

李敬民, 陈 斯, 唐海晨, 等. 基于 P-ISSA-GRU 模型的养殖水体溶解氧含量预测[J]. 江苏农业学报, 2025, 41(9): 1781-1790.  
doi: 10.3969/j.issn.1000-4440.2025.09.013

## 基于 P-ISSA-GRU 模型的养殖水体溶解氧含量预测

李敬民, 陈 斯, 唐海晨, 杨增汪  
(江苏师范大学物理与电子工程学院, 江苏 徐州 221116)

**摘要:** 为了解决养殖水体溶解氧(DO)含量预测精度低的难题,本研究提出了一种基于改进的麻雀搜索算法(ISSA)优化门控循环单元(GRU)的养殖水体溶解氧含量预测模型(P-ISSA-GRU)。通过皮尔逊(Pearson)相关系数法确定水质中各种因子与溶解氧含量的相关系数,选取强关联因子为模型输入特征;通过引入 Tent 混沌映射改进种群初始化,自适应动态权重因子 $\omega$ 动态改变权重系数以及高斯扰动(GP)改进最优位置更新,增强了麻雀搜索算法(SSA)在寻找全局最优解和局部最优解的能力,加快了其收敛速度;通过 ISSA 优化 GRU 网络,进行模型参数的优化搜索,构建了非线性溶解氧含量预测模型(P-ISSA-GRU)。试验结果表明,P-ISSA-GRU 模型与其他 5 个常用的模型相比显示出更好的预测效果,均方误差(MSE)为 0.152 (mg/L)<sup>2</sup>、平均绝对误差(MAE)为 0.311 mg/L、均方根误差(RMSE)为 0.390 mg/L、决定系数( $R^2$ )为 0.984。因此,本研究建立的 P-ISSA-GRU 模型与传统模型相比在一定程度上提高了对养殖水体溶解氧含量的预测性能。

**关键词:** 溶解氧含量预测; 皮尔逊相关系数; 改进的麻雀搜索算法(ISSA); 门控循环单元(GRU)

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 1000-4440(2025)09-1781-10

## Prediction of dissolved oxygen content in aquaculture water based on P-ISSA-GRU model

LI Jingmin, CHEN Si, TANG Haichen, YANG Zengwang  
(School of Physics and Electronic Engineering, Jiangsu Normal University, Xuzhou 221116, China)

**Abstract:** In order to solve the problem of low prediction accuracy of dissolved oxygen (DO) content in aquaculture water, this study proposed a gated recurrent unit (GRU) prediction model based on an improved sparrow search algorithm (ISSA). The correlation coefficients of various factors in water and dissolved oxygen content were determined by Pearson correlation coefficient method, and the strong correlation factors were selected as the input features of the model. By introducing Tent chaotic mapping to improve population initialization, adaptive dynamic weight factor  $\omega$  to dynamically change the weight coefficient, Gaussian perturbation (GP) to improve the optimal location updating, the ability of sparrow search algorithm (SSA) to find global and local optimal solutions was enhanced, and its convergence speed was increased. A nonlinear dissolved oxygen content prediction model (P-ISSA-GRU) was constructed by optimizing the GRU network using ISSA for parameter search. The experimental results showed that the P-ISSA-GRU model exhibited superior prediction performance compared to five other commonly used models. The mean square error (MSE) was 0.152 (mg/L)<sup>2</sup>, the mean absolute error (MAE) was 0.311 mg/L, the root mean square error (RMSE) was 0.390 mg/L, and the coefficient of determination ( $R^2$ ) was 0.984. Therefore, compared to traditional models, the P-ISSA-GRU model developed in this study demonstrates improved predictive performance for dissolved oxygen content in aquaculture water.

收稿日期: 2024-11-26

基金项目: 国家自然科学基金项目(61975070); 徐州市重点研发计划项目(KC21087)

作者简介: 李敬民(1997-), 男, 江苏邳州人, 硕士研究生, 主要研究方向为机器学习、智能水产养殖。(E-mail) 2694760723@qq.com

通讯作者: 陈 斯, (E-mail) chensism@126.com

**Key words:** prediction of dissolved oxygen content; Pearson correlation coefficient; improved sparrow search algorithm (ISSA); gated recurrent unit (GRU)

溶解氧是衡量池塘水体生态参数变化的重要指标,也是影响养殖池塘环境的关键因子<sup>[1-2]</sup>。溶解氧含量对水中生物的生长发育、新陈代谢和摄食量有显著影响,无论是含量过高还是过低都可能导致生物死亡<sup>[3]</sup>。因此,开展养殖池塘溶解氧含量监测,加强溶解氧含量预测模型研究<sup>[4]</sup>,及时掌握水体中溶解氧含量的动态变化,对提升水质预测精度、增加水产养殖效益具有重要意义。

池塘水体中的溶氧量分布规律独特,与昼夜变化、季节变化等因素密切相关,而且这些影响因素之间存在复杂的耦合关系<sup>[5]</sup>。为提高运算速度与精度,需明确对溶氧量起关键作用的因子<sup>[6]</sup>。在自然科学领域,Pearson 相关系数分析法常被用于度量两个变量间的相关程度,进而提高预测模型的准确性<sup>[7-8]</sup>。

