

基于 IBAS 和 LSTM 网络的池塘水溶解氧含量预测

孙龙清^{1,2} 吴雨寒^{1,3} 孙希蓓^{1,3} 张松^{1,3}

(1. 中国农业大学国家数字渔业创新中心, 北京 100083;

2. 农业农村部精准农业技术集成科研基地(渔业), 北京 100083;

3. 中国农业大学信息与电气工程学院, 北京 100083)

摘要: 为了提高池塘水体中溶解氧含量(DO)预测精度,本文提出了一种基于改进的天牛须搜索算法(Improved beetle antennae search algorithm, IBAS)和长短期记忆网络(Long short-term memory, LSTM)相结合的溶解氧含量预测模型。为了降低模型输入维度,提高模型计算效率,采用皮尔逊(Pearson)相关系数分析法得出各因子与溶解氧含量之间的相关性,提取强关联因子作为模型输入特征;为了使天牛须搜索算法(Beetle antennae search algorithm, BAS)在全局搜索和局部搜索中达到平衡,提高算法的收敛速度,提出衰减因子指数递减策略改进天牛须搜索算法,将衰减因子 γ 与迭代次数相联系并呈指数函数递减;通过IBAS优化LSTM网络,得到最优参数组合策略,建立P-IBAS-LSTM非线性溶解氧含量预测模型。并利用该模型对江苏省宜兴市水产养殖研究中心某池塘水体溶解氧含量进行验证,预测2 h后的溶解氧含量。在与常见的7种模型对比中发现,本文所提出的方法在各项指标中都取得了最优的性能,均方误差(MSE)为0.644 2 mg²/L²、均方根误差(RMSE)为0.802 6 mg/L、平均绝对误差(MAE)为0.530 6 mg/L。实验结果表明本文所提出的模型预测精度更高,泛化性能更强,可以满足实际对溶解氧含量准确预测的需求,并为池塘养殖中水质预警控制提供参考。

关键词: 池塘水环境; 溶解氧含量预测; 长短期记忆网络; 改进天牛须搜索算法; 因子筛选; 参数优化

中图分类号: TP391 **文献标识码:** A **文章编号:** 1000-1298(2021)S0-0252-09

Dissolved Oxygen Prediction Model in Ponds Based on Improved Beetle Antennae Search and LSTM Network

SUN Longqing^{1,2} WU Yuhan^{1,3} SUN Xibei^{1,3} ZHANG Song^{1,3}

(1. National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China

2. Precision Agricultural Technology Integration Research Base (Fishery),
Ministry of Agriculture and Rural Affairs, Beijing 100083, China

3. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: To improve the prediction accuracy of dissolved oxygen content in ponds, a novel long short-term memory (LSTM) optimized by an improved beetle antennae search algorithm (IBAS) was proposed. Firstly, Pearson correlation coefficient was used to obtain the linear correlation between each factor and dissolved oxygen. The key impact factors of dissolved oxygen were selected by Pearson correlation coefficient as the input feature, which can reduce the input dimension, eliminate the correlations of original variable, and improve the calculation efficiency of the model. Secondly, to balance the global search and local search, and improve the convergence speed of beetle antennae search algorithm (BAS), an IBAS with exponential decreasing strategy of attenuation factor was proposed, which linked the attenuation factor η with the number of iterations. Finally, LSTM network was optimized by IBAS to get the best parameter combination strategy to construct a P-IBAS-LSTM prediction model between dissolved oxygen and these factors. Based on the presented model, the dissolved oxygen was predicted for an experimental pond during April 28th to September 8th, 2020 in the Research Center of Yixing City, Jiangsu Province. In the case of the same data, the mean squared error (MSE), root mean square error (RMSE), and the average absolute error (MAE) of the P-IBAS-

LSTM were 0.644 2 mg²/L², 0.802 6 mg/L, 0.530 6 mg/L, respectively. The experimental results showed that the proposed model of P-IBAS-LSTM had higher performance and stronger generalization performance when compared with common prediction models, which could meet the actual needs of predicting dissolved oxygen accurately and help farmers make decisions in ponds.

Key words: pond water environment; dissolved oxygen content prediction; long short-term memory network; improved beetle antennae search algorithm; factor selection; parameter optimization

0 引言

溶解氧含量(DO)是池塘养殖水环境的一个重要的参数化指标,是影响池塘水质的关键因素之一^[1-2]。溶解氧含量会直接影响到水生生物的生长发育、新陈代谢和摄食量^[3],比如池塘水中充足的溶解氧是保证河蟹正常新陈代谢的前提条件,溶解氧浓度与河蟹食欲也有很大关联,适宜的溶解氧浓度会提高饲料的利用率,促进河蟹的生长和发育,从而降低河蟹的病死率,提高河蟹的品质^[4]。因此,对池塘溶解氧进行监测,提前预测溶解氧含量,掌握溶解氧含量的变化趋势,对减少养殖者生产损失,提高其经济效益有关键性作用。

池塘水中溶解氧含量的变化是不规律的,过程是动态复杂的,溶解氧含量易受到其他水质参数(如水温、pH 值、浊度等)、气象参数(如风速、太阳辐射、空气温度等)等因素的影响^[5],具有非线性和不稳定的特点。而这些参数之间也存在很强的耦合性,因此,获取溶解氧含量与外部因素之间的相关性,选出对溶解氧含量有关键性影响的因子,不仅可以减少冗余信息,还能提高算法的计算速率和预测准确度^[6]。Pearson 相关系数目前应用最广泛,可以较好地衡量 2 个随机变量之间的相关性,并对最终预测结果产生重要影响^[7-8]。

目前很多研究学者都对溶解氧进行了预测,并取得一定成果^[9-11]。文献[12-16]方法都实现了对溶解氧含量的预测,但在时间维度上仍然不能很好地表达特征之间的相关性。长短期记忆网络将时序概念引入网络结构中,在处理时序数据方面具有明显的优势,更具备挖掘长距离时序数据信息的能力^[17],已经应用在多个领域^[18-20]。

LSTM 模型的预测效果与其参数的设定有很大关联,网络参数的设置对模型最终预测准确度产生重要影响。利用智能优化算法对 LSTM 的超参数进行寻优,可以较迅速地找到最优参数组合,减少人工手动调参的复杂性和不确定性,在一定程度上提高模型预测精度^[21]。天牛须搜索算法(BAS)^[22]是一个智能优化算法,具有收敛速度快,计算量小等优点,现在已广泛应用于多个领域^[23-24]。YANG

