

基于两层分解算法和改进 SVM 的油田采出水处理效果预测研究

徐磊^{1,2}, 侯磊^{1,2*}, 朱振宇^{1,2}, 徐震³, 雷婷^{1,2}, 李雨^{1,2}, 李强³, 陈秀芹³, 王九玲³, 陈星燃⁴

1 中国石油大学(北京)机械与储运工程学院, 北京 102249

2 中国石油大学(北京)石油工程教育部重点实验室, 北京 102249

3 中国石化胜利油田有限公司桩西采油厂, 东营 257237

4 北京交通大学(威海校区)土建学院, 威海 264200

* 通信作者, houleicup@126.com

收稿日期: 2020-09-14

摘要 准确的水质预测是评估油田联合站采出水处理效果的重要依据, 为水质预警提供科学依据。传统方法存在主观性强和耗时性长等缺点, 现有部分研究借助于机器学习方法, 但对数据噪声和数据非线性考虑不足。本研究提出一种基于两层分解算法与改进支持向量机相结合的预测方法。通过两层分解算法消除冗余噪声, 提取初始数据主要特征。利用分层抽样对原始数据集进行划分, 避免传统随机抽样引起的样本偏差。采用改进粒子群算法优选支持向量机参数, 提高全局收敛能力。针对桩西采油厂联合站 4 个案例, 依据相对误差、平均绝对百分比误差和决定系数 3 个评价指标对提出的预测方法展开准确性评价, 基于 4 个案例 3 个指标值的平均值分别为 -0.38%、5.23% 和 0.82%。相比于现有主流机器学习方法, 提出的预测方法具有较高的预测精度。

关键词 油田联合站; 水质预测; 机器学习; 两层分解算法; 改进支持向量机

Prediction of oilfield produced water treatment based on a two-layer decomposition technique and modified SVM

XU Lei^{1,2}, HOU Lei^{1,2}, ZHU Zhenyu^{1,2}, XU Zhen³, LEI Ting^{1,2}, LI Yu^{1,2}, LI Qiang³, CHEN Xiuqin³, WANG Jiuling³, CHEN Xingran⁴

1 College of Mechanical and Transportation Engineering, China University of Petroleum-Beijing, Beijing 102249, China

2 MOE Key Laboratory of Petroleum Engineering, China University of Petroleum-Beijing, Beijing 102249, China

3 Zhuangxi Oil Production Company of Sinopec Shengli Oilfield, Dongying 257237, China

4 School of Civil Engineering, Beijing Jiaotong University(Weihai), Weihai 264200, China

Abstract The accurate prediction of the produced water quality is an important basis for evaluation of the treatment effect of the produced water at the oilfield joint station, which can provide a scientific basis for early warning of water quality. In the traditional method, we can see that the prediction of the oilfield produced water quality is mainly based on the experience of experts,

引用格式: 徐磊, 侯磊, 朱振宇, 徐震, 雷婷, 李雨, 李强, 陈秀芹, 王九玲, 陈星燃. 基于两层分解算法和改进 SVM 的油田采出水处理效果预测研究. 石油科学通报, 2021, 03: 505-515

XU Lei, HOU Lei, ZHU Zhenyu, XU Zhen, LEI Ting, LI Yu, LI Qiang, CHEN Xiuqin, WANG Jiuling, CHEN Xingran. Prediction of oilfield produced water treatment based on a two-layer decomposition technique and modified SVM. Petroleum Science Bulletin, 2021, 03: 505-515. doi: 10.3969/j.issn.2096-1693.2021.03.041

however, there is no doubt that this method has a strong personal subjectivity so it is difficult to reach an accurate prediction of the quality of the produced water. There is also a part of existing studies to measure whether the produced water quality is up to the relevant standard. However, this method has the disadvantage of taking a long time so that it is not conducive for the efficient development of on-site work. Now there is a part of the existing research with the help of machine learning algorithms, but the situation of data noise and data non-linearity is not fully considered in these methods. In response to the above problems, a novel method for water quality prediction is put forward in this paper, which is based on the combination of the two-layer decomposition method and the modified support vector machine (SVM) algorithm. Through the two-layer decomposition method put forward above, the redundant noise in the prediction process can be eliminated effectively, and at the same time the major features of the original data can be extracted. The method of stratified sampling is used to divide the original dataset so as to avoid the sample deviation caused by the method of traditional random sampling. A modified particle swarm algorithm is applied to optimize the parameters of the SVM so that the global convergence ability can be improved by this algorithm. On the basis of the four cases of the Zhuangxi oil production plant joint station, the prediction accuracy of this method is evaluated in the light of three evaluation indexes: the relative error, the average absolute percentage error and the determination coefficient. On the basis of the average values of these 4 cases on the three indicators are -0.38%, 5.23% and 0.82%, respectively. Compared with the existing mainstream machine learning algorithms, we can see that the method in this paper has higher prediction accuracy.

Keywords oilfield joint station; water quality prediction; machine learning; two-layer decomposition; modified support vector machine

doi: 10.3969/j.issn.2096-1693.2021.03.041

0 引言

随着国内油田的持续开发,部分油田开采后期采出液含水率达到90%以上^[1]。随着国家对环境保护越来越重视,采出水排放标准也越来越严格。因此,油田采出水处理达标已成为国内油田开发过程中的重要任务。油田采出水水质预判主要是凭借专家经验,但该方法具有强烈的个人主观性,难以做到水质的准确预测。也存在一部分研究通过对现场采出水水质展开化验来测算水质是否达标,但该方法耗时较长,不利

于现场工作高效开展。

近年来,随着软计算技术的快速发展,各类机器学习方法在水质预测领域已被广泛应用。相比传统方法,机器学习能够替代从事一些枯燥机械化的工作,客观精确做出一些智能决策,提高工作效率^[2]。图1为主流机器学习方法在水质预测领域应用的热力图,统计了2000年—2020年6大类机器学习方法在水质领域的实践案例,数据来源于Web of Science数据库。通过统计分析可得,机器学习方法在水产养殖、江河湖泊、化工厂、长江、海水、水厂等多个领域运