近年来,许多研究人员采用深度学习法对水体溶解氧含量进行预测,并获得一些成果<sup>[9-10]</sup>。施珮等<sup>[11]</sup>利用局部回归网络构建了溶解氧含量的预测模型。刘双印等<sup>[12]</sup>提出了蚁群算法(ACA)优化最小二乘支持向量回归机(LSSVR)的河蟹养殖水体溶解氧含量预测方法,取得较好的预测效果。任妮等<sup>[13]</sup>提出了一种基于粒子群优化算法(PSO)和长短时记忆神经网络(LSTM)的蟹塘溶解氧质量浓度预测模型,采用 PSO 算法优化 LSTM 模型参数后对蟹塘溶解氧质量浓度进行预测,提高了预测的准确性和稳定性。还有研究人员将 K 均值(K-means)聚类算法和极限学习机相结合进行水体溶氧量预测模型的研究,发现改进后的模型有效减少了各样点间的相互干扰,提高了对水体溶解氧含量预测的精确性<sup>[14]</sup>。虽然上述方法能够对水体溶解氧含量进行预测,但在时间序列上未能充分捕捉特征之间的动态相关性。一些研究人员将 LSTM 模型应用于水体溶解氧含量的预测<sup>[9-10,15]</sup>,取得了不错的结果。将主成分分析法和 LSTM 相结合应用到溶氧量的预测中,试验结果表明,LSTM 模型在处理时序数据时表现出色,但存在计算复杂度高、训练速度慢、内存占用大以及调整超参数难度高等问题<sup>[5,13]</sup>。相比之下,门控循环单元(GRU)作为 LSTM 模型的一种简化变体,结构更为简洁,训练速度更快,且在调试方面更便捷。GRU 模型的预测效果与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。麻雀搜索算法(SSA)是一种群体智能优化算法,具有收敛速度快,全局搜索能力强等优点。利用麻雀搜索算法对 GRU 模型的超参数进行寻优,可以迅速找到最优参数组合,减少人工手动调整参数的复杂性和不确定性,在一定程度上提高了模型的预测精度。

本研究提出了一种基于皮尔逊相关系数分析(P)与改进麻雀搜索算法(ISSA)相结合的方法,用于优化 GRU 预测模型。皮尔逊分析法可以帮助筛选影响溶解氧含量的主要因素,简化模型结构。然后,利用改进的麻雀搜索优化算法对门控循环单元(GRU)的关键参数进行优化,从而得到最优解。最后,建立 P-ISSA-GRU 组合预测模型。采用徐州市龙辰水产养殖专业合作社养殖池塘采集的水质数据和气象数据为样本进行模型训练,以期解决目前现有养殖水体溶解氧预测模型鲁棒性差、信息冗余和预测精度低的缺点。

## 1 材料与方法

### 1.1 数据获取

本研究的数据包括水质参数和气象数据两部分。其中水质参数数据来源于徐州市龙辰水产养殖专业合作社 2 号养殖池塘,该池塘面积约  $1.2 \times 10^4 \text{ m}^2$ ,主要养殖河蟹,池塘内配备了溶解氧含量(DO)、pH、水温(WT)、氨态氮含量(AN)和电导率(EC)传感器,用于实时监测水质变化。气象数据则来自临近养殖池塘的气象站,通过该气象站实现对大气环境数据的在线监测,在线监测系统结构如图 1 所示,采集数据类型如表 1 所示。

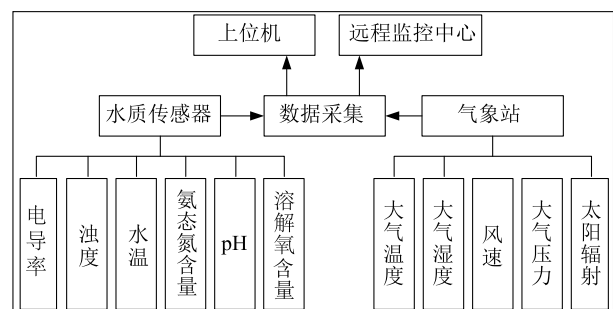


图 1 在线监测系统结构图

Fig.1 Structural diagram of online monitoring system

表 1 本研究采集的数据类型

Table 1 Types of data collected in this study

类型	参数	采样周期(min)
水质数据	溶解氧含量(mg/L)	30
	水温(°C)	
	pH	
	氨态氮含量(mg/L)	
	电导率(mS/cm)	
	浊度(NTU)	
气象数据	大气温度(°C)	10
	大气湿度(%)	
	风速(m/s)	
	大气压力(kPa)	
	太阳辐射(W/m <sup>2</sup> )	

表 2 2023 年 8 月 20 日采集的部分原始数据

Table 2 Partial raw data collected on August 20, 2023

采集时刻	溶解氧含量(mg/L)	pH	水温(°C)	氨态氮含量(mg/L)	电导率(mS/cm)	浊度(NTU)	相对湿度(%)	气温(°C)	大气压(kPa)	风速(m/s)	太阳辐射(W/m <sup>2</sup> )
00:10:08	8.02	7.02	19.20	0.5	347	21.8	72.4	35.60	100.35	1.3	46
00:40:08	8.47	7.05	18.93	0.5	347	21.6	72.8	35.80	100.24	1.3	18
01:10:08	7.35	7.08	18.74	0.6	347	21.3	72.2	35.60	100.26	1.5	24
01:40:08	7.28	7.05	18.88	0.6	348	20.8	72.1	35.70	100.28	1.5	36
17:00:12	6.72	7.17	21.42	0.5	343	19.7	89.5	28.40	99.70	0.8	9
17:30:12	7.33	7.17	21.92	0.5	343	19.4	88.0	28.30	99.75	0.6	16
18:00:12	7.52	7.22	22.04	0.5	343	20.2	87.6	28.50	99.77	0.6	25

电导率和太阳辐射为整数采集值。

对于异常数据采用均值平滑法进行处理。均值平滑法的公式如下:

$$X_n = \frac{X_{n-1} + X_{n+1}}{2} \quad (2)$$

式中,  $X_n$  表示均值平滑法求得的第  $n$  时刻的数据值,  $X_{n-1}$ 、 $X_{n+1}$  分别表示第  $n-1$ 、第  $n+1$  时刻数据。