等^[25]将 BAS 应用于桥梁传感器位置部署的优化问题中,有效覆盖了桥梁的敏感监控点。WANG 等^[26]利用 BAS 来优化 SVM 模型中的惩罚因子和核函数,提出 BAS-SVM 方法,可以准确识别出风力发电机滚动轴承的不同状态。FEI 等^[27]提出 BASA-MkRVR 模型,用来预测电力变压器中溶解气体含量,并通过实验证明了该模型的有效性和优越性。WU 等^[28]采用 BAS 优化无人机的路径,可以快速地实现全局路径规划。XU 等^[29]首先利用灰色关联度分析方法选出关联度较高的特征作为最终特征,再用 BAS 优化极限学习机(Extreme learning machine),实现对锂离子电池容量的预测。

本文通过 Pearson 相关系数分析法对所采集的数据进行因子筛选,降低模型输入向量的维度。提出衰减因子指数递减策略改进 BAS,并优化 LSTM 神经网络,建立 P-IBAS-LSTM 非线性溶解氧预测模型。采用江苏省宜兴市河蟹集约化养殖池塘所采集到的水质和气象参数展开实验,以期探索出较高准确度、适应性的池塘养殖溶解氧含量预测模型,解决信息冗余、预测准确度差及鲁棒性差的问题。

1 材料和方法

1.1 数据来源

数据来源于中国农业大学农业物联网工程技术研究中心——江苏省宜兴市高腾镇水产养殖水质监测系统,如图 1 所示。水产养殖基地配有 DO 传感器、pH 值传感器、温度传感器、氨氮传感器、电导率

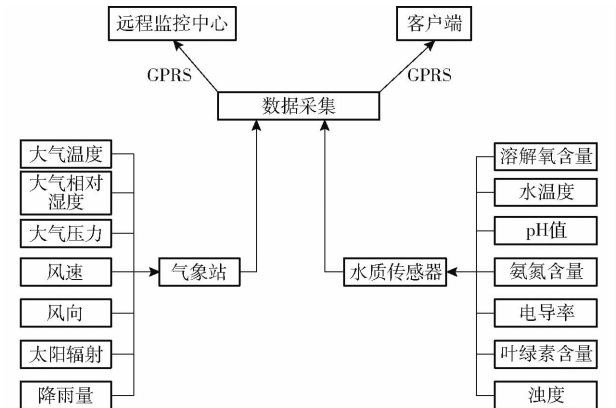


图 1 在线监测系统结构图

Fig. 1 Online monitoring system structure diagram

传感器、叶绿素传感器、小型气象站等,采集数据类型如表 1 所示。池塘旁边配有传感器监测显示屏,工作人员可以实时监测观察数据。本文数据选择的是宜兴市水产养殖管理基地中 3 号塘口,面积约 $1.3 \times 10^4 \text{ m}^2$,养殖种类为河蟹,池塘内有增氧机和投饲机,传感器和气象站同时将所采集到的数据上传至水产养殖智能管理系统。

表 1 采集的数据类型
Tab.1 Type of data collected

类型	参数	简称	单位	采集周期/(次·h ⁻¹)
水质数据	溶解氧含量	DO	mg/L	6
	pH 值	pH		
	水温度	WT	℃	
	氨氮含量	AN	mg/L	
	电导率	EC	μS/cm	
	叶绿素含量	C	μg/L	
	浊度	T	NTU	
气象数据	大气相对湿度	AH	%	60
	大气温度	AT	℃	
	大气压力	AP	kPa	
	风速	WS	m/s	
	风向	WD	(°)	
	太阳辐射	SR	W/m ²	
	降雨量	R	mm	

1.2 Pearson 相关系数

Pearson 相关系数用于描述 2 个随机变量之间的依赖性,反映 2 个变量之间的相关性。使用 Pearson 相关系数会生成一个相关系数矩阵,去除相关性较弱和不相关的数据,选出重要特征。Pearson 相关系数 ρ 计算公式见文献[30]。 $\rho > 0$ 时,2 个变量呈正相关关系; $\rho = 0$ 时,2 个变量无线性相关关系; $\rho < 0$ 时,2 个变量呈负相关关系。

ρ 的绝对值越大表示两者相关性越大,越接近于 1 或 -1,相关性就越强,越接近于 0,相关性就越弱。

1.3 长短期记忆神经网络

LSTM 是一种递归神经网络,具有记忆的网络模型,能有效地学习长期依赖关系,在时间序列预测方面能够达到较好的效果^[31]。常见的 LSTM 记忆单元如图 2 所示。LSTM 通过引入门结构,即输入门 i_t 、遗忘门 f_t 、输出门 o_t 3 种门控单元作为内

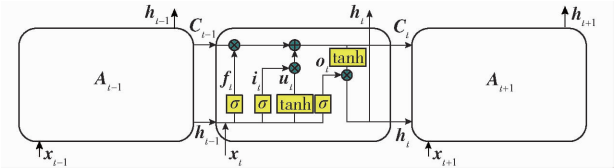


图 2 LSTM 结构图

Fig.2 Structure diagram of LSTM cell

部机制来调节信息流,实现对网络长期记忆,从而解决梯度消失和梯度爆炸问题。

遗忘门 f_t 主要是决定 C_{t-1} 中需要记住和丢弃的信息,通过上一个单元输出向量 h_{t-1} 和当前单元输入向量 x_t ,来决定当前记忆单元需要留下和丢弃的信息。输入门 i_t 控制哪些信息更新至新的细胞状态 C_t 中。确定当前单元新增了多少信息,确定当前单元状态需要输出的一部分信息,最后将 C_t 与 o_t 相结合确定本单元最终输出信息^[32]。LSTM 整体运算公式为

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f)$$

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i)$$

$$u_t = \tanh(W_{ux}x_t + W_{uh}h_{t-1} + b_u)$$

$$C_t = f_t C_{t-1} + i_t u_t$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o)$$

$$h_t = o_t \tanh(C_t)$$

式中 h_t ——记忆单元格输出向量

t ——时刻

σ ——sigmoid 函数

W_{fx} 、 W_{fh} ——遗忘门权值矩阵

W_{ix} 、 W_{ih} ——输入门权值矩阵

W_{ux} 、 W_{uh} ——候选向量的权值矩阵

W_{ox} 、 W_{oh} ——输出门权值矩阵

b_f ——遗忘门偏差矩阵

b_i ——输入门偏差矩阵

b_u ——候选向量的偏差矩阵

b_o ——输出门偏差矩阵

u_t —— t 时刻的候选向量

1.4 衰减因子指数递减策略改进天牛须搜索算法 (IBAS)

