \quad (2)$$

$$l_i = m_{ik}. \quad (3)$$

step 3: 模块集成. 筛选的过程会重复到分量或者残差值小于预定值,或者残差变成单调函数. 当所有的IMF从原始数据信号中提取出来后,剩余的残余项 r 具有周期长、趋势明显的特点. 通过该筛选过程,原

始数据被分为 I 个经验模态 $l_i (i = 1, 2, \dots, I)$ 和一个残余项 r , Q 为最后的集成结果, 其中 I 的值由数据自身波动性质确定, 有

$$Q = \sum_{i=1}^I l_i + r. \quad (4)$$

虽然 EMD 分解能够较为准确地提取出数据的有效信息, 但它也存在不足, 例如端点效应. 真实数据由于收集时间有限, 可能存在数据的端点值并非极值的情况, 这种情况会使上下包络线在数据两端发散, 从而随着数据的分解过程污染数据, 影响分解结果的有效性. 国内外很多学者提出了缓解 EMD 端点效应的方法, 本文采用 Rilling 等^[16] 提出的极值镜像延拓方法. 该方法首先判断端点值与极值的关系, 然后镜像延拓靠近两端的极值, 通过增加两端极值点的方式生成上下包络线, 再重复 EMD 的分解步骤, 以此缓解端点效应问题.

1.2 最小二乘支持向量机 LSSVM

支持向量机 (support vector machines, SVM) 作为一项经典的机器学习算法, 其核心思想是在特征空间中学习超平面, 使得间隔最大^[17]. 在 SVM 的基础上, Suykens 等^[18] 发展了最小二乘支持向量机 (least squares support vector machines, LSSVM), 以改进 SVM 训练时间长特点. 假设有训练集 $U = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 利用 LSSVM 学习回归模型 $f(x) = W \cdot X + b$, 使得模型输出结果尽可能接近真实值 y . W 和 b 分别为模型的法向量和截距, X 为特征向量. 回归问题可以表示为

$$\begin{aligned} \min_{W, b} (W, b) &= \frac{1}{2} \|W\|^2 + \frac{\gamma}{2} \sum_{n=1}^N \tau_n^2; \\ \text{s.t. } y_n (W \cdot \varphi(x_n) + b) &= 1 - \tau_n, \quad n = 1, 2, \dots, N. \end{aligned} \quad (5)$$

其中: γ 为误差项 τ_n 的正则化参数, $\varphi(x_n)$ 为原始 x_n 从低维空间到高维空间的映射. 为了计算该模型, 引入拉格朗日乘子 α_n , 有

$$\begin{aligned} \psi(W, b, \tau, \alpha) &= \\ \frac{1}{2} \|W\|^2 + \frac{\gamma}{2} \sum_{n=1}^N \tau_n^2 - \sum_{n=1}^N \alpha_n \{ & [y_n (W \cdot \varphi(x_n) + \\ b)] - 1 + \tau_n \}. \end{aligned} \quad (6)$$

通过对 W 、 b 、 τ_n 以及 α_n 分别进行偏导可以求解式 (6), 最后的输出结果为

$$f(x) = W \cdot \varphi(x) + b = \sum_{n=1}^N \alpha_n K(x_n, x) + b, \quad (7)$$

其中 $K(x_n, x)$ 为核函数, 利用核函数的技巧可以将

原本只能应用在线性问题上的 LSSVM 扩展到更高维的特征空间. 常见的核函数有线性核、多项式核、S 型核和径向基函数核 (radial basis function, RBF). RBF 核相较于其他核函数而言复杂度更小, 因此本文采用 RBF 核将原问题映射到高维空间, 即

$$K(x_n, x) = \varphi(x_n) \cdot \varphi(x) = \exp\left(\frac{-\|x_n - x\|^2}{2\sigma^2}\right). \quad (8)$$

其中: $\varphi(x_n)$ 为映射函数, σ 为核参数.

1.3 粒子群算法优化的 LSSVM

LSSVM 解决回归问题的过程中不可避免的问题就是正则参数 γ 和核参数 σ 的确定. 所幸粒子群优化算法 (particle swarm optimization, PSO) 作为一种智能算法, 在参数优化方面已经得到了有效验证^[19-20]. 具体步骤如下.

step 1: 数据处理. 从原始数据序列中划分出训练集和测试集, 使用训练集对模型进行训练, 然后使用测试集对模型进行评估, 训练集和测试集的详细生成过程见 1.4 节.

step 2: 初始化. 随机产生正则参数 γ 和核参数 σ 的候选解并定义为初始种群的粒子, 赋予其初始位置和随机速度.

step 3: 迭代. 粒子会移动到更优的位置以获得更优的目标函数. 迭代过程如下所示:

$$\begin{aligned} v_g^{s+1} &= v_g^s + d_1 \times \text{rand}_1 \times (p_{\text{local}} - e_g^s) + \\ & d_2 \times \text{rand}_2 \times (p_{\text{global}} - e_g^s), \\ e_g^{s+1} &= e_g^s + v_g^{s+1}. \end{aligned} \quad (9)$$

其中: v_g^{s+1} 为第 g 个粒子在 $s+1$ 次迭代的速度, d_1 和 d_2 为决定搜索能力的学习因子, rand_1 和 rand_2 是 $[0, 1]$ 之间的随机数, 粒子通过追踪局部最优解 p_{local} 和全局最优解 p_{global} 确定后续位置, e_g^s 为第 g 个粒子在 s 次迭代的位置. 每个粒子单独进行寻优, 粒子本身寻优过程的最优解作为局部最优解 p_{local} , 所有粒子共享最优位置后的最优解为全局最优解 p_{global} . γ 和 σ 的寻优迭代过程相同且同时进行, 均遵守式 (9) 的计算过程.

step 4: 目标评估. 每次迭代均对 LSSVM 模型进行训练并计算出预测值, 本文选用均方误差 (mean square error, MSE) 作为每次迭代的评估标准, 如果下一次迭代的 MSE 结果比上一次的 MSE 值优越, 则更新最优解 p_{local} 和 p_{global} . MSE 为

$$\text{MSE} = \frac{1}{H} \sum_{h=1}^H (\hat{y}_h - y_h)^2. \quad (10)$$

其中: \hat{y}_h 为每次迭代中测试集的预测结果; y_h 为测

试集的真实值; H 为测试集的样本容量, $h = 1, 2, \dots, H$.

step 5: 停止迭代. LSSVM 的搜索过程会在达到最大迭代次数或者满足终止条件时停止, 如果没有达到这些要求则返回 step 3.