为了消除各样本间数据差异性带来的不利影响,加快优化算法的求解,将原始数据归一化,映射到  $[0, 1]$  范围内。归一化过程如下:

$$X = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (3)$$

式中  $X_{\max}$  是样本变量  $X$  的最大值,  $X_{\min}$  是样本变量  $X$  的最小值。

### 1.3 模型的构建

1.3.1 皮尔逊(Pearson)相关系数 Pearson 相关系数( $\rho$ )用于衡量两个变量之间线性关系的紧密程

实时在线采集水质和气象数据,数据收集时间为 2023 年 8 月 6 日至 2023 年 9 月 8 日,水质数据每 30 min 采集 1 次,气象数据每 10 min 采集 1 次,因为数据采集的时间不同步,故将气象数据转换为 30 min 的平均值,累计共 1 603 条数据。2023 年 8 月 20 日采集的部分样本数据如表 2 所示。

### 1.2 数据预处理

现实中由于传感器老化、线路故障、网络延迟和恶劣天气等因素影响,容易出现数据丢失和不规律的情况。本研究利用线性插值和均值平滑修复池塘水质环境数据。线性插值法公式为:

$$X_{k+i} = X_k + \frac{i(X_{k+j} - X_k)}{j} \quad (0 < i < j) \quad (1)$$

式中  $X_{k+i}$  表示第  $k+i$  时刻所缺失的数据;  $X_k$ 、 $X_{k+j}$  表示第  $k$  和第  $k+j$  时刻数据;  $i$ 、 $j$  表示时刻。

度,  $\rho$  值的绝对值越大,表示两个变量间相关性越强<sup>[16]</sup>。应用 Pearson 相关系数后,会生成一个相关矩阵,用于剔除低相关或无关的变量,并筛选出关键特征。Pearson 相关系数( $\rho$ )计算公式<sup>[17]</sup>如下:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (4)$$

式中,  $\rho_{X,Y}$  表示变量  $X$  和  $Y$  之间的 Pearson 相关系数,  $\text{cov}(X,Y)$  表示  $X$ 、 $Y$  的协方差。  $\sigma_X$  和  $\sigma_Y$  分别表示  $X$  和  $Y$  的标准差。

当  $\rho > 0$  时,两个变量呈正相关;当  $\rho = 0$  时,表明它们之间没有线性相关性;当  $\rho < 0$  时,表示两个变量间存在负相关关系<sup>[17]</sup>。  $\rho$  值为 0.81~1.00 属于极强相关;  $\rho$  值为 0.61~0.80 属于强相关;  $\rho$  值为 0.41~0.60 属于中等相关;  $\rho$  值为 0.21~0.40 属于弱相关性;  $\rho$  值为 0.01~0.20 属于极弱相关。

本研究采用 Pearson 相关系数法减少参数的维

度,从中选择最具预测性的有效变量用于溶解氧含量的预测。从表 3 可见,水温、pH 和浊度与溶解氧含量的相关系数分别为 0.85、0.74 和 0.77,这表明它们与溶解氧含量之间的关联度较高,对溶解氧含

量的影响也更显著,这一结果与曾一川等<sup>[18]</sup>的研究结果相一致。而电导率、氨态氮含量与溶解氧含量的相关系数分别为 0.42 和 0.69,表明它们与溶解氧含量呈中等程度的相关性。

表 3 采用 Pearson 相关系数法对各参数进行选择

Table 3 Parameter selection using Pearson correlation coefficient method

参数	DO	pH	WT	AN	EC	T	AT	AH	WS	AP	SR
DO	1.00										
pH	0.74	1.00									
WT	0.85	-0.25	1.00								
AN	0.69	-0.30	0.55	1.00							
EC	0.42	0.31	-0.28	-0.10	1.00						
T	0.77	-0.11	0.17	0.29	-0.08	1.00					
AT	0.31	-0.22	0.42	0.51	-0.30	0.18	1.00				
AH	0.36	-0.30	0.05	-0.04	-0.26	-0.16	-0.11	1.00			
WS	0.13	-0.08	-0.06	-0.06	0.03	0.01	-0.04	-0.16	1.00		
AP	0.26	0.25	-0.14	-0.15	0.40	0.01	-0.30	-0.30	0.03	1.00	
SR	0.20	0.16	-0.44	-0.03	0.11	-0.02	-0.50	-0.22	0.11	-0.05	1.00

DO:溶解氧含量,pH:酸碱度,WT:水温,AN:氨态氮含量,EC:电导率,T:浊度,AT:大气温度,AH:大气湿度,WS:风速,AP:大气压力,SR:太阳辐射。

基于上述的 Pearson 相关性特征选择,本研究选取水温度、浊度、pH、电导率和氨态氮作为模型输入特征,预测 2 h 后养殖水体的溶解氧含量。

1.3.2 门控循环单元 (GRU) 门控循环单元 (GRU) 是长短时记忆网络 (LSTM) 改动较大的变体形式<sup>[19]</sup>,该方法既保留了 LSTM 算法求解梯度消失和梯度暴发的性能,又比 LSTM 算法在时间收敛、参数更新等方面具有更好的性能。GRU 的内部单元结构与 LSTM 有所不同,LSTM 有 3 个门结构,但 GRU 只有 2 个门结构<sup>[20]</sup>。GRU 的 2 个门分别为更新门和重置门,更新门用于对前一时间状态信息的保留,重置门用于将新的输入信息与前一时刻的记忆信息相结合。GRU 单元结构如图 2 所示,图中  $h_t$  和  $x_t$  为隐含层的输出与输入,  $r_t$  和  $z_t$  为复位门和更新门,  $g_t$  为时刻  $t$  的候选隐藏状态,  $h_{t-1}$  为前一次隐含层的输出。

GRU 神经网络的输入:

$$r_t = \sigma(W_r X_t + U_r h_{t-1} + b_r) \quad (5)$$

$$z_t = \sigma(U_z h_{t-1} + W_z X_t + b_z) \quad (6)$$

$$g_t = \tanh[U_c(r_t h_{t-1}) + W_c X_t + b_c] \quad (7)$$

$$h_t = (1 - z_t) g_t + z_t h_{t-1} \quad (8)$$

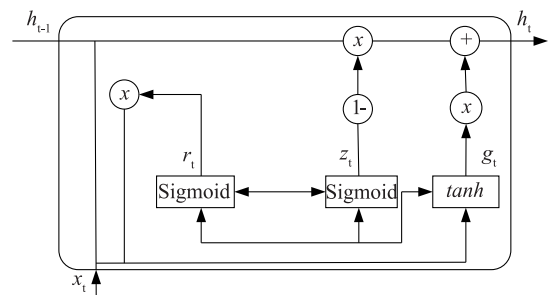


图 2 门控循环单元 (GRU) 结构

Fig.2 Structure of the gated recurrent unit (GRU)