图1 机器学习方法在水质预测领域的热力图

Fig. 1 Heat map of machine learning methods in the field of water quality prediction

用最为普遍,在油田采出水处理领域处于起步阶段。

近 5 年来,众多机器学习方法中神经网络^[3]和支持向量机^[4]在水质领域运用最为广泛,二者对数据的非线性处理能力通常要优于其它机器学习方法。秦文虎等^[5]构建基于缺失值填补算法和长短时记忆网络相结合的水质预测模型。以太湖水水质监测数据为样本,对模型进行精度检验。结果表明,相较于对比模型,所提出的模型预测精度更高。AHMED^[6]建立一种基于 WDT—ANFIS 的水质混合预测模型,通过两个实际案例的验证,得出该模型预测效果较理想。NOORI^[7]利用提出的 SWAT—ANN 混合模型对美国亚特兰大市区流域的水质进行预测,研究表明,建立的混合模型对于未知流域水质预测具有很大的潜力。但神经网络在运用过程存在容易过拟合、易陷入局部极小、数据量较少或一般时无法充分利用样本信息。

与传统神经网络相比,基于结构风险最小化原则的支持向量机能够在数据量不多、数据呈现较强非线性特征时依旧保持不错的预测效果。李建文等^[8]利用提出的 EEMD—SVR 模型对天津某渔业养殖池塘内溶解氧和 PH 展开预测,研究表明该模型具有较好的预测效果,能够满足实际渔业养殖水质精细化管理需要。白云等^[9]利用提出的 VMD—LSSVM 混合模型对河水水质展开预测,与对比模型相比,该方法具有更高的预测精度,能够为河水水质污染预控提供有效技术支持。ZHANG^[10]建立基于 WT—PSO—SVM 的径流预测模型,以黄河上游唐乃海站 1956 年—2008 年的数据为依托,研究表明,混合模型比单一具有更好的预测精度。KISI^[11]利用 LSSVM 建立水质预测模型,研究证明, LSSVM 比 ANN 和 SRC 的预测效果更好。但以上基于支持向量机的混合方法无法准确捕获数据分解后高频率序列的特征,参数优化过程容易出现局部最优解和早熟收敛等问题。

鉴于目前油田现场通常依据专家经验对水质进行预判,具有个人主观性,很难做到水质的客观准确监测;亦或对采出水水质指标展开化验来测算水质是否达标,该方法耗时较长,不利于提高现场工作效率。部分研究借助于机器学习方法对水质展开了预测,但忽略了数据噪声,对数据的非线性也考虑不足。为解决以上问题,本文提出一种两层分解算法与改进支持向量机相结合的预测方法,利用两层分解算法准确捕获数据分解后的高频率序列特征,采用改进支持向量机解决参数优化过程中容易出现局部最优解和早熟收敛等问题。并结合胜利油田桩西联合站采出水数据对该方法进行准确性评价。

1 基本方法介绍

1.1 两层分解算法(CEEMDAN—VMD)

实际现场采集的数据含有噪声,前人学者对此做了诸多研究, HUANG^[12]提出经验模态分解(EMD)方法,将原始信号分解为多个内涵模态分量(IMF),其本质是滤波和消噪过程,但模型和 EMD 无法很好地融合。针对此问题, WU 和 HUANG^[13]提出集成经验模态分解(EEMD)方法,相较于 EMD, EEMD 对原始信号增加了有限幅度的白噪声,以划分时频空间中的频率范围,减少模式混叠的机会,提高分解算法稳定性,但依旧无法完全抵消所增加的噪声。因此,为了提高 EEMD 的性能,提出一种高级的自适应噪声完整集成经验模式分解算法(CEEMDAN)^[14]。CEEMDAN 对 EEMD 进行改进,采用添加有限次自适应白噪声的方法,有效解决 EMD 中模态混合的问题,高效消除噪声,克服 EEMD 因添加噪声完备性较差和 CEEMD 在 EMD 分解过程中添加一对相反的白噪声而增加计算量的缺点,实现了较高的预测性能。

但 CEEMDAN 分解得到的序列中通常含有高频子序列,高频子序列是制约预测性能的又一障碍,为了进一步提高预测性能,提出两层数据分解(CEEMDAN—VMD)方法。相比其它二次分解算法,变分模态分解(VMD)能够克服 EMD 等方法存在端点效应和模态分量混叠的问题,降低复杂度高和非线性强的时间序列非平稳性,将高频子序列分解为低频子序列,获取较平稳的序列分量,充分削弱原始数据的非平稳性,成功解决高频率序列难以准确预测的难题^[15]。

1.2 改进粒子群算法的支持向量机(MPSO—SVM)

支持向量机(SVM)是由 VAPNIK 在 1995 年^[4]提出的。SVM 主要依赖于结构风险最小化原则,能够有效避免过拟合问题,在数据不足的情况下依旧保持较好的预测性能^[16]。SVM 的预测效果在很大程度上取决于核函数,相关实验表明, RBF 核函数在大多数情况下比其他核函数有更好的性能,需要考虑的参数也更少,因此,考虑采用 RBF 核函数。

超参数的合理选取直接影响预测结果,为此学者们进行了大量研究,提出将 SVM 与优化算法相结合的混合模型。其中,粒子群(PSO)算法是由 KENNEDY 和 EBERHART 在 1995 年^[17]提出的,是一种基于群体行为的全局优化算法。相比于遗传算法(GA)、果蝇