1.4 分解集成预测模型框架

通过建立分解集成预测模型提高碳排预测的精度. 首先收集共 F 个时间段的原始数据 $a_t (t = 1, 2, \dots, F)$, 其中 a_t 为碳排在 t 时间节点的取值. 利用 EMD 将原始数据进行分解, 对分解模块进行重组获得 PSO-LSSVM 模型的输入输出部分, 同时将重组数据分为测试集和训练集, 实现预测模型的训练和检验过程. 具体步骤如下.

step 1: 将原始数据集 $a_t (t = 1, 2, \dots, F)$ 作为 EMD 的输入, 输出不同频率的模块分量 $l_i (i = 1, 2, \dots, I)$ 和残差值序列 residual, l_i 为第 i 个 IMF 分量, 有

$$L = \begin{bmatrix} l_i \\ \text{residual} \end{bmatrix} = \begin{bmatrix} l_{i1} \dots l_{iF} \\ r_1 \dots r_F \end{bmatrix}, \quad i = 1, 2, \dots, I. \quad (11)$$

step 2: 数据分解后, 利用滑动窗口对 $l_i (i = 1, 2, \dots, I)$ 及残差项进行重新排序, 以 l_i 为例, 相应的汉克尔矩阵 (Hankel Matrix) 为

$$\text{windowize}(l_i^T) = \begin{bmatrix} L_{i1} & L_{i2} \\ L_{i3} & L_{i4} \end{bmatrix} = \begin{bmatrix} l_{i1} & \dots & l_{i(\text{delays})} & l_{i(\text{delays}+1)} \\ \vdots & \ddots & \vdots & \vdots \\ l_{i(F-H-\text{delays})} & \dots & l_{i(F-H-1)} & l_{i(F-H)} \\ l_{i(F-H-\text{delays}+1)} & \dots & l_{i(F-H)} & l_{i(F-H+1)} \\ \vdots & \ddots & \vdots & \vdots \\ l_{i(F-\text{delays})} & \dots & l_{i(F-1)} & l_{iF} \end{bmatrix}. \quad (12)$$

l_i 经过转置后重组生成 $(F - \text{delays}) \times (\text{delays} + 1)$ 的汉克尔矩阵, 其中 H 为测试集样本容量. PSO-LSSVM 模型输入输出的确定依据了时间序列自相关性的特点, 以 delays 为延迟期数, 预测 $\text{delays} + 1$ 时间节点的取值, delays 的取值通过重复实验, 选取 PSO-LSSVM 模型预测结果中最优的 MSE 值进行确定.

step 3: 将式 (12) 中的汉克尔矩阵分解为 $(F - H - \text{delays}) \times \text{delays}$ 的训练集输入部分 L_{i1} 和 $(F - H - \text{delays}) \times 1$ 的训练集输出部分 L_{i2} , 以及 $H \times \text{delays}$ 的测试集输入部分 L_{i3} 和 $H \times 1$ 的测试集输出部分 L_{i4} . 利用训练集输入部分 L_{i1} 和训练集输出部分 L_{i2} 对 PSO-LSSVM 模型进行训练, 测试集相应的

输入部分 L_{i3} 和输出部分 L_{i4} 用来检验模型的预测表现. 最终得到每个分解模块的预测结果合集 \hat{L} , 有

$$\hat{L} = \begin{bmatrix} \hat{l}_{i(F-H+1)} \dots \hat{l}_{iF} \\ \hat{r}_{F-H+1} \dots \hat{r}_F \end{bmatrix}, \quad i = 1, 2, \dots, I. \quad (13)$$

step 4: 最后集成各个模块的预测结果, 有

$$\hat{Q} = (\hat{y}_{(F-H+1)} \dots \hat{y}_F), \quad \hat{y}_s = \sum_{i=1}^I \hat{l}_{is} + \hat{r}_s, \quad s = (F - H + 1), \dots, F. \quad (14)$$

其中: \hat{Q} 为最终预测结果序列; I 表示数据经 EMD 分解后共生成 I 个 IMF 和一个残差项; \hat{l}_{is} 为第 i 个模块分量在第 s 个时间节点的预测值; $s = (F - H + 1), \dots, F$ 表示从 $F - H + 1$ 个时间节点至第 F 个时间节点; \hat{r}_s 为残差项在第 s 个时间节点的预测值.

分解集成预测模型的框架如图 1 所示.

