

基于 APSO-WLSSVR 的水质预测模型

徐龙琴¹, 刘双印^{1,2,3,4*}

- (1. 广东海洋大学信息学院, 广东 湛江 524088;
2. 中国农业大学中欧农业信息技术研究中心, 北京 100083;
3. 中国农业大学北京市农业物联网工程技术研究中心, 北京 100083;
4. 中国农业大学先进农业传感技术北京市工程研究中心, 北京 100083)

摘要:为解决传统预测方法和标准最小二乘支持向量回归机(least squares support vector regression, LSSVR)在水质预测中存在预测精度低、鲁棒性差等问题,提出了自适应粒子群优化加权最小二乘支持向量回归机(adaptive particle swarm optimization weighted least squares support vector regression, APSO-WLSSVR)的水质预测模型。根据样本对模型重要性不同为各样本赋予不同权重,建立了加权最小二乘支持向量回归机(weighted least squares support vector regression, WLSSVR),实现对样本数据“重近轻远”的优化选择,避免标准 LSSVR 算法因没有考虑样本重要性差异致使预测精度低的问题;采用自适应粒子群优化算法对模型参数组合进行优化选择,克服了标准 LSSVR 算法因试凑法获取参数的盲目性和人为因素的影响。为验证该模型的性能,对江苏省宜兴市集约化河蟹养殖水质进行预测,并与其他预测方法对比分析,结果表明该模型预测精度明显提高,还具有较好的鲁棒性和泛化能力,能够满足集约化水产养殖水质管理的实际需要。

关键词:加权最小二乘支持向量回归机;自适应粒子群优化算法;水质预测;参数优化;集约化水产养殖

中图分类号:TP309 **文献标志码:**A

Water quality prediction model based on APSO-WLSSVR

XU Long-qin¹, LIU Shuang-yin^{1,2,3,4*}

- (1. College of Information, Guangdong Ocean University, Zhanjiang 524088, China;
2. China-EU Center for ICT in Agriculture, China Agricultural University, Beijing 100083, China;
3. Beijing ERC for Internet of Things in Agriculture, China Agricultural University, Beijing 100083, China;
4. Beijing ERC for Advanced Sensor Technology in Agriculture, China Agricultural University, Beijing 100083, China)

Abstract: In order to solve the problem of low prediction accuracy, the bad robustness of the traditional forecasting methods and the standard least squares support vector regression (LSSVR) in water quality prediction, the adaptive particle swarm optimization weighted least squares support vector regression (APSO-WLSSVR) model for water quality prediction was proposed. Different weights were set for various samples according to its different importance for the model. A weighted least squares support vector regression model (WLSSVR) was established, which could avoid low prediction accuracy of the standard LSSVR, and ignore the importance of the samples differences. The particle swarm optimization algorithm was adopted to optimize and choose the model parameters to reduce the blindness and the impact of human factors of the standard LSSVR trial method when obtaining the parameters. In order to verify the performance of the model, the water quality of intensive farming river crab in Yixing, Jiangsu Province, was predicted, which was also compared with other forecast methods. The results showed that the model prediction accuracy was obviously improved, and also had good robustness

收稿日期:2012-05-02

基金项目:“十二五”国家科技支撑计划资助项目(2011BAD21B01);国家自然科学基金资助项目(61100115,61101211);广东省科技计划资助项目(2012A020200008);湛江市科技计划资助项目(2010C3113011)

作者简介:徐龙琴(1977-),女,陕西汉中,讲师,硕士,主要研究方向为机器学习,智能计算,数据库安全等。E-mail:xlqlw@126.com

* 通讯作者:刘双印(1977-),男,山东单县人,副教授,博士研究生,主要研究方向为智能计算,智能信息系统,农业信息化技术等。

E-mail:hdsyx1q@126.com

and generalization ability, which could met the practical needs of the intensive aquaculture water quality management.

Key words: least squares support vector regression; self-adaptive particle swarm optimization algorithm; water quality prediction; parameters optimization; intensive aquaculture

0 引言

集约化水产养殖水质预测是水产品健康养殖、养殖水域规划、水质调节与应对突发水质事件等方面实现科学决策的重要依据,其水质的优劣决定着水产品的产量和养殖户的收益。准确预测养殖水质,及时洞察水质变化趋势,提前辨识养殖水域环境质量的异常变化,是集约化水产养殖亟需解决的重要问题之一^[1]。因此对集约化水产养殖水质预测和水质变化趋势的研究,具有重要的现实意义和实用价值。

