

基于 EMD - IGA - SELM 的池塘养殖水温预测方法

施珮^{1,2} 袁永明¹ 匡亮³ 李光辉² 张红燕¹

(1. 中国水产科学研究院淡水渔业研究中心, 无锡 214081; 2. 江南大学物联网工程学院, 无锡 214122;
3. 江苏信息职业技术学院物联网工程学院, 无锡 214153)

摘要: 为了有效指导工厂化水产养殖, 提高水产养殖过程中水体温度预测的精度和稳定性, 在分析水体温度影响因素的基础上, 提出基于经验模态分解(EMD)、改进遗传算法(IGA)和改进极限学习机(SELM)相结合的水温预测模型(EMD-IGA-SELM)。首先, 通过综合天气指数的计算完成异常和缺失数据的校正; 利用皮尔森相关分析计算各影响因子与水温之间的相关度, 从而确定预测模型的输入输出量; 选择 Softplus 函数代替 Sigmoid 函数组成 SELM, 并引入混沌序列改进标准遗传算法, 获得 SELM 的最佳初始权值和阈值; 最后, 采用 EMD 方法将原始水温时序数据进行多尺度分解, 在各分量中对 IGA-SELM 训练建模, 并叠加求和各分量预测值, 从而完成水温序列的预测。将 EMD-ELM 和 GA-BP 模型的预测结果与 EMD-IGA-SELM 进行对比, 结果表明, EMD-IGA-SELM 取得了较好的预测精度, 评价指标平均绝对误差、平均绝对百分比误差和均方根误差分别为 0.123 3℃、0.004 3 和 0.147 8℃, 能够满足水产养殖的生产需要, 可为池塘水质管理和调控提供决策支持。

关键词: 水产养殖; 水温预测; 极限学习机; 改进遗传算法; 经验模态分解

中图分类号: S969.33 **文献标识码:** A **文章编号:** 1000-1298(2018)11-0312-08

Water Temperature Prediction in Pond Aquaculture Based on EMD - IGA - SELM Neural Network

SHI Pei^{1,2} YUAN Yongming¹ KUANG Liang³ LI Guanghui² ZHANG Hongyan¹

(1. Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi 214081, China

2. School of IoT Engineering, Jiangnan University, Wuxi 214122, China

3. School of IoT Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China)

Abstract: In order to guide the intensive aquaculture effectively and improve the accuracy and stability of water temperature prediction, based on the analysis of water temperature factors, a prediction model (EMD-IGA-SELM) was proposed with the combination of empirical mode decomposition (EMD), improved genetic algorithm (IGA) and improved extreme learning machine (SELM). Firstly, the outlier and missing data were corrected with the calculation of composite meteorological index. Secondly, the Pearson correlation was utilized to explore the relationships between affecting factors and water temperature, and construct the input and output of prediction model. Then, Softplus function was used as activation function of SELM to replace Sigmoid. The best weight and threshold of SELM were obtained from the IGA, which introduced the chaotic sequence to traditional GA. Finally, EMD algorithm was applied to decompose the original water temperature time series into a series of intrinsic mode function (IMF). IGA-SELM prediction models were trained in each IMF sequence, and the predicted values were calculated by the sum of predicted value in each IMF sequence. The experimental results showed that EMD-IGA-SELM had better prediction accuracy, and the mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) of GA-SELM were 0.123 3℃, 0.004 3 and 0.147 8℃, respectively. Research results met the practical needs of the aquaculture and provided decision support for water quality management and control.

Key words: aquaculture; water temperature prediction; extreme learning machine; improved genetic algorithm; empirical mode decomposition

收稿日期: 2018-04-16 修回日期: 2018-08-22

基金项目: 国家自然科学基金项目(61174034)、中央级公益性科研院所基本科研业务费专项(2016HY-ZD1404)和现代农业产业技术体系专项(CARS-46)

作者简介: 施珮(1988—), 女, 助理研究员, 江南大学博士生, 主要从事渔业信息化和机器学习研究, E-mail: ship@ffrc.cn

0 引言

水温是水产养殖中影响鱼类生长状况、生长质量的关键因素之一,在工厂化养殖中水温的作用更为突出^[1]。高密度的养殖环境、大量的饵料和狭小的养殖空间对水质、水温的要求更为严格。水温的突变会对水体中 pH 值、溶解氧、氨氮等产生影响,影响鱼类适宜的生存环境^[2]。

为了对水体温度进行准确预测,近年来很多学者进行了相关研究^[3-6,8-12],其预测方法包括传统数理统计法和人工智能方法。基于水温机理的数理方法对影响水温变化的机理和因素进行了分析,研究不同水层条件、时空分布、流速、流量等对水温的影响^[3-4],利用传热传质理论、水文气象理论、二维和三维模型等构建水温预测模型^[5-6],这些方法虽然能够对水库、水槽的水温进行有效预测,但预测模型复杂,参数多且难获取,在水产养殖上难以应用。基于数据挖掘的非机理方法一般通过大量数据构建数学模型,获得水温与相关影响因子的关系,捕捉水温变化规律。而针对大量数据样本构建的数学模型较多使用机器学习算法,包括贝叶斯算法、支持向量机、神经网络等^[7]。贝叶斯算法能够处理不确定性的信息数据,但需要有处理目标的先验知识,获取目标的先验分布^[8-9];支持向量机(SVM)能够较好地拟合处理非线性系统问题,且泛化性能较强,但算法复杂,参数较多且难以确定^[10-11]。传统的神经网络与 SVM 相似,适用于处理非线性问题,其缺陷也表现在算法复杂度和参数设置及核函数确定等问题上^[12]。极限学习机(ELM)不同于传统的神经网络,其结构较为简单,拥有快速学习能力和强泛化能力^[13],已广泛应用于各类预测、分类和识别问题中^[14-16]。然而单一的 ELM 预测精度有限,在初始参数和激活函数上仍有优化和改进的空间。