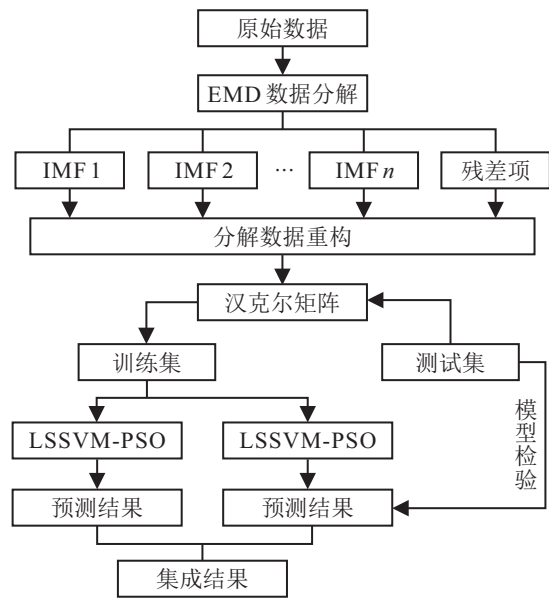


图 1 分解集成预测模型框架

1.5 评估标准

为了评估模型的预测效果, 选用平均绝对百分误差 (mean absolute percentage error, MAPE)、平均绝对误差 (mean absolute error, MAE) 评估模型的预测效果, 并比较分解集成预测模型和其他模型的预测表现, 利用 Pearson 相关系数 (PCC) 评估不同模型的稳定性. \hat{y}_s 为第 s 个时间节点的测试样本预测值, y_s 为第 s 个时间节点的测试样本真实值, $s = (F - H + 1), \dots, F$, 有

$$\text{MAPE} = \frac{100\%}{H} \sum_{s=F-H+1}^F \left| \frac{\hat{y}_s - y_s}{y_s} \right|, \quad (15)$$

$$\text{MAE} = \frac{1}{H} \sum_{s=F-H+1}^F |\hat{y}_s - y_s|, \quad (16)$$

$$\text{PCC} =$$

$$\frac{\sum_{s=F-H+1}^F (y_s - \bar{y}_s)(\hat{y}_s - \bar{\hat{y}}_s)}{\sqrt{\sum_{s=F-H+1}^F (y_s - \bar{y}_s)^2} \sqrt{\sum_{s=F-H+1}^F (\hat{y}_s - \bar{\hat{y}}_s)^2}} \quad (17)$$

2 实证分析

2.1 数据描述

本文选取美国、中国、日本、印度等12个国家的碳排数据作为实证案例,数据来源于世界能源统计年鉴 (<https://www.bp.com>). 为了得到更准确的对比结果,以验证本文的分解集成预测模型在提高碳排

预测精确度上的有效性,选取碳排数据的时间范围与Qiao等^[21]的研究保持一致,收集1965年~2017年各国的碳排放数据,其中1965年~2012年为训练集数据,2013年~2017年为测试集数据. 实验使用64位Windows处理系统,利用Matlab软件进行数据分解和模型预测的计算.

2.2 碳排预测结果

通过EMD数据分解技术,12组碳排序列被分解为从高到低频率的不同模块,图2展示了各个国家的原始碳排放序列以及分解后的序列.