因养殖池塘水质受生物、物理、化学、水文气象以及人类活动等多种因素的交叉影响^[2],生态环境因子之间作用机理复杂,具有多变量、滞后性、非线性、不完整和模糊不确定性等特点,采用传统方法构建精准的非线性预测模型非常困难,且效果不理想^[3]。为此,国内外专家采用模糊推理^[4-6]、灰色理论^[7]、时间序列^[8-10]、回归分析^[11-12]、神经网络^[13-16]、支持向量机^[17-19]等方法,对多种水质参数回归预测取得了一定的成效,但也存在易陷入局部极小点、早熟、欠学习与过学习、算法结构复杂、维数灾难、计算效率低、泛化能力不高、可操作性差等缺陷,尤其对包含“不良数据”的小样本进行预测时,其预测结果很难满足实际应用的需要。

最小二乘支持向量回归机(least squares support vector regression, LSSVR)作为标准支持向量回归机(SVR)的一种扩展^[20],将 SVR 算法中不等式约束变成等式约束、求解二次规划问题转化为求解线性方程组的问题^[20-21],在相同规模数据的条件下,与 SVR 相比,LSSVR 具有结构简单、学习能力强,计算复杂度小,求解速度快、收敛精度高,应用领域更为广泛的特点^[22-23],能较好地规避传统预测方法存在的缺陷。同 SVR 一样,LSSVR 算法的拟合精度和性能仍受制于核函数选择及参数率定,获得最佳的参数组合,提高算法性能是亟需探索和完善的课题之一。自适应粒子群优化(adaptive particle swarm optimization, APSO)算法是一种以信息共享机制为基础的群体智能优化技术,具有调节参数少、并行处理、收敛快速、鲁棒性好、简单易实现等优点^[24],还可避免标准 PSO 算法早熟的不足,比较适合于

LSSVR 参数组合优化问题。

结合集约化水产养殖生态环境的实际情况,在研究前人成果的基础上,将群智能体并行全局搜索的 APSO 算法与加权支持向量回归机(weighted least squares support vector regression, WLSSVR)有机结合,提出了基于 APSO-WLSSVR 的水质预测模型。首先,根据近期数据较远期历史数据对预测结果贡献不同应赋予不同权重的“重近轻远”思想,引入权重参数将标准 LSSVR 改进为 WLSSVR 算法,以提高算法拟合精度;利用 APSO 算法对 WLSSVR 参数进行自动寻优,以获得 WLSSVR 性能最佳的参数组合;将该模型应用于江苏省宜兴市集约化河蟹养殖某池塘水质预测中,与标准 LSSVR 和 BP 神经网络预测模型进行对比分析。结果表明,本研究算法具有较高的泛化性能与预测精度,可为集约化水产养殖水质预测提供一种新途径。

1 加权最小二乘支持向量回归机

1.1 最小二乘支持向量回归机

给定训练集 $T = \{(x_k, y_k) \mid x_k \in \mathbf{R}^n, y_k \in \mathbf{R}\}_{k=1}^n$, 由统计学理论,LSSVR 对应的优化问题^[20]如下:

$$\begin{aligned} \min_{\omega, b, \xi} J(\omega, \xi) &= \frac{1}{2} \|\omega\|^2 + \frac{\psi}{2} \sum_{k=1}^n \xi_k^2, & (1) \\ \text{s. t. } y_k &= \omega^T \varphi(x_k) + b + \xi_k, \quad (k=1, 2, \dots, n), & (2) \end{aligned}$$

式中, J 为损失函数,估计误差 $\xi_k \in \mathbf{R}$,偏置量 $b \in \mathbf{R}$,惩罚系数 $\psi \in \mathbf{R}^+$, b 和 ψ 调整误差的影响,非线性函数 $\varphi(\cdot)$ 将系统输入向量映射到高维特征空间。为求解上述约束优化问题,将模型通过 Wolfe 对偶变换到对偶空间,得到 Lagrange 多项式函数:

$$L(\omega, b, \xi; \alpha) = J(\omega, \xi) - \sum_{k=1}^n \alpha_k \{ \omega^T \phi(x_k) + b + \xi_k - y_k \}, \quad (3)$$

式中,Lagrange 乘子 $\alpha = [\alpha_1, \dots, \alpha_n]^T$,也常称作支持向量值。根据 KKT 条件,分别对上式参数 $\omega, b, \xi_k, \alpha_k$ 求偏导数并令其等于 0,消去 ω 和 ξ_k ,整理得到线性方程组如下:

$$\begin{bmatrix} \Omega + \psi^{-1} E & I \\ I^T & 0 \end{bmatrix} * \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix}, \quad (4)$$