在前人研究的基础上^[12,17-18],本文提出基于 EMD(经验模态分解)-IGA(改进遗传算法)-SELM(改进极限学习机)的池塘水温预测模型。该模型利用 EMD 对水温原始数据进行不同尺度的分解,挖掘数据特征,利用改进的遗传算法获取最优初始参数,避免组合优化算法的抖振问题和寻优过程的早熟问题,使用模型新的激活函数,在 EMD 分量中进行模型训练和预测,叠加后最终获得水温预测结果。

1 材料与方法

1.1 研究区域

实验地点无锡市位于东经 120.18°、北纬 31.34°。在该养殖区域内选取长 110 m、宽 45 m、水

深约 1.5 m 的池塘作为实验池塘。在池塘内搭建 3 个高密度水泥槽,每个水槽长 9 m、宽 3 m,投放 5 000 尾罗非鱼鱼苗,鱼苗长度约 3 cm。在高密度水泥槽中使用工厂化循环水养殖技术,并采用微孔曝气增氧和气提式推水装置进行增氧。

1.2 数据采集

实验数据采集装置使用淡水渔业研究中心智能渔业物联服务中心研制的水产物联服务远程监控系统平台采集工厂化池塘养殖中的水质数据和自动气象站监测数据,其系统架构图如图 1 所示。该系统对溶解氧含量、水温、pH 值、气温、气压、湿度、雨量等水产养殖的水质环境数据进行在线监测。通过系统的感知层采集数据,经系统传输层传输至系统应用层。水下传感器放置深度为 0.5 m,气象环境监测传感器集成在自动气象站中,所有获取的数据在应用层进行分析和处理,为用户的控制决策提供依据。

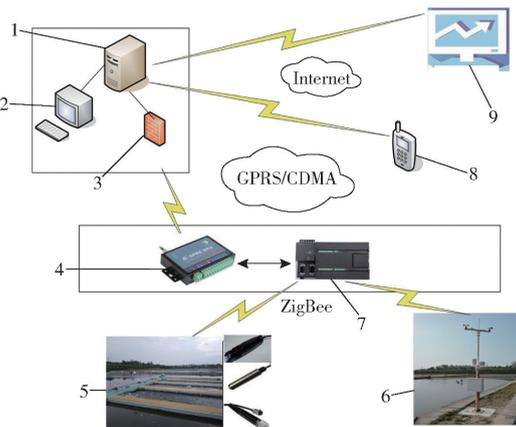


图 1 系统架构图

Fig. 1 Structure diagram of monitoring system

1. 服务器 2. 监测系统 3. 防火墙 4. DTU 数据传输单元
5. 池塘水质传感器节点 6. 自动气象站 7. PLC 控制器 8. 手机终端 9. PC 终端

1.3 研究方法

1.3.1 经验模态分解

EMD 通过将信号分解成一系列简单的本征模态分量(Intrinsic mode function, IMF)和残量 r ,各 IMF 分量之间相互独立,从而获取信号的特征,并广泛地应用在非平稳信号的分析处理中^[19]。在实际应用中,对池塘水温序列 $\{x(t) | t=1, 2, \dots, T\}$ (t 表示时间序号)的 EMD 算法步骤如下:

(1) 对于 $x(t)$ 序列,首先获得它的极大值点和极小值点。利用三次样条插值法将所有极大、极小值连接,形成上包络线 $x_{\max}(t)$ 和下包络线 $x_{\min}(t)$ 。

(2) 计算 $x_{\max}(t)$ 和 $x_{\min}(t)$ 的均值 $m(t)$ 和差值 $h(t)$ 。将 $h(t)$ 作为一个获取的新数据序列,重复 k 次步骤(1),计算得到 $h_k(t)$,当满足 R_k 小于设定阈

值时,则 $h_k(t)$ 成为一个 IMF, 公式为

$$R_k = \frac{\sum_{t=1}^T |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=1}^T |h_{k-1}(t)|^2} \quad (1)$$

(3) 计算得到第 1 个 IMF 分量后,从原始水温序列 $x(t)$ 中减去它,获得差值序列 $r_1(t)$ 。重复上述步骤,依次获得各 IMF 分量和余项 $r_n(t)$ 。 $r_n(t)$ 为一个单调函数,且小于设定的阈值,公式为

$$x(t) = \sum_{i=1}^n F_i(t) + r_n(t) \quad (2)$$

式中 $F_i(t)$ ——各 IMF 分量

1.3.2 改进的遗传算法

遗传算法 (Genetic algorithm, GA) 是一种模拟自然界遗传机制和生物进化论而形成的并行随机搜索最优化方法^[20]。该算法主要通过选择、交叉、变异等遗传操作,使群体逐代进化,直到满足进化终止条件才结束。传统的遗传算法在全局寻优过程中都是基于交叉操作和变异操作,且变异操作是在交叉操作的基础上进行的。本文将混沌系统的随机性^[21]引入交叉和变异操作中,将生物进化看作为随机性和反馈作用的结果,提出改进的遗传算法 (Improved genetic algorithm, IGA),改进内容如下,以期避免遗传算法的早熟问题。

