

基于PCA-MCAFA-LSSVM的养殖水质pH值预测模型*

刘双印^{1,2} 徐龙琴¹ 李振波^{2,3} 李道亮^{3,4}

(1. 广东海洋大学信息学院, 湛江 524025; 2. 中国农业大学农业部农业信息获取技术重点实验室, 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:为解决水质预测传统方法精度低、鲁棒性差等问题,提出了基于主成分分析(PCA)、改进文化鱼群算法(MCAFA)和最小二乘支持向量机(PCA-MCAFA-LSSVM)的养殖水质pH值预测模型。该模型通过主成分分析提取养殖生态环境指标的主成分,降低模型输入向量维数,利用改进文化鱼群算法对最小二乘支持向量机超参数进行组合优化,以自动获取最优超参数建立非线性养殖水质pH值预测模型。应用该模型对宜兴市河蟹养殖某池塘2011年9月1日~9月4日在线监测的水质数据进行了预测分析,试验结果表明:该模型取得较好的预测效果,与分别用蚁群算法或遗传算法优化LSSVM的方法相比,PCA-MCAFA-LSSVM模型有93.05%的测试样本绝对误差小于8%,最大绝对误差仅为11.61%,均方根误差、平均相对误差绝对值和运行时间分别为0.0474、0.0041和4.367s,且均优于其他预测方法。PCA-MCAFA-LSSVM算法不仅计算速度快、预测精度高,还能够为河蟹养殖水质调控管理提供决策依据。

关键词: 养殖水质 pH值预测 文化鱼群算法 最小二乘支持向量机 参数优化 主成分分析

中图分类号: TP391 **文献标识码:** A **文章编号:** 1000-1298(2014)05-0239-08

引言

水产养殖水质预测是养殖区域水环境规划、现代渔业健康养殖精准化管理的重要基础性工作,准确的水质预测对养殖水质科学化调控,防范水质恶化和水产品疾病爆发具有十分重要的经济价值和现实意义。而在集约化河蟹养殖池塘中水质pH值过高或过低都会影响河蟹的新陈代谢及其他生理功能,胁迫河蟹正常生长,为此,根据集约化河蟹养殖水质调控管理的应用需求,选定pH值为水质预测对象。水质指标pH值变化受众多因素交叉影响,作用机理复杂,如何提高其预测的准确性一直是国内外学者关注和研究的热点。

目前常用的水质预测方法有水质模拟法、专家评估法、基于数理统计的多元回归法、灰色理论法等^[1],这些方法对于大样本、低维数的趋势性因素线性回归估计比较准确,但对具有非线性、多参数、模糊不确定性、高维数等特点的养殖水质模拟效果不理想。神经网络法具有较好的非线性预测能力,

但预测过程中存在易陷于局部极值、过学习、不适于高维数、小样本等不足^[2-5]。最小二乘支持向量机是支持向量回归机的一种扩展^[6],能够克服上述预测方法存在的缺陷,以任意精度逼近任意函数,在模式识别、回归估计领域取得了一些成效,但最小二乘支持向量机算法性能仍然受制于超参数(C, σ^2)选择的是否合理。最近,一些学者采用遗传算法^[7]、粒子群算法^[8-9]、蚁群算法^[10]等群智能算法对支持向量机超参数进行寻优,改善了算法的学习能力和预测精度。

因pH值预测需要多种生态环境因子作为输入,但因子间存在一定的信息重叠或冗余,若不进行关键因子的筛选,就会影响pH值的预测结果。而主成分分析法(Principal components analysis, PCA)具有在不损失或尽量少损失原有指标信息的情况下,能够对数据降维,消除各个自变量相关性的作用^[11]。此外,改进文化鱼群算法在解决多目标优化问题时具有较强的全局搜索能力。为此,本文将主成分分析法、改进文化鱼群算法与最小二乘支持向

收稿日期: 2013-06-24 修回日期: 2013-07-17

*“十二五”国家科技支撑计划资助项目(2011BAD21B01)、广东省科技计划资助项目(2012A020200008、2011B040200034、2012B091100431)、广东省自然科学基金资助项目(S2013010014629、S2012010008261)、广东省省部产学研结合专项资金资助项目(2012B090500008)和宁波市农业重点科技攻关资助项目(2011C11006)

作者简介: 刘双印,副教授,中国农业大学博士生,主要从事智能计算、智能信息处理、农业信息化技术研究, E-mail: hdslyxlq@126.com
通讯作者: 李道亮,教授,博士生导师,主要从事农业先进传感与智能信息处理研究, E-mail: dliangl@cau.edu.cn

量机相结合,建立基于改进文化鱼群优化最小二乘支持向量机的养殖水质 pH 值预测模型(PCA - MCAFA - LSSVM)。采用主成分分析法筛选出 pH 值关键影响因子,改进文化鱼群算法对最小二乘支持向量机超参数进行组合优化,以最佳超参数组合构建河蟹养殖生态环境因子 pH 值、溶解氧、水温、氧化还原电位与未来某一时刻的 pH 值之间的非线性关系模型,并应用模型对江苏省宜兴市集约化河蟹养殖某池塘水质进行实证分析,验证算法的有效性。