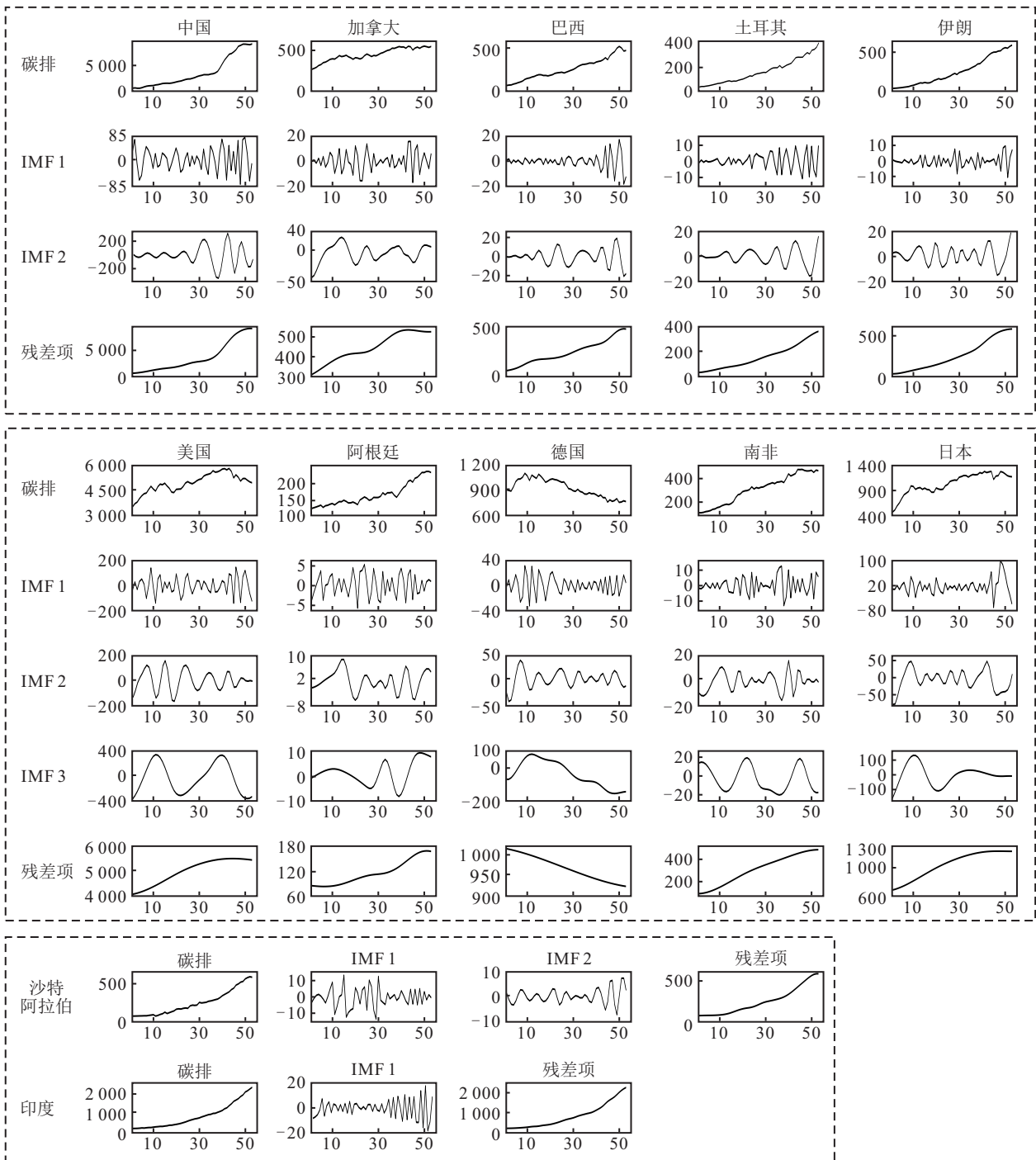


图2 EMD分解结果

由图2可见,各国原始的碳排具有较为明显的发展趋势,例如中国碳排的上升趋势,德国碳排的下降趋势.但原始碳排数据仍然存在波动性和不稳定性,EMD分解后的IMFs表达了碳排数据短期受到的干扰.残差项的趋势与原始数据相符,但相较于原始数据而言更加平滑稳定.对每个分解模块分别利用PSO-LSSVM模型进行预测,以中国的碳排为例,EMD分解后获得两个IMF模块以及残差项,对分解模块进行重组后获得汉克尔矩阵,由式(12)确定PSO-LSSVM模型的输入输出部分,得到分解模块的预测结果.最优的delays值经过多次计算选定(例如:中国碳排数据IMF1的最优delays值为15).各国家分解模块的训练集和测试集预测结果分别如表1和表2所示.

表1 训练集各分解模块预测结果评估值

| 国家 | | IMF 1 | IMF 2 | IMF 3 | 残差项 |
|-------|------|---------|---------|---------|--------|
| 中国 | MAPE | 73.441 | 10.729 | - | 0.135 |
| | MAE | 12.931 | 4.554 | - | 3.079 |
| 美国 | MAPE | 1.872 | 59.378 | 40.56 | 0.243 |
| | MAE | 0.708 | 7.504 | 9.395 | 11.248 |
| 印度 | MAPE | 111.814 | - | - | 3.527 |
| | MAE | 2.435 | - | - | 19.618 |
| 加拿大 | MAPE | 89.346 | - | - | 46.199 |
| | MAE | 5.762 | - | - | 1.729 |
| 巴西 | MAPE | 79.561 | 291.153 | - | 9.9 |
| | MAE | 2.749 | 1.788 | - | 15.917 |
| 阿根廷 | MAPE | 97.432 | 5.456 | 13.677 | 0.485 |
| | MAE | 2.27 | 0.047 | 0.258 | 0.523 |
| 德国 | MAPE | 85.114 | 32.361 | 8.017 | 0.011 |
| | MAE | 4.864 | 1.494 | 1.635 | 0.102 |
| 土耳其 | MAPE | 68.365 | 16.832 | - | 1.524 |
| | MAE | 1.358 | 0.481 | - | 1.793 |
| 伊朗 | MAPE | 5.594 | 44.279 | - | 3.047 |
| | MAE | 0.071 | 1.949 | - | 5.043 |
| 沙特阿拉伯 | MAPE | 150.45 | 81.74 | - | 1.577 |
| | MAE | 1.814 | 0.63 | - | 2.501 |
| 南非 | MAPE | 97.985 | 50.434 | 7.075 | 0.075 |
| | MAE | 1.472 | 2.928 | 0.616 | 0.221 |
| 日本 | MAPE | 56.846 | 190.285 | 108.232 | 0.026 |