式中 $I^T = [1, \dots, 1]^T$, $y = [y_1, \dots, y_n]^T$, E 是 $n \times n$ 维

的单位矩阵; Ω 为 $n \times n$ 的 Hessian 矩阵, 其第 i 列 j 行的元素为 $\Omega_{ij} = \varphi(x_i)^T \cdot \varphi(x_j) = K(x_i, x_j)$ 为满足 Mercer 定理的核函数, 原空间的核函数取代多维特征空间的点积, 降低了计算复杂度。因高斯 RBF 核函数的参数少、容易选择、操作简单易于实现、拟合效果较好, 本研究采用高斯 RBF 核函数作为 LSSVR 模型的核函数。

通过求解(4)式线性方程组得到参数 α_k 和 b , 则在点 x 处的 LSSVR 拟合数学模型为

$$\hat{y}(x) = f(x, a) = \sum_{k=1}^n \alpha_k K(x, x_k) + b. \quad (5)$$

1.2 加权最小二乘支持向量回归机 (WLSSVR)

虽然 LSSVR 模型具有结构简单、计算复杂度低等优点, 但由于采用平方和误差作为损失函数, 并将各时段监测数据重要性同等对待, 使 LSSVR 拟合性能受到一定的影响。为此, 结合水产养殖水质预测过程中近期监测样本较远期样本对预测值影响较大的实际情况, 故采用“重近轻远”思想对近期监测数据应赋予更高的权重 Θ_k 。采用线性加权函数对每个样本 (x_k, y_k) 的估计误差 ξ_k 进行修正, 即为每个样本赋予不同的权重, 以提高拟合精度。其线性加权函数表达式为

$$\Theta(k) = \frac{k}{n(n+1)/2}, \quad 0 < \Theta(k) < 1. \quad (6)$$

优化问题式(1)经过加权转化为^[20]

$$\min_{\omega^*, b^*, \xi^*} J^*(\omega^*, \xi^*) = \frac{1}{2} \|\omega^*\|^2 + \frac{\psi}{2} \sum_{k=1}^n \Theta(k) \xi_k^{*2}, \quad (7)$$

$$\text{s.t. } y_k = \omega^{*T} \varphi(x_k) + b^* + \xi_k^*, \quad (k=1, 2, \dots, n). \quad (8)$$

上述优化问题可归结为以下求解线性方程组:

$$\begin{bmatrix} \Omega + \Theta_\psi & I \\ I^T & 0 \end{bmatrix} * \begin{bmatrix} \alpha^* \\ b^* \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (9)$$

式中, 对角矩阵 $\Theta_\psi = \text{diag}\left\{\frac{1}{\psi\Theta(1)}, \dots, \frac{1}{\psi\Theta(n)}\right\}$ 。由式(9)即可计算出 α^* 和 b^* , 最终获得 WLSSVR 回归估计模型:

$$\hat{y}(x) = \sum_{k=1}^n \alpha_k^* K(x, x_k) + b^*. \quad (10)$$

2 自适应粒子群优化算法 (APSO)

粒子群优化算法是由 Kenney J 和 Eberhart R C 提出的一种基于群体并行搜索全局寻优的算法^[24], 具有原理简单、调节参数少、收敛速度快等特点。采用式(11)和式(12)对粒子速度和位置进行更新:

$$v_{kj}^{t+1} = \tilde{\omega} v_{kj}^t + \mu_1 \cdot r_1 \cdot (\text{pbest}_{kj}^t - x_{kj}^t) + \mu_2 \cdot r_2 \cdot (\text{pgbest}_j^t - x_{kj}^t), \quad (11)$$

$$x_{kj}^{t+1} = x_{kj}^t + v_{kj}^{t+1}, \quad (12)$$

式中, $\tilde{\omega}$ 为惯性权重因子, μ_1 与 μ_2 分别为个体学习因子和社会学习因子, r_1 与 r_2 为介于(0, 1)之间的随机数, t 为当前迭代次数, v_{kj}^t 、 x_{kj}^t 和 pbest_{kj}^t 分别为粒子 p_k 在第 t 次迭代中第 j 维所处的速度、个体位置与个体最优位置; pgbest_j^t 为群体粒子在第 j 维的全局最优位置。

研究发现, 因惯性权重因子 $\tilde{\omega}$ 选择不当, 不能动态调节速度, 是直接影响标准 PSO 算法的收敛效果和性能的关键, 还使算法易出现“早熟”、“振荡”和不收敛等问题而导致搜索失败^[25]。为克服标准 PSO 算法的不足, 给出了一种非线性动态自适应粒子群优化算法 (APSO), 使惯性权重因子 $\tilde{\omega}$ 能够随粒子适应度不同而自动调节, 其计算表达式为

$$\tilde{\omega} = \begin{cases} \tilde{\omega} - \left| \frac{(\tilde{\omega} - \tilde{\omega}_{\min})(f_k - f_{\text{ave}1})}{f_{\text{best}} - f_{\text{ave}1}} \right|, & (f_{\text{ave}} \leq f_k); \\ \tilde{\omega} + \left| \frac{(\tilde{\omega} - \tilde{\omega}_{\min})(f_k - f_{\text{ave}2})}{f_{\text{best}} - f_{\text{ave}2}} \right|, & (f_{\text{ave}} > f_k). \end{cases} \quad (13)$$