1 试验

1.1 试验区域

试验区域为中国农业大学-宜兴市水产养殖物联网应用示范基地。该示范基地覆盖 4 个乡镇约 667 hm²河蟹养殖池塘。每个河蟹养殖池塘大小为 3.4 hm²的标准池塘,具有水泵、增氧机、用于监测养殖生态环境数据的多种智能传感器,以及基于物联

网的集约化水产养殖水质远程在线监控系统。

1.2 数据获取与数据源

以宜兴市高塍镇水产养殖物联网应用示范基地河蟹养殖某池塘水质为研究对象,采用中国农业大学研制的基于物联网的集约化水产养殖生态环境远程在线监控系统获取河蟹养殖生态环境数据,其系统架构如图 1 所示。该系统主要由数据感知层、传输层、信息智能处理和应用层组成,分别实现数据获取、无线传输、智能信息处理、水质调控及精细喂养辅助决策等功能^[10]。

采用上述远程在线监控系统每间隔 10 min 对 pH 值、水温、氧化还原电位、溶解氧、浊度、叶绿素、电导率、太阳辐射、气温、气压等河蟹养殖生态环境数据在线采样一次,将 2011 年 9 月 1 日至 9 月 4 日在线采集的 576 个样本数据作为研究的数据源,从中选择 432 个样本作为训练集,144 个样本作为测试集,分别用来训练和验证水质预测模型性能。

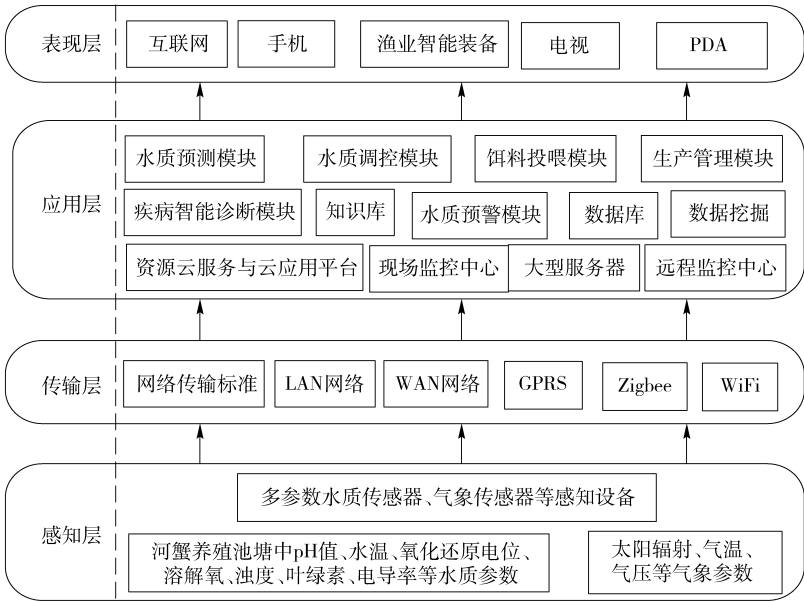


图 1 远程在线监控系统拓扑结构框图

Fig. 1 Topology structure diagram of the remote on-line monitoring system

1.3 建模方法

采用主成分分析法、改进文化鱼群算法和最小二乘支持向量机相结合的方法,建立基于 PCA - MCAFA - LSSVM 的 pH 值预测模型。即运用文献[12]的数据修复方法对采集的养殖生态环境数据进行分析,奇异值剔除、消除不规则的噪声数据和修补缺失数据等预处理,以减少“不良数据”对水质预测结果的影响;采用 SPSS 软件对河蟹养殖生态环境数据进行主成分分析,筛选出 pH 值关键影响因子,作为预测模型输入向量,以对未来某一时刻的 pH 值进行预测估计;以筛选的主成分指标数据为基础,通过改进文化鱼群算法对 LSSVM 预测模型参数

组合优化训练,获得最佳预测模型,并对未来某一时刻 pH 值进行预测。同时分别以蚁群算法优化最小二乘支持向量机(ACO - LSSVM)模型和遗传算法优化最小二乘支持(GA - LSSVM)模型进行预测性能对比分析。

1.4 养殖水质参数 pH 值关键影响因子筛选

为消除数据间的多重共线性,采用主成分分析法对养殖生态环境数据降维和分析,提取养殖水质参数 pH 值的关键影响因子。其 pH 值的关键影响因子筛选过程为:①集约化河蟹养殖生态环境的原始数据标准化处理。②建立标准化后的养殖生态环境指标的相关系数矩阵。③计算相关矩阵的特征值