| | MAE | 7.706 | 5.284 | 13.779 | 0.24 |

表2 测试集各分解模块预测结果评估值

| 国家 | | IMF 1 | IMF 2 | IMF 3 | 残差项 |
|-------|------|---------|---------|--------|--------|
| 中国 | MAPE | 105.72 | 7.566 | - | 0.024 |
| | MAE | 40.709 | 7.157 | - | 2.264 |
| 美国 | MAPE | 102.849 | 75.397 | 1.073 | 0.07 |
| | MAE | 70.398 | 2.635 | 1.645 | 1.888 |
| 印度 | MAPE | 24.668 | - | - | 2.929 |
| | MAE | 3.242 | - | - | 30.419 |
| 加拿大 | MAPE | 237.093 | 4.492 | - | 0.033 |
| | MAE | 1.056 | 0.342 | - | 0.138 |
| 巴西 | MAPE | 111.856 | 14.008 | - | 0.127 |
| | MAE | 6.445 | 0.692 | - | 0.444 |
| 阿根廷 | MAPE | 80.791 | 16.247 | 0.24 | 0.048 |
| | MAE | 0.629 | 0.395 | 0.012 | 0.067 |
| 德国 | MAPE | 155.035 | 535.066 | 0.099 | 0.014 |
| | MAE | 5.243 | 0.482 | 0.08 | 0.095 |
| 土耳其 | MAPE | 55.348 | 11.207 | - | 0.065 |
| | MAE | 20145 | 0.587 | - | 0.149 |
| 伊朗 | MAPE | 88.596 | 73.983 | - | 0.014 |
| | MAE | 3.611 | 2.211 | - | 0.056 |
| 沙特阿拉伯 | MAPE | 24.447 | 797.288 | - | 0.032 |
| | MAE | 0.386 | 0.798 | - | 0.118 |
| 南非 | MAPE | 74.687 | 55.76 | 1.994 | 0.047 |
| | MAE | 3.534 | 0.248 | 0.125 | 0.206 |
| 日本 | MAPE | 91.292 | 35.192 | 41.882 | 0.015 |
| | MAE | 24.412 | 6.309 | 0.181 | 0.191 |

对测试集的预测结果进行分析可以得到,除了印度碳排数据的分解项,PSO-LSSVM模型在低频和残差项的预测结果均比高频率序列的预测结果更精准.IMFs决定了数据的波动,是数据不稳定的原因.残差项决定了数据的整体趋势,是数据的主要组成部分.因此对IMFs的精准预测有利于提高预测精度,而对残差项的精准预测是整体预测结果准确的保证.印度的碳排数据经过EMD分解后只得到一项IMF和残差项,同时印度作为排放大国,其碳排基数相较于一般国家而言较大,故其残差项的预测结果相较于其他国家的预测值而言,拥有较大的MAPE和MAE,其他国家的预测结果都较为理想.