式中, $\tilde{\omega}$ 和 $\tilde{\omega}_{\min}$ 分别为惯性权重因子及其最小值; f_k 、 f_{best} 与 f_{ave} 分别当前迭代每个粒子的适应度、粒子群中最优粒子的适应度和粒子的平均适应度; $f_{\text{ave}1}$ 为当前迭代中 f_k 优于 f_{ave} 的所有粒子适应度的均值, $f_{\text{ave}2}$ 为当前迭代中 f_k 劣于 f_{ave} 的所有粒子适应度的均值。根据粒子个体适应度值对粒子群体进行划分, 对 f_k 优于 f_{ave} 较趋向全局最优的粒子赋予较小的惯性权重因子, 以便进行局部寻优; 而对 f_k 劣于 f_{ave} 的粒子, 易出现早熟收敛, 可通过式(13)线性增大其惯性权重, 使该粒子跳出局部极小值, 更快地趋向较好的搜索空间。通过动态调节惯性权重使 APSO 算法不仅保障了粒子多样性, 还有效提高算法全局搜索性能。

3 基于 APSO-WLSSVR 的短期水质预测

3.1 基于 APSO 的 WLSSVR 参数组合优化

文献[26-27], 惩罚系数 ψ 与 RBF 核函数参数 δ^2 直接影响着 WLSSVR 模型的拟合精度和泛化性能, 为提高 WLSSVR 模型的性能, 获得最佳的 ψ 、 δ^2 参数组合是关键^[28]。本研究提出采用 APSO 算法对 WLSSVR 模型参数组合进行优化选择, 其程序流程图如图 1 所示。

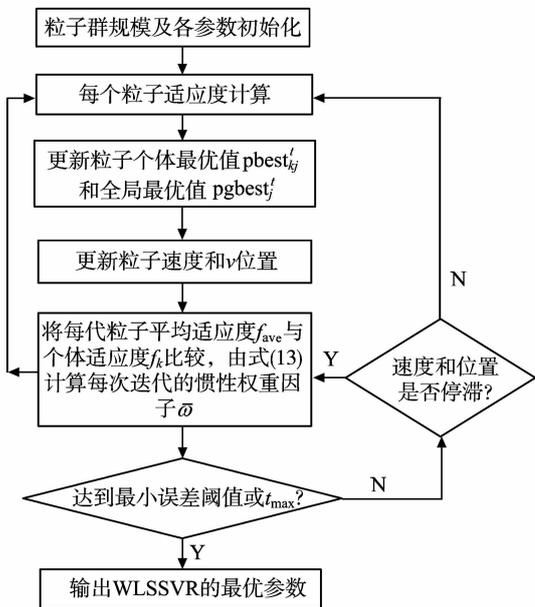


图1 基于 APSO 的 WLSSVR 参数组合优化流程图
Fig.1 Flowchart of parameters optimization by APSO-WLSSVR

在上述优化算法中粒子适应度函数 $f()$ 表达式为

$$f(\psi, \sigma^2) = 1 - \sqrt{\frac{1}{n} \sum_{k=1}^n (\hat{y}_k - y_k)^2}, \quad (14)$$

其中, y_k 和 \hat{y}_k 分别为真实值和预测值。

3.2 基于 APSO-WLSSVR 短期水质预测

集约化水产养殖水质预测是一个复杂的多变量非线性预测问题,也可以被看作一个函数拟合问题。以水质和气象智能传感器采集的生态环境数据集为基础,把对养殖对象有重要影响的水质参数作为函数输出值 \hat{y} ,将监测的历史数据中对该参数有影响的其他生态环境因子如:溶解氧(DO)、水温(WT)、气压(AP)、太阳辐射(SR)、空气湿度(AH)、风速(WV)、风向(WD)等作为函数输入向量 X ,对 APSO-WLSSVR 预测模型进行优化训练,最终构建一个能够处理预测对象与其各影响因子之间复杂的非线性水质预测模型。基于 APSO-WLSSVR 的水质预测模型结构图如图2所示。

具体的水质预测步骤如下:

Step 1 首先对历史水产养殖生态环境收集与分析,通过修补缺失数据或去除异常噪声数据等方式对数据集进行预处理,减少不良数据的影响,然后数据归一化预处理,建立样本训练集和测试集;