及相应的特征向量。④计算贡献率和累计贡献率。⑤计算主成分载荷矩阵,筛选出 pH 值的关键影响因素。

根据上述步骤采用 SPSS 统计分析软件对 pH 值的关键影响因素进行筛选,得到主成分的特征值与各自贡献率,如表 1 所示。由表 1 可知,前 3 个因子的累计方差贡献率达到 90.315%,符合累计方差贡献率大于等于 85% 作为提取主成分的原则,因此,可确定前 3 个因子可代替原变量。

表 1 生态环境因子的特征值与累计方差贡献率

Tab.1 Eigenvalue and cumulative proportion of ecologic environmental factors

主因子	特征值	贡献率/%	累计方差贡献率/%
1	4.967	49.674	49.674
2	2.596	25.965	75.631
3	1.468	14.658	90.315
4	0.514	5.137	95.452
5	0.222	2.221	97.672
6	0.111	1.114	98.787
7	0.061	0.607	99.394
8	0.035	0.347	99.741
9	0.019	0.186	99.927
10	0.007	0.073	100

因部分养殖生态环境因子数据间的相关性较强,采用方差极大斜交旋转,计算得到因子在各主因子上的因子荷载,如表 2 所示。由表 2 可知,pH 值和溶解氧对第 1 主因子的相关系数绝对值都超过了 0.85,相对其他因子贡献较大;水温对第 2 主因子贡献最大;氧化还原电位对第 3 主因子贡献最大。因此,本文选用 pH 值、溶解氧、氧化还原电位和水温作为主成分指标,也与养殖领域专家靠经验选择 pH 值的关键影响因素基本一致,并以此构建未来某一时刻 pH 值预测模型的输入样本。

表 2 生态环境因子方差极大斜交旋转因子载荷矩阵

Tab.2 Oblique rotation factor loading matrix of ecologic environmental factors

因子	主因子 1	主因子 2	主因子 3
pH 值	0.965	0.111	0.164
水温	0.423	0.896	0.731
氧化还原电位	-0.291	-0.120	-0.937
溶解氧	0.861	0.298	0.354
浊度	0.116	-0.786	0.094
叶绿素	0.739	-0.478	-0.161
电导率	0.349	0.278	0.817
太阳辐射	0.257	0.549	0.354
气温	0.353	0.578	0.446
气压	0.079	0.461	-0.746

2 改进文化鱼群算法优化最小二乘支持向量机的软测量模型

2.1 最小二乘支持向量机

设训练集 $T = \{(x_i, y_i) | x_i \in \mathbf{R}^n, y_i \in \mathbf{R}\}_{i=1, \dots, n}$ 。在非线性映射函数 $\varphi(x)$ 构成的高维特征空间中,为求解最优拟合函数 $y(x) = \mathbf{w}^T \varphi(x) + b$,可将 LSSVM 的约束优化问题转换为带有惩罚参数 C 的线性方程组^[13-14]

$$\begin{bmatrix} \boldsymbol{\Omega} + C^{-1}\mathbf{E} & \mathbf{I} \\ \mathbf{I}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}$$

(1)

式中, $\boldsymbol{\Omega} = \boldsymbol{\phi}(x_i)^T \boldsymbol{\phi}(x_j) = \mathbf{K}(x_i, x_j)$, $\mathbf{E}$ 是 $n \times n$ 维的单位矩阵, $\mathbf{I}^T = [1, \dots, 1]^T$, $\mathbf{a} = [a_1, a_2, \dots, a_n]$, $\mathbf{y} = [y_1, \dots, y_n]^T$,通过求解线性方程组(1),即可得到在点 x 处的 LSSVM 拟合决策函数

$$\hat{y} = \text{sign} \left(\sum_{i=1}^n \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \right)$$

(2)

本文选择 RBF 径向基核函数作为 LSSVM 的核函数,最后获得 LSSVM 回归方法的决策表达式^[13]

$$\hat{y} = \sum_{i=1}^n a_i \exp \left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2} \right) + b$$

(3)

式中 $\hat{y}$ ——决策函数输出值
 $\mathbf{x}$ ——任意一个输入样本向量
 $\mathbf{x}_i$ ——RBF 核函数中心
 σ^2 ——RBF 核函数的宽度系数

2.2 改进文化鱼群算法

2.2.1 人工鱼群算法

人工鱼群算法(Artificial fish-fwarm algorithm, AFSA)是一种基于模拟鱼群行为的仿生型优化算法^[15],其基本思想是利用鱼的觅食、聚群和追尾行为,从构造单条鱼的底层行为做起,通过鱼群中各个体的局部寻优,达到全局最优,并具有对初值和参数选择不敏感、全局性、快速简单、易实现等特点^[16]。虽然人工鱼群在参数全局优化、参数辨识与估计等领域取的成效,但在标准人工鱼群算法中人工鱼为获得较优位置常采用试探性方式进行搜索,不能有效利用已获取的先验知识,存在较大的盲目性和随机性,致使算法收敛速度慢,后期易陷入局部最优的现象发生,制约着人工鱼群算法的推广应用。