最终的预测结果为每个模块预测值的集成,集成结果与真实结果对比如图3所示.由图3可见,除日本外,其他国家的碳排发展趋势都能较准确地被描

述. 分析日本的碳排数据可以发现,其2008年~2009年的碳排迅速下滑,从1299.7(million tonnes)跌至1130.0(million tonnes),主要原因之一是2008年的金融危机对日本经济的重大影响. 直至2012年,日本的碳排才逐渐上升至2008年的水准,即1296.1(million tonnes),而2013年以后的碳排下降趋势得益于节能和可再生能源的使用. 本文并未考虑外部因素的影响,仅考虑碳排数据的自相关性,因此对日本2013年~2017年的碳排预测,即测试集的数据选取范围,

延续了2009年~2012年的上升趋势,使得日本的碳排预测结果较为不准确,出现PCC为负值的情况. 随着未来碳排数据的更新,这种情况会得到有效缓解,其他国家的预测结果可以验证所提出的分解集成预测模型的有效性.

为了进一步验证分解集成模型 EMD-PSO-LSSVM 对未来碳排放量预测过程的有效性,对2018年~2025年的碳排放量进行样本外预测,结果如表3所示.

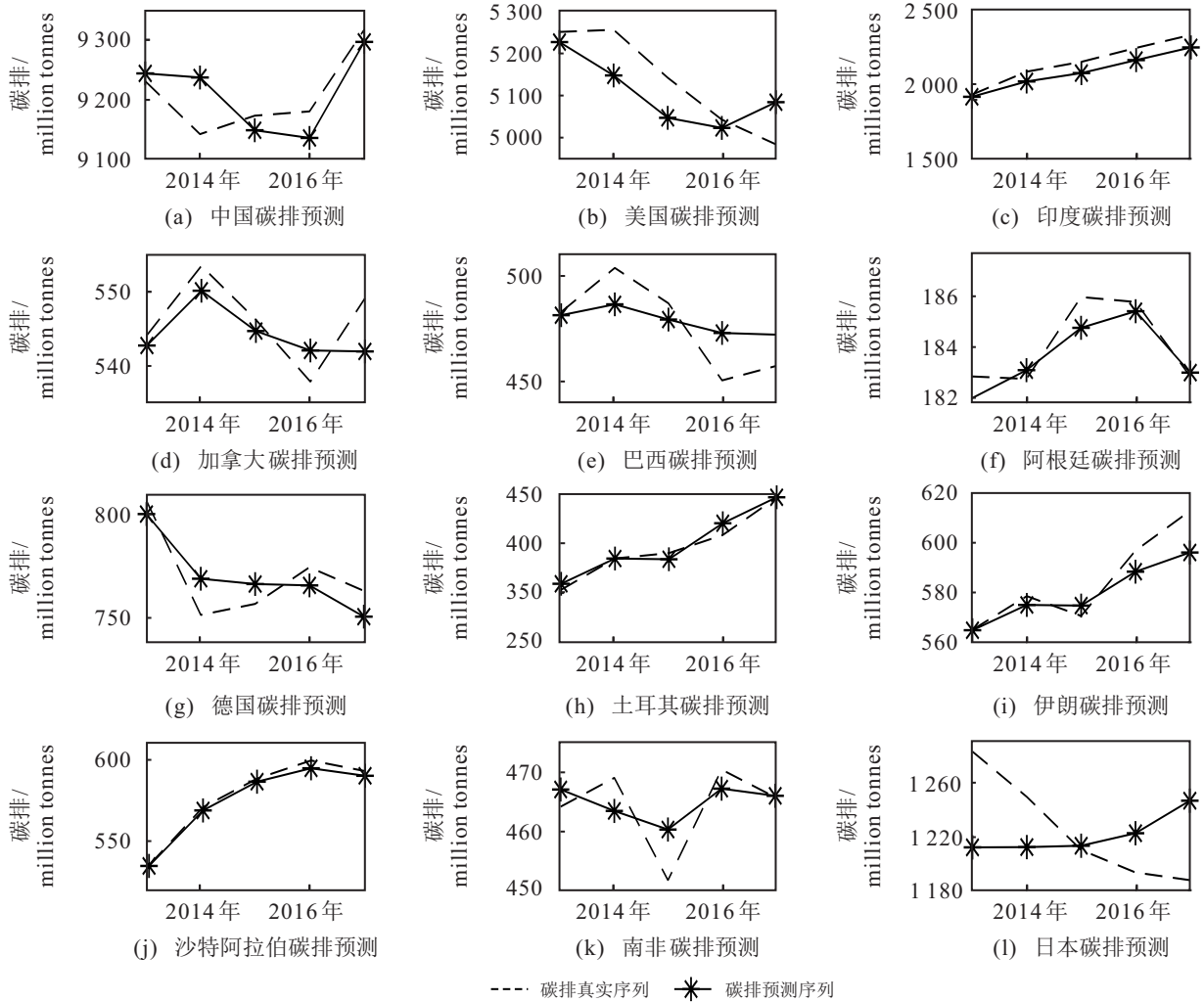


图3 碳排预测结果和真实数据的比较

表3 各国家2018年~2025年碳排预测结果

| | 中国 | 美国 | 印度 | 加拿大 | 巴西 | 阿根廷 | 德国 | 土耳其 | 伊朗 | 沙特阿拉伯 | 南非 | 日本 |
|-------|---------|---------|---------|--------|--------|--------|--------|--------|--------|--------|--------|---------|
| 2018年 | 9322.23 | 5367.58 | 2287.69 | 533.13 | 337.59 | 157.73 | 742.48 | 388.83 | 572.8 | 558.18 | 473.91 | 1252.21 |
| 2019年 | 9347.47 | 5441.42 | 2284.94 | 536.14 | 334.44 | 155.93 | 732.49 | 387.38 | 579.83 | 558.83 | 478.80 | 1228.10 |
| 2020年 | 9011.23 | 5608.26 | 2288.76 | 541.00 | 337.41 | 154.64 | 733.96 | 390.63 | 585.24 | 561.84 | 482.52 | 1221.92 |
| 2021年 | 8715.19 | 5661.83 | 2284.96 | 545.85 | 342.22 | 154.42 | 731.50 | 392.28 | 594.52 | 568.03 | 482.30 | 1239.68 |
| 2022年 | 8951.29 | 5624.30 | 2287.52 | 550.16 | 344.72 | 151.79 | 741.35 | 389.42 | 603.24 | 573.29 | 484.45 | 1256.35 |
| 2023年 | 9453.90 | 5683.17 | 2286.56 | 553.72 | 343.34 | 152.59 | 760.95 | 382.50 | 583.21 | 572.32 | 480.03 | 1253.97 |
| 2024年 | 9769.36 | 5876.35 | 2285.82 | 556.14 | 338.83 | 157.48 | 754.17 | 380.59 | 579.62 | 566.08 | 475.96 | 1255.21 |
| 2025年 | 9740.82 | 5795.29 | 2287.44 | 557.43 | 332.80 | 165.75 | 735.27 | 385.54 | 578.15 | 562.62 | 470.28 | 1251.71 |

观察表3各国家样本外预测结果可以看出,中国的未来碳排放量预测值存在先下降后上升的趋势,这是因为中国的碳排放量在2014年~2017年也存在一个先下降后上升的趋势.在模块分解过程中,中国碳排序列IMF2模块的预测值延续了这种走势,使得最终的碳排未来预测值趋势与现实的上升趋势存在差别,该结果验证了碳排序列的不稳定特征对预测可能会造成影响.但中国碳排放量分解模块残差项的未来预测值呈上升趋势([9 120.17, 9 148.18, 9 176.12, 9 198.81, 9 213.28, 9 218.57, 9 215.61, 9 206.81]),与实际情况更加相符.残差项作为碳排序列的主要组成部分,对其进行精准地预测能够保证获得较为准确的总碳排预测值.另一方面,本文主要从碳排序列本身进行研究,并未考虑碳排的外部因素对其造成的影响,这也是预测结果出现偏差的原因.最后,针对2018年~2025年的碳排预测为样本外预测,在总数据量

较小的情况下可能会因为训练集的选取范围不同而造成结果的偏差,随着数据量的逐渐增大,这种情况将会得到改善.除此之外,美国、加拿大未来8年碳排放量呈较为明显的上升趋势,其他国家未来8年的碳排放量趋于平缓.这里只是举例显示预测的结果,理论上样本外预测到本例涵盖的年限这么远需要十分谨慎.

3 多种预测方法的对比分析

分析表4和表5给出的测试集MAE和PCC预测结果对比以及测试集预测结果的MAPE值可以看出,大部分集成预测结果能够保证MAPE和MAE的值在较低的水平,且PCC值较高,表明模型稳定性较好.同时,将分解集成预测模型与现有的预测模型进行对比.MAE代表预测的精度,数值越低精度越好,PCC作为不同模型稳定性的评估标准,其绝对值越接近1表明预测值与真实值的相关性越强.