以上公式中  $r_t$ 、 $z_t$ 、 $g_t$ 、 $h_t$  分别表示复位门、更新门、候选隐藏状态、当前隐藏状态,  $\sigma$  为 sigmoid 激活函数,  $\tanh$  为候选隐态激活函数,  $W_r$ 、 $U_r$  为复位门的权值,  $W_z$ 、 $U_z$  为更新门的权值,  $W_c$ 、 $U_c$  为候选隐藏状态的权值,  $b_r$ 、 $b_z$ 、 $b_c$  为偏置向量,  $X_t$  表示当前时间步的输入数据。

1.3.3 麻雀搜索算法 (SSA) 麻雀搜索算法 (SSA) 是一种模拟麻雀觅食行为的智能优化算法<sup>[21]</sup>,它依据适应度动态调整群体中麻雀个体的位置。在算法流程里,发现者角色凭借更大的搜索范围,承担起在种群中探索最优解的任务,通过主动拓展搜索空间,

为种群寻优提供更广泛的探索方向。其位置更新公式如下:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{\max}}\right) & R_2 < ST \\ X_{i,j}^t + Q \cdot L & R_2 \geq ST \end{cases} \quad (9)$$

公式(9)中的  $X_{i,j}^{t+1}$  表示第  $t+1$  代迭代中,第  $i$  只麻雀在第  $j$  维的位置。 $X_{i,j}^t$  表示第  $t$  代迭代时,第  $i$  只麻雀在第  $j$  维空间的位置。 $\alpha$  为一个在  $(0 \sim 1)$  区间的随机数, $R_2$  为警戒值, $ST$  为安全阈值, $iter_{\max}$  表示最大迭代次数。 $Q$  为正态分布的随机数, $L$  表示一个  $1 \times d$  的矩阵,该矩阵内每个元素均为 1。

为了提高适应度,跟随者会根据发现者的觅食行为更新自己的位置,具体更新公式如下:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{\text{worst}}^t - X_{i,j}^t}{i^2}\right) & i > \frac{n}{2} \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1}| \cdot A^+ \cdot L & i \leq \frac{n}{2} \end{cases} \quad (10)$$

公式(10)中  $X_{i,j}^{t+1}$  表示第  $t+1$  代中,第  $i$  只麻雀在第  $j$  维的位置。 $n$  表示种群中所有麻雀的个数, $X_p^{t+1}$  为训练迭代过程中表现最优的发现者所处的位置, $X_{\text{worst}}^t$  为训练迭代过程中表现欠缺的全局位置。 $X_{i,j}^t$  表示第  $t$  代时,第  $i$  只麻雀在第  $j$  维空间的位置。矩阵  $A$  中的每个元素是在 1 或者 -1 中间随机挑选的, $A^+$  表示矩阵  $A$  的伪逆,且  $A^+ = A^T(AA^T)^{-1}$ ,其中  $A^T$  是矩阵  $A$  的转置, $(AA^T)^{-1}$  表示先计算  $A$  与其转置的乘积再对结果求普通逆矩阵。

当种群中的麻雀在觅食的时候,会随机产生一些承担搜索食物和防止入侵任务的麻雀,它们的位置更新公式如下:

$$X_{i,j}^{t+1} = \begin{cases} X_{\text{best}}^t + \beta |X_{i,j}^t - X_{\text{worst}}^t| & f_i > f_g \\ X_{i,j}^t + K \cdot \left(\frac{|X_{i,j}^t - X_{\text{worst}}^t|}{(f_i - f_w) + \varepsilon}\right) & f_i \leq f_g \end{cases} \quad (11)$$

公式(11)中  $X_{i,j}^{t+1}$  表示第  $t+1$  代中,第  $i$  只麻雀在第  $j$  维的位置。 $X_{\text{best}}^t$  表示训练迭代过程中表现最好的全局位置, $\beta$  用于步长的控制, $K$  的取值是随机产生的,但是值在  $[-1, 1]$  之间。 $X_{i,j}^t$  表示第  $t$  代时,第  $i$  只麻雀在第  $j$  维空间的位置。 $X_{\text{worst}}^t$  为训练迭代过程中表现欠缺的全局位置。 $\varepsilon$  是一个非常小的固定值,用于防止当  $f_i = f_w$  时,分母为 0 的情况。 $f_i, f_g, f_w$  分别代表当前适应度、最佳适应度和最差适应度。

### 1.3.4 改进的麻雀搜索算法 (ISSA)

1.3.4.1 Tent 混沌映射改进种群初始化 在群体智能算法中,传统的种群初始化方式常利用确定性系统生成随机序列进行优化操作,这种方式容易导致群体初始分配不均衡。混沌映射具备随机性和遍历性等特性,能够有效克服这一问题。将混沌映射引入算法初始化过程,替代传统的随机参数生成方式,可显著增强麻雀搜索算法生成优质初始解的能力。而高质量的初始种群有助于提升算法的收敛速度与计算精度<sup>[22]</sup>。基于此,本研究在种群初始化阶段引入 Tent 混沌映射,旨在改进种群初始化效果,为后续优化过程奠定良好基础,其表达式为:

$$X_{n+1} = \begin{cases} \frac{X_n}{\alpha} & (0 \leq X_n < \alpha) \\ \frac{1-X_n}{1-\alpha} & (\alpha \leq X_n \leq 1) \end{cases} \quad (12)$$