BAS 是文献[23]提出来的一种新的全局优化算法,具有速度快,简单易实现的特点。BAS 是根据天牛觅食过程所受到的启发,其搜索过程如图 3 所示。天牛根据左右两须接收到的食物气味强度来判断食物所在的位置,若左须接收到的食物气味强度较强,则天牛飞向左边,否则,天牛飞向右边。而在实际应用中,食物气味强度即为适应度函数,通过模仿天牛觅食过程,根据天牛位置不断更新,最终得到适应度函数的最优值。

利用改进后的天牛须搜索算法来优化 LSTM 参数,采用均方根误差作为适应度函数。适应度函数 $f(x)$ 定义为

$$f(x) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

式中 N ——样本数

$\hat{y}_i$ ——第 i 个样本的预测值

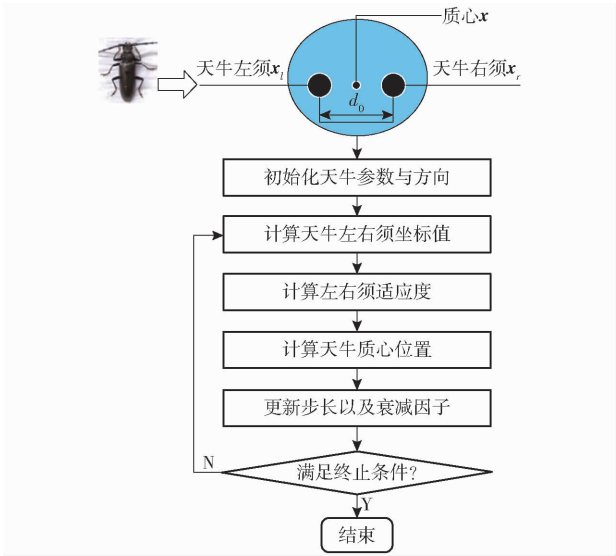


图 3 IBAS 寻优过程
Fig. 3 IBAS optimization process

y_i ——第 i 个样本的真实值
初始化天牛质心向量 $\mathbf{x}$, 步长 l_{step} 为 1, 步长与两须之间距离的比率 $c = 3$, 迭代次数 $T = 150$, 随机生成天牛右须指向左须的一个 k 维的方向向量并归一化, 其中以 LSTM 的隐藏层神经元个数 (2 层)、迭代次数、批处理规模、学习率作为优化对象, 所以天牛空间位置的维度 $k = 5$, 天牛初始位置 $\mathbf{x}$ 和方向向量 $\mathbf{r}_{\text{dir}}$ 为

$$\mathbf{x} = \text{rank}(k, 1) \tag{8}$$

$$\mathbf{r}_{\text{dir}} = \frac{\text{rank}(k, 1)}{\|\text{rank}(k, 1)\|} \tag{9}$$

式中 rank ——随机生成向量函数
计算天牛左右须的坐标, 分别为

$$\mathbf{x}_l = \mathbf{x} + d_0 \frac{\mathbf{r}_{\text{dir}}}{2} \tag{10}$$

$$\mathbf{x}_r = \mathbf{x} + d_0 \frac{\mathbf{r}_{\text{dir}}}{2} \tag{11}$$

其中
$$d_0 = \frac{l_{\text{step}}}{c} \tag{12}$$

式中 $\mathbf{x}_l$ ——天牛左须坐标

$\mathbf{x}_r$ ——天牛右须坐标

d_0 ——天牛两须之间的距离

然后根据适应度函数 $f(x)$ 来推进天牛须在空间区域内的搜索。根据得出的左、右须坐标 $\mathbf{x}_l$ 和 $\mathbf{x}_r$, 得出左、右须的适应度为

$$f_{\text{left}} = f(\mathbf{x}_l) \tag{13}$$

$$f_{\text{right}} = f(\mathbf{x}_r) \tag{14}$$

式中 $f_{\text{left}}, f_{\text{right}}$ ——天牛左、右须适应度

天牛下一步的质心位置公式为

$$\mathbf{x}^i = \begin{cases} \mathbf{x}^{i-1} + l_{\text{step}} \mathbf{r}_{\text{dir}} (\mathbf{x}_l - \mathbf{x}_r) & (f_{\text{left}} < f_{\text{right}}) \\ \mathbf{x}^{i-1} - l_{\text{step}} \mathbf{r}_{\text{dir}} (\mathbf{x}_l - \mathbf{x}_r) & (f_{\text{left}} > f_{\text{right}}) \end{cases} \tag{15}$$

式中 $\mathbf{x}^i, \mathbf{x}^{i-1}$ ——第 $i, i-1$ 次迭代时天牛质心位置

为了提高天牛在空间搜索能力上的适应性, 在保证模型精度的基础上, 提高迭代的效率。本文提出衰减因子指数递减策略改进天牛须算法, 将衰减因子 γ 与当前迭代次数 i 联系在一起, 使步长 l_{step} 成为一个可变量。由初始迭代时大步长搜索大致范围, 全局搜索能力强; 迭代后期 γ 较小, 小步长局部精细搜索, 从而达到算法精度和全局收敛速度之间的平衡。步长更新如下

$$l_{\text{step}} = \gamma l_{\text{step}} \tag{16}$$

其中
$$\gamma = \frac{1}{e^{1 + 10 \frac{i}{T}}} \tag{17}$$

式中 T ——总迭代次数

1.5 P-IBAS-LSTM 溶解氧预测模型实现过程

溶解氧数据作为一种时间序列, 其受到多种因素的共同影响, 具有不稳定、周期不确定特点, 为了更加准确地预测溶解氧含量, 本文选择在时间序列中表现较好的 LSTM 模型, 构建溶解氧含量预测模型。LSTM 中超参数控制着 LSTM 网络结构, 并对最终预测结果有重要影响, 为了更加匹配网络拓扑结构与溶解氧特征, 本文将 IBAS 与 LSTM 相融合, 对 LSTM 寻优获取最优超参数组合, 构建 P-IBAS-LSTM 溶解氧含量预测模型。