表4 不同预测模型的测试集预测对比结果

| 国家 | | EMD-PSO-LSSVM | LSO-GA-LSSVM | BP-NN | RBF-NN | ELMAN-NN | LS-SVM | SMO-SVM |
|-------|-----|---------------|--------------|----------|---------|----------|--------|---------|
| 中国 | MAE | 37.41 | 214.57 | 1 778.77 | 1 002.6 | 442.26 | 704.04 | 635.37 |
| | PCC | 0.67 | 0.35 | 0.31 | 0.35 | 0.34 | 0.34 | 0.24 |
| 美国 | MAE | 69.19 | 92.66 | 145.34 | 133.47 | 97.07 | 98.41 | 98.35 |
| | PCC | 0.74 | 0.54 | 0.50 | 0.45 | 0.44 | 0.44 | 0.44 |
| 印度 | MAE | 64.21 | 49.28 | 199.74 | 427.32 | 82.84 | 486.11 | 442.75 |
| | PCC | 0.99 | 0.98 | 0.78 | 0.90 | 0.98 | 0.94 | 0.94 |
| 加拿大 | MAE | 3.55 | 8.65 | 10.66 | 10.12 | 11.94 | 14.94 | 13.33 |
| | PCC | 0.72 | 0.73 | 0.72 | 0.52 | 0.53 | 0.64 | 0.59 |
| 巴西 | MAE | 12.80 | 22.24 | 119.82 | 81.46 | 29.32 | 81.67 | 73.81 |
| | PCC | 0.97 | 0.25 | 0.07 | 0.23 | 0.20 | 0.24 | 0.24 |
| 阿根廷 | MAE | 0.58 | 2.63 | 4.75 | 16.60 | 29.99 | 29.20 | 20.00 |
| | PCC | 0.93 | 0.63 | 0.61 | 0.45 | 0.60 | 0.58 | 0.54 |
| 德国 | MAE | 9.22 | 17.01 | 36.19 | 20.94 | 28.30 | 21.04 | 27.34 |
| | PCC | 0.81 | 0.50 | 0.09 | 0.36 | 0.09 | 0.41 | 0.26 |
| 土耳其 | MAE | 4.92 | 21.84 | 77.59 | 59.46 | 48.68 | 73.38 | 49.43 |
| | PCC | 0.98 | 0.97 | 0.25 | 0.59 | 0.82 | 0.95 | 0.22 |
| 伊朗 | MAE | 6.66 | 17.69 | 46.52 | 87.58 | 21.61 | 80.51 | 89.28 |
| | PCC | 0.98 | 0.69 | 0.69 | 0.61 | 0.67 | 0.12 | 0.56 |
| 沙特阿拉伯 | MAE | 2.4 | 9.41 | 88.65 | 99.20 | 36.32 | 100.25 | 81.07 |
| | PCC | 0.99 | 0.99 | 0.98 | 0.85 | 0.90 | 0.84 | 0.82 |
| 南非 | MAE | 4.14 | 7.81 | 17.45 | 7.94 | 8.14 | 8.13 | 8.26 |
| | PCC | 0.76 | 0.50 | 0.47 | 0.43 | 0.40 | 0.40 | 0.42 |
| 日本 | MAE | 39.92 | 12.54 | 46.87 | 39.03 | 27.96 | 13.73 | 20.43 |
| | PCC | -0.68 | 0.94 | 0.85 | 0.90 | 0.69 | 0.93 | 0.88 |
| 平均 | MAE | 21.25 | 39.69 | 214.36 | 165.48 | 72.04 | 142.62 | 129.95 |
| | PCC | 0.74 | 0.67 | 0.53 | 0.55 | 0.56 | 0.57 | 0.51 |

注:本文主要引用Qiao等^[21]的狮子群优化和遗传算法优化的LSSVM模型(LSO-GA-LSSVM)、BP神经网络(BPNN)、RBF神经网络(RBFNN)、ELMAN神经网络(ELMANFNN)、LSSVM模型序列最小优化的SVM(SMOSVM)的预测结果,并将其与分解集成预测模型进行比较.经部分读者指出,Qiao等^[21]的MAPE值可能存在问题,故本文仅针对MAE以及PCC值进行比较,EMD-PSO-LSSVM模型预测的MAPE值将单独给出.

表5 EMD-PSO-LSSVM模型测试集预测的MAPE值

| 国家 | EMD-PSO-LSSVM |
|-------|---------------|
| 中国 | 0.41 |
| 美国 | 1.35 |
| 印度 | 2.93 |
| 加拿大 | 0.66 |
| 巴西 | 2.72 |
| 阿根廷 | 0.31 |
| 德国 | 1.21 |
| 土耳其 | 1.45 |
| 伊朗 | 1.11 |
| 沙特阿拉伯 | 0.41 |
| 南非 | 0.90 |
| 日本 | 0.32 |
| 平均 | 1.15 |

由预测的对比结果可以看出,所提出的分解集成预测模型能够有效提高碳排放量预测的精度.大部分国家的实例预测结果均得到大幅度提升,验证了解析集成预测模型的有效性.在文献[21]预测精度最高的LSO-GA-LSSVM模型基础上,分解集成预测模型(即EMD-PSO-LSSVM模型)能够将平均MAE值提高46.46% $(|21.25 - 39.69|/39.69 = 46.46\%)$.与BPNN相比,EMD-PSO-LSSVM模型能够将平均MAE值提高精度90.09% $(|21.25 - 214.36|/214.36 = 90.09\%)$ 左右.从模型的稳定性标准平均PCC值看,EMD-PSO-LSSVM相较于LSO-GA-LSSVM而言提高了10.45% $(|0.74 - 0.67|/0.67 = 10.45\%)$,表明EMD-PSO-LSSVM能够在保证一定精度的情况下拥有更好的稳定性.因此,更优的MAE值和更稳定的性质表明了EMD-PSO-LSSVM模型的优越性.

预测及对比结果表明了解析集成预测模型在碳排放预测问题上的有效性.EMD分解方法能够有效降低数据的波动性和非线性,使得碳排放数据更加平稳,从而得到更精确的预测结果.另一方面,PSO优化后的LSSVM模型能够较为精确地模拟出碳排的发展趋势.各国的碳排走势能够反映其节能减排措施的有效性.利用本文的分解集成预测模型对碳排放量进行预测,可以为当前的减排措施提供计划方向或示警,同时为各国拟定低碳政策、颁布减排措施提供决策参考.

4 结论

碳排是评判环境质量的重要指标之一,碳排预测可以为未来的减排计划提供参考.为了缓解碳排放数据的波动性和非线性,本文结合数据分解方法EMD和PSO优化的LSSVM,构建碳排的分解集成预测模

型.首先,利用EMD将碳排数据分解为不同频率的信号模块以及反应趋势的残差项;然后,利用PSO-LSSVM模型预测每个模块;最后,将多个模块的预测值进行集成得到最后的预测结果.在实验过程中,发现利用PSO算法对LSSVM进行参数寻优时,参数的范围以及滞后值的选取较为重要,不恰当的参数范围或滞后值可能会使寻优过程陷入局部最优,或者直接影响预测精度.结合上述结果分析,本文的研究结论如下:

1) EMD分解技术能够将碳排放数据分解为更稳定的序列模块,缓解了碳排放数据的非线性和波动性问题,从而有利于提高碳排放预测的精度.对残差项的精准预测能够保证总体预测值的偏差较小.