(1) 交叉操作

不同于传统的单点交叉,本文的交叉操作以“门当户对”为原则,按照适应度函数值的大小对父代个体进行排序,适应度值小的与小的配对,适应度值大的与大的配对,从而完成父代配对。利用混沌序列的随机性确定交叉点位置,完成交叉项交叉操作。以本文使用的 logistics 混沌序列为例,若对染色体 $(O_1, O_2) \{O_1 = o_1^1, o_2^1, \dots, o_p^1; O_2 = o_1^2, o_2^2, \dots, o_p^2\}$ 上的基因进行配对,则需利用 logistics 混沌序列 $x(u+1) = 4x(u)(1-x(u))$ 产生一个 $2 \sim p-1$ 之间的正整数 a ,并以 a 为交叉点对 (O_1, O_2) 染色体的相应的基因进行单点交叉,得到新的染色体 (O'_1, O'_2) 。其中 $O'_1 = o_1^1, o_2^1, \dots, o_a^2, \dots, o_p^1; O'_2 = o_1^2, o_2^2, \dots, o_a^1, \dots, o_p^2$ 。

(2) 变异操作

本文中,根据设定的变异率,随机获取 $2 \sim p-1$ 之间的两个正整数 c, d 。对 c, d 对染色体上相应位置的基因进行变异操作,利用混沌序列把 c, d 位置上的基因换成新的基因值,获得新的染色体。

改进的遗传算法将变异操作与交叉操作拨离开,使二者独立并列进行,并在具体遗传操作中,引入混沌序列来确定交叉点,利用单点交叉的低改动

性,削弱和避免遗传算法组合优化问题中的寻优抖动问题。利用混沌序列完成染色体中的多基因变异过程,进而避免遗传算法的寻优早熟问题。

1.3.3 改进激活函数的极限学习机

ELM 是一种前馈神经网络学习算法,算法具有很好的全局搜索能力^[22],且算法的参数一经确认,则训练过程中无需调整。与其他机器学习算法相比,ELM 具有学习效率高、泛化性能好等优点。

在本文的池塘水温训练样本 (x_i, y_i) 中,设 ELM 有 u 个输入节点, L 个隐含层节点, q 个输出节点,激活函数为 $g(x)$,则 $x_i = [x_{i1} \ x_{i2} \ \dots \ x_{iu}]$,网络输出可表示为

$$\sum_{j=1}^L \beta_j g(\mathbf{w}_j x_i + b_j) = y_i \quad (i=1, 2, \dots, N) \quad (3)$$

式中 \mathbf{w}_j ——第 j 个隐含层节点与输入节点的权值向量

b_j ——第 j 个隐含层节点的阈值

β_j ——第 j 个隐含层节点与输出节点间的权值向量

N ——样本个数

激活函数 $g(x)$ 是 ELM 中影响网络性能的关键因素,适宜的激活函数能够提高 ELM 的精准度和泛化性。Sigmoid 函数是 ELM 中传统的隐含层激活函数,它是一种采用双侧抑制的判别函数。然而当遇到广义 Hop-world 问题时,其函数逼近值为单调的,则双侧抑制方式会增加废运算^[23],此时则需要单边抑制来完成值的判别。修正线性函数作为一种新型激活函数正在被广泛地应用于深度学习领域^[24]。其产生的修正线性单元 (Rectified linear units, ReLU) 被定义为

$$g(x) = \max(0, x) \quad (4)$$

ReLU 函数为分段函数,该函数形式简单、运算快,泛化性较 Sigmoid 更好,但函数的稀疏性会减小函数的预测能力,降低网络平均性能^[25]。文献[26]中提出了 Softplus 函数,该函数为 ReLU 的非线性平滑表示。Softplus 函数是非线性连续可微的,且相较于 Sigmoid 函数更接近生物学激活模型,能较好地避免 ReLU 的强制稀疏性,提高网络的平均性能。在本研究中,选用 Softplus 函数作为 ELM 的激活函数,函数定义为

$$g(x) = \ln(1 + e^x) \quad (5)$$

2 EMD-IGA-SELM 的池塘养殖水温预测模型

2.1 数据预处理

由于水质传感器常年放置在水下环境中,受到

水体腐蚀作用和其他环境因素作用,会使得传感器在数据采集过程中发生偏差,传感器精度受到影响。同时,网络的传输也会产生数据的延迟和丢失,这些问题都会使得采集到的数据发生丢失和数据异常,针对这类问题,需要进行数据预处理来提高数据质量。

2.1.1 数据校正

对短时间内发生丢失问题的数据,使用线性插值法^[27]来完成缺失数据的插补,公式为

$$x_{k+i} = x_k + \frac{i(x_{k+j} - x_k)}{j} \quad (0 < i < j) \quad (6)$$

式中 x_k, x_{k+j} ——第 k 和第 $k+j$ 时刻采集的传感器数据

x_{k+i} ——第 $k+i$ 时刻传感器丢失数据

基于相似时间段内天气条件相似概率较高,水体的温度在相似天气条件下也具有一定规律,故以相似时间相似天气为评价对象,将监测系统采集的气象指标:大气压强、空气相对湿度、空气温度、空气中 CO_2 浓度、照度、光合有效辐射度、辐射照度、风速和风向等 9 项指标建立评估体系,采用因子分析法评估综合天气指数

$$W_{\text{index}} = \frac{\sum F_{ij} W_j}{\sum W_j} \quad (7)$$

其中 $F_{ij} = \sum x_{iM} a_{Mj}$

式中 W_{index} ——综合天气指数

W_j ——因子分析中提取的公共因子的方差贡献率

F_{ij} ——各公共因子得分系数

x_{iM} ——第 i 个公共因子在第 M 个天气指标上的实际值

a_{Mj} ——第 M 个天气指标在第 j 个公共因子上的得分系数

利用水质监测数据和气象环境数据在时间上的连续性,设定这些采集数据任意时刻前后差值超过 10% 时,判定传感器发生偏差,数据需要进行误差校正^[28]。遵循综合天气指数临近原则,将这些发生偏差的数据替换为综合天气指数相近条件下对应的水质数据,并保证替换后的数据前后差值不超过 10%。另外,对长时间丢失的数据,选择相似时刻、天气指数临近的对应水质数据插补丢失数据,同时确保插补数据满足差值不超过 10% 的条件。