算法(FOA)等其它优化算法, PSO具有参数少、收敛快的优点。然而, 粒子群算法容易出现局部最优解和早熟收敛的问题。为解决这一缺陷, 引入一个惯性权重 ω , 较大的 ω 能够使粒子跳出极值点执行全局搜索, 较小的 ω 能够使粒子进行精细搜索。因此, 为了在全局搜索和精细搜索之间获得平衡, 需要采用一些方法来调整惯性权值, 考虑惯性权值对粒子群优化算法搜索能力的影响, 提出以下改进:

$$\omega(k) = \omega_{\min} + 0.6 \sin(\pi k / k_{\max}) \quad (1)$$

式中: ω_{\min} 为最小惯性权重; k_{\max} 为最大迭代次数。

根据 ω 的变化, 解决过程分粒子本身的局部优化, 全局优化和局部搜索最优粒子3个阶段。

2 数据预处理

2.1 数据来源与分析

研究数据来自胜利油田桩西联合站, 以日数据为时间单位采集2019年9月1日—2020年3月28日的水质数据, 对205组有效数据进行归一化处理。桩西采出水处理站于1993年10月建成投产, 主要负责桩西联合站采出水和海四联采出水处理, 1998年、2004年增设二、三级过滤, 2012年整体改造设计规模为 $1.5 \times 10^4 \text{ m}^3/\text{d}$ 。采出水工艺概括为“三段沉降、三

级过滤”, 图2为桩西联合站采出水系统工艺流程图。

实际现场采集的参数有絮凝剂A剂、絮凝剂B剂、杀菌剂、缓蚀剂、滤料清洗剂、1号分水器界面、2号分水器界面、来液温度、来液水质、各级水质、进站水量和处理后外输水量等, 共计37种参数。其中现场重点关注本站可人为调节的参数、方便水质预警后的水质优化。结合现场实际需求选取37类参数中的絮凝剂A剂、絮凝剂B剂、杀菌剂、缓蚀剂、1号分水器界面和2号分水器界面作为预测模型的输入参数, 二级滤罐回注含油、二级滤罐回注悬浮、三级滤罐注水含油和三级滤罐注水悬浮分别作为A、B、C和D共4个案例中评价水质状况的输出参数, 部分参数如表1所示。

2.2 分层抽样

4个案例的输出参数分为二级滤罐回注含油、二级滤罐回注悬浮、三级滤罐注水含油和三级滤罐注水悬浮。训练集和测试集按7:3的比例进行划分。由于实际样本量不大, 如果采用随机抽样方法划分整个数据集, 得到的训练集和测试集的分布规律可能会与原始数据集分布规律呈现较大偏差。因此, 考虑采用分层抽样方法^[18-19]。

根据4个案例输出参数的分布规律, 将4个案例输出参数的数据均划分为4个区间。以案例A的二级

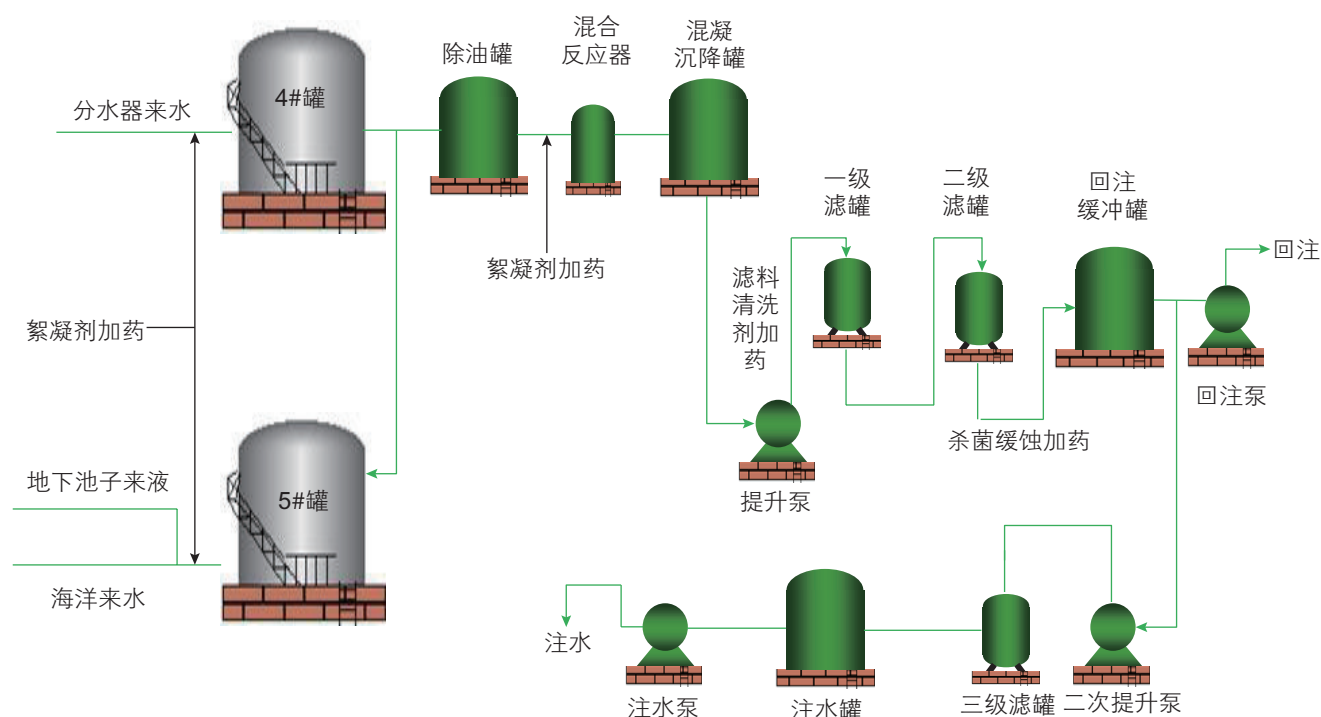


图2 桩西联合站采出水系统工艺流程图

Fig. 2 Process flow chart of the produced water treatment system of Zhuangxi Joint Station