基于上文对天牛搜索算法的改进, 建立的非线性溶解氧含量预测模型流程如图 4 所示, 具体步骤为:

(1) 设计实验方案, 采集水质数据和气象数据, 对实验数据进行预处理, 并采用 Pearson 相关系数对数据进行特征选择, 确定最终模型输入特征, 降低模型输入向量的维度。

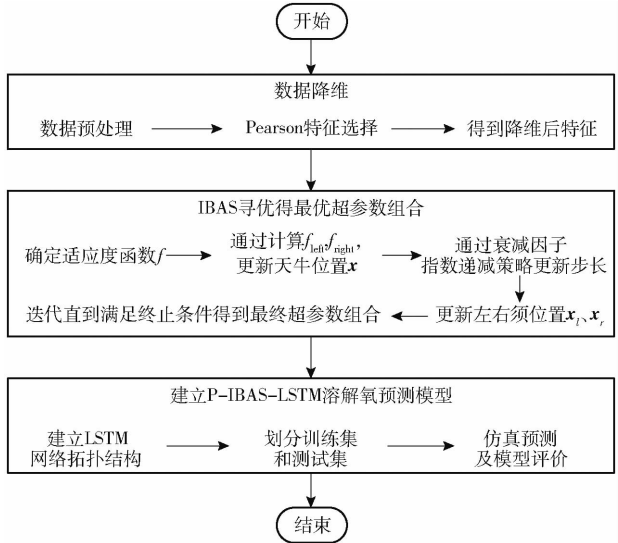


图 4 P-IBAS-LSTM 溶解氧预测模型实现流程图
Fig. 4 Dissolved oxygen prediction flow chart based on P-IBAS-LSTM

(2)提出衰减因子指数递减策略对天牛须搜索算法进行改进,根据迭代次数不断增加,衰减因子不断减小,改变天牛步长,从而改变天牛在空间的位置,有效平衡天牛全局搜索和局部搜索,优化LSTM 获取最佳超参数组合,构建最终溶解氧预测模型。

(3)对数据集进行标准化,按照比例 9:1 分配训练集和测试集,对模型进行训练。

(4)将测试样本输入预测模型,预测溶解氧含量。

(5)通过模型评价指标,与其他模型进行对比。

1.6 模型评价指标

选择均方误差 (MSE)、平均绝对误差 (MAE)、均方根误差 (RMSE) 作为性能评价指标对本文模型

和现有的传统模型进行对比和客观评价。MSE、RMSE、MAE 值越小,表示模型预测精度就越高。

2 实验分析

2.1 数据集和数据预处理

在线实时采集水质参数和气象数据,即溶解氧含量、pH 值、温度、氨氮含量、电导率、叶绿素含量、浊度 7 个水质参数,大气相对湿度、大气温度、大气压力、风速、风向、太阳辐射以及降雨量 7 个气象参数。数据收集时间为 2020 年 4 月 28 日至 2020 年 9 月 8 日,水质数据每 10 min 采集一次,累计共 17 663 条数据,气象数据每 1 min 采集一次,累计共 156 299 条数据。2020 年 6 月 18 日部分样本数据,如表 2 所示。

表 2 2020 年 6 月 18 日采集的部分原始数据

Tab.2 Part of original environmental data collected on June 18, 2020

采集时刻	DO/ (mg·L ⁻¹)	pH 值	WT/ ℃	AN/ (mg·L ⁻¹)	EC/ (μS·cm ⁻¹)	C/ (μg·L ⁻¹)	T/ NTU	AH/ %	AT/ ℃	AP/ kPa	WS/ (m·s ⁻¹)	WD/ (°)	SR/ (W·m ⁻²)	R/ mm
00:40:19	8.70	8.88	28.55	0.4	339	0	19.6	68.4	31.7	99.70	1.2	51	23	0
00:50:19	8.49	8.80	28.61	0.4	339	0	21.7	68.8	31.7	99.67	0	39	3	0
01:00:19	8.42	8.88	28.59	0.5	339	0	20.4	68.2	31.7	99.67	1.2	37	33	0
01:10:19	5.84	8.93	28.59	0.5	338	0	20.6	68.1	31.8	99.65	0	37	44	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
18:30:38	7.43	8.61	25.63	0.5	342	0	20.9	91.3	21.4	100.25	0	99	37	0.2
18:40:33	7.58	8.52	25.58	0.5	342	0	20.3	90.8	21.1	100.26	0.5	102	4	0
18:50:17	7.49	8.51	25.54	0.5	342	0	21.0	90.3	21.1	100.29	0	103	33	0

在采集数据时,由于网络不稳定或者断电等影响,所采集的样本数据会存在一定的异常值和缺失值,为保证模型预测性能,需对数据进行预处理,包括以下几个步骤:

(1)数据清洗:利用 Dixon 检验法找出数据中存在的异常值,并用对应特征序列的全局平均值代替异常值。

(2)数据转换:由于本文水质数据采集频率为每 10 min 采集一次,而气象数据每 1 min 采集一次,大部分气象数据在 10 min 内变化很小,因此,将气象数据转换为 10 min 的平均值。

(3)数据同步及融合:水质数据和气象数据并没有同步采集,将转换后的气象数据与时间最接近的水质数据进行匹配,实现数据的同步及融合。

(4)填充数据:采用线性插值法对缺失的数据进行填充,若丢失数据较多,则采用天气类型相同或相邻几天中同一时刻的数据填补。线性插值法公式为

$$x_{k+i}=x_k+\frac{i(x_{k+j}-x_k)}{j} \quad (0<i<j) \quad (18)$$

式中 x_k 、 x_{k+j} ——第 k 、 $k+j$ 时刻数据

i 、 j ——时刻
 x_{k+i} ——第 $k+i$ 时刻所缺失的数据

对于出现异常的数据,采用均值平滑法进行填充,均值平滑法公式为

$$x_m=\frac{x_{m-1}+x_{m+1}}{2} \quad (19)$$

式中 x_{m-1} 、 x_{m+1} ——第 $m-1$ 、 $m+1$ 时刻数据
 x_m ——第 m 时刻异常的数据

2.2 Pearson 特征选择分析

池塘水产养殖中溶解氧的变化会受到其他水质参数和气象数据的影响,为了提升模型计算速度和预测精度以及更加客观地进行特征选择,本文使用 Pearson 模型降低参数的维度,共同选取有效参数对溶解氧含量进行预测。