2.1.2 数据归一化

为了消除不同度量标准的数据间的量纲差异问题,对所有经过校正的数据进行归一化处理,在 M 个指标的数据集中,利用 Z -score 数据标准化方法对

N 组数据进行对应标准化处理,即

$$Z_{mn} = \frac{x_{MN} - \bar{x}_N}{S_N} \quad (M = 1, 2, \dots, i; N = 1, 2, \dots, j) \quad (8)$$

式中 Z_{mn} ——归一化后数据

x_{MN} ——校正数据

\bar{x}_N —— x_{MN} 的数据平均值

S_N —— x_{MN} 的数据标准差值

2.2 池塘水温预测模型

2.2.1 预测指标体系构建

由于水体环境因子和气象环境因子对水温的变化均具有一定的影响,本文选择监测指标中水温、pH 值、大气压强、空气相对湿度、空气温度、空气中 CO_2 浓度、照度、光合有效辐射度、辐射照度、风速和风向等 11 项指标进行关联分析。在实验的 1008 组数据中,采用 Person 相关分析对这些指标与水温的相关性进行分析。经计算后可得,水温与 pH 值、大气压强等 10 项指标的相关系数分别为 0.512、0.597、-0.445、0.601、0.435、0.112、0.181、0.105、0.203 和 -0.134。由此可知水体环境因子和气象环境因子与水体温度均具有一定的关联性,故选择这 10 项指标构建池塘水温的预测模型。

2.2.2 预测模型设计

由于传统的 ELM 网络模型的初始参数存在随机性,且在激活函数的选择上并未考虑适宜性,而改进的遗传算法能够解决寻优抖振和早熟问题,确定最佳权值、阈值,Softplus 函数提高 ELM 泛化性。同时,针对原始监测数据构建的预测模型易同质化监测数据的不同尺度信息特征,不能充分发现时序数据的时频特征,影响预测模型的性能。本文将 EMD、IGA 和 ELM 3 种算法结合起来,提出基于 EMD-IGA-SELM 的池塘养殖水温预测模型。

在建模的过程中首先采用 EMD 对池塘水温时序序列进行自适应地多尺度分解,获得不同频率的信号分量和余项;然后在分解后的各分量中采用 IGA-SELM 进行建模预测;最后将各分量预测模型的预测值进行叠加获取最终预测值。其预测流程图如图 2 所示,基于 EMD-IGA-SELM 的池塘水温预测步骤如下:

(1) 池塘水温时序数据分解。利用 EMD 对 $x(t)$ 进行分解,获得 n 个 IMF 分量和一个余项 r_n 。

(2) 构建训练、测试样本集。在 IMF 分量中,构建各分量训练样本集、测试样本集的输入量和输出量。

(3) 构建各分量的优化极限学习机训练、预测模型。在改进的遗传算法中,计算改进遗传算法中

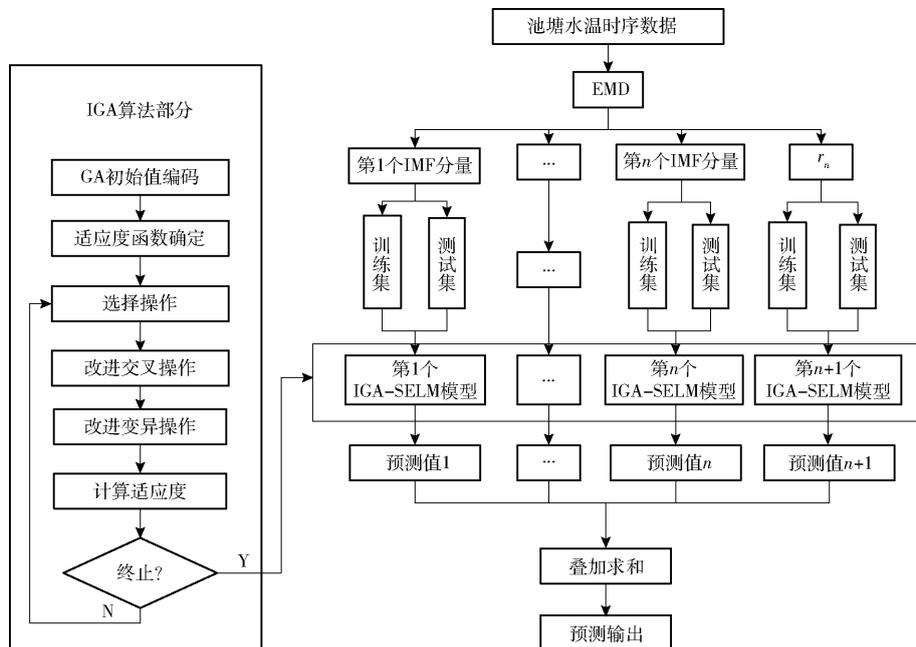


图2 EMD-IGA-SELM 预测流程图

Fig. 2 Flow chart of EMD-IGA-SELM prediction model

的适应度函数;设置初始种群规模 $size$ 和最大进化代数 $maxgen$, 对种群中的个体进行选择、改进的交叉和变异等遗传操作, 最终确定全局最优的适应度; 利用最优适应度获得最优权值 a_{best} 和阈值 b_{best} ; 设定 ELM 网络的激活函数为 $Sofplus$, 由 a_{best} 和 b_{best} 计算 ELM 的输出矩阵 H 和输出权值 β , 确定 SELM 网络结构; 利用 IGA 方法对每个 SELM 模型参数迭代寻优, 在每个 IMF 分量和余项 r_n 内建立最优参数的 SELM 拟合预测模型, 获得各分量的预测结果, 公式为