滤罐回注含油数据为例, 随机抽样得到的训练集与分层抽样得到的训练集的偏差如表 2 所示, 能够大致看出分层抽样误差相对较小。进一步依据 4 个案例输出参数的数据分别计算随机抽样和分层抽样的平均绝对百分比误差(MAPE), 统计数据见表 3。经分析, 4 个案例随机抽样和分层抽样 MAPE 的对比依次为 19.87% 和 0.68%、21.41% 和 2.24%、48.00% 和 10.73%、12.16% 和 1.77%。能够看出, 分层抽样的训练集与初始样本保持较好的一致性。

通过对 4 个案例数据展开抽样研究可得, 简单随机抽样得到的训练集与初始样本的分布规律不一致, 分层抽样得到的训练集与初始样本的分布规律保持较

好的一致性。与简单随机抽样相比, 分层抽样能够避免明显的抽样偏差, 保证预测结果的有效性。因此, 采用分层抽样的方法来划分训练集和测试集。

3 混合模型预测框架

根据两层分解算法(CEEMDAN-VMD)和改进粒子群算法的支持向量机(MPSO-SVM)建立混合预测模型, 混合模型的主要实验参数如表 4 所示, 通过实验验证对所提出的混合模型的预测效果进行准确性评价, 混合预测系统流程如图 3 所示。数据预处理技术和模型的建立均采用 Python 3.6.6 编程语言实现。

表 1 水质流程部分数据

Table 1 Partial data of water quality during operation

采集参数	数据采集日期					
	20190901	20190927	20200112	20200131	20200228	20200323
絮凝剂 A 剂/(kg)	400	900	800	800	1200	800
絮凝剂 B 剂/(kg)	600	400	600	600	600	1000
杀菌剂/(kg)	600	600	500	300	500	500
缓蚀剂/(kg)	200	200	200	100	200	200
1#分水器界面	1.30	2.24	2.24	2.26	2.01	2
2#分水器界面	2.40	2.38	2.31	2.32	2.23	2.02
二级滤罐回注含油/(mg/L)	5.60	5.10	4.20	3.60	8.40	6.80
二级滤罐回注悬浮/(mg/L)	4.50	4.30	4.60	3.90	5.90	4.30
三级滤罐注水含油/(mg/L)	3.60	3.50	2.28	3.00	4.70	4.50
三级滤罐注水悬浮/(mg/L)	2.70	2.60	3.60	2.90	3.70	3.40

表 2 基于二级滤罐回注含油数据的随机抽样和分层抽样对比

Table 2 Comparison of random sampling and stratified sampling based on the oil content of secondary filter tank during reinjection

分类	初始样本	随机抽样	分层抽样	随机抽样相对误差/%	分层抽样相对误差/%
1.0	0.32	0.31	0.32	4.81	-0.20
2.0	0.34	0.29	0.34	13.75	-0.63
3.0	0.16	0.24	0.16	-50.29	-0.19
4.0	0.18	0.16	0.18	10.64	1.70

表 3 4 份数据随机抽样和分层抽样的平均绝对百分比误差

Table 3 The mean absolute percentage error of random sampling and stratified sampling based on four data

分类	随机抽样 MAPE/%	分层抽样 MAPE/%
二级滤罐回注含油	19.87	0.68
二级滤罐回注悬浮	21.41	2.24
三级滤罐注水含油	48.00	10.73
三级滤罐注水悬浮	12.16	1.77

表4 实验参数值

Table 4 Experimental parameters

算法	参数	数值
CEEMDAN	最大迭代次数	1000
	噪声标准差	0.20
VMD	适度带宽限制	200
	模态数	4
	种群规模	100
MPSO	最大迭代次数	100
	加速因子	2
	精度	10^{-8}
	惩罚因子	4.19
SVM	核参数	5.36
	核函数	RBF

4 预测结果分析

4.1 预测性能评价指标

预测精度是评价性能好坏的重要指标。相关文献中记载了多种评价预测精度的指标^[20-22]。在各种指标中,相对误差(RE)、平均绝对百分比误差($MAPE$)和决定系数(R^2) 3个指标的绝对值通常在0~1范围内。其中, RE 表示预测值与真实值的偏差程度,利用 RE 的箱型图能够直观反映模型的预测效果。 $MAPE$ 在 RE 的基础上增加了绝对值, $MAPE$ 通过百分比衡量模型性能,其值在10%以内说明预测效果较好。 R^2 越大,