公式(12)中  $n$  代表映射的迭代次数, $X_n$  代表第  $n$  次映射的函数值, $\alpha$  是 Tent 混沌映射的一个控制参数,取值范围在  $(0, 1)$ , $X_{n+1}$  为第  $n+1$  次映射后的函数值。本研究使用 Tent 混沌映射取代 SSA 中的伪随机数生成器,使得初始解的分布更加均衡,避免了算法陷入局部最优解,从而提升了整体优化性能。这种改进使得算法在求解全局最优化问题时更加高效。

1.3.4.2 引入自适应动态权重因子  $\omega$  在发现者的更新公式(9)中引入自适应动态权重因子  $\omega$ ,如公式(13)所示,可以根据迭代次数的不同动态改变权重系数,从而提高算法的初始搜索范围,提升全局搜索能力,有效避免其陷入局部最优解。后期可以通过调整搜索范围来增加搜索的灵活性,并在一定程度上确保算法最优解的准确性,发现者引入权重因子如公式(14)所示:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{\max}}\right) \cdot \omega & R_2 < ST \\ X_{i,j}^t + Q \cdot L \cdot \omega & R_2 \geq ST \end{cases} \quad (13)$$

$$\omega = \frac{e^{2(1-\frac{t}{iter_{\max}})} - e^{2(1+\frac{t}{iter_{\max}})}}{e^{2(1-\frac{t}{iter_{\max}})} + e^{2(1+\frac{t}{iter_{\max}})}} \quad (14)$$

公式(14)中  $\omega$  代表自适应动态权重因子, $t$  表示当前迭代次数, $iter_{\max}$  是算法预先设定的最大迭代次数,代表算法迭代过程的总步数上限, $e$  代表自然常数。

1.3.4.3 高斯扰动改进最优位置更新 为提升算法的搜索能力和位置更新过程中的随机性,本研究采用高斯变异优化策略,在每个维度上根据公式(15)进行变异操作:

$$X_{\text{Gbest}}^t = X_{\text{best}}^t + \text{Gaussian}(\mu, \sigma^2) \cdot X_{\text{best}}^t \quad (15)$$

公式(15)中,  $X_{\text{Gbest}}^t$  表示经过高斯变异操作后得到的新位置,  $X_{\text{best}}^t$  表示在群体搜索过程中当前迭代时刻  $t$  的全局最优位置,  $\text{Gaussian}(\mu, \sigma^2)$  表示高斯分布的随机数生成函数,其中  $\mu$  是均值,取值为 0,  $\sigma^2$  是方差,取值为 1。

为防止在优化迭代中出现停滞或局部最优解,在每次迭代过程中使用贪婪选择策略比较两个个体的适应度值,并将适应度好的个体位置进行更新,具体的个体位置更新方法如公式(16)所示:

$$X_{\text{best}} = \begin{cases} X_{\text{best}} & f(X_{i,j}^{t+1}) < f(X_{\text{best}}) \\ X_{\text{Gbest}} & f(X_{i,j}^{t+1}) \geq f(X_{\text{best}}) \end{cases} \quad (16)$$

公式(16)中  $X_{\text{best}}$  表示当前迭代过程中经过贪婪选择策略更新后得到的最优个体位置,  $X_{\text{Gbest}}$  表示经过高斯变异操作后得到的位置,  $X_{i,j}^{t+1}$  表示第  $i$  个个体在第  $j$  个维度上,经过 1 次迭代(从第  $t$  次迭代到第  $t+1$  次迭代)后更新得到的位置,  $f(\cdot)$  是适应度函数,用于计算输入位置(个体)对应的适应度值。

改进后的麻雀搜索算法步骤如下:

1. 确定种群规模,然后使用混沌映射对种群进行初始化配置。

2. 通过评估每个个体的适应度值,从中选择最佳适应度值、最差适应度值及其种群中麻雀的位置。

3. 从最佳适应度的麻雀个体中选择部分发现者,依据公式(15)对发现者的位置进行调整,并采用动态步长进行搜索。

4. 从剩余的麻雀中挑选出部分加入者,依据公式(10)对加入者的位置进行调整。

5. 从麻雀中选择部分个体作为警戒者,依据公式(11)对警戒者的位置进行调整。

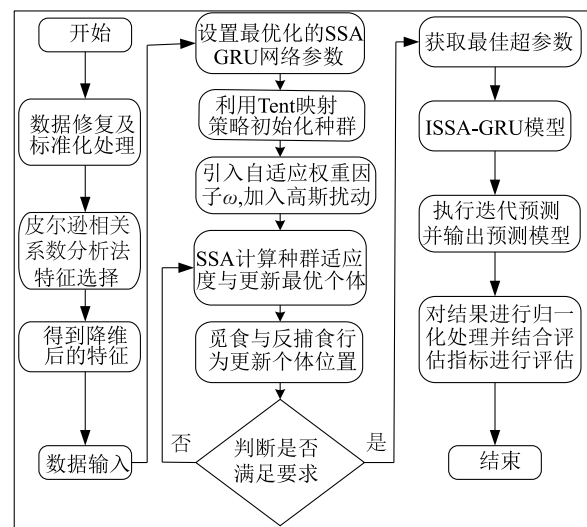
6. 利用高斯变异扰动策略,获得更新后的最佳位置解。

7. 依据公式(16)通过贪婪选择策略判断是否采用公式(9)中的最优解作为目标来更新位置。

8. 检查是否满足预先设定的条件,如果条件未达成,则返回步骤 2 重新开始搜索。

9. 算法执行完毕,获得最优可行解。

1.3.5 预测模型的构建 养殖水体中溶解氧含量是随着时间动态变化的,且易受到各种因素的影响,由于 GRU 模型运算速度快,而且模型中超参数对网络结构起关键作用,会直接影响预测结果的准确性。因此,本研究将改进的麻雀搜索算法(ISSA)用于优化 GRU 模型的超参数,构建 P-ISSA-GRU 溶解氧含量预测模型。基于对麻雀搜索算法的改进,研究设计了一个非线性溶解氧含量的预测模型,其流程图如图 3 所示。



