

基于 GA-SVM 的渤海湾富营养化模型

向先全¹, 陶建华²

(1.天津大学环境科学与工程学院, 天津 300072; 2.天津大学机械工程学院, 天津 300072)

摘要: 为了更好地模拟和认知渤海湾富营养化的复杂行为, 通过研究遗传算法(GA)和支持向量机(SVM)的结合形式, 即参数寻优和特征选择, 以渤海湾水质实测资料为依据, 叶绿素a的质量浓度为输出, 建立了GA-SVM的富营养化模型。无特征选择时, 用遗传算法对支持向量机的参数(惩罚参数和核参数)进行自适应地优选, 预测模型的均方误差可达到1.831 μg/L, 具有较好的认知、泛化能力。再利用遗传算法二进制编码及启发式寻优的优点, 对所建模型的输入空间进行特征选择, 提取出代表性的特征变量:DO%、pH值、水温、COD、盐度以及氨氮。特征提取后预测模型的均方误差可达到1.363 μg/L, 模型性能有了很大提高。分析表明, COD、盐度及氨氮可作为人为控制的首要指标。

关键词: 富营养化模型; 支持向量机; 遗传算法; 参数寻优; 特征选择; 渤海湾

中图分类号: X171;TP181 文献标志码: A 文章编号: 0493-2137(2011)03-0215-06

Eutrophication Model of Bohai Bay Based on GA-SVM

XIANG Xian-quan¹, TAO Jian-hua²

(1. School of Environmental Science and Engineering, Tianjin University, Tianjin 300072, China;
2. School of Mechanical Engineering, Tianjin University, Tianjin 300072, China)

Abstract: For better simulating and cognizing the complex eutrophication behaviors of Bohai Bay, the combining forms of genetic algorithm and support vector machine (parameter optimization and feature selection) have been researched to establish GA-SVM eutrophication model for Bohai Bay based on the field measured data, and chlorophyll_a content has been selected as the model output. Firstly, with GA self-adaptive optimizing for penalty parameter and kernel parameter, the root mean square error (RMSE) of SVM test model was 1.831 μg/L, indicating preferable generalization performance. Then, with GA-based feature selection for the established SVM model, RMSE of SVM test model was 1.363 μg/L, showing great improvement for model performance. The representative features were extracted such as DO%, pH, water temperature, COD, salinity, and ammonia-nitrogen, the latter three of which could be considered as prior indexes for artificial control of eutrophication based on further analysis.

Keywords: eutrophication model; support vector machine; genetic algorithm; parameters optimization; feature selection; Bohai Bay

沿海地区经济的快速发展, 使得作为陆地产生污水最后归宿的海洋, 尤其是近岸海域受到了越来越严重的污染。渤海湾是位于渤海西侧的一个半封闭淤泥质浅水海湾, 海水交换能力和自净能力很弱^[1], 近岸海域水体富营养化严重, 赤潮频繁发生。2008年国家海洋环境质量公报指出: 连续5年的监测结果表明, 持续的城市化进程和陆源排污未得到有效控制,

致使渤海湾水体始终处于严重的富营养化状态, 生态系统处于亚健康状态。

国内外对富营养化模型开展了广泛研究, 主要包括以下几类: ① 基于机理或假设的确定性模型, 如简单的营养物平衡模型、复杂的生态-水质-水动力模型、复杂的生态结构动力学模型^[2]; ② 多元统计方法, 如 Handan 等^[3]通过主成分分析和多元线性回归

收稿日期: 2010-01-07; 修回日期: 2010-09-17。

基金项目: 国家自然科学基金资助项目(10872144)。

作者简介: 向先全(1984—), 男, 博士研究生, xxquan@tju.edu.cn。

通讯作者: 陶建华, jhtao@tju.edu.cn。

方法预测叶绿素 a 与 16 个物理、化学、生物指标的关系;③ 不确定性富营养化模型, Malmaeus 等^[4]结合 Monte-Carlo 模拟和灵敏度分析, 评价模型参数中的不确定因素对模型预测结果的影响, Chen 等^[5]提出了基于模糊逻辑理论预测藻类生物量的富营养化模型;④ 非线性富营养化模型, 黑箱模型人工神经网络^[6-7]在富营养化建模和评价中被广泛应用。