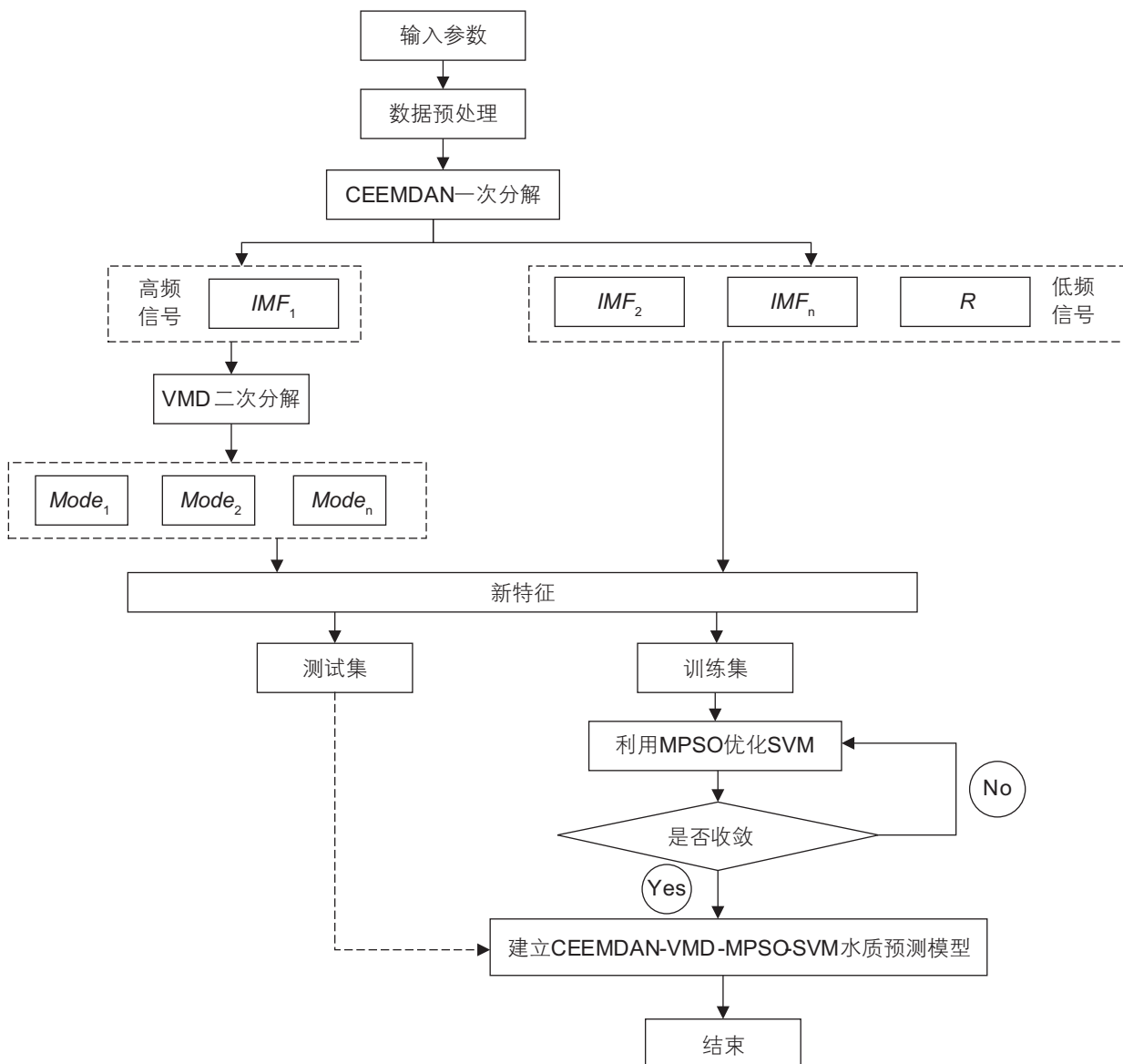


图3 基于CEEMDAN-VMD-MPSO-SVM的混合预测模型流程图

Fig. 3 Flow chart of the hybrid prediction model based on CEEMDAN-VMD-MPSO-SVM

因变量能够被自变量解释的比重越大, R^2 越接近 1, 模型拟合越好, 其值在 0.80 以上说明预测效果较好。上述 3 个物理量均为无量纲单位, 便于直观评价混合模型的预测性能, 因此, 采用 RE 、 $MAPE$ 和 R^2 三个指标来评价模型的预测能力, 各评价指标公式如下:

$$RE = (y_i - \hat{y}_i) / y_i \times 100\% \quad (2)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (4)$$

式中, y_i , \bar{y}_i 和 \hat{y}_i 分别表示初始值、平均值和预测值。

4.2 预测结果对比

将建立的混合预测模型 CEEMDAN-VMD-MPSO-SVM 与 SVM、PSO-SVM、MPSO-SVM、VMD-MPSO-SVM、CEEMDAN-MPSO-SVM 等 5 种模型展开预测性能对比, 验证混合模型的预测性能。

以案例 A 的测试集为例, CEEMDAN-VMD-MPSO-SVM 预测曲线和真实数据曲线如图 4 所示, 与 5 种对比模型相对误差绝对值的箱型图如图 6 所示。经图 4 分析可得, 该混合模型对历史数据的拟合较好, 相对误差 (RE) 主要在 $[-10\%, 10\%]$ 范围内, 保证了较