2.2.2 文化算法

文化算法(Culture algorithm, CA)是 Reynolds 于 1994 年提出一种新型进化算法^[17],它将知识机制引入到种群进化过程中提取有用的信息,指导种群搜索过程,以加快种群的进化速度和提高其对环境适应能力。文化算法框架主要由种群空间、信念空间和通信协议三元素构成,其中种群空间和信念

空间分别基于种群和信念文化进行进化,通信协议又由接收函数 $\text{accept}()$ 、选择函数 $\text{select}()$ 、影响函数 $\text{influence}()$ 、更新函数 $\text{update}()$ 、目标函数 $\text{objective}()$ 等构成,用来明确种群和信念空间进行信息交换所需遵循的规则^[18],其算法的基本框架如图 2 所示^[19]。

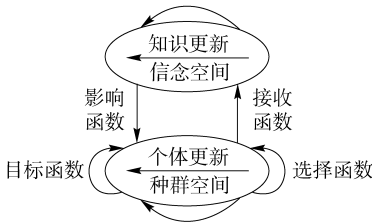


图 2 文化算法框架
Fig. 2 Framework of cultural algorithm

2.2.3 改进文化鱼群算法

为弥补标准人工鱼群算法搜索过程中存在的问题,本文以文化算法的框架模型为基础,将文化算法、人工鱼群算法和能有效摆脱局部极值束缚的柯西变异相结合,构建了改进的文化鱼群算法(Modified cultural artificial fish-swarm algorithm, MCAFA)。算法主要思想是在人工鱼群进化过程中,随着食物浓度和公告板的不断更新,将持续获取的全局最优解作为情景知识或规范知识保存并更新信念空间,利用信念空间知识对人工鱼群体进化的步长和方向进行指导^[17],引入柯西变异算子改变人工鱼的状态,尽可能地搜索其他区域的可能解。通过柯西变异算子、人工鱼种群空间和信念空间的双重进化和互相作用,该算法不仅可避免人工鱼群搜索过程的盲目性和随机性,加快其收敛速度,还克服算法后期易陷入局部极值的缺陷,从而保持鱼群的多样性,有效提高了算法的全局搜索性能。

文化鱼群算法的信念空间采用 $\langle S, N \rangle$ 结构对表示^[17],情景知识 $S = \langle S^k | S^k = \{s_1^k, \dots, s_n^k\} \rangle$ 为一组由历代鱼群所产生的最优人工鱼集合,为人工鱼个体提供学习的典范,第 k 步种群中第 i 个最优个体记作 s_i^k , n 为 S 集合的规模。规范知识 $N = [N_1, \dots, N_m]$ 描述了优化问题的解空间^[18],为人工鱼个体提供进化指导原则和行为准则, m 为变量的个数;每个变量的规范知识 N_i 描述为 $\langle I_i, L_i, U_i \rangle (i = 1, \dots, m)$, 其中, $I_i = [L_i, u_i] = \{x_i | L_i \leq x_i \leq u_i, x_i \in \mathbf{R}\}$, 由问题的变量 x_i 取值范围对上限 u_i 和下限 L_i 进行初始化, U_i 和 L_i 分别为第 i 个人工鱼个体的上限 u_i 和 下限 L_i 所对应的人工鱼群适应度值。因篇幅所限,情景知识和规范知识的更新过程、种群空间与信念空间的交互设计、人工鱼的觅食行为、追尾行为、群聚行为不再赘述,详见文献[20-22]。

在文化鱼群算法后期随着搜索速度减缓,鱼群表现出强烈的趋同性。当连续两代位置最优值无变化或连续两代人工鱼个体食物浓度相对误差 τ 小于 0.1,且算法仍不满足结束条件时,则表明算法鱼群丧失了多样性,陷入早熟状态。则对人工鱼父代个体 x_i^k 执行变异操作生成子代个体 $x_{i,q}^{m(k)}$,其表达式为^[21]

$$x_{i,q}^{m(k)} = \begin{cases} l_q^k + (u_q^k - l_q^k)N(0,1) & (x_{i,q}^k \notin S) \\ x_{i,q}^k + x_{i,q}^k \text{Cauchy}(0,1) & (x_{i,q}^k \in S) \end{cases} \quad (4)$$

式中 $N(0,1)$ ——服从标准正态分布的随机数
 $\text{Cauchy}(0,1)$ ——服从柯西分布的随机数
 $x_{i,q}^k$ ——第 k 代时第 i 个人工鱼的第 q 个变量值
 u_q^k, l_q^k ——第 k 代人工鱼第 q 个变量取值的上限和下限

由此可见,改进文化鱼群算法将柯西变异算子与进化过程隐含知识有效结合,在人工鱼群进化过程中由信念空间知识引导人工鱼群体进化的步长和搜索方向,实现在较大空间的快速搜索;而在算法后期的平坦易陷入极值区域,采用柯西变异算子快速跳出局部最优和保持鱼群多样性,增强了算法的全局搜索能力,使组合优化问题求解的精度得到提高。