由图 5 可明显看出,pH 值和氨氮含量与溶解氧含量的相关系数分别是 0.78 和 0.63,对溶解氧含量影响最大,相关性最强;电导率、水温度与溶解氧含量相关系数分别是 0.42 和 0.40,属于中等相关。

在气象参数中,大气相对湿度、大气温度对溶解氧含量影响较大;大气压力、浊度、叶绿素含量、风向、风速、太阳辐射对溶解氧含量影响较小,降雨量

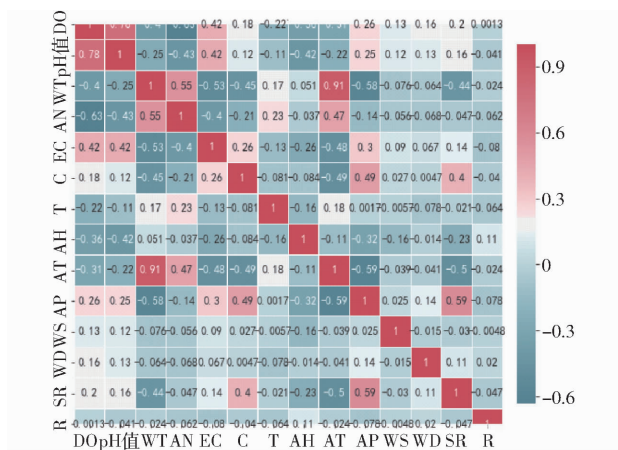


Fig. 5 Pearson feature selection

对溶解氧含量影响几乎可以忽略不计。因此,本文选择 pH 值、氨氮含量、电导率、水温度、大气温度、大气相对湿度作为模型输入特征,这与水产养殖专家凭借经验所选取的对溶解氧含量有关键影响的因子基本一致,以此作为输入样本,预测 2 h 后的溶解氧含量。

2.3 寻优结果分析

P-IBAS-LSTM 模型结构由输入层、2 层隐藏层、输出层组成,其中模型采用 ReLU 激活函数,Adam 优化算法,模型框架在 Keras 框架下实现。为了减少人为因素对模型的影响,对模型超参数取值范围设置如下:迭代次数为 $[50, 200]$,批处理规模 $[1, 300]$,学习率为 $[0.0001, 0.01]$,神经元个数为 $[1, 100]$ 。模型最终寻优结果,迭代次数为 150,批处理为 128,学习率为 0.0023,第 1 层神经元个数为 52,第 2 层神经元个数为 16。

为了对比普通 BAS 算法和本文 IBAS 算法的收敛效果,在迭代次数为 150 时,BAS-LSTM 与本文 P-IBAS-LSTM 适应度随迭代次数变化情况如图 6 所示,由图 6 可以看出,本文改进后的天牛须算法在大概 45 次时已达到收敛,普通天牛须算法大概在 72 次时收敛,并且 IBAS 能够更好地找到最优值,说

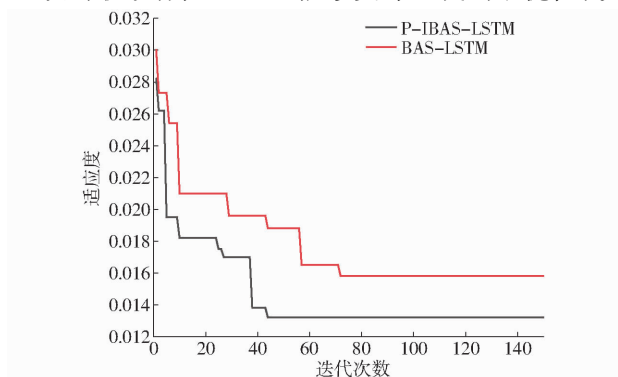


Fig. 6 Changing curves of fitness function

明本文算法能够更好地进行局部搜索。

2.4 模型对比

通过对数据进行预处理后,共有 17 663 条数据,对数据按 9:1 划分为训练集和测试集,训练集有 15 895 条数据,测试集有 1 766 条数据。每 10 min 采集一次数据,需要预测 2 h 后溶解氧含量,因此,时间步长设为 12。

为了验证本文所提 P-IBAS-LSTM 模型的预测性能,本文选取经过特征选择后的随机森林(P-RF)、反向传播神经网络(P-BP)、循环神经网络(P-RNN)、单一的 LSTM 模型(P-LSTM)、普通天牛须优化的 LSTM 模型(P-BAS-LSTM),以及未经特征选择的 BAS-LSTM 和 IBAS-LSTM 这 7 种模型与本文模型(P-IBAS-LSTM)进行预测比较,结果见图 7。通过与其他 7 种模型对比发现,本文模型预测曲线与溶解氧真实值曲线更加贴合,其中从图 7e 与图 7f,图 7g 与图 7h 对比发现特征选择对溶解氧预测有重要作用。通过图 7e 与图 7g,图 7f 与图 7h 对比,预测性能均有明显提升,可体现出 IBAS 的有效性。结果表明本文模型与其他模型相比具有更好的拟合效果和泛化能力。

本文模型 P-IBAS-LSTM 以及其他 7 个模型的性能对比如表 3 所示,与 P-BP、P-RF、P-RNN、P-LSTM、BAS-LSTM、P-BAS-LSTM、IBAS-LSTM 相比,本文模型 MSE 分别下降了 40.4%、98.3%、16.6%、12.7%、11%、10.4%、5.7%;RMSE 分别下降了 22.8%、87.1%、8.7%、6.5%、5.5%、5.3%、2.9%;MAE 分别下降了 36.3%、90.2%、8.2%、8.8%、8.5%、3.5%、2.2%。基于上述分析,本文模型各个指标都表现出明显优势,表明所提出的 P-IBAS-LSTM 模型溶解氧预测精度更高,泛化能力更强,能够较理想地预测 2 h 后的水产养殖溶解氧含量,达到较为理想的预测效果。