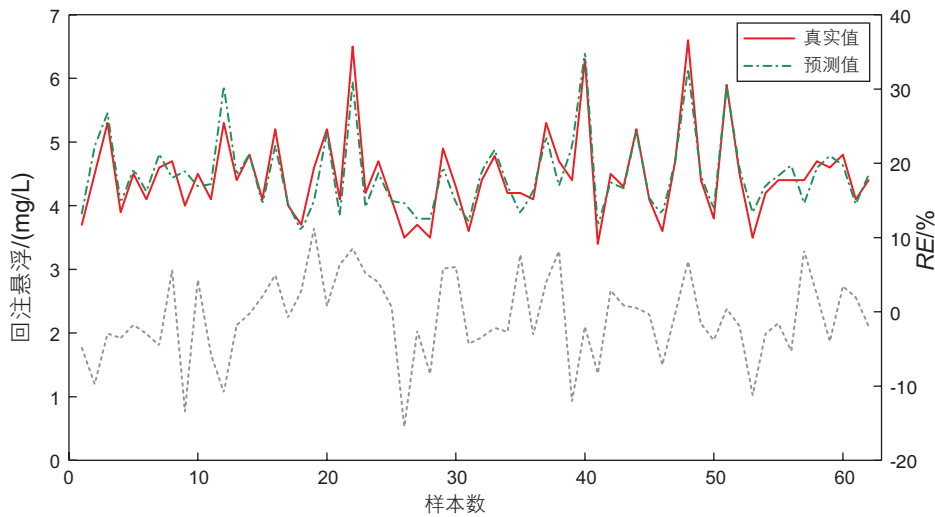


图 4 建立的模型与实际曲线拟合对比图

Fig. 4 Curve fitting of the proposed model and the actual data

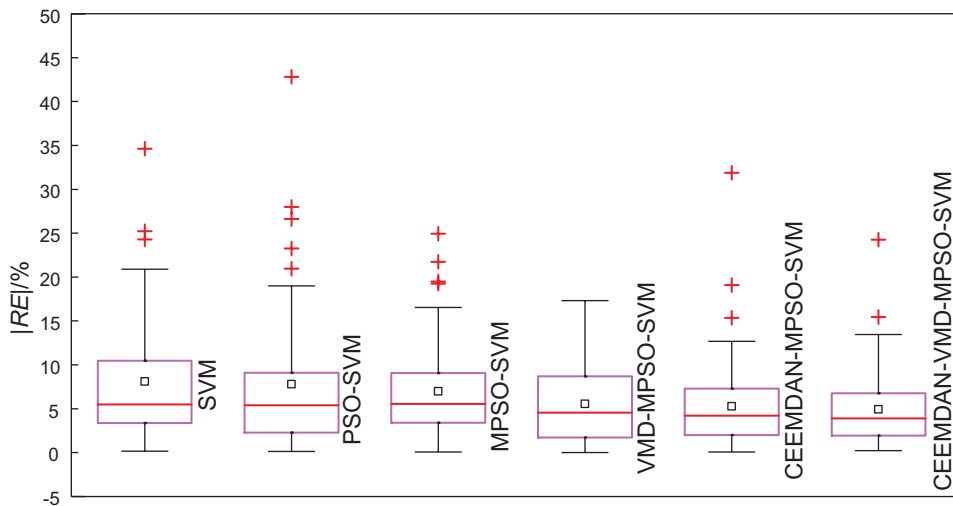


图 5 建立的模型与 5 种对比模型相对误差绝对值的箱型图

Fig. 5 Box plot of the absolute value for the relative error between the proposed model and five comparison models

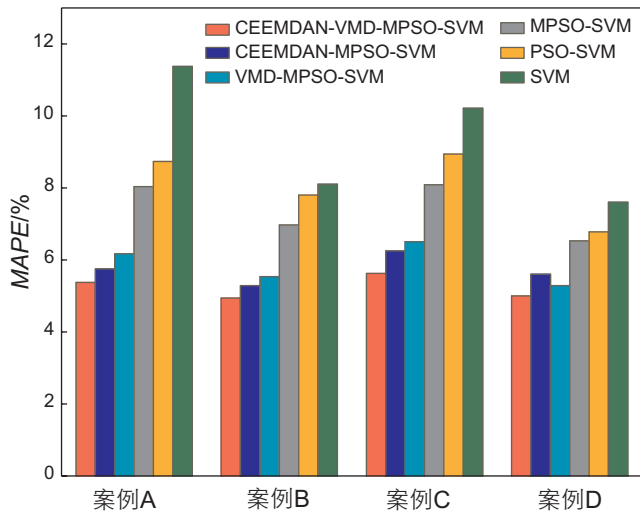


图6 6种模型在4个案例上MAPE的直方图
Fig. 6 Histogram of MAPE for six models based on four cases

高的拟合度。通过图5进一步对比多个模型的箱型图,分析可得,建立的混合模型的预测效果要优于5种对比模型的预测效果。

为了进一步全面比较建立的混合模型的预测效果,表5总结了基于4个案例不同模型预测得到的MAPE和 R^2 值。图6为6种模型在4个案例上MAPE的直方图,图7为6种模型在4个案例上 R^2 的雷达图。

通过6个预测模型的对比可得,CEEMDAN-VMD-MPSO-SVM模型能够在MAPE上获得最小值,在 R^2 获得最大值,预测效果最好。为了深入评价所提出混合模型各子模块的性能,开展了进一步对比研究。