Step 2 初始化 APS-WLSSVR 模型的各项参数,将训练样本集输入 APSO-WLSSVR 预测模型中,采用 APSO 算法对 WLSSVR 模型参数组合进行优化选择,运用寻优获得最佳参数 (ψ, δ^2) 建立 WLSSVR 回

归估计模型;

Step 3 再用样本测试集对 WLSSVR 进行训练和性能测试;

Step 4 采用训练优化好的 WLSSVR 预测模型对未来某一时刻的水质进行短期预测。

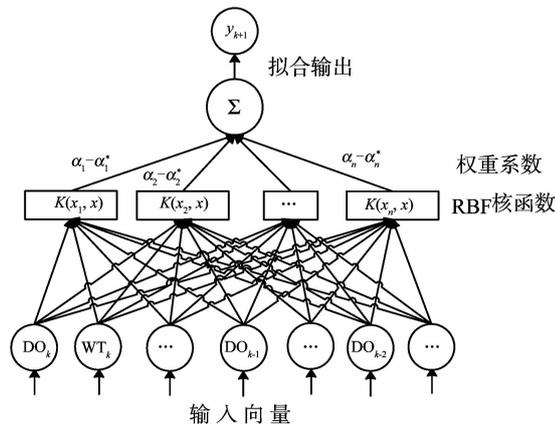


图2 基于 APSO-LSSVR 的水质预测模型结构图
Fig.2 The structure chart of water quality prediction model based on APSO-WLSSVR

4 实验结果与分析

4.1 研究对象及数据源

以江苏省宜兴市河蟹养殖某池塘为试验点,采用北京市农业物联网工程技术研究中心研制的基于 WSN 的集约化水产养殖在线监控系统,每间隔 10 min 对河蟹养殖池塘生态环境数据采样一次,将获取的数据作为验证预测模型的样本集,每个样本由 DO、WT、AP、AH、WV、WD、SR 等 7 项特征指标组成。采样周期为 2012 年 7 月 20 日至 7 月 23 日,共采集 432 个样本,将前 2 天的 288 个样本作为训练集,后一天的 144 个样本作为测试集,对未来某时刻的养殖水质指标进行定量预测。其监测的养殖生态环境原始数据变化曲线如图3所示。

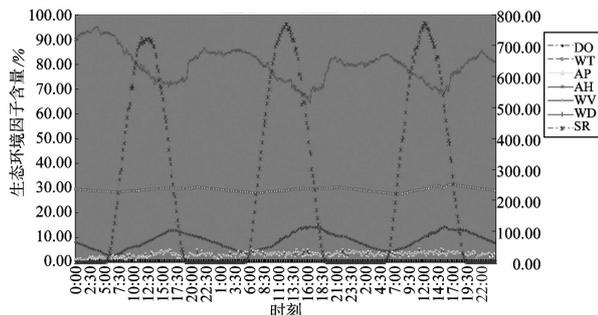


图3 原始数据变化曲线图

Fig.3 The variation curve of the original water quality data

4.2 数据修复及归一化预处理

良好的数据集是准确预测的前提和基础,因此水质预测除采用性能良好的回归估计模型外,还需

要重视原始数据的收集与整理。养殖生态环境数据采集的过程因侵蚀致传感器设备老化性能下降、数据传输丢包、人工增氧与饵料投喂等多种因素的影响,使历史数据中某时刻的数据可能出现数据缺失,产生异常波动或非真实的“不良数据”。数据预处理就是在使用历史数据进行回归估计之前,先对其进行加工修复,消除不规则的噪声数据和修补缺失数据,减少“不良数据”对预测结果的影响。

4.2.1 缺失数据的修补处理

若缺失数据的前后时间间隔不大,采用如式(15)线性插值的方法将其补上。

$$x_{k+i} = x_k + \frac{i \cdot (x_{k+j} - x_k)}{j}, \quad 0 < i < j, \quad (15)$$

式中, x_k 和 x_{k+j} 分别为已知 k 时刻和 $k+j$ 时刻的水质参数值, x_{k+i} 为 $k+i$ 时刻缺失的水质参数取值。若时间间隔较大,则采用天气类型相同或相近相邻几天中同时刻的数据来填补。

4.2.2 错误的“不良数据”处理

因水质数据具有连续性和时序性,前后相邻时段监测的数据一般不会发生跳变,通过比较,若该时刻水质数据变化范围在其前后水质监测值的 $\pm 10\%$ 以外,则认为该数据有误,可采用如式(16)均值平滑法进行水平处理。