为了进一步验证 P-IBAS-LSTM 预测效果,8 种模型的预测误差情况见图 8。从图 8 可以看出,与其他 7 种模型相比,本文模型 P-IBAS-LSTM 与参照曲线更加贴合,误差波动幅度整体上比其他模型更小,其中图 8h 与图 8f、8g 相比,分别体现出 IBAS、特征选择对模型预测性能有较明显的提升,表明本文模型预测值与实际值误差较小,预测效果更稳定。

2.5 讨论

研究表明,提出的模型 P-IBAS-LSTM 在池塘养殖溶解氧预测中表现出良好的性能。为了提高水产养殖溶解氧预测模型准确度,可以从以下几方面进行研究:①溶解氧含量在每个季节都

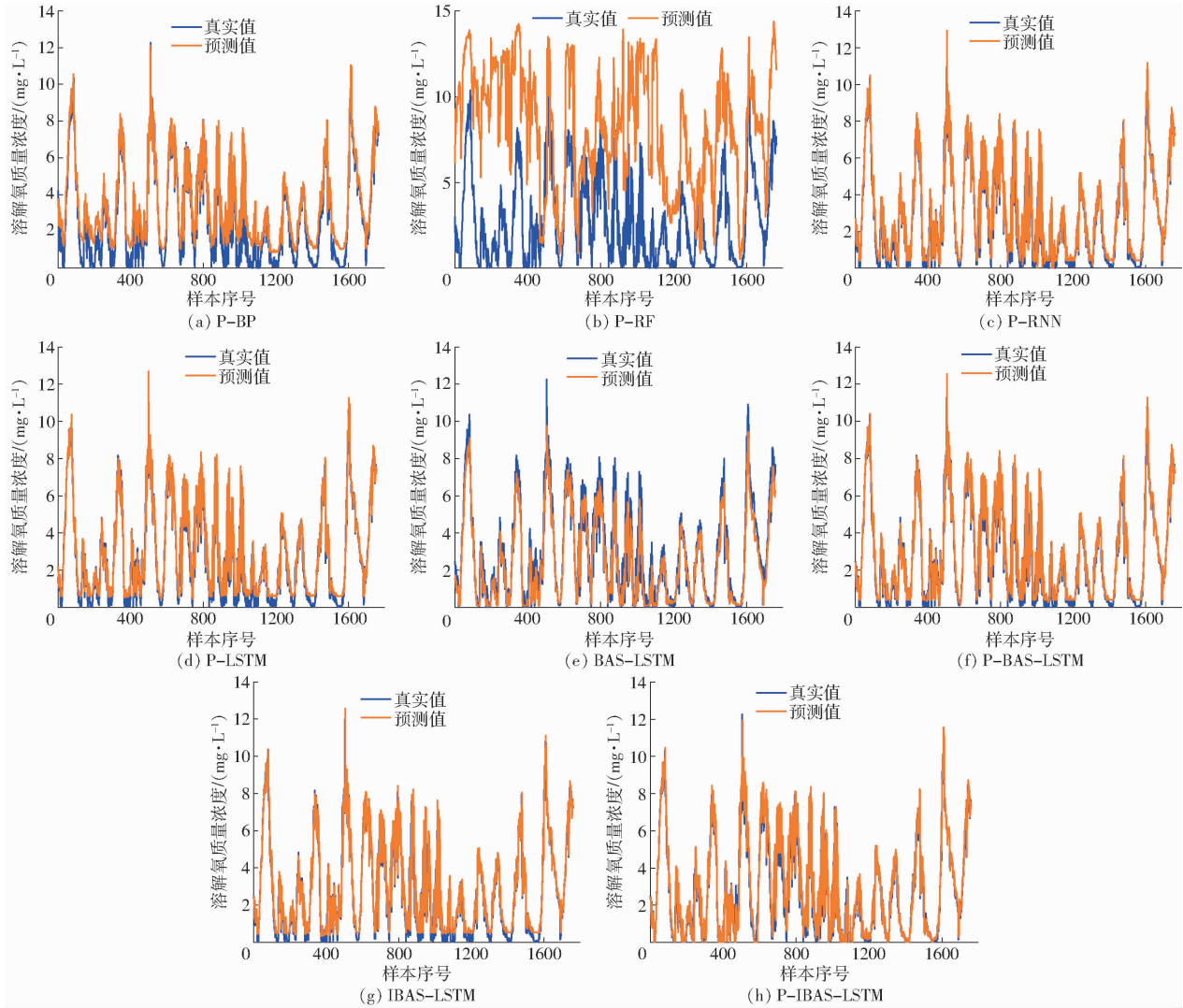


图7 8种模型预测结果

Fig.7 Prediction results of eight models

表3 8种模型预测性能比较

Tab.3 Comparison of prediction results of eight models

模型	MSE/ ($\text{mg}^2 \cdot \text{L}^{-2}$)	RMSE/ ($\text{mg} \cdot \text{L}^{-1}$)	MAE/ ($\text{mg} \cdot \text{L}^{-1}$)
P-BP	1.081 6	1.040 0	0.833 6
P-RF	38.672 6	6.218 7	5.393 8
P-RNN	0.772 8	0.879 1	0.578 1
P-LSTM	0.737 5	0.858 8	0.581 9
BAS-LSTM	0.723 9	0.849 2	0.579 8
P-BAS-LSTM	0.718 9	0.847 3	0.549 9
IBAS-LSTM	0.683 6	0.826 8	0.542 8
P-IBAS-LSTM	0.644 2	0.802 6	0.530 6

有不同的变化趋势,另外,溶解氧含量除了受水质和气象因素影响,还会受到池塘土壤的影响,比如土壤水分和土壤温度等,为了更加客观分析溶解氧在不同季节的变化,就需要采集不同季节下尽可能完整的数据。②目前对溶解氧含量预测的研究大多处于时间维度的研究,但是在池塘中不同位置,溶解氧含

量之间也有一定的相关性,接下来可以从三维时空方面考虑,可以进一步提高模型预测准确度。③仿生算法在优化模型方面具有明显优势,比如麻雀搜索算法(Sparrow search algorithm, SSA)^[33],并已取得广泛应用,未来可以探索SSA在池塘养殖溶解氧预测模型优化的可行性。

3 结论

(1)使用Pearson相关系数筛选出对溶解氧含量有关键性影响的特征,消除变量之间的冗余性,实现降低模型输入维度,有利于提高模型计算速率及预测精度。

(2)提出衰减因子指数递减策略改进天牛须搜索算法,将衰减因子 γ 与迭代次数相联系并呈指数函数递减,通过与普通天牛须算法相对比,适应度收敛更快,可看出IBAS在全局搜索和局部搜索中能够达到平衡,并提高算法的收敛速度。