近岸海域水生态系统是一个具有多因素耦合的复杂系统, 生态要素间的关系错综复杂, 表现出极大的随机性、不确定性和非线性, 系统内各因素之间的相互作用及其动态变化过程未被完全知晓, 制约着确定型的生态水动力学发展。随着现代化监测技术的发展和监测手段的多样化, 在大量监测数据基础上利用多元统计、神经网络等方法建立的富营养化模型, 得到了广泛的应用。但这些以传统渐进统计学为理论基础的经验风险最小化原则是基于样本无限大的假设, 并不能很好地实现由贝叶斯决策理论导出的期望风险最小化原则, 这在神经网络学习中表现得尤为突出(过学习问题)。在统计学习理论基础之上发展起来的支持向量机(support vector machine, SVM)算法具有严格的理论基础, 采用结构风险最小化准则(而非传统的经验风险最小化原则), 把学习问题转化为一个二次规划问题来获得全局最优解, 从而克服了神经网络的不足, 在小样本情况下具有良好外推能力。近年来支持向量机开始被应用于富营养化模型的研究, 并取得了一定的效果^[8-9]。

但是, 作为一种学习机器, SVM 也存在一些有待完善的地方。其参数选取便是亟待完善的问题之一, 惩罚系数 C、核函数以及核函数的相关参数等的选取在一定程度上对模型的预测精度存在很大影响。另一方面, 如何从原始特征中去挑选出一些最有代表性的特征, 也是建立 SVM 模型时需要面临的问题。因此, 笔者提出基于遗传算法的支持向量机富营养化模型, 用于自适应地特征选择和参数寻优, 克服以上存在的不足, 提高了富营养化模型的精度, 并对富营养化的影响因子进行了分析。

1 理论基础

1.1 支持向量机

支持向量机的基本思想可以概括为:首先通过非线性变换将输入空间变换到一个高维空间, 然后在这个新空间中求取最优线性分类面, 而这种非线性变换是通过定义适当的内积函数实现的。目前支持向量

机已经发展了很多种, 如 C-SVC、v-SVC、ε-SVR、v-SVR 等, 笔者建立的富营养化模型主要是基于支持向量回归(support vector regression, SVR)方法进行的, 采用的是 ε-SVR。

介绍 ε-SVR 的算法原理。设样本集 $\{\mathbf{x}_i, y_i\} \dots \{\mathbf{x}_m, y_m\} \subseteq \{X \subseteq R^n \times Y \subseteq R\}$, \mathbf{x}_i 为样本输入, y_i 为样本输出。则函数逼近问题就是要找出一个函数 $f(\mathbf{x})$, 使之通过样本训练后尽量逼近 y_i , 并具有一定的推广能力。引入 ε 不敏感损失函数

$$L_\epsilon(y, f(\mathbf{x})) = \max(0, |y - f(\mathbf{x})| - \epsilon) \quad (1)$$

不敏感参数 ϵ 反映回归模型对输入变量所含噪声的敏感程度, 可用于控制模型拟合精度。首先考虑为线性回归函数 $f(\mathbf{x}) = \boldsymbol{\omega} \cdot \mathbf{x} + b$, 根据结构风险最小化原理, 函数估计问题就是寻找使下面风险函数最小的 $f(\mathbf{x})$, 即

$$R_{\text{reg}} = \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^m L_\epsilon(y_i, f(\mathbf{x}_i)) \quad (2)$$

式中: $\frac{1}{2} \|\boldsymbol{\omega}\|^2$ 反映了模型的复杂度; $\sum_{i=1}^m L_\epsilon(y_i, f(\mathbf{x}_i))$ 反映了训练误差即经验风险; C 为设定的误差惩罚参数, 用于控制对错分样本惩罚的程度, 实现错分样本数和模型复杂度之间的折中。