GRU: 门控循环单元; SSA: 麻雀搜索算法。

图 3 P-ISSA-GRU 溶解氧含量预测模型流程图

Fig.3 Flowchart of the P-ISSA-GRU dissolved oxygen content prediction model

从图 3 可知, P-ISSA-GRU 模型流程主要有以下 5 个步骤:

1. 设计试验方案,并对采集的数据进行预处理,为了减少模型的输入维度,提高算法的计算效率,本研究使用 Pearson 相关系数对水质数据和气象数据进行特征提取,以确定与溶解氧含量相关性较强的几个因素作为最终模型的输入特征。

2. 采用改进的麻雀搜索算法寻找 GRU 模型的最佳超参数。

3. 对数据集进行归一化处理,并按照 7:3 的比例将数据集划分为训练集和测试集,随后对模型进行训练。

4. 将待测样本的数据导入预测模型中,进行溶

解氧含量预测。

5.通过模型评估指标对预测结果进行评估,并与其他模型的表现进行对比。

1.3.6 模型评价指标 本研究选择均方误差 ( $MSE$ )、平均绝对误差 ( $MAE$ )、均方根误差 ( $RMSE$ ) 以及决定系数 ( $R^2$ ) 作为模型性能的评估标准,分别与传统模型的表现进行对比分析。具体来说, $MSE$ 、 $MAE$ 、 $RMSE$  的值越小,表示模型预测结果与实际值之间的误差越小。 $R^2$  值越大,说明模型的预测结果与实际数据之间的拟合程度越高,预测精度也就越高。以下是各项评价指标的计算公式:

$$MSE = \frac{1}{n} \sum_{i=1}^n [y_i - g(x_i)]^2 \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - g(x_i)| \quad (18)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [y_i - g(x_i)]^2} \quad (19)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n [y_i - g(x_i)]^2}{\sum [y_i - \bar{y}]^2} \quad (20)$$

以上 4 个公式中, $n$  为样本数量, $y_i$  为养殖水体溶解氧含量的真实值, $g(x_i)$  表示模型对第  $i$  个样本输入  $x_i$  时的预测值, $[y_i - g(x_i)]$  表示第  $i$  个样本的预测误差,反映真实值和预测值之间的差异, $|y_i - g(x_i)|$  表示对第  $i$  个样本预测误差取绝对值, $\bar{y}$  为真实结果的平均值。

## 2 结果与分析

### 2.1 寻优结果

P-ISSA-GRU 模型的结构包括输入层、两个隐藏层和输出层,采用 ReLU 激活函数,并使用 Adam 优化算法。该模型在 MATLAB 2022 环境中实现。

麻雀搜索算法的初始参数:种群大小为 30,最大迭代次数为 50 次。通过该算法优化,确定 GRU 神经网络的理想参数为:学习率 0.05,最优迭代次数 26 次,第 1 层神经元个数 30 个,第 2 层神经元个数 28 个。为了解改进后算法的收敛效果,在模型迭代 50 次的时候,绘制 SSA-GRU 和 ISSA-GRU 的适应度随迭代次数变化的情况(图 4)。从图 4 可以看出,改进后的麻雀搜索算法(ISSA 大约在第 26 次迭代达到收敛)收敛速度比普通麻雀搜索算法(SSA 大约在第 38 次迭代才能收敛)收敛速度快。此外,

ISSA 在收敛过程中表现出更强的寻找最优解的能力,这表明本研究提出的 ISSA 算法在局部搜索方面具有更好的性能。

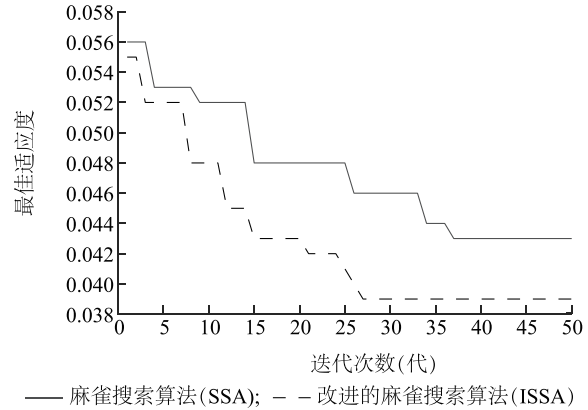


图 4 改进前后麻雀搜索算法适应度函数的变化曲线

Fig.4 Change curves of the fitness function of the sparrow search algorithm before and after improvement

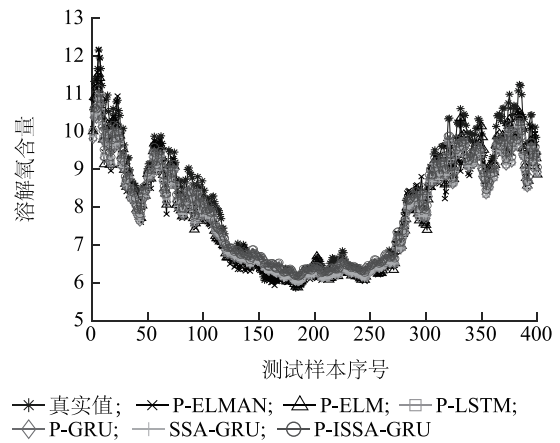
### 2.2 P-ISSA-GRU 模型预测性能

在对数据进行预处理后,最终得到 1 603 个数据,并按照 7 : 3 的比例将其划分为训练集和测试集,其中训练集 1 203 个数据用于训练模型,测试集 400 个数据用于测试模型的性能。每 30 min 采集 1 次数据,预测 2 h 后养殖水体溶解氧含量,故时间步长设置为 4。