(3)提出P-IBAS-LSTM非线性溶解氧预测模

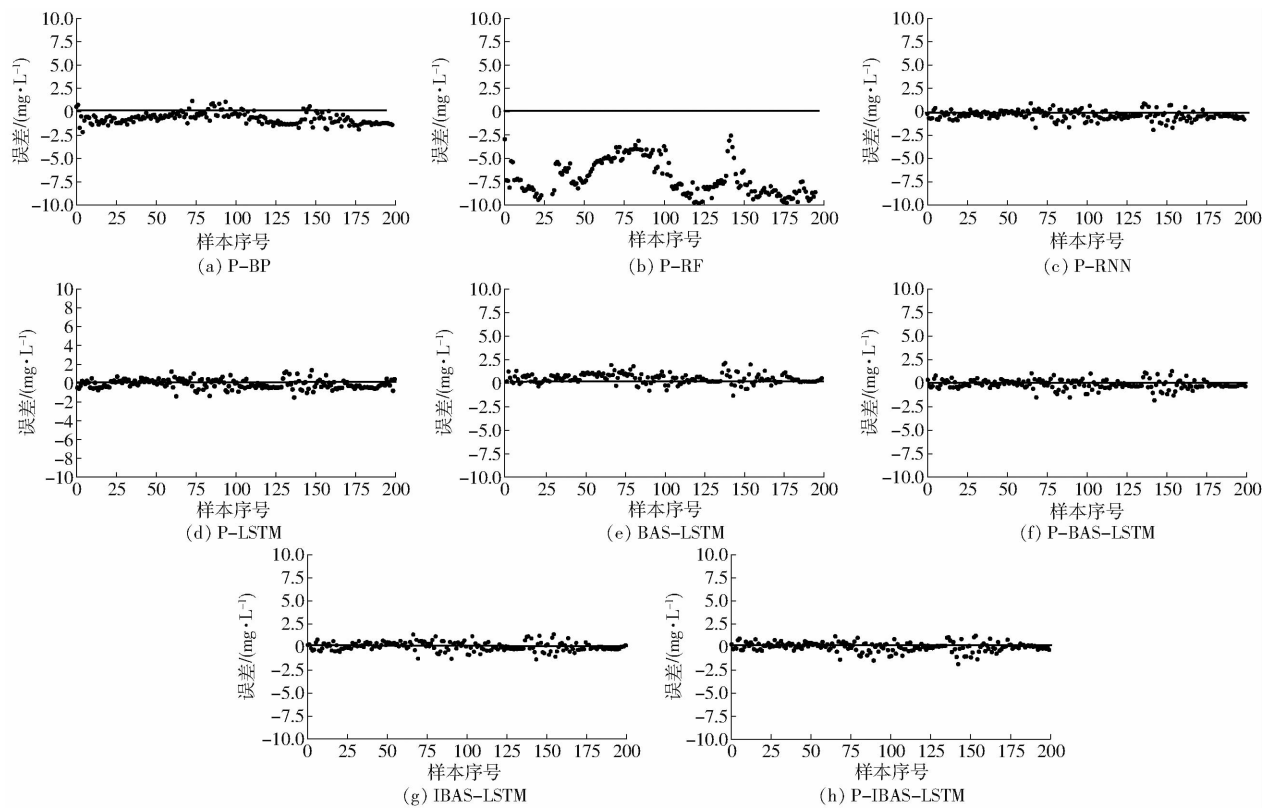


图 8 8 种模型误差曲线
Fig.8 Error curves of eight models

型,利用 IBAS 优化 LSTM 网络得最优超参数组合,克服了人工经验的盲目性和不准确性,能够较为准确地预测池塘养殖 2 h 后的溶解氧含量,可以较好地拟合池塘养殖中溶解氧与水质因子和气象因素之间的非线性关系。

(4) 提出的 P-IBAS-LSTM 模型,性能评价指标均方误差(MSE)、均方根误差(RMSE)、平均绝对误差(MAE)分别为 0.644 2 mg²/L²、0.802 6 mg/L、0.530 6 mg/L,对比其他几种模型,在预测精度和鲁棒性方面均有较大提升。

参 考 文 献

[1] XIAO R C, WEI Y G, AN D, et al. A review on the research status and development trend of equipment in water treatment processes of recirculating aquaculture systems[J]. Reviews in Aquaculture, 2019, 11(3): 863-895.

[2] 吴静,李振波,朱玲,等.融合 ARIMA 模型和 GAWNN 的溶解氧含量预测方法[J].农业机械学报,2017,48(增刊):205-210,204.

WU Jing, LI Zhenbo, ZHU Ling, et al. Hybrid model of ARIMA model and GAWNN for dissolved oxygen content prediction [J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(Supp.): 205-210, 204. (in Chinese)

[3] CAO X K, LIU Y R, WANG J P, et al. Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network[J]. Aquacultural Engineering, 2020, 91(2): 102122.

[4] LIU S Y, XU L Q, JIANG Y, et al. A hybrid WA-CPSO-LSSVR model for dissolved oxygen content prediction in crab culture[J]. Eng. Appl. Artif. Intell, 2014, 29: 114-124.

[5] LIU S Y, XU L Q, LI D L, et al. Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization[J]. Elsevier B. V., 2013, 95: 82-91.

[6] HUAN J, LI H, LI M B, et al. Prediction of dissolved oxygen in aquaculture based on gradient boosting decision tree and long short-term memory network; a study of Changzhou Fishery Demonstration Base, China [J]. Computers and Electronics in Agriculture, 2020, 175: 105530.

[7] JEBLI I, BELOUADHA F, KABBAJ M I, et al. Prediction of solar energy guided by Pearson correlation using machine learning [J]. Energy, 2021, 224(11): 120109.

[8] 徐岩,程姝,薛艳静.基于故障暂态电流 Pearson 相关系数的直流配电网保护[J].华北电力大学学报(自然科学版), 2021, 48(4): 11-19.