将式(2)风险最小化问题转化为求解约束优化问题, 并采用对偶理论和 Lagrange 乘子法进行变换, 该优化问题的对偶形式为

$$\left\{ \begin{array}{l} \min \frac{1}{2} \sum_{i,j=1}^m (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(\mathbf{x}_i \cdot \mathbf{x}_j) + \\ \epsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) - \sum_{i=1}^m y_i(\alpha_i^* - \alpha_i) \\ \text{s.t. } \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0 \quad 0 \leq \alpha_i, \alpha_i^* \leq C \end{array} \right. \quad (3)$$

利用最优化理论中的二次型规划方法求解可得拉格朗日乘子 α 和 α^* , α 和 α^* 只有一小部分不为 0, 它们对应的样本就是支持向量(support vector, SV)。再根据凸二次规划的 Karush-Kuhn-Tucker(KKT) 条件求得偏置 b 。由此可得支持向量回归机的决策函数

$$f(\mathbf{x}) = \sum_{i=1}^m (\alpha_i^* - \alpha_i)(\mathbf{x}_i, \mathbf{x}) + b \quad (4)$$

当样本集为非线性可分问题时, 引入了核函数将输入样本通过非线性函数 Φ 映射到一个高维的 Hilbert 空间, 在该空间中构造并求解最优分类面。所谓核函数就是在输入空间中存在一个函数 $K(\mathbf{x}, \mathbf{x}')$, 满足 $K(\mathbf{x}, \mathbf{x}') = (\Phi(\mathbf{x}), \Phi(\mathbf{x}'))$ 。非线性求解将式(3)中的 $(\mathbf{x}_i, \mathbf{x}_j)$ 替换为 $K(\mathbf{x}_i, \mathbf{x}_j)$ 即可。

由统计学习理论可知,只要任意函数满足 Mercer 条件,就可作为核函数。应用核函数可避免直接计算 $(\Phi(\mathbf{x}) \Phi(\mathbf{x}'))$,从而解决由于模式升维而引起的“维数灾难”问题,使得在没有增加复杂度情况下,智能学习在高维特征空间成为可能。目前常见核函数有以下几种:多项式核 $K(\mathbf{x}, \mathbf{x}_i) = (\gamma \mathbf{x} \cdot \mathbf{x}_i + \gamma_0)^d$;径向基核 (radial basis function, RBF) $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2)$;Sigmoid 核 $K(\mathbf{x}, \mathbf{x}_i) = \tanh(\gamma \mathbf{x} \cdot \mathbf{x}_i + \gamma_0)$ 。

1.2 遗传算法

遗传算法 (genetic algorithm, GA) 是一类具有很强鲁棒性的优化算法,隐含并行性和全局搜索特性是其两大显著特征。利用遗传算法求解问题,首先必须对每个可行解进行编码,从而将解空间变换为染色体空间,并定义染色体的适应度,使得较优个体的适应度也较高。然后再对种群施加遗传算子如选择、交叉、变异等操作,使得群体不断向着最优解的方向进化。

2 研究对象

选定渤海湾近岸的一个监控区作为研究对象。模型建立与验证的数据采用该监控区的实测数据资料。在渤海湾 $N38^{\circ} 49'20'' \sim 38^{\circ} 55'00'', E117^{\circ} 37'00'' \sim 117^{\circ} 50'00''$ 的范围内共设 6 个站位。分别在 2006 和 2007 年 6~10 月期间,每 2 周一次,对监控区的表层海水进行了共 20 多次的水质监测,监测指标包括表层水温、透明度、pH 值、盐度、化学需氧量 (chemical oxygen demand, COD)、溶解氧 (dissolved oxygen, DO)、溶解氧饱和度 (DO%)、活性硅酸盐 (Si)、磷酸盐 (P)、硝基氮、亚硝基氮、氨氮以及叶绿素 a 的质量浓度等 13 项。