$$f = \frac{1}{N} \sum_{i=1}^N (y_i(t) - \hat{y}_i(t))^2 \quad (9)$$

$$H(a_1, a_2, \dots, a_L, b_1, b_2, \dots, b_L, x_1, x_2, \dots, x_N) = \begin{bmatrix} g(a_1 x_1 + b_1) & \dots & g(a_L x_1 + b_L) \\ \vdots & & \vdots \\ g(a_1 x_N + b_1) & \dots & g(a_L x_N + b_L) \end{bmatrix}_{N \times L} \quad (10)$$

$$\tilde{\beta} = H^+ Y \quad (11)$$

式中 $y_i(t)$ —— t 时刻的实际值

$\hat{y}_i(t)$ —— t 时刻的预测值

H^+ ——输出矩阵 H 的广义逆

$\tilde{\beta}$ ——隐含层与输出神经元的权值向量

Y ——输出矩阵

(4) 预测结果输出。对各 IMF 分量和余项 r_n 的预测结果进行累加求和, 获得池塘水温的最终预测结果。

2.2.3 模型参数设置

在 EMD-IGA-SELM 预测模型中, 分别对改进的遗传算法部分和 ELM 神经网络参数进行如下

设置。

(1) 遗传算法部分。设置 IGA 算法的初始种群规模为 10, 迭代次数为 50。依据多次运行均方误差最小的原则, 确定交叉概率为 0.1, 变异概率为 0.1。

(2) ELM 神经网络部分。对 ELM 参数进行设置, 输入节点数为 10, 输出节点数为 1。为了避免出现“欠适配”、“过适配”的问题, 依据“试错法”, 最终设置各 IMF 分量 ELM 网络的隐含层节点数分别为 28、19、23、31、29、31, 由此获得各分量 ELM 神经网络的结构模型。

3 实验结果与分析

3.1 实验数据源

以无锡市南泉罗非鱼养殖基地的池塘水温为研究对象, 基于 1.2 节中物联服务水质监控系统, 每 10 min 采集一组数据, 将获取的所有气象站数据和水体环境数据作为预测模型的数据样本。实验数据起始时间为 2016 年 7 月 1—7 日, 将前 6 d 的 864 组数据作为水温预测的训练集, 最后 1 d 的 144 组数据组成预测集。在构建的预测体系中, 确定 pH 值、大气压强、空气相对湿度、空气温度、空气中 CO_2 浓度、照度、光合有效辐射度、辐射照度、风速和风向为输入变量, 水温为输出量, 系统运行环境为 Matlab 2014a Microsoft Windows 7, 处理器为 3.4 GHz Core (TM), 内存为 4.0 GB。

3.2 结果分析

3.2.1 基于 EMD 的水温多尺度分解

按照 2.2.2 节的 EMD-IGA-SELM 模型的步

骤,以 1 008 组水温时序数据为对象进行 EMD 分解,获得 5 个 IMF 分量和 1 个余项,分解图如图 3 所示。

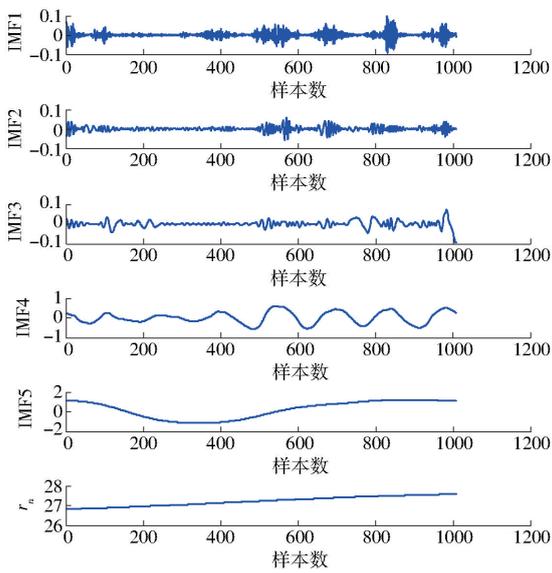


图 3 水温 EMD 分解图

Fig. 3 Decomposition of water temperature by EMD

从图 3 可以发现,池塘水温时序序列有明显的多尺度特征,5 个 IMF 分量呈现高低变化不同波动尺度的信息。其中,IMF1 的频率较高,能体现出原始时序数据的随机噪声信息; r_n 余项频率较低,变化平稳,体现水温时序的周期性和趋势性信息,反映水温总体变化特征。

3.2.2 预测结果分析

在 EMD 分解的基础上,采用 IGA-SELM 分别对各 IMF 分量和余项构建预测模型,得到 EMD-IGA-SELM 的池塘水温预测结果。同时,为了对预测模型的性能有清晰的了解,本文设置了对照实验。将 IGA-SELM、GA-SELM 和 GA-ELM 作为参照模型,分析和对比 EMD、IGA 和 SELM 对预测性能的影响。在 IGA-SELM 中设置其隐含层节点数为 28,激活函数为 Softplus,IGA 算法的参数设置同 2.2.3 节中遗传算法部分;GA-SELM 和 GA-ELM 均使用标准遗传算法,GA-ELM 中激活函数设置为 Sigmoid,其他参数设置相同。各模型的预测值与实际值对比结果如图 4 所示。