XU Yan, CHENG Shu, XUE Yanjing, et al. Protection of DC distribution network based on Pearson correlation coefficient of fault transient current[J]. Journal of North China Electric Power University (Natural Science Edition), 2021, 48(4): 11-19. (in Chinese)

[9] 施珮,匡亮,袁永明,等.基于改进极限学习机的水体溶解氧预测方法[J].农业工程学报,2020,36(19):225-232.

- SHI Pei, KUNAG Liang, YUAN Yongming, et al. Dissolved oxygen prediction for water quality of aquaculture using improved ELM network[J]. Transactions of the CSAE, 2020, 36(19): 225 – 232. (in Chinese)
- [10] 朱成云, 刘星桥, 李慧, 等. 工厂化水产养殖溶解氧预测模型优化[J]. 农业机械学报, 2016, 47(1): 273 – 278.
ZHU Chengyun, LIU Xingqiao, LI Hui, et al. Optimization of prediction model of dissolved oxygen in industrial aquaculture [J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(1): 273 – 278. (in Chinese)
- [11] YANG T, QIU W, MA You, et al. Fuzzy model-based predictive control of dissolved oxygen inactivated sludge processes[J]. Neurocomputing, 2014, 136: 88 – 95.
- [12] SHI P, LI G H, YUAN Y M, et al. Prediction of dissolved oxygen content in aquaculture using clustering-based softplus extreme learning machine[J]. Computers and Electronics in Agriculture, 2019, 157: 329 – 338.
- [13] REN Q, WANG X Y, LI W S, et al. Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network[J]. Aquacultural Engineering, 2020, 90: 102085.
- [14] CAO W J, HUAN J, LIU C, et al. A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction[J]. Aquacultural Engineering, 2019, 84: 50 – 59.
- [15] LIU Y Q, ZHANG Q, SONG L H, et al. Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction[J]. Computers and Electronics in Agriculture, 2019, 165: 104964.
- [16] 樊宇星, 任妮, 田港陆, 等. 基于 DeepAR – RELM 的池塘溶解氧时空预测方法研究[J]. 农业机械学报, 2020, 51(增刊 1): 405 – 412.
FAN Yuxing, REN Ni, TIAN Ganglu, et al. Spatio-temporal prediction method of dissolved oxygen in ponds based on DeepAR – RELM[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(Supp. 1): 405 – 412. (in Chinese)
- [17] 陈英义, 方晓敏, 梅思远, 等. 基于 WT – CNN – LSTM 的溶解氧含量预测模型[J]. 农业机械学报, 2020, 51(10): 284 – 291.
CHEN Yingyi, FANG Xiaomin, MEI Siyuan, et al. Dissolved oxygen prediction model based on WT – CNN – LSTM [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(10): 284 – 291. (in Chinese)
- [18] 宋刚, 张云峰, 包芳勋, 等. 基于粒子群优化 LSTM 的股票预测模型[J]. 北京航空航天大学学报, 2019, 45(12): 1 – 11.
SONG Gang, ZHANG Yunfeng, BAO Fangxun, et al. Stock prediction model based on particle swarm optimization LSTM[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(12): 1 – 11. (in Chinese)
- [19] 滕金保, 孔韦韦, 田乔鑫, 等. 基于 LSTM – Attention 与 CNN 混合模型的文本分类方法[J]. 计算机工程与应用, 2021, 57(14): 126 – 133.
TENG Jinbao, KONG Weiwei, TIAN Qiaoxin, et al. Text classification method based on LSTM – attention and CNN hybrid model[J]. Computer Engineering and Applications, 2021, 57(14): 126 – 133. (in Chinese)
- [20] 陈英义, 程倩倩, 方晓敏, 等. 主成分分析和长短时记忆神经网络预测水产养殖水体溶解氧[J]. 农业工程学报, 2018, 34(17): 183 – 191.
CHEN Yingyi, CHENG Qianqian, FANG Xiaomin, et al. Principal component analysis and long short-term memory neural network for predicting dissolved oxygen in water for aquaculture[J]. Transactions of the CSAE, 2018, 34(17): 183 – 191. (in Chinese)
- [21] BARZEGAR R, AALAMI M T, ADAMOWSKI J. Short-term water quality variable prediction using a hybrid CNN – LSTM deep learning model[J]. Stochastic Environmental Research and Risk Assessment, 2020, 34(8): 1 – 19.
- [22] JIANG X Y, LI S. BAS: beetle antennae search algorithm for optimization problems[J]. International Journal of Robotics and Control, 2018, 1(1): 1.
- [23] HUANG Q Y, XIE L F, YIN G F, et al. Acoustic signal analysis for detecting defects inside an arc magnet using a combination of variational mode decomposition and beetle antennae search[J]. ISA Transactions, 2020, 102: 347 – 364.
- [24] KATSIKIS V N, MOURTAS S D, STANIMIROVIC P S, et al. Time-varying minimum-cost portfolio insurance under transaction costs problem via beetle antennae search algorithm (BAS) [J]. Applied Mathematics and Computation, 2020, 385: 114 – 124.
- [25] YANG J H, PENG Z R. Beetle-swarm evolution competitive algorithm for bridge sensor optimal placement in SHM[J]. IEEE Sensors, 2020, 20: 8244 – 8255.
- [26] WANG Z P, QU J F, FANG X Y, et al. Prediction of early stabilization time of electrolytic capacitor based on ARIMA – Bi_LSTM hybrid model[J]. Neurocomputing, 2020, 403: 63 – 79.
- [27] FEI S W, HE C X. Prediction of dissolved gases content in power transformer oil using BASA-based mixed kernel RVR model [J]. International Journal of Green Energy, 2019, 16(8): 1 – 5.
- [28] WU Q, SHEN X D, JIN Y Z, et al. Intelligent beetle antennae search for UAV sensing and avoidance of obstacles[J]. Sensors, 2019, 19(8): 1758.
- [29] XU T T, PENG Z, WU L F. A novel data-driven method for predicting the circulating capacity of lithium-ion battery under random variable current[J]. Energy, 2021, 218(C): 119530.
- [30] XU H H, DENG Y. Dependent evidence combination based on shearman coefficient and Pearson coefficient[J]. IEEE Access, 2018, 6: 11634e40.
- [31] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735 – 1780.
- [32] WU D, JIANG Z K, XIE X F, et al. LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT[J]. IEEE Transactions on Industrial Informatics, 2019, 16(8): 5244 – 5253.
- [33] XUE J K, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science & Control Engineering, 2020, 8(1): 22 – 34.