3 方 法

3.1 确定模型的输入输出

富营养化是指水体中营养物质增加,水生生物特别是浮游藻类大量繁殖,使生物量的种群、种类以及数量发生改变,破坏了水体的生态平衡。藻类中叶绿素 a (Chla) 是叶绿素主要成分,其含量高低与水体藻类的种类、数量等密切相关,其浓度影响水色、水质及水中初级生产力,是表征水体富营养化现象及其程度的最重要的指示剂之一。分析叶绿素 a 的含量与动态,可以快速简便地了解浮游植物生物量状况及其变化趋势,在一定程度上反映水体富营养化状况。因此,确定叶绿素 a 的含量作为富营养化预测模型的输

出,即

$$\rho_{\text{Chla}} = f(T, \text{SD}, \text{pH}, S, \text{COD}, \text{DO}, \text{DO}\%), \\ \rho_{\text{Si}}, \rho_{\text{P}}, \rho_{\text{NO}_2}, \rho_{\text{NO}_3}, \rho_{\text{NH}_3} \quad (5)$$

式中: ρ_{Chla} 为叶绿素 a 的质量浓度; T 为海水表层水温; SD 为海水透明度; pH 为酸碱值; S 为盐度; COD 为化学需氧量; DO 为溶解氧; DO% 为溶解氧的饱和度; ρ_{Si} 为活性硅酸盐的质量浓度; ρ_{P} 为磷酸盐的质量浓度; ρ_{NO_2} 为亚硝基氮的质量浓度; ρ_{NO_3} 为硝基氮的质量浓度; ρ_{NH_3} 为氨氮的质量浓度。

为了对原始数据中的冗余信息进行处理,提取更有用的信息,首先对数据进行平滑处理,以剔除奇异值;其次,为了消除各属性数据的单位差异以及数值量纲上的差异,对原始数据进行归一化处理(防止计算溢出),使之落入一个小小的特定的区域 [0, 1],对样本的输入和输出变量分别用下式进行归一化处理

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

式中: \bar{x}_i 代表归一化后的数据; x_i 代表输入或输出变量; x_{\max} 、 x_{\min} 分别代表变量的最大值和最小值。为了检验 SVM 模型的预测能力,把样本集分为训练集和测试集,测试集不参加模型的训练。选取 2 号站位对模型进行验证,其余站位监测数据为学习样本数据。

3.2 参数寻优

由于有多种核函数,设计 SVM 一个重要步骤就是选择核函数和核参数。Vapnik^[10]的研究表明,SVM 的性能与所选用的核函数类型关系不大,而核参数和误差惩罚因子 C 是影响 SVM 性能的主要因素。RBF 核函数是一个普适的核函数,只有一个可控参数,通过参数的选择,它可以适用于任意分布的样本且输出权值由算法自动确定。同时考虑到富营养化的高度非线性和复杂性,选择了 RBF 核函数,此时,SVM 回归方法中的参数主要有不灵敏参数 ε 、惩罚常数 C 和 RBF 核参数 γ 。

一般情况下 ε 越大,支持向量数目就越少,解就越稀疏,超平面的复杂程度越低。文献[11]表明,对于 ε 的不同取值,预测误差随参数对 (γ, C) 具有相似的变化趋势,表明 ε 的选取在某种程度上独立于 (γ, C) 的选择。因此,可以先确定 ε ,把三参数优化问题转化为二参数优化问题。核参数 γ 的改变实际上是隐含地改变映射函数从而改变样本数据子空间分布的复杂程度(维数)。如果数据子空间维数很高,则得到的最优分类面就可能比较复杂,经验风险小但置信范围大;反之亦然。要想得到推广能力良好的 SVM 回归

器,需选择合适的核函数 γ 将数据投影到合适的特征空间. 在确定的数据子空间中, C 的取值小表示对经验误差的惩罚小, 学习机器的复杂度小而经验风险值较大; 反之亦然. 前者称为“欠学习”现象, 而后者则为“过学习”. 当 C 超过一定值时, SVM 的复杂度达到了数据子空间允许的最大值, 此时经验风险和推广能力几乎不再变化.

参数调整通常以最小化推广误差的估计来实现, 交叉验证误差是推广误差一种近似无偏估计. 所谓交叉验证是将学习样本先随机地分成 k 个同样大小数量的子集, 用 $k-1$ 个子集作为训练样本, 预测那个没有参加训练的子集, 共需进行 k 次. 这样全部数据中每个样本点都被预测一遍, 准确率是比较稳定的.