图 4 显示,各预测模型均能在不同程度上较好地实现池塘水温的预测,水温的预测变化趋势与实际值变化趋势较为一致,但预测效果存在一定的差异。就整体预测结果而言,EMD-IGA-SELM 的预测结果比其他 3 种模型的拟合效果更好,变化起伏更小,预测效果更稳定,全局看来无较大起伏的波动点。EMD-IGA-SELM 与 IGA-SELM 的预测趋势较为相近,二者曲线变化方向较一致,这两个预测模

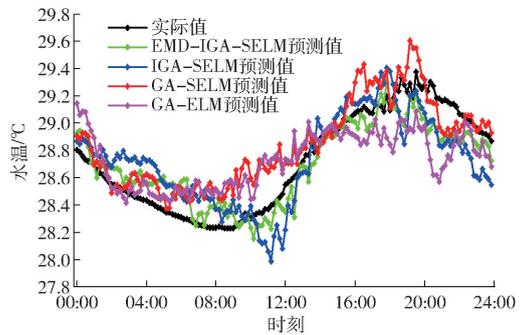


图 4 各模型预测值与实际值对比

Fig. 4 Comparison of original data and predictive values of three models

型均使用 IGA 优化算法,且激活函数相同。而 GA-SELM 与 GA-ELM 也有较大程度的一致趋势,二者在 ELM 均使用标准算法,但激活函数不同。从图中还可以发现,各预测模型均不同程度的在 08:00—13:00 和 18:00—22:00 时段内呈现相对较大的起伏,联系实际情况可知,这两个时段分别为光合作用变化较大的时间段,池塘内浮游植物和微生物发生光合作用,水温的变化幅度最大。由于水温随空气温度的变化具有延时性,在变化幅度较大的时间段中,水温预测的难度相对较大。

为了对这 4 个模型进行综合的性能对比,本文选择均方根误差(Root mean square error, RMSE)、平均绝对百分比误差(Mean absolute percent error, MAPE)、平均绝对误差(Mean absolute error, MAE)和运行时间^[22]4 项指标进行比较,其预测性能结果如表 1 所示。

表 1 预测模型的预测性能对比

Tab. 1 Performances comparison of four predictive models

参数	EMD-IGA-SELM	IGA-SELM	GA-SELM	GA-ELM
MAE/°C	0.123 3	0.142 0	0.154 3	0.205 0
MAPE	0.004 3	0.004 9	0.005 4	0.007 1
RMSE/°C	0.147 8	0.168 3	0.187 6	0.243 2
运行时间/s	5.52	0.97	2.09	3.64

从表 1 可以发现,4 个模型的预测准确率达到 99% 以上(MAPE 均小于 0.01)。EMD-IGA-SELM 模型的精度指标 MAE、MAPE、RMSE 分别为 0.123 3°C、0.004 3 和 0.147 8°C。该模型预测精度与其他方法相比有很大的改进,仅 MAPE 相比另外 3 个模型分别降低 12.25%、20.37% 和 39.44%。在运行时间上,EMD-IGA-SELM 较其他 3 种模型分别多耗时 4.55、3.43、1.88 s。水温时序数据的 EMD 分解降低了不同尺度特征信息间的相互干扰,提高了预测模型的精度,但是数据 EMD 分解的过程也消耗了时间。改进的遗传算法可以有效地避免遗传算

法的早熟问题,提高算法的效率。Softplus 激活函数能提高 ELM 的精度和效率,适宜使用在水温的预测模型中。

为了验证本文 EMD-IGA-SELM 的预测性能,分别采用已经被应用的 EMD-ELM 和 GA-BP 模型对池塘水温进行预测。其中,GA-BP 中 BP 的网络结构为 10-6-1,训练次数为 1 000,训练目标为 0.1,激活函数为 Sigmoid;EMD-ELM 中,ELM 的神经元设置、EMD 分解后的 ELM 网络设置同 2.2.3 节,激活函数为 Sigmoid。由此获得 3 个预测算法的性能,如表 2 所示。

表 2 不同水温预测模型的预测性能对比

Tab.2 Performances comparison of three existing models

参数	EMD-IGA-SELM	EMD-ELM	GA-BP
MAE/°C	0.123 3	0.185 3	0.191 4
MAPE	0.004 3	0.006 4	0.006 7
RMSE/°C	0.147 8	0.210 0	0.235 2
运行时间/s	5.52	2.01	15.51

由表 2 可以发现,EMD-IGA-SELM 的预测精度与 EMD-ELM、GA-BP 相比有明显的提升,在 MAE 指标上,EMD-IGA-SELM 比 EMD-ELM 和 GA-BP 分别降低了 33.46%、35.58%;在 MAPE 上,EMD-IGA-SELM 比 EMD-ELM 和 GA-BP 分别降低了 32.81%、35.82%;EMD-IGA-SELM 的 RMSE 比 EMD-ELM 和 GA-BP 分别降低了 29.62%、37.16%。EMD-IGA-SELM 的运行时间较 EMD-ELM 慢了 3.51 s,较 GA-BP 快了 9.99 s。

综上所述,EMD-IGA-SELM 的综合性能较好,有一定的优势,能够解决 ELM 模型中参数随机的问题,可以充分发掘水温数据的多尺度特征。将 EMD 与 IGA、SELM 结合在一起可以有效地提高水温预测模型的精度,克服单一模型存在的低精度问题,为工厂化的池塘养殖提供水温预测的方法,为实际生产中的水质监控和管理提供依据。并且在实际生产过程中,可以根据系统预测精度和响应时效要求选择符合生产需求的预测模型完成水温的预测。