当 $|x_k - x_{k-1}| > \vartheta_1$ 或 $|x_k - x_{k+1}| > \vartheta_2$ 时,

$$x_k = \frac{x_{k-1} + x_{k+1}}{2}, \quad (16)$$

ϑ_1 和 ϑ_2 分别为相邻数据误差阈值。

因水质数据具有周期性,不同日期在天气状况类似的前后几天应具有近似的水质特征,类似日期的同一时刻的水质监测值应维持在一定范围内。通过比较,若某时刻水质数据变化范围在与其天气状况类似的前一天、前两天相同时刻的水质数据的 $\pm 10\%$ 以外,则认为该数据异常,可采用如式(17)均值法进行垂直处理。

$$|x_{(d,k)} - \bar{x}'_k| > \vartheta_3, \quad x_{(d,k)} = \begin{cases} \bar{x}'_k + \vartheta_3; \\ \bar{x}'_k - \vartheta_3. \end{cases} \quad (17)$$

其中, $x_{(d,k)}$ 为第 d 天 k 时刻的水质数据, ϑ_3 为天气状况相似同一时刻数据误差阈值, \bar{x}'_k 为待处理数据在天气状况类似最近几天同一时刻水质参数的平均值。

4.2.3 数据归一化处理

为消除因水质样本数据的数量级和量纲不同对预测模型性能的影响,必须对其进行归一化处理,文中采用公式(18)对河蟹养殖生态环境数据进行归一化处理,使其取值统一于某种共同的数值特性范围内。

$$x' = \frac{x - x_{\max}}{x_{\max} - x_{\min}}, \quad (18)$$

式中 x 和 x' 分别为归一化前后的数据, x_{\min} 与 x_{\max} 分别为原始数据的最小值和最大值。

4.3 算法实现及性能对比分析

采用 Matlab7.13 编制仿真程序, APSO-WLSSVR 算法初始化:种群规模为 $m = 50$, $t_{\max} = 150$, $\mu_1 = \mu_2 = 1.49$, $\tilde{\omega} \in [0.9, 0.4]$, 粒子初始速率 v 为 0, 最小误差阈值 $h = 0.005$, $\psi \in [0.05, 200]$, $\delta \in [0.001, 10]$ 。将预测点前 10、20 和 30 min 的生态环境参数值作为输入项并行输入到模型中,下一时刻的溶解氧预测值作为输出项。根据 3.1 节的 APSO 对 WLSSVR 参数组合优化选择的步骤进行寻优,获得 WLSSVR 的最优参数组合 $\sigma^2 = 2.816$, $\psi = 121.72$ 。将参数代入 APSO-WLSSVR 模型中进行水质预测。

为了验证 APSO-WLSSVR 水质预测模型性能,选择标准 LSSVR 算法,三层 7-5-1 网络结构且激励函数为 S 型的 BP 神经网络进行对比分析;分别采用平均绝对误差 (MAE)、平均绝对百分比误差 (MAP)、均方根误差 (RMSE) 等指标对上述算法性能进行评价。以生态环境数据训练集对上述模型进行优化训练,测试集对模型拟合性能进行检验,3 种模型拟合预测曲线与真实值曲线对比图如图 4 所示,模型性能评价结果统计见表 1。

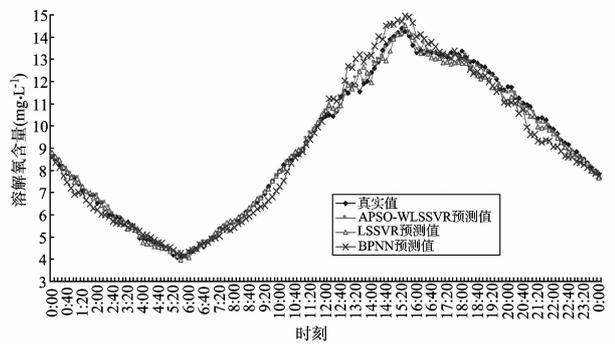


图4 APSO-WLSSVR、标准的LSSVR和BPNN算法预测结果比较

Fig. 4 Comparisons of the results predicted by BPNN, standard LS SVR and APSO-WLSSVR

由图4得知,本研究提出的 APSO-WLSSVR 短期水质预测模型预测曲线与真实值曲线拟合的最好,能够较好地描述水产养殖生态环境因子与预测目标之间的非线性关系;标准 LSSVR 预测曲线拟合效果次之,其结果可能与模型训练优化过程中没有考虑样本重要性的差异,以及采用试凑法人为选择模型参数有关;BP 神经网络预测曲线拟合最差,这与不适宜高维数和小样本特性的 BP 神经网络自身的缺陷有关。