通过SVM与PSO-SVM和MPSO-SVM对比

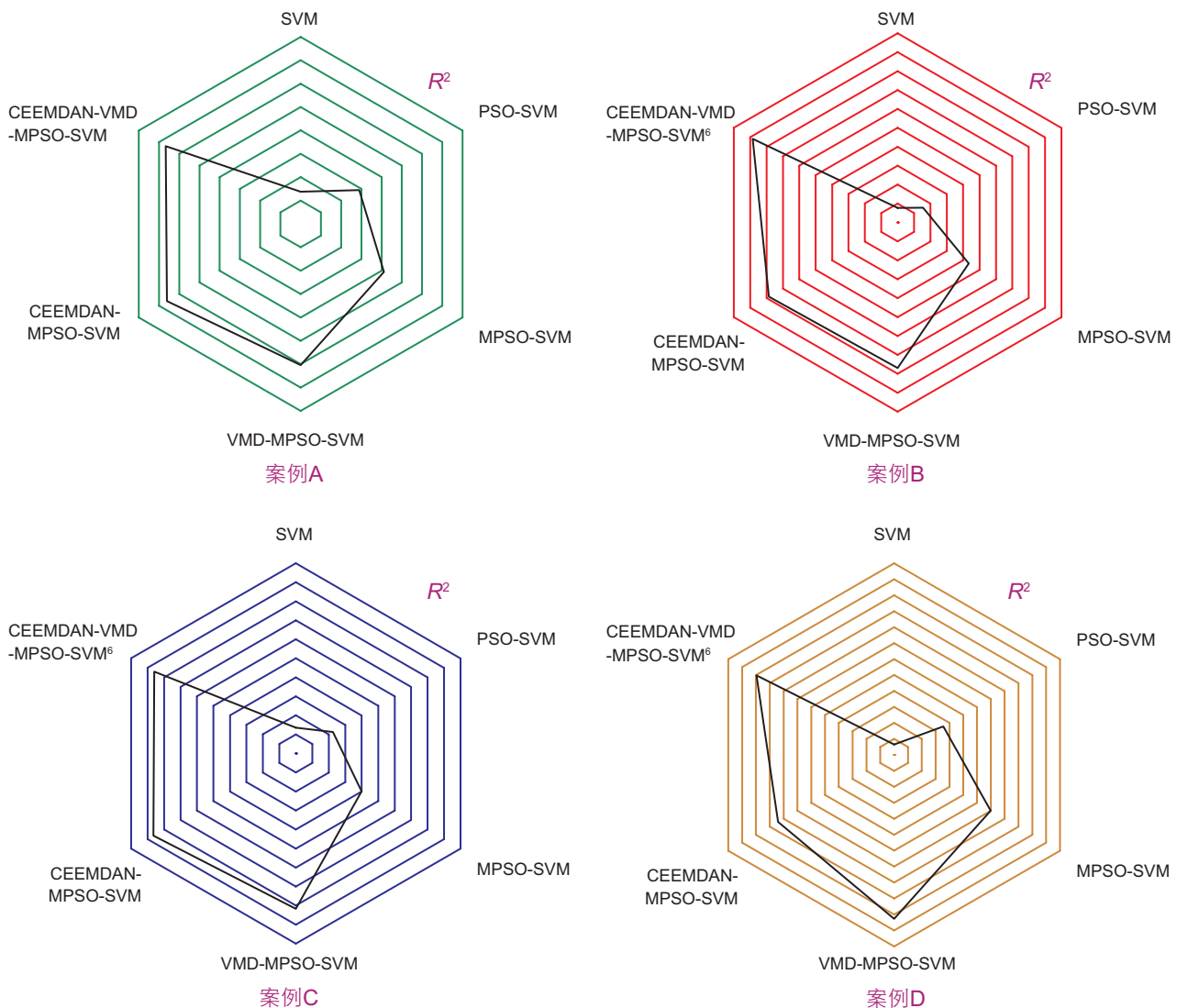


图7 6种模型在4个案例上 R^2 的雷达图
Fig. 7 Radar chart of R^2 for six models based on four cases

分析,发现PSO-SVM和MPSO-SVM的预测性能明显优于SVM,证明优化算法对模型预测性能能够带来一定的提升。通过PSO-SVM与MPSO-SVM对比分析,发现MPSO-SVM预测性能优于PSO-SVM,证明PSO的改进进一步提升了PSO-SVM的预测性能。

通过MPSO-SVM与VMD-MPSO-SVM和CEEMDAN-MPSO-SVM预测结果的对比,发现VMD和CEEMDAN单层分解算法的添加均能够提高模型的预测精度,证明了单一分解算法能够在一定程度上捕获数据信号特征。

通过CEEMDAN-VMD-MPSO-SVM与CEEMDAN-MPSO-SVM和VMD-MPSO-SVM的对比,发现两层分解算法模型的预测效果要优于单层分解算法模型的预测效果。证明CEEMDAN-VMD作为一种强大的数据分解算法,相比单一数据分解算法,能够进一步有效消除冗余噪声、捕获数据主要特征、降低模型预测难度。

建立的混合模型在4个案例上的预测精度均优于对比模型,证明该模型在油田水质预测领域具有良好的适用性,能够为水质预警等方面提供科学依据。

以CEEMDAN-VMD-MPSO-SVM为例,进一步分析该模型在案例A、B、C和D上的预测效果,图8为该模型在4个案例上MAPE和R²值的趋势线图。分析可得MAPE在4个案例上的波动性较小,预测值范围为[4.94%, 5.62%],保持在10%以内,预测效果良好。但R²针对4个案例大致呈现出下降的趋势,且针对案例D,R²值小于0.8,预测效果不理想。由2.1节可知,4个案例的输出值分别为二级滤罐回注含油、二级滤罐回注悬浮、三级滤罐注水含油和三级滤罐注水悬浮,依次对应图2流程中的先后顺序,表明从二级滤罐到三级滤罐过程增加了不确定性,越靠近流程尾部,越难准确预测,以其它对比模型为例依旧能够得出类似的结论,因此,在水质预警过程中更应该注重提高二级滤罐指标的预测精度。

表5 基于4个案例不同模型预测得到的MAPE和R²值

Table 5 MAPE and R² values obtained by different models based on four cases

模型	案例A		案例B		案例C		案例D	
	MAPE/(%)	R ²	MAPE/(%)	R ²	MAPE/(%)	R ²	MAPE/(%)	R ²
SVM	11.38	0.67	8.11	0.47	10.21	0.44	7.60	0.23
PSO-SVM	8.73	0.74	7.79	0.51	8.94	0.48	6.78	0.38
MPSO-SVM	8.04	0.81	6.97	0.60	8.09	0.62	6.53	0.55
VMD-MPSO-SVM	6.17	0.90	5.54	0.81	6.51	0.78	5.29	0.61
CMMEDAN-MPSO-SVM	5.75	0.93	5.29	0.83	6.25	0.79	5.60	0.62
CEEMDAN-VMD-MPSO-SVM	5.37	0.94	4.94	0.84	5.62	0.84	5.00	0.70

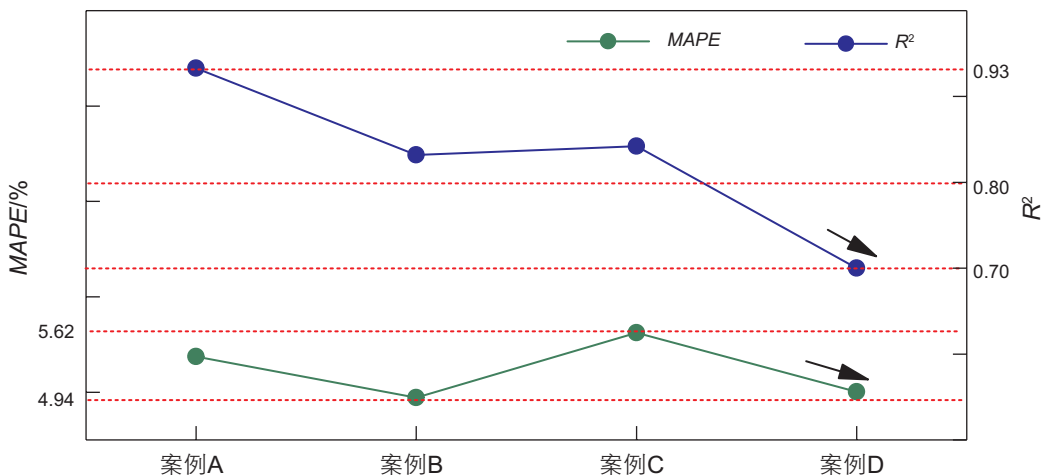


图8 所提出模型在4个案例上MAPE和R²值

Fig. 8 The MAPE and R² values of the proposed model on four cases

5 结论

(1)采用分层抽样的方法对水质数据进行划分,使训练集和测试集的分布规律与原始数据集的分布规律保持较好的一致性,有效避免了随机抽样引起的较大偏差,保证预测结果的客观性和可靠性。