4 结论

(1)采用改进的遗传算法对 ELM 的输入权值和隐含层阈值进行优化,建立 IGA-SELM 预测模型,在多次实验的基础上确定预测模型的隐含层节点数。改进的遗传算法引入混沌序列,有效地解决了 ELM 优化模型中寻优抖振,避免出现早熟问题。获得的最佳权值和阈值能够避免 ELM 随机参数的不稳定性,提高预测模型的准确性。

(2)使用 Softplus 激活函数替换传统 ELM 中的 Sigmoid 函数,不仅提高了 ELM 泛化能力,还提高了预测精度和效率。

(3)采用经验模态分解对水温时序数据进行多尺度分解,获得多个 IMF 分量和余项,构建各分量中的预测模型,捕捉到水温数据特征,虽增加了一定的时间成本,但提高了水温预测模型的精度,可根据需求供不同系统选择使用。

参 考 文 献

- 孙中之,柳学周,徐永江,等.半滑舌鲷工厂化人工育苗工艺技术研究[J].中国水产科学,2007,14(2):244-248.
SUN Zhongzhi, LIU Xuezhou, XU Yongjiang, et al. Technique and technology of productive scale seedling rearing of *Cynoglossus semilaevis* Günther[J]. Journal of Fishery Sciences of China, 2007, 14(2):244-248. (in Chinese)
- 王丽娜,陈辉煌,刘乐冕,等.亚热带分层水库固氮微生物时空分布格局[J].生态学报,2016,36(18):5827-5837.
WANG Lina, CHEN Huihuang, LIU Lemian, et al. Spatiotemporal patterns of diazotrophic microorganisms in a subtropical stratified reservoir[J]. Acta Ecologica Sinica, 2016, 36(18):5827-5837. (in Chinese)
- 唐笑,程永光.基于 Fluent 的水库水温多维预测[J].武汉大学学报(工学版),2010,43(1):59-63.
TANG Xiao, CHENG Yongguang. Multidimensional prediction of reservoir water temperature based on Fluent[J]. Engineering Journal of Wuhan University, 2010, 43(1):59-63. (in Chinese)
- 陈武,刘德仁,董元宏,等.寒区封闭引水渡槽中水温变化预测分析[J].农业工程学报,2012,28(4):69-75.
CHEN Wu, LIU Deren, DONG Yuanhong, et al. Prediction analysis on water temperature in closed aqueduct in cold regions[J]. Transactions of the CSAE, 2012, 28(4):69-75. (in Chinese)
- 胡平,刘毅,唐忠敏,等.水库水温数值预测方法[J].水利学报,2010,41(9):1045-1053.
HU Ping, LIU Yi, TANG Zhongmin, et al. Numerical prediction of water temperature in reservoir[J]. Journal of Hydraulic Engineering, 2010, 41(9):1045-1053. (in Chinese)
- 陆俊卿,张小峰,强继红,等.水库水温数学模型及其应用[J].水力发电学报,2008,27(5):123-129.
LU Junqing, ZHANG Xiaofeng, QIANG Jihong, et al. Numerical model for temperature simulation of water in reservoir[J]. Journal of Hydroelectric Engineering, 2008, 27(5):123-129. (in Chinese)
- 胡星,王千祥.深度学习在缺陷修复者推荐中的应用[J].计算机科学与探索,2017,11(5):700-707.
HU Xing, WANG Qianxiang. Application of deep learning in recommendation of bug reports assignment[J]. Journal of Frontiers of Computer Science and Technology, 2017, 11(5):700-707. (in Chinese)
- BAL G, RIVOT E, BAGLINIÈRE J L, et al. A hierarchical bayesian model to quantify uncertainty of stream water temperature forecasts[J]. Plos One, 2014, 9(12):e115659.
- SOHRABI M M, BENJANKAR R, TONINA D, et al. Estimation of daily stream water temperatures with a bayesian regression approach[J]. Hydrological Processes, 2017, 31(9):1719-1733.