在实际应用中, SVR 参数的确定方法主要有经验确定、网格搜索等. 经验确定是按照样本的分布规律, 对各个参数给出经验估计值, 这需要使用者有较深厚的 SVR 理论基础, 对于非专业人员并不适合; 网格搜索是将 3 个参数所处的空间划分成网格, 在每一个网格点上逐一实验以确定最优参数, 计算量随着参数数量的增加呈指数增长, 对于实际中很多大型回归分析问题由于计算量太大而不实用. 而遗传算法具有很强的全局搜索能力, 且不依赖于特定的求解模型, 故笔者采用遗传算法对支持向量机的参数进行智能选取. 具体实现如下.

(1) 编码及初始种群 采用浮点数方式进行编码, 设定 γ 、 C 的搜索空间分别为: $(0, 50)$ 、 $(0, 50)$, 初始种群由完全随机的方法产生.

(2) 适应度函数 为了使相近回归精度下的最优分类面结构尽可能简单, 适应度函数除考虑交叉验证的均方根误差外, 还加入了一个复杂度控制项. 在实际计算中, 适应度通常按照目标函数值从小到大的顺序来取值的, 此时的目标函数为

$$O(C, \gamma) = E_{\text{sv}} / n \quad (7)$$

式中: E 是 SVM 在训练样本集的交叉验证均方根误差; n_{sv} 和 n 分别为支持向量数和训练样本数.

(3) 遗传操作(选择、交叉、变异) 用随机遍历抽样与最优个体保留策略相结合的方法对个体进行选择; 交叉采用具有突变特征的线性重组; 变异操作采用遗传算法育种器的变异算子.

3.3 特征选择

从一组特征中挑选出一些最有效的、最有代表性的特征以达到降低特征空间维数的目的, 这就是特征选择. 笔者利用遗传算法对支持向量机的输入空间进行特征选择, 以期了解在渤海湾富营养化中哪些特

征更具有代表性. 其方法简述为:

(1) 编码及初始种群 采用二进制编码, 每一位二进制数字对应一个特征. 把训练集上所有特征(即式(5)中的 12 个自变量), 分别对应染色体上不同的位(共 12 位). 当某位的数字为 1 时, 表示这个特征被选中, 否则没被选中. 用随机方法生成初始种群.

(2) 适应度函数 根据染色体各个位的二进制信息, 把对应的特征组成训练集, 作为 SVM 的输入, 调用遗传算法优化 SVM 参数过程建立 SVM 模型. 为寻找富营养化模型最有效的特征, 应使训练模型误差和预测模型误差都尽可能小, 同时两误差不应差别过大, 因此将计算适应度值的目标函数确定为

$$O = E_{\text{train}}^2 + E_{\text{test}}^2 \quad (8)$$

即为训练模型和预测模型均方根误差的平方之和.

(3) 遗传操作(选择、交叉、变异) 用随机遍历抽样与最优个体保留策略相结合方法对个体进行选择, 种群交叉和变异分别采用单点交叉、离散变异算子.