2.3 改进文化鱼群算法优化 LSSVM 模型

在 LSSVM 模型中惩罚参数 C 用来调节 LSSVM 置信范围和经验风险的比例, C 越大模型的拟合精度越高,而泛化能力越差^[23]。而 RBF 核函数宽度系数 σ^2 实际上是隐含地改变映射函数,从而改变样本在特征空间分布的复杂程度^[24]。参数 C 和 σ^2 很大程度上决定着 LSSVM 模型的学习能力和泛化性能,为此,合理选择最佳的参数组合 (C, σ^2) 是提高 LSSVM 模型拟合性能的关键。

利用改进文化鱼群算法优化 LSSVM 超参数的算法具体实现步骤如下:

(1)文化鱼群 (C, σ^2) 初始化,包括人工鱼的可视域 V 、拥挤度因子 δ 、人工鱼的最大移动步长 $S_{\max}$ 、最大试探次数 $n_{\max_try_number}$ 、最大进化代数 $E_{\max}$ 等参数设置;并在参数 $C \in [0, 500]$ 和 $\sigma^2 \in [0, 80]$ 的可行域内,随机生成规模为 N 的人工鱼群体空间,每条人工鱼代表不同输入权值的最小二乘支持向量机。

(2)优选最小二乘支持向量回归机模型超参数的目的是为获得较高的预测精度,为此,文化人工鱼的食物浓度函数定义为

$$f(x) = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5)$$

式中 $\hat{y}_i, y_i$ ——第 i 代预测值和真实值

文化人工鱼食物浓度越小,获得 LSSVM 超参数组合越好。

(3)养殖生态环境样本数据读取,计算并比较初始鱼群中人工鱼个体当前位置的食物浓度值,取食物浓度为最小值者进入公告板,按照信念空间的结构初始化信念空间。

(4)由文化机制的影响函数对各个人工鱼个体分别模拟觅食、群聚和追尾行为,动态调整人工鱼的位置。

(5)计算人工鱼新位置上的食物浓度,若新位置上的人工鱼食物浓度值优于父代,则用子代替父代;由接收函数选择优异的人工鱼,利用文化算法更新信念空间^[22],并把相应的知识存入公告板。

(6)如果满足连续两代公告板最优值无变化或连续两代人工鱼个体食物浓度相对误差 $\tau = \frac{f(x_i^k) - f(x_i^{k-1})}{f(x_i^{k-1})} < 0.1$ 的条件,则采用式(4)对步骤(3)中的人工鱼个体执行柯西变异操作。

(7)循环迭代终止条件判断,若达到最大进化代数 $E_{\max}$ 或食物浓度小于指定的精度阈值,则算法结束,输出当代最优个体,解码得到 LSSVM 最佳超参数 (C, σ^2) 组合,获得最优基于 MCAFA-LSSVM 的预测模型;否则转到步骤(4)继续执行。

3 实证分析

3.1 实证环境及算法实现

基于 PCA-MCAFA-LSSVM 预测模型的开发环境为:酷睿 i5 3210M (4 核),内存 4 GB,Win7 64 位操作系统,编程语言 Matlab7.14 试验仿真环境。文化鱼群算法初始化为:人工鱼种群规模 $N = 100$,可视域 $V = 0.7$,最大移动步长 $S_{\max} = 0.5$,拥挤度因子 $\delta = 3.618$,最大进化代数 $E_{\max} = 500$,最大试探次数 $n_{\max_try_number} = 20$ 。通过 MCAFA 算法对 LSSVM 模型超参数进行优选,获得最佳超参数 $C = 184.36$ 、 $\sigma^2 = 2.079$;将最佳超参数组合代入 PCA-MCAFA-LSSVM 算法中建立 pH 值预测模型。

3.2 实证结果分析

为验证 PCA-MCAFA-LSSVM 模型性能,选择相同种群规模和最大迭代次数的 PCA-ACO-LSSVM^[10]和 PCA-GA-LSSVM^[7](交叉和变异概率分别为 0.6 和 0.03)进行对比分析,以相同数据源为基础,由训练集对上述模型进行优化训练,测试集对模型预测性能进行验证,采用均方根误差(RMSE)、平均相对误差绝对值(MAPE)和运行时间 t 分别作为预测方法的评价指标,各个预测算法拟合曲线和实测值曲线对比图如图 3 所示。

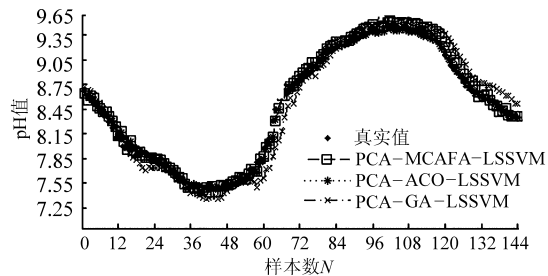


图3 不同模型的预测结果比较

Fig.3 Comparison of forecasting results for different models

由图 3 可知,不同预测模型预测结果总体趋势与实测值比较吻合;而本文提出的 PCA-MCAFA-LSSVM 的预测曲线比 PCA-ACO-LSSVM 和 PCA-GA-LSSVM 的预测曲线更接近实测值曲线,PCA-ACO-LSSVM 预测效果次之,PCA-GA-LSSVM 预测曲线与实测值曲线拟合度最差,从而说明改进文化鱼群算法优化 LSSVM 超参数时,能获得较好的超参数组合,使基于 PCA-MCAFA-LSSVM 的水质预测方法具有更好的泛化性能,能够较好地描述河蟹养殖生态环境因子与 pH 值之间复杂的非线性关系。