- 10 HE Z, WEN X, LIU H, et al. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region[J]. *Journal of Hydrology*, 2014, 509(379-386): 379-386.
- 11 MALINOWSKI P, SULOWICZ M. Use of SVM neural networks to prediction of the cooling water temperature in the PKN Orlen refinery[J]. *Przemysl Chemiczny*, 2010, 89(9):1236-1240.
- 12 陈英义, 程倩倩, 成艳君, 等. 基于 GA-BP 神经网络的池塘养殖水温短期预测系统[J/OL]. *农业机械学报*, 2017, 48(8):172-178. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170819&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2017.08.019.
- CHEN Yingyi, CHENG Qianqian, CHENG Yanjun, et al. Short-term prediction system of water temperature in pond aquaculture based on GA-BP neural network[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(8):172-178. (in Chinese)
- 13 FENG Y, PENG Y, CUI N, et al. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data[J]. *Computers & Electronics in Agriculture*, 2017, 136:71-78.
- 14 邹伟东, 张百海, 姚分喜, 等. 基于改进型极限学习机的日光温室温湿度预测与验证[J]. *农业工程学报*, 2015, 31(24):194-200.
- ZOU Weidong, ZHANG Baihai, YAO Fenxi, et al. Verification and forecasting of temperature and humidity in solar greenhouse based on improved extreme learning machine algorithm[J]. *Transactions of the CSAE*, 2015, 31(24):194-200. (in Chinese)
- 15 林怡, 季昊巍, NICO S, 等. 基于鱼群算法的极限学习机影像分类方法优化[J/OL]. *农业机械学报*, 2017, 48(10):156-164. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20171019&flag=1. DOI:10.6041/j.issn.1000-1298.2017.10.019.
- LIN Yi, JI Haowei, NICO S, et al. Optimization of ELM classification model for remote sensing image based on artificial fish-swarm algorithm[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(10):156-164. (in Chinese)
- 16 MAHMOOD S F, MARHABAN M H, ROKHANI F Z, et al. Fasta-ELM: a fast adaptive shrinkage/thresholding algorithm for extreme learning machine and its application to gender recognition[J]. *Neurocomputing*, 2016, 219:312-322.
- 17 徐龙琴, 张军, 李乾川, 等. 基于 EMD 和 ELM 的工厂化育苗水温组合预测模型[J/OL]. *农业机械学报*, 2016, 47(4):265-271. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20160435&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2016.04.035.
- XU Longqin, ZHANG Jun, LI Qianchuan, et al. Combined prediction model of water temperature in industrialized cultivation based on empirical mode decomposition and extreme learning machine[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(4):265-271. (in Chinese)
- 18 徐大明, 周超, 孙传恒, 等. 基于粒子群优化 BP 神经网络的水产养殖水温及 pH 预测模型[J]. *渔业现代化*, 2016, 43(1):24-29.
- XU Daming, ZHOU Chao, SUN Chuanheng, et al. Prediction model of aquaculture water temperature and pH based on BP neural network optimized by particle swarm algorithm[J]. *Fishery Modernization*, 2016, 43(1):24-29. (in Chinese)
- 19 王守相, 王亚旻, 刘岩, 等. 基于经验模态分解和 ELM 神经网络的逐时太阳能辐照量预测[J]. *电力自动化设备*, 2014, 34(8):7-12.
- WANG Shouxiang, WANG Yamin, LIU Yan, et al. Hourly solar radiation forecasting based on EMD and ELM neural network[J]. *Electric Power Automation Equipment*, 2014, 34(8):7-12. (in Chinese)
- 20 刘斌, 王传武, 杨为民, 等. 基于并行改进遗传算法的三维电阻率反演方法[J]. *岩土工程学报*, 2014, 36(7):1252-1261.
- LIU Bin, WANG Chuanwu, YANG Weimin, et al. 3D resistivity inversion using improved parallel genetic algorithm[J]. *Chinese Journal of Geotechnical Engineering*, 2014, 36(7):1252-1261. (in Chinese)
- 21 梁俊毅, 张建龙, 马雪瑞, 等. 基于多混沌算子遗传算法的混合动力汽车控制策略优化[J]. *上海交通大学学报*, 2015, 49(4):442-449.
- LIANG Junyi, ZHANG Jianlong, MA Xuerui, et al. Control strategy optimization for hybrid electric vehicle based on multi-chaotic operators genetic algorithm[J]. *Journal of Shanghai Jiao Tong University*, 2015, 49(4):442-449. (in Chinese)
- 22 陆慧娟, 魏莎莎, 宋夫华, 等. 一种 Fibonacci 优化理论的改进 ELM 分类方法[J]. *小型微型计算机系统*, 2015, 36(12):2745-2748.
- LU Huijuan, WEI Shasha, SONG Fuhua, et al. ELM algorithm based on Fibonacci method optimization algorithm[J]. *Journal of Chinese Computer Systems*, 2015, 36(12):2745-2748. (in Chinese)
- 23 徐圆, 黄兵明, 贺彦林. 基于改进 ELM 的递归最小二乘时序差分强化学习算法及其应用[J]. *化工学报*, 2017, 68(3):916-924.
- XU Yuan, HUANG Bingming, HE Yanlin. Recursive least-squares TD(λ) learning algorithm based on improved extreme learning machine[J]. *CIESC Journal*, 2017, 68(3):916-924. (in Chinese)
- 24 傅隆生, 冯亚利, ELKAMIL T, 等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. *农业工程学报*, 2018, 34(2):205-211.
- FU Longsheng, FENG Yali, ELKAMIL T, et al. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks[J]. *Transactions of the CSAE*, 2018, 34(2):205-211. (in Chinese)
- 25 邵丽鹏, 郑辉. 基于 PReLU-Softplus 非线性激励函数的卷积神经网络[J]. *沈阳工业大学学报*, 2018, 40(1):54-59.
- GAO Lipeng, ZHENG Hui. Convolutional neural network based on PReLU-Softplus nonlinear excitation function[J]. *Journal of Shenyang University of Technology*, 2018, 40(1):54-59. (in Chinese)
- 26 GLOTOT X, BORDES A, BENGIO Y. Deep sparse rectifier networks[J]. *Journal of Machine Learning Research*, 2011, 15:315-323.
- 27 杨贵军, 李静华. 基于 PMM 多重插补法的线性模型系数估计量的模拟研究[J]. *数量经济技术经济研究*, 2014(10):139-150.
- YANG Guijun, LI Jinghua. On estimators of coefficients of linear regression model based on PMM multiple imputation[J]. *The Journal of Quantitative & Technical Economics*, 2014(10):139-150. (in Chinese)
- 28 徐龙琴, 刘双印. 基于 APSO-WLSSVR 的水质预测模型[J]. *山东大学学报(工学版)*, 2012, 42(5):80-86.
- XU Longqin, LIU Shuangyin. Water quality prediction model based on APSO-WLSSVR[J]. *Journal of Shandong University (Engineering Science)*, 2012, 42(5):80-86. (in Chinese)