4 结果及分析

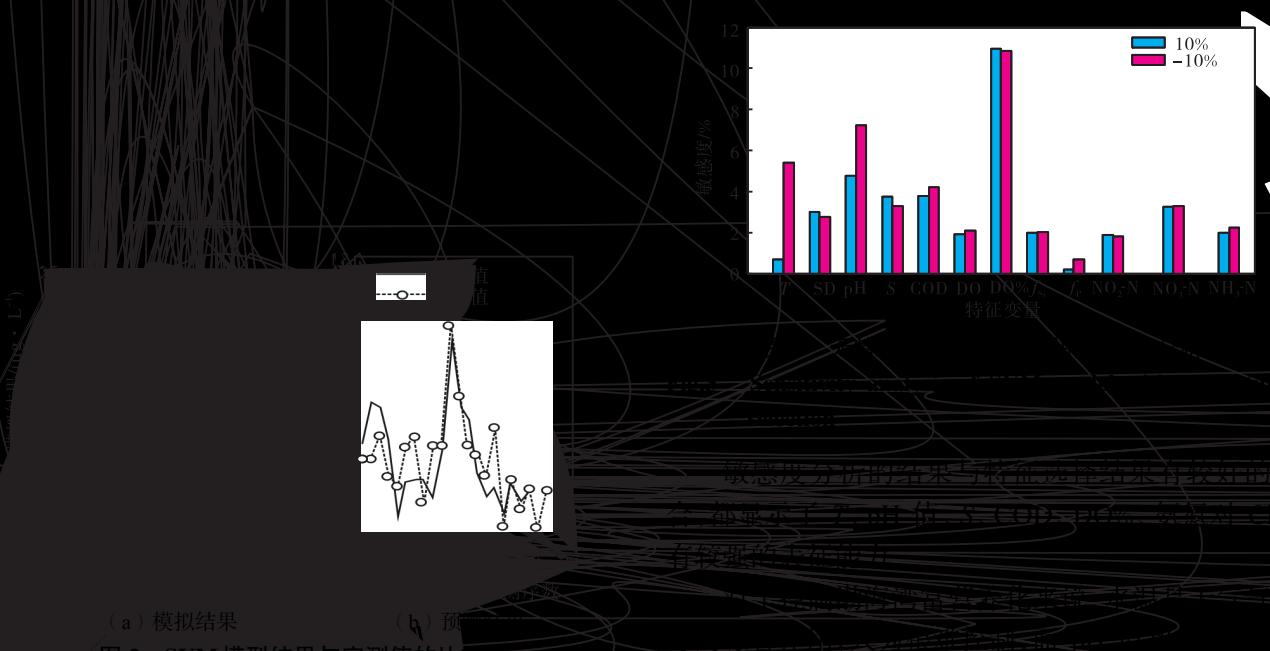
对原始监测数据进行前处理(去噪和归一化)之后, 首先考虑在无特征选择时, 利用遗传算法进行参数优选, 建立渤海湾富营养化的 SVM 模型; 其次, 利用遗传算法进行特征选择, 分析对富营养化产生作用的主要特征因子. 为了便于比较分析模型性能, 采用以下两个统计量来评价模型的拟合程度和预测效果, 这两个统计量是均方根误差 σ 和确定系数 r^2 , 即

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(\mathbf{x}_i) - y_i)^2} \quad (9)$$

$$r^2 = \frac{\left(\sum_{i=1}^n f(\mathbf{x}_i) y_i - \sum_{i=1}^n f(\mathbf{x}_i) \sum_{i=1}^n y_i \right)^2}{\left(\left(\sum_{i=1}^n f(\mathbf{x}_i) \right)^2 - \left(\sum_{i=1}^n f(\mathbf{x}_i) \right)^2 \right) \left(\left(\sum_{i=1}^n y_i \right)^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)} \quad (10)$$

4.1 GA 优化 SVM 参数

取交互验证的次数 k 为 10, 选取遗传算法操作过程的相关参数如下: 种群个体数为 40, 最大遗传代数为 100, 遗传代沟为 0.9. 单参数 ε 的选取相当于一个一维优化问题, 给定 ε 的取值范围并离散化, 取不同的离散值使遗传算法目标函数值最小, 经计算选 ε 为 0.01. 遗传算法的进化过程如图 1 所示, 经 40 次遗传迭代后, 种群目标函数最小值几乎达到稳定, 可以认为达到了近似最优解, 而种群目标函数的平均值上下波动表明种群是有活力的.