(2)提出了基于支持向量回归的改进粒子群优化算法,提高了粒子群的搜索能力、避免了局部最优解和早熟收敛,能够有效提高全局最优解的收敛速度和能力。

(3)首次将两层数据分解算法应用到油田水质领域,与其他分解算法相比,两层分解算法具有2种子

分解算法的综合性能,能够消除预测过程中的冗余噪声,有效捕捉原始数据集的主要特征。

(4)混合预测模型综合了数据分解算法和优化算法的优势,针对4个预测案例,对比常规SVM模型、MAPE下降幅度分别达到6.001%、3.164%、4.590%和2.6%, R^2 提升幅度分别达到0.264、0.364、0.404和0.467,建立的水质模型拥有较高的预测精度,同时弥补了机器学习方法在该领域的空白。

(5)提出的混合模型能够用于油田联合站采出水处理效果预测研究,准确预测处理后的水质是否达标,为水质预警提供科学的依据,保证水质安全,进一步能够有效降低电耗和药耗。

参考文献

- [1] 任永忠,陈苏宁,刘智金,等.油田外排污水处理技术及研究进展[J].安全与环境工程,2011,18(2):45-49. [REN Y Z, CHEN S N, LIU ZHI J, et al. Treatment Technology of Oilfield Discharged Wastewater and Development Tendency [J]. Safety and Environmental Engineering, 2011, 18(2): 45-49.]
- [2] Li Y, Yang L, Yang B, et al. Application of interpretable machine learning models for the intelligent decision[J]. Neurocomputing, 2019, 333(14): 273-283.
- [3] MOSHE L, VLADIMIR Y, ALLAN P, et al. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural Networks 1993, 6(6): 861-867.
- [4] VAPNIK V. The nature of statistical learning theory springer NewYork google scholar. 1995.
- [5] 秦文虎,陈溪莹.基于长短时记忆网络的水质预测模型研究[J].安全与环境学报,2020,20(1):328-334. [QIN W H, CHEN X Y. Water quality forecast and prediction model based on long-and short-term memory network [J]. Safety and Environmental Engineering, 2020, 20(1): 328-334.]
- [6] AHMED A N, OTHMAN F B, AFAN H A, et al. Machine learning methods for better water quality prediction[J]. Journal of Hydrology, 2019, 578: 124084.
- [7] NOORI N, KALIN L, ISIK S. Water quality prediction using SWAT-ANN coupled approach[J]. Journal of Hydrology, 2020, 590: 125220.
- [8] 李建文,孟连子,刘娜,等.基于EEMD-SVR的渔业养殖水质预测模型[J].天津理工大学学报,2018,34(5):14-19. [LI J W, MENG L Z, LIU N, et al. The prediction model of fishery water quality based on EEMD-SVR [J]. Journal of Tianjin University of Technology, 2018, 34(5): 14-19.]
- [9] 白云,李勇.基于VMD-LSSVR的河水水质预测方法[J].安全与环境学报,2020,20(3):1162-1168. [BAI Y, LI Y. River water quality prediction method based on VMD-LSSVR [J]. Journal of Safety and Environment, 2020, 20(3): 1162-1168.]
- [10] ZHANG F P, DAI H C, TANG D S. A Conjunction Method of Wavelet Transform-Particle Swarm Optimization-Support Vector Machine for Streamflow Forecasting[J]. Journal of Applied Mathematics, 2014, 2014: 1-10.
- [11] Kisi O. Modeling discharge-suspended sediment relationship using least square support vector machine[J]. Journal of Hydrology, 2012, 456: 110-120.
- [12] 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. Proceedings A 1998; 454(1971): 903-995.
- [13] Wu Z, Huang N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adaptive Data Anal 2009; 1(1): 1-41.
- [14] Torres M E, Colominas M A, Schlotthauer G, et al. A complete ensemble empirical mode decomposition with adaptive noise. In: Acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on. IEEE; 2011.
- [15] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2014; 62(3): 531-544.
- [16] Shawe-Taylor J, Bartlett P L, Williamson R C, et al. Structural risk minimization over data-dependent hierarchies. IEEE Transactions on Information Theory 1998; 44(5): 1926-1940.

- [17] Kennedy J, Eberhart R, Particle swarm optimization neural networks, 1995. proceedings, IEEE Int. Conf. 4 (1995) 1942–194.
- [18] Hens A B, Tiwari M K. Computational time reduction for credit scoring: An integrated approach based on support vector machine and stratified sampling method. *Expert Systems with Applications* 2012; 39(8): 6774–6781.
- [19] Y Y M, Wu Q, Huang J Z, et al. Stratified sampling for feature subspace selection in random forests for high dimensional data. *Pattern Recognition* 2013; 46(3): 769–787.
- [20] Shamshirband S, Rabczuk T, Chau K W. A Survey of Deep Learning Techniques: Application in Wind and Solar Energy Resources. *IEEE Access* 2019;7: 164650–164666.
- [21] Ahmad T, Chen H, Guo Y, et al. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy and Buildings* 2018;165: 301–320.
- [22] Mathew J, Griffin J, Alamaniotis M, et al. Prediction of welding residual stresses using machine learning: Comparison between neural networks and neuro-fuzzy systems. *Applied Soft Computing* 2018;70: 131–146.

(责任编辑 韩晔华 编辑 马桂霞)