表1 不同预测模型性能比较
Table 1 The performance comparison for different forecast models

方法	RMSE	MAE	MAPE
APSO-WLSSVR	0.175 3	0.128 4	0.014 6
标准 LSSVR	0.233 5	0.179 3	0.023 3
BPNN	0.518 8	0.408 0	0.043 5

由表1可知,APSO-WLSSVR模型与BPNN相比,性能评价指标RMSE、MAE、MAPE分别下降了66.21%、68.53%和66.44%,这表明在小样本数据集的情况下,APSO-WLSSVR比BPNN具有更好的性能和预测精度。APSO-WLSSVR模型与标准LSSVR相比,性能评价指标RMSE、MAE、MAPE分别下降了24.93%、28.39%和37.34%,足以体现采用APSO对WLSSVR参数自动寻优,有效避免了采用试凑法确定WLSSVR参数的盲目性和拟合效果差等问题;此外,以“重近轻远”思想为指导,充分考虑样本重要性的差异将标准LSSVR改进为WLSSVR,有效提高了APSO融合WLSSVR模型的预测精度和泛化性能。总之,本研究中提出的APSO-WLSSVR短期水质预测模型不仅预测精度高、鲁棒性强、泛化性能好,还很适宜于集约化水产养殖水质管理的实际需要,也为其他领域的短期预测提供一种新思路。

5 结论

(1)为解决传统水质预测方法和标准LSSVR算法在回归估计中存在的问题,提出了APSO-WLSSVR的短期水质预测模型,在模型构建过程中,以“重近轻远”思想为指导,将标准LSSVR改进为WLSSVR算法,为不同重要性的训练样本中设置不同的权重系数,实现样本的优化选择;同时采用基于群智能体并行搜索的APSO算法对WLSSVR参数组合进行自适应寻优,避免了标准LSSVR算法因试凑法获取参数的盲目性、人为主观因素的影响,致使算法拟合精度低等问题。将该模型应用于江苏省宜兴市集约化河蟹养殖某池塘水质预测中,与标准WLSSVR和BP神经网络预测模型进行对比分析,结果表明,APSO-WLSSVR算法在一定程度上提高了模型的预测精度,还具有较好的鲁棒性和泛化能力。

(2)良好的数据集是准确预测的前提和基础。数据采集过程中因传感器设备受侵蚀老化、数据传输丢包、人工增氧与投喂等因素造成数据缺失或产生异常波动的“坏数据”,有必要对数据集进行预处

理,消除不规则的噪声数据和修补缺失数据,减少“不良数据”对预测结果的影响,从数据层面来提高预测精度。

(3)通过水质预测,可提前辨识集约化河蟹养殖池塘未来短期内水质变化趋势,为用户提供应对突发水质恶化事件赢得时间,在减少养殖风险,提高生产效益方面,有一定的实用价值。

(4)实用高效的数据修复预处理技术,利用新型机器学习方法优化WLSSVR参数组合,提高模型的预测精度和效率是今后的进一步研究的方向。

参考文献:

- [1] 徐成斌,马溪平,孟雪莲,等. 光合细菌的分离鉴定及在河蟹养殖中的应用[J]. 辽宁大学学报:自然科学版, 2009,36(1):77-81.
XU Chengbin, MA Xiping, MENG Xuelian, et al. Isolation and identification of photosynthetic bacteria and its application in eriocheir sinensis aquiculture[J]. Journal of Liaoning University: Natural Sciences Edition, 2009, 36(1):77-81.
- [2] HATZIKOS E, HÄTÖNEN J, BASSILIADES N, et al. Applying adaptive prediction to sea-water quality measurements[J]. Expert Systems with Application, 2009(36): 6773-6779.
- [3] DELLANA S, WEST D. Predictive modeling for wastewater applications: linear and nonlinear approaches [J]. Environmental Modelling and Software, 2009(24):96-106.
- [4] MAHAPATRA S S, NANDA S K, PANIGRAH B K. A cascaded fuzzy inference system for Indian river water quality prediction [J]. Advances in Engineering Software, 2011(42):787-796.
- [5] José Juan Carbajal Hernández, Luis P Sánchez Fernández, Oleksiy Pogrebnyak. Assessment and prediction of water quality in shrimp culture using signal processing techniques[J]. Aquacult Int, 2011(19):1083-1104.
- [6] YAGIZ S, GOKCEOGLU C. Application of fuzzy inference system and nonlinear regression models for predicting rock brittleness [J]. Expert Systems with Applications, 2010, 37(3): 2265-2272.
- [7] 王小艺,赵晓平,刘载文,等. 基于灰色理论的湖库水体富营养化预测方法研究[J]. 计算机仿真,2011,28(1): 17-19.
WANG Xiaoyi, ZHAO Xiaoping, LIU Zaiwen, et al. Research on lake eutrophication forecasting methods based on grey theory[J]. Computer Simulation, 2011, 28(1): 17-19.
- [8] TAN G H, YAN J Z, GAO C, et al. Prediction of water