为了评估本研究提出的 P-ISSA-GRU 模型对养殖水体溶解氧含量的预测性能,将 P-ISSA-GRU 模型与经过特征选择的局部回归网络模型(P-ELMAN)、极限学习机模型(P-ELM)、单一长短时记忆网络模型(P-LSTM)、单一门控循环单元模型(P-GRU)以及未经特征选择的 SSA-GRU 模型进行对比,预测结果如图 5 所示。通过与其他 5 种模型比较可以发现,P-ISSA-GRU 模型的预测值曲线与溶解氧含量的真实值曲线更加贴合,表现出更好的预测效果。

P-ISSA-GRU 模型与其他 5 个模型的性能对比如图 6 所示。从 SSA-GRU 模型和 P-ISSA-GRU 模型的对比结果可以看出,特征选择对溶解氧含量预测具有重要影响。P-ISSA-GRU 模型与 P-ELMAN 模型、P-ELM 模型、P-LSTM 模型相比预测性能均有不同程度提升;P-ISSA-GRU 模型与 P-GRU 模型相比也有明显的性能提升,进一步验证了 ISSA 算法的

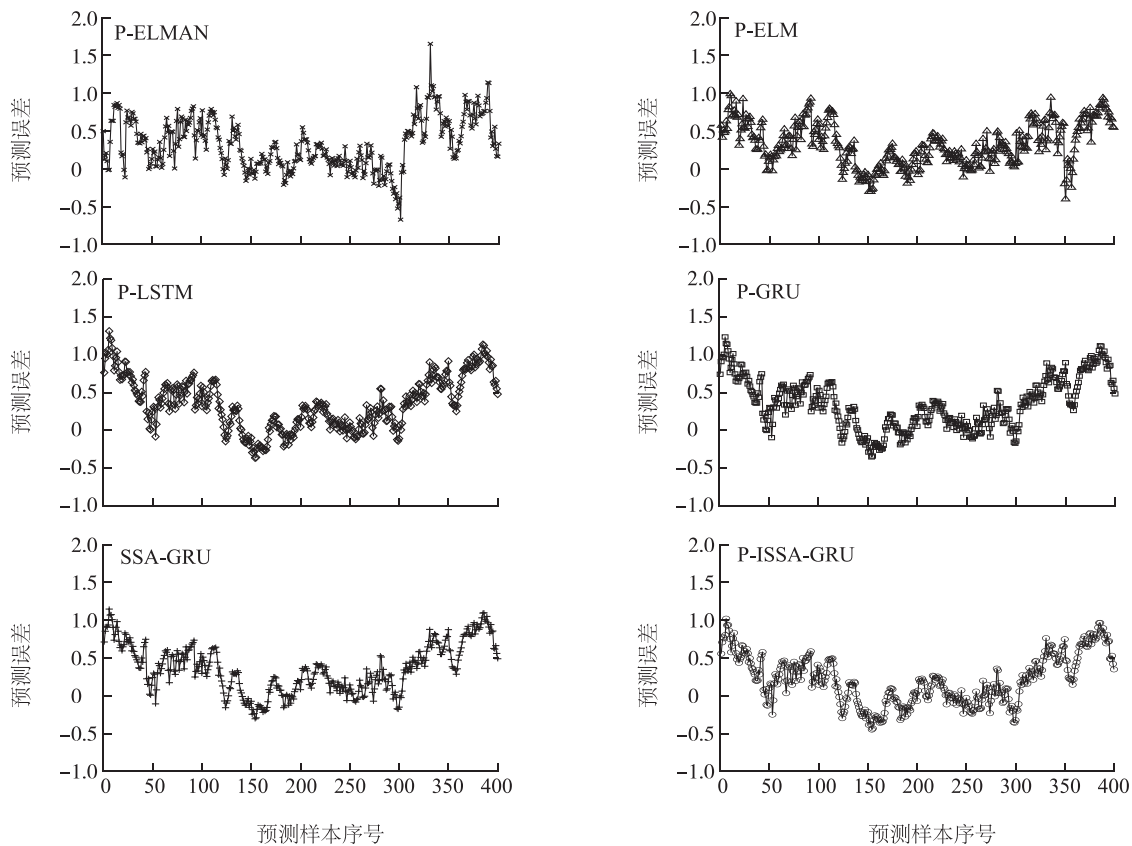
有效性。将 P-ISSA-GRU 模型预测误差曲线与 SSA-GRU 模型进行对比,可以看出 P-ISSA-GRU 模型预测误差曲线变化幅度较小,多在 0.2 附近。6 种模型预测误差的对比结果表明,P-ISSA-GRU 模型具有更好的拟合效果和泛化能力。6 种模型各类型误差值见表 4,与 P-ELMAN、P-ELM、P-LSTM、P-GRU、SSA-GRU 模型相比,P-ISSA-GRU 模型的 *MSE* 分别下降了 30.9%、19.6%、39.0%、36.4%、35.0%;*MAE* 分别下降了 15.5%、12.6%、22.4%、20.9%、20.9%;*RMSE* 分别下降了 16.8%、10.3%、21.8%、20.3%、19.4%;同时,P-ISSA-GRU 模型的决定系数( $R^2$ )达到 0.984。通过以上分析可见,P-ISSA-GRU 模型在各项指标上均表现出显著优势,表明该模型具有更高的溶解氧含量预测精度和更强的泛化能力,能够更加准确地预测 2 h 后的养殖水体溶解氧含量。



P-ELMAN:局部回归网络模型,P-ELM:极限学习机模型,P-LSTM:单一长短时记忆网络模型,P-GRU:单一门控循环单元模型,SSA-GRU:麻雀搜索算法-门控循环单元神经网络模型,P-ISSA-GRU:非线性溶解氧含量预测模型。

图 5 6 种模型的预测结果对比

Fig.5 Comparison of prediction results of six models



P-ELMAN:局部回归网络模型,P-ELM:极限学习机模型,P-LSTM:单一长短时记忆网络模型,P-GRU:单一门控循环单元模型,SSA-GRU:麻雀搜索算法-门控循环单元神经网络模型,P-ISSA-GRU:非线性溶解氧含量预测模型。