控制。因此 COD、盐度及氨氮可作为人为控制的首要指标。COD 最能表征出人类活动(生活污水和工业废水)对水体的污染,需进行有效地管理控制;盐度对藻类生长繁殖、鱼类产卵等活动有重要的影响,而近岸海域盐度主要受流域径流的影响,因此,在不加重污染的前提下,需加强利用流域径流这一环节来调节近岸海域的盐度;氨氮中的氮元素与浮游植物体内氨基酸中氮的价态相同,易被浮游植物吸收利用,可侧重于通过调查研究渤海湾中氨氮的输入、分布及转化特性,从而有效控制渤海湾浮游植物的生长繁殖。

5 结 论

(1) 通过对支持向量机各参数响应的研究,利用自适应的遗传算法对惩罚常数 C 和 RBF 核参数 γ 进行启发式寻优,GA 适应度函数考虑了最优分类面的结构简单性。以渤海湾实测资料为依据,Chla 质量浓度的预测均方误差控制在 $1.831 \mu\text{g/L}$ 内,表明所建立的 GA-SVM 模型具有较好的认知、泛化能力。

(2) 为提取富营养化模型中更具有代表性的特征变量,利用遗传算法二进制编码及启发式寻优的优点,综合考虑 SVM 富营养化模型的训练误差和预测误差,对渤海湾富营养化影响因子进行特征选择,结合无特征选择时 GA-SVM 模型的敏感性分析,提取出代表性的特征变量有:海水表层水温、pH 值、盐度、化学需氧量、溶解氧的饱和度以及氨氮。特征提取后预测均方误差最小可达到 $1.363 \mu\text{g/L}$,模型性能有了很大提高。

(3) 通过特征选择和敏感性分析,结合可操作性,在渤海湾富营养化的控制中,可考虑将 COD、盐度及氨氮作为人为控制的首要指标。

参考文献:

- [1] Sun Jian,Tao Jianhua. Relation matrix of water exchange for sea bays and its application[J]. *China Ocean Engineering*,2006,20(4):529-544.
- [2] 卢小燕,徐福留,詹巍,等. 湖泊富营养化模型的研究现状与发展趋势[J]. 水科学进展,2003,14(6):792-798.
Lu Xiaoyan,Xu Fuli, Zhan Wei, et al. Current situation and development trends in lake eutrophication models[J]. *Advances in Water Science*,2003,14(6):792-798 (in Chinese).
- [3] Handan Camdevren,Nilsun Demir,Arzu Kanik,et al. Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs[J]. *Ecological Modelling*,2005,181:581-589.
- [4] Malmaeus J M,Hakanson L. A dynamic model to predict suspended particulate matter in lakes[J]. *Ecological Modelling*,2003,167(3):247-262.
- [5] Chen Qiuwen,Mynett A E. Modelling algal blooms in the Dutch coastal waters by integrated numerical and fuzzy cellular automata approaches[J]. *Ecological Modelling*,2006,199(1):73-81.
- [6] Yao Z H,Fei M R,Li K,et al. Recognition of blue-green algae in lakes using distributive genetic algorithm based neural networks[J]. *Neurocomputation*,2007,70(4/5/6):641-647.
- [7] Velo-Suarez L,Gutierrez-Estrada J C. Artificial neural network approaches to one-step weekly prediction of dinophysis acuminata blooms in Huelva[J]. *Harmful Algae*,2007(6):361-371.
- [8] 冯剑丰,王洪礼,李胜朋. 基于支持向量机的浮游植物密度预测研究[J]. 海洋环境科学,2007,26(5):438-441.
Feng Jianfeng,Wang Hongli,Li Shengpeng. Research on prediction of phytoplankton's density using support vector machines[J]. *Marine Environmental Science*,2007,26(5):438-441 (in Chinese).
- [9] Behzad M,Asghari K,Eazi M,et al. Generalization performance of support vector machines and neural networks in runoff modeling[J]. *Expert Systems with Applications*,2009,36:7624-7629.
- [10] Vapnik V N. 统计学习理论的本质[M]. 张学工,译. 北京:清华大学出版社,2000.
Vapnik V N. *Statistical Learning Theory*[M]. Zhang Xuegong , Trans. Beijing : Tsinghua University Press , 2000 (in Chinese).
- [11] 刘靖旭,蔡怀平,谭跃进. 支持向量回归参数调整的一种启发式算法[J]. 系统仿真学报,2007,19(7):1540-1547.
Liu Jingxu,Cai Huiping,Tan Yuejin. Heuristic algorithm for tuning hyperparameters in support vector regression[J]. *Journal of System Simulation*,2007,19(7):1540-1547 (in Chinese).