- quality time series data based on least squares support vector machine[J]. *Procedia Engineering*, 2012(31):1194-1199.
- [9] 朱跃龙,李士进,范青松,等. 基于小波神经网络的水文时间序列预测[J]. *山东大学学报:工学版*, 2011, 41(4):119-124.
ZHU Yuelong, LI Shijin, FAN Qingsong, et al. Wavelet-neural network model based complex hydrological time series prediction[J]. *Journal of Shandong University: Engineering Science*, 2011, 41(4):119-124.
- [10] WU J, LU J, WANG J Q. Application of chaos and fractal models to water quality time series prediction[J]. *Environmental Modelling & Software*, 2009(24):632-636.
- [11] ABAURREA J, ASÍN J, CEBRIÁN A C, et al. Trend analysis of water quality series based on regression models with correlated errors [J]. *Journal of Hydrology*, 2011(400):341-352.
- [12] FARUK D Ö. A hybrid neural network and ARIMA model for water quality time series prediction[J]. *Engineering Applications of Artificial Intelligence*, 2010(23):586-594.
- [13] PALANI S, LIONG SY, TKALICH P. An ANN application for water quality forecasting[J]. *Marine Pollution Bulletin*, 2008(56):1586-1597.
- [14] EYNARD J, GRIEU S, POLIT M. Wavelet-based multi-resolution analysis and artificial neural networks for forecasting temperature and thermal power consumption[J]. *Engineering Applications of Artificial Intelligence*, 2011, 24(3):501-516.
- [15] HAN H G, CHEN Q L, QIAO J F. An efficient self-organizing RBF neural network for water quality prediction [J]. *Neural Networks*, 2011(24):717-725.
- [16] WEST D, DELLANA S. An empirical analysis of neural network memory structures for basin water quality forecasting [J]. *International Journal of Forecasting*, 2011, 27:777-803.
- [17] SINGH K P, BASANT N, GUPTA S. Support vector machines in water quality management [J]. *Analytica Chimica Acta*, 2011(703):152-162.
- [18] LIU Lixia, ZHUANG Yiqi, LIU Xueyong. Tax forecasting theory and model based on SVM optimized by PSO [J]. *Expert Systems with Applications*, 2011(38):116-120.
- [19] HONG Weichang, DONG Yucheng, CHEN Liyue, et al. SVR with hybrid chaotic genetic algorithms for tourism demand forecasting [J]. *Applied Soft Computing*, 2011(11):1881-1890.
- [20] SUYKENS J. Least squares support vector machines [M]. Singapore: World Scientific Publishing Co Pte Ltd, 2002:71-89.
- [21] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers [J]. *Neural Process Lett*, 1999(9):293-300.
- [22] CHUA K S. Efficient computations for large least square support vector machine classifiers [J]. *Pattern Recogn Lett*, 2003(24):75-80.
- [23] TAN Guohua, YAN Jianzhuo, GAO Chen, et al. Prediction of water quality time series data based on least squares support vector machine [J]. *Procedia Engineering*, 2012(31):1194-1199.
- [24] KENNEDY J, EBERHART R C, SHI Y. *Swarm intelligence* [M]. San Francisco, CA: Morgan Kaufmann Publishers Inc, 2001.
- [25] 韩江洪,李正荣,魏振春. 一种自适应粒子群优化算法及其仿真研究[J]. *系统仿真学报*, 2006, 18(10):2969-2971.
HAN Jianghong, LI Zhengrong, WEI Zhenchun. Adaptive particle swarm optimization algorithm and simulation [J]. *Journal of System Simulation*, 2006, 18(10):2969-2971.
- [26] QU Jian, ZUO Ming J. An LSSVR-based algorithm for online system condition prognostics [J]. *Expert Systems with Applications*, 2012(39):6089-6102.
- [27] K de Brabanter, KARSMAKERS P, J de Brabanter, et al. Confidence bands for least squares support vector machine classifiers: a regression approach [J]. *Pattern Recognition*, 2012(45):2280-2287.
- [28] 王雪松,程玉虎,郝名林. 一种支持向量机参数选择的改进分布估计算法 [J]. *山东大学学报:工学版*, 2009, 39(3):7-10.
WANG Xuesong, CHENG Yuhu, HAO Minglin. Parameters selection of a support vector machine using an improved estimation of the distribution algorithm [J]. *Journal of Shandong University: Engineering Science*, 2009, 39(3):7-10.

(编辑:胡春霞)