图 6 6 种模型预测误差的对比

Fig.6 Comparison of prediction errors of six models

表 4 6 种模型的预测性能

Table 4 Prediction performance of six models

模型	误差类型			$R^2$
	$MSE$ ( $\text{mg/L}$ ) <sup>2</sup>	$MAE$ ( $\text{mg/L}$ )	$RMSE$ ( $\text{mg/L}$ )	
P-ELMAN	0.220	0.368	0.469	0.965
P-ELM	0.189	0.356	0.435	0.979
P-LSTM	0.249	0.401	0.499	0.984
P-GRU	0.239	0.393	0.489	0.984
SSA-GRU	0.234	0.393	0.484	0.984
P-ISSA-GRU	0.152	0.311	0.390	0.984

P-ELMAN:局部回归网络模型;P-ELM:极限学习机模型;P-LSTM:单一长短时记忆网络模型;P-GRU:单一门控循环单元模型;SSA-GRU:麻雀搜索算法-门控循环单元神经网络模型;P-ISSA-GRU:非线性溶解氧含量预测模型。 $MSE$ :均方误差; $MAE$ :平均绝对误差; $RMSE$ :均方根误差; $R^2$ :决定系数。

### 3 结论

本研究采用 Pearson 相关系数法筛选出对养殖水体溶解氧含量具有关键影响的特征,减少了参数的维度与训练时间,有效避免了各参数间的非关联性和冗余性,从而保证了模型的运算速度和预测的准确性。采用 Tent 混沌映射改进种群初始化,而后引入自适应动态权重因子  $\omega$  更新种群的发现者位置,加入高斯扰动改进麻雀搜索算法。与传统麻雀搜索算法(SSA)相比,改进后的麻雀搜索算法(ISSA)适应度收敛速度更快。可以看出,ISSA 在全局搜索和局部搜索之间实现了良好的平衡,从而有效提高了算法的收敛速度。本研究提出的 P-ISSA-GRU 非线性溶解氧含量预测模型,利用 ISSA 优化 GRU 网络的超参数组合,避免了人为经验的盲目性,从而实现了池塘养殖 2 h 后水体溶解氧含量更加准确的预测。在此过程中,GRU 时间序列模型的预测精度与 LSTM 模型相当,但由于 GRU 参数较少,收敛速度更快,因此有效克服了传统模型预测的延迟问题,并提升了预测的准确性。本研究结果表明,P-ISSA-GRU 模型具备较高的预测精度和优秀的泛化能力,克服了传统模型在预测精度和鲁棒性方面的不足。该模型能够为养殖池塘水质预警与调控提供可靠的技术支持,具有重要的应用价值。

### 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 Aquacul-

ture, 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(1):205-210,204.
- [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, 91:102122.
- [4] 徐龙琴,李乾川,刘双印,等.基于集合经验模态分解和人工蜂群算法的工厂化养殖 pH 值预测[J].农业工程学报,2016,32(3):202-209.
- [5] 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.
- [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 Chang Zhou fishery demonstration base, China[J]. Computers and Electronics in Agriculture, 2020, 175:105530.
- [7] NEAL C, NEAL M, WICKHAM H, et al. The water quality of a tributary of the Thames, the Pang, southern England[J]. Science of the Total Environment, 2000, 251:459-475.
- [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.
- [9] 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]. Computers and Electronics in Agriculture, 2013, 95:82-91.
- [10] YANG T, QIU W, MA Y, et al. Fuzzy model-based predictive control of dissolved oxygen in activated sludge processes[J]. Neurocomputing, 2014, 136:88-95.
- [11] 施珮,袁永明,张红燕,等. GRNN 和 Elman 神经网络在水体溶解氧预测中的应用[J].江苏农业科学,2017,45(23):217-221.
- [12] 刘双印,徐龙琴,李道亮,等.基于蚁群优化最小二乘支持向量回归机的河蟹养殖溶解氧预测模型[J].农业工程学报,2012, 28(23):167-175.
- [13] 任妮,鲍形,刘杨,等.基于粒子群优化算法和长短时记忆神经网络的蟹塘溶解氧预测[J].江苏农业学报,2021, 37(2):426-434.
- [14] 宦娟,刘星桥.基于 K-means 聚类和 ELM 神经网络的养殖水质溶解氧预测[J].农业工程学报,2016,32(17):174-181.
- [15] TA X X, WEI Y G. Research on a dissolved oxygen prediction method for recirculating aquaculture systems based on a convolution neural network[J]. Computers and Electronics in Agriculture, 2018, 145:302-310.
- [16] XU H H, DENG Y. Dependent evidence combination based on shearman coefficient and Pearson coefficient[J]. IEEE Access, 2017, 6:11634-11640.

- [17] 孙龙清,吴雨寒,孙希蓓,等. 基于 IBAS 和 LSTM 网络的池塘水溶解氧含量预测[J]. 农业机械学报,2021,52(1):252-260.
- [18] 曾一川,曾会国,袁伟皓,等. 长江口入海通道水质综合分析与模型预测[J]. 环境工程,2022,40(5):95-102,108.
- [19] 王佳钰,郝思鹏,李森文,等. 基于 ES-GRU-LSTM 的风电场群功率预测[J]. 计算技术与自动化,2022,41(3):37-41.
- [20] 黄春桃,范东平,卢集富,等. 基于深度学习模型的广州市大气 PM<sub>2.5</sub> 和 PM<sub>10</sub> 浓度预测[J]. 环境工程,2021,39(12):135-140.
- [21] 王海瑞,鲜于建川. 改进麻雀搜索算法在分布式电源配置中的应用[J]. 计算机工程与应用,2021,57(20):245-252.
- [22] 马晓宁,李笑含. 基于 Tent 混沌映射的可复制的鲸鱼算法[J]. 计算机仿真,2022,39(8):363-368.

(责任编辑:黄克玲)