将未经主成分分析法处理的河蟹养殖生态环境数据作为数据源,对 MCAFA-LSSVM 模型进行训练优化和预测,其预测结果如图 4 所示。由图 4 知,MCAFA-LSSVM 模型在大部分时刻都能取得较好的预测结果,但也存在预测值与实测值严重偏离的点。

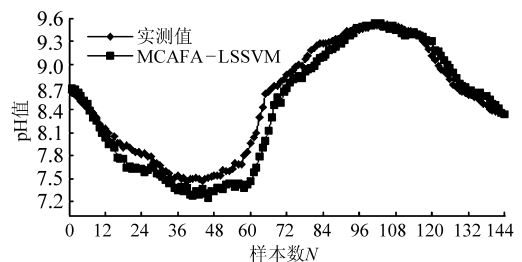


图4 MCAFA-LSSVM 模型预测结果

Fig.4 Forecasting results of MCAFA-LSSVM models

图 5 描述了基于 PCA-MCAFA-LSSVM 的 pH 值预测模型的绝对误差变化曲线,其中绝对误差小于 8% 的样本达到 93.05%,最大绝对误差为 11.61%,说明 PCA-MCAFA-LSSVM 预测模型具有较高的预测精度。

由表 3 可知,在预测条件基本相同的情况下,PAC-MCAFA-LSSVM 方法与 PCA-ACO-LSSVM 相比,其评价指标 RMSE、MAPE 和运行时间 t 分别下降了 19.5%、21.2% 和 1.312 s;PCA-MCAFA-LSSVM 方法与 PCA-GA-LSSVM 相比,其评价指标 RMSE、MAPE 和运行时间 t 分别下降了 58.5%、59.4% 和 2.945 s。PAC-MCAFA-LSSVM 方法与

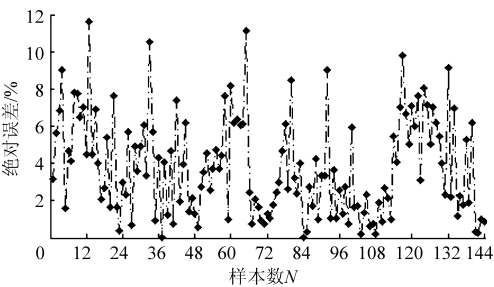


图5 PCA-MCAFA-LSSVM 模型预测绝对误差

Fig.5 Absolute errors forecasted by the
PCA-MCAFA-LSSVM model

MCAFA-LSSVM 相比,其评价指标 RMSE、MAPE 和运行时间 t 分别下降了 38.1%、40.58% 和 6.506 s。对上述数据分析可知:① 在同种条件下,不同的最小二乘支持向量机超参数组合优化方法获得的超参数使养殖水质 pH 值预测模型精度和性能存在一定的差异。② 养殖生态环境指标数据经过 PCA 处理后,预测模型输入向量维数减少且主要影响因素分布更加集中,因此预测精度最高,在算法运行时间上,未经 PCA 处理的运算时间是降维后的 2.49 倍。③ 对于各项预测模型评价指标,本文提出的 PCA-MCAFA-LSSVM 算法远远优于 PCA-GA-LSSVM 算法和 MCAFA-LSSVM 算法,略高于 PCA-ACO-LSSVM 算法,说明经过 PCA 方法筛选后的养殖生态

表3 不同模型预测结果精度分析
Tab.3 Precision analysis of forecast
results for different models

预测模型	RMSE	MAPE	运行时间 t/s
MCAFA-LSSVM	0.076 1	0.006 9	10.873
PCA-GA-LSSVM	0.114 3	0.010 1	7.312
PCA-ACO-LSSVM	0.058 9	0.005 2	5.679
PCA-MCAFA-LSSVM	0.047 4	0.004 1	4.367

环境指标数据用于 MCAFA 算法优化 LSSVM 模型能获得较好的超参数组合,使本文算法具有更高的预测精度、泛化性能和较快的计算速度,能够为集约化河蟹水质预警和水质调控管理提供决策依据。

4 结论

(1) 针对传统水质预测方法预测精度低、鲁棒性差等问题,提出了基于 PCA-MCAFA-LSSVM 的养殖水质 pH 值预测模型,并对江苏省宜兴市集约化河蟹养殖池塘水质 pH 值进行建模和预测。结果表明,该预测模型具有较好预测效果,能够满足河蟹养殖水质 pH 值预测及水质精准化调控管理的实际需要。