2) 分解集成预测模型在12个国家碳排放数据上的应用结果表明了其在碳排放预测问题上的普适性.

3) 与其他模型相比,所提出的分解集成预测模型能够在保证模型稳定的情况下提高模型的预测精度,验证了其在碳排放预测问题上的有效性和优越性.

虽然本文的分解集成预测模型能够稳定提升碳排放的预测准确性,但其依然存在一些不足,例如未能选取最新的碳排放数据,不过由于碳排放数据的年更性,本文选取的案例仍能够为未来的碳排放预测提供参考.其次是本文未考虑外界因素对碳排放的影响,今后的研究将会进一步考虑碳排放的各类影响因子,例如能源使用情况、经济发展情况、城镇化水平、大流行病等影响下的碳排放预测研究.另外,针对高频率序列的高度波动性和非线性也会进行深层次的探讨.

参考文献(References)

- [1] Zhou P, Wang M. Carbon dioxide emissions allocation: A review[J]. *Ecological Economics*, 2016, 125: 47-59.
- [2] Walther G R, Post E, Convey P, et al. Ecological responses to recent climate change[J]. *Nature*, 2002, 416(6879): 389-395.
- [3] Zhao X B, Du D. Forecasting carbon dioxide emissions[J]. *Journal of Environmental Management*, 2015, 160: 39-44.
- [4] Fang D B, Zhang X L, Yu Q, et al. A novel method for carbon dioxide emission forecasting based on improved Gaussian processes regression[J]. *Journal of Cleaner Production*, 2018, 173: 143-150.
- [5] Wang Z X, Ye D J. Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models[J]. *Journal of Cleaner Production*, 2017, 142: 600-612.
- [6] Wang S J, Li C F, Yang L Z. Decoupling effect and

- forecasting of economic growth and energy structure under the peak constraint of carbon emissions in China[J]. *Environmental Science and Pollution Research*, 2018, 25(25): 25255-25268.
- [7] 赵亚涛, 南新元, 贾爱迪. 基于情景分析法的煤电行业碳排放峰值预测[J]. *环境工程*, 2018, 36(12): 177-181. (Zhao Y T, Nan X Y, Jia A D. Prediction of carbon emission peak in coal-fired power industry based on scenario analysis[J]. *Environmental Engineering*, 2018, 36(12): 177-181.)
- [8] Li Z, Li Y B, Shao S S. Analysis of influencing factors and trend forecast of carbon emission from energy consumption in China based on expanded STIRPAT model[J]. *Energies*, 2019, 12(16): 3054.
- [9] Ge X L, Wang Y, Zhu H Q, et al. Analysis and forecast of the Tianjin industrial carbon dioxide emissions resulted from energy consumption[J]. *International Journal of Sustainable Energy*, 2017, 36(7): 637-653.
- [10] Wen L, Cao Y. Influencing factors analysis and forecasting of residential energy-related CO₂ emissions utilizing optimized support vector machine[J]. *Journal of Cleaner Production*, 2020, 250: 119492.
- [11] Ofosu-Adarkwa J, Xie N M, Javed S A. Forecasting CO₂ emissions of China's cement industry using a hybrid Verhulst-GM(1, N) model and emissions' technical conversion[J]. *Renewable and Sustainable Energy Reviews*, 2020, 130: 109945.
- [12] 徐勇戈, 宋伟雪. 基于FCS-SVM的建筑业碳排放预测研究[J]. *生态经济*, 2019, 35(11): 37-41. (Xu Y G, Song W X. Carbon emission prediction of construction industry based on FCS-SVM[J]. *Ecological Economy*, 2019, 35(11): 37-41.)
- [13] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. *Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, 1998, 454(1971): 903-995.
- [14] 刘金培, 郭艺, 陈华友, 等. 基于非结构数据流行学习的碳价格多尺度组合预测[J]. *控制与决策*, 2019, 34(2): 279-286. (Liu J P, Guo Y, Chen H Y, et al. Multi-scale combined forecast of carbon price based on manifold learning of unstructured data[J]. *Control and Decision*, 2019, 34(2): 279-286.)
- [15] Liu M X, Zhang G X. Characteristics of multi-national carbon emissions based on empirical mode decomposition[C]. *The 6th International Conference on Business Intelligence and Financial Engineering*. Hangzhou, 2013: 458-462.
- [16] Rilling G, Flandrin P, Goncalves P. On empirical mode decomposition and its algorithms[C]. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*. Grado, 2003: 8-11.
- [17] Vapnik V N. *The nature of statistical learning theory*[M]. New York: Springer, 1995: 138-145.
- [18] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. *Neural Processing Letters*, 1999, 9(3): 293-300.
- [19] Aziz N A A, Mohemmed A W, Alias M Y, et al. Particle swarm optimization for constrained and multiobjective problems: A brief review[J]. *Management and Artificial Intelligence*, 2011, 6: 146-150.
- [20] Kennedy J, Eberhart R. Particle swarm optimization[C]. *Proceedings of ICNN'95-International Conference on Neural Networks*. Perth, 1995: 1942-1948.
- [21] Qiao W B, Lu H F, Zhou G F, et al. A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer[J]. *Journal of Cleaner Production*, 2020, 244: 118612.

作者简介

张雯(1996—), 女, 博士生, 从事机器学习、计量经济、数据分析的研究, E-mail: zhangwentonia@163.com;

吴志彬(1982—), 男, 教授, 博士生导师, 从事决策理论与方法、机器学习及应用、运营管理与商业分析等研究, E-mail: zhibinwu@scu.edu.cn;

徐玖平(1962—), 男, 教授, 博士生导师, 从事管理科学、系统工程、数据科学等研究, E-mail: xujiuping@scu.edu.cn.

(责任编辑: 郑晓蕾)