(2) 为解决 LSSVM 模型超参数选择的盲目性和人为主观影响等问题,提出了改进文化鱼群优化 LSSVM 模型超参数 (C, σ^2) 的组合优化方法,有效增强了算法的全局搜索能力,提高了参数组合优化问题求解的精度,实现了 LSSVM 最优超参数 (C, σ^2) 的自动获取,改善了 LSSVM 算法的学习能力和预测精度。

(3) 通过预测试验表明,经过主成分分析降维后的预测模型输入向量比未经 PCA 处理的输入向量在预测精度和运行时间上都有明显的优势,说明主成分分析保留了对 pH 值预测最有效的生态环境信息,同时剔除了增加运算时间和干扰预测性能的噪声数据。从预测性能和运行时间等方面综合考虑,所提 PCA-MCAFA-LSSVM 模型预测效果最好,能够更好地拟合高密度河蟹养殖生态环境影响因子与 pH 值之间的非线性关系,为集约化河蟹养殖水质 pH 值的预测提供了一条新途径,并在减少河蟹养殖风险,提高生产效益方面,有一定的实用价值。

参 考 文 献

1 Dellana S, West D. Predictive modeling for wastewater applications: linear and nonlinear approaches [J]. Environmental Modelling and Software, 2009,24(1):96-106.

2 Faruk D Ö. A hybrid neural network and ARIMA model for water quality time series prediction [J]. Engineering Applications of Artificial Intelligence, 2010,23(4):586-594.

3 Palani S, Liong S Y, Tkalic P. An ANN application for water quality forecasting[J]. Marine Pollution Bulletin, 2008,56(9):1586-1597.

4 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(7):717-725.

5 West D, Dellana S. An empirical analysis of neural network memory structures for basin water quality forecasting[J]. International Journal of Forecasting, 2011,27(3):777-803.

6 Suykens J A K, van Gestel T, de Brabanter J, et al. Least squares support vector machines[M]. Singapore: World Scientific, 2002:71-76.

7 Ibrahim B A, Ahmet A. A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm[J]. Information Sciences, 2013,233:25-35.

8 Liao R J, Zheng H B, Grzybowski S, et al. Particle swarm optimization-least squares support vector regression based forecasting

model on dissolved gases in oil-filled power transformers[J]. Electric Power Systems Research, 2011,81(12):2074 – 2080.

9 刘伟,王建平,刘长虹,等. 基于粒子群寻优的支持向量机番茄红素含量预测[J]. 农业机械学报,2012,43(4):143 – 147.
Liu Wei, Wang Jianping, Liu Changhong, et al. Lycopene content prediction based on support vector machine with particle swarm optimization[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012,43(4):143 – 147. (in Chinese)

10 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.
Liu Shuangyin, Xu Longqin, Li Daoliang, et al. Dissolved oxygen prediction model of eriocheir sinensis culture based on least squares support vector regression optimized by ant colony algorithm [J]. Transactions of the CSAE, 2012, 28(23):167 – 175. (in Chinese)

11 李慧,祁力钧,张建华,等. 基于 PCA – SVM 的棉花出苗期杂草类型识别[J]. 农业机械学报,2012,43(9):184 – 189.
Li Hui, Qi Lijun, Zhang Jianhua, et al. Recognition of weed during cotton emergence based on principal component analysis and support vector machine [J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(9):184 – 189. (in Chinese)

12 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):79 – 85.
Xu Longqin, Liu Shuangyin. Water quality prediction model based on APSO – WLSSVR [J]. Journal of Shandong University: Engineering Science, 2012, 42(5):79 – 85. (in Chinese)

13 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.

14 张秀芝,王龙山,于忠党. 基于最小二乘支持向量回归的小零件精密测量技术[J]. 农业机械学报,2009,40(8):189 – 192.
Zhang Xiuzhi, Wang Longshan, Yu Zhongdang. Precision measurement technology for small parts using LSSVR[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(8):189 – 192. (in Chinese)

15 李晓磊,邵之江,钱积新. 一种基于动物自治体的寻优模式:鱼群算法[J]. 系统工程理论与实践,2002,22:32 – 38.
Li Xiaolei, Shao Zhijiang, Qian Jixin. An optimizing method based on autonomous animats: fish-swarm algorithm[J]. Systems Engineering-Theory & Practice, 2002, 22:32 – 38. (in Chinese)

16 刘双印. 基于改进 AFSA 算法的 BP 神经网络的研究[J]. 计算机工程与设计,2009,30(20):4719 – 4721.
Liu Shuangyin. Research on BP neural networks based on improved artificial fish-swarm algorithm [J]. Computer Engineering and Design, 2009, 30(20):4719 – 4721. (in Chinese)

17 Robert R G. An introduction to cultural algorithms[C] //Proceedings of the 3th Annual Conference on Evolution Programming. Singapore: World Scientific Publishing, 1994: 131 – 136.

18 Chung C J. Knowledge-based approaches to self-adaptation in cultural algorithms [D]. Detroit: Wayne State University,1997:20 – 50.

19 Gao X Z, Jokinen T, Wang X, et al. A new harmony search method in optimal wind generator design [C] //The XIX International Conference on Electrical Machines, Rome, Italy, 2010: 1 – 6.

20 黄海燕,顾幸生. 基于文化算法的神经网络及其在建模中的应用[J]. 控制与决策,2008,23(4):477 – 480.
Huang Haiyan, Gu Xingsheng. Neural network based on cultural algorithms and its application on modeling [J]. Control and Decision, 2008, 23(4):477 – 480. (in Chinese)

21 吴莹,黄显林,高晓智,等. 一种文化鱼群算法及其在电机参数辨识中的应用[J]. 电机与控制学报,2012,16(5):102 – 108.
Wu Ying, Huang Xianlin, Gao Xiaozhi, et al. Cultural artificial fish-swarm optimization algorithm and application in the parameters identification of rotor system[J]. Electric Machines and Control, 2012,16(5):102 – 108. (in Chinese)

22 郭一楠,刘丹丹,程健,等. 自适应混合变异文化算法[J]. 电子学报, 2011,39(8):1913 – 1918.
Guo Yinan, Liu Dandan, Cheng Jian, et al. Adaptive cultural algorithm adopting mixed mutation[J]. Acta Electronica Sinica, 2011,39(8):1913 – 1918. (in Chinese)

23 成鹏,汪西莉. SVR 参数对非线性函数拟合的影响[J]. 计算机工程, 2011,37(3):190 – 194.
Cheng Peng, Wang Xili. Influence of SVR parameter on non-linear function approximation[J]. Computer Engineering, 2011,37(3):190 – 194. (in Chinese)

24 Vapnik V. Statistical learning theory [M]. New York: Wiley-Interscience, 1998:421 – 432.

Abstract: In order to solve the problem of low prediction accuracy and bad robustness of the traditional forecasting methods in water quality, this paper put forward the prediction model for pH value of aquaculture water quality based on the principal component analysis (PCA) and least squares support vector machine (LSSVM), which the hyper-parameters is optimized by modified cultural artificial fish-swarm algorithm(MCAFA). The dimension of aquiculture ecologic environmental data was reduced by principal component analysis; double evolutionary mechanism of cultural algorithm for reference was applied and LSSVM was taken as an artificial fish; belief space was used to guide the shoal evolution step size, global search direction and Cauchy mutation to improve the diversity of the artificial fish swarm; so the optimal hyper-parameters nonlinear pH value prediction model was automatically obtained. Based on the prediction model, the water quality on-line monitoring was predicted for a high-density aquaculture pond from September 1, 2011 to September 4, 2011 in Yixing city, Jiangsu province. Experimental results show that the PCA – MCAFA – LSSVM prediction model has good prediction effect than the ant colony algorithm-LSSVM and genetic algorithm-LSSVM. The absolute error of the 93.05% test samples is less than 8%, and the max absolute error is only 11.61%; the root-mean-square error, average absolute relative error and the running time are 0.047 4, 0.004 1 and 4.367 s respectively, which are better than those from the other models. It is obvious that PCA – MCAFA – LSSVM prediction model has low computational complexity and high forecast accuracy. It can provide the decision basis for the water quality controlling in the high density eriocheir sinensis culture.

Key words: Aquaculture water quality pH value forecasting Cultural artificial fish-swarm algorithm
Least squares support vector regression Parameter optimization Principal component analysis



(上接第 232 页)

Prototype System of Shape Measurements of Animal Based on 3D Reconstruction

Guo Hao¹ Ma Qin¹ Zhang Shengli² Su Wei¹ Zhu Dehai¹ Gao Yunbing¹

(1. Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture,
China Agricultural University, Beijing 100083, China

2. College Animal Science & Technology, China Agricultural University, Beijing 100083, China)

Abstract: In order to improve the accuracy, efficiency and automatic degree of existing acquisition techniques of animal's conformation, based on two depth cameras, a real time system which can reconstruct the animal for body measurement, an automatic ball-target-based method for the extrinsic calibration of reconstruction system which is based on RANSAC were put forward. The extrinsic parameters can be used for registering point clouds acquired synchronously from depth camera in real time. And then, interactive measuring method which is optimized in picking mode for body measurement was used to get shape traits of animal. Xtion PRO was chosen as data acquisition equipment, and taking the pig specimen as an example, point clouds data acquired from high precision laser scanner was applied to evaluation of the system. The results show that the ball-target-based method can get the extrinsic parameters of depth camera automatically; the average error of registered data is less than 7.50 mm; the speed of reconstructing the whole pig can reach in excess of 15 f/s; body measurement error is less than 4%. So the system can be applied to body measurements in agricultural field.

Key words: Body measurement 3D Reconstruction Point cloud Depth camera