第56卷第2期
2016年3月

大连理工大学学报
Journal of Dalian University of Technology

Vol. 56, No. 2
Mar. 2016

过程建模与控制

文章编号：1000-8608(2016)02-0185-08

基于 ELMR-SVMR 的海水水质预警模型研究

张　颖*，李　梅，高倩倩，施　佳

（上海海事大学 信息工程学院，上海　201306）

摘要： 海水水质预警对海洋环境监控和保护有着重要的意义. 极限学习机回归和支持向量机回归（extreme learning machine regression-support vector machine regression，ELMR-SVMR）集成预警模型采用预测与评价相结合的方式，实现对未来海水富营养化状况综合预警的目的. 其中 SVMR 方法用于构建海水富营养化评价模型；ELMR 用于对未来一段时间的水质状况进行综合预测，ELMR 的预测结果作为评价模型的输入变量. 集成模型的可靠性直接影响预警的有效性，ELMR-SVMR 的可靠性通过分析 ELMR 预测误差对 SVMR 评价模型的灵敏度得到. 将各参变量的预测误差结果作为评价模型的灵敏度影响参量，通过灵敏度计算可获得对 ELMR-SVMR 模型的灵敏度评价. 通过与其他方法的实验对比及分析，验证了该区域范围内所建 ELMR-SVMR 预警模型的有效性，为探索建立有效的海水水质预警模型提供了一种新途径.

关键词： 集成预警模型；极限学习机；支持向量机；海水富营养化；灵敏度分析

中图分类号： TP391　　**文献标识码：** A　　**doi：** 10.7511/dllgxb201602012

0　引　言

2014 年 3 月 21 日，国家海洋局发布的《2013 年中国海洋环境状况公报》显示，我国近岸海域频繁发生水污染、生态受损、赤潮等环境问题. 研究发现，正在进行监测的近岸河口及海湾等海洋生态系统多达 81%受到了不同程度的污染. 正如联合国海洋污染科学问题专家组（GESAMP）在 2001 年报告中所指出的，“无论是从影响范围还是破坏结果来看，海水富营养化无疑是目前人类影响海洋的最大环境问题”. 水环境的监测与治理是国际上研究的重点与难点，实现水环境监测的重点是水质预测模型的建立. 海水富营养化是海水中营养物质过剩引起生态恶化的改变过程，是由多种因素引起的一系列反应. 因此，海水富营养化综合水质预警模型的研究，对近岸海域赤潮灾害防治具有重要的科学价值和实际意义[1].

近海海水水质富营养化趋势预测预警其实质是基于对影响水质富营养化的主要环境理化因子的分析预测，通过水质评价模型获得对水质评价等级的预测，进而起到对水质富营养化变化预警的目的. 针对海洋环境进行评价研究的基于数据驱动的方法主要有人工神经网络方法、模糊综合评判法、机器学习法等. 支持向量机（SVM）方法近来被较多地应用于环境理化因子的软测量及评价[2-3]. 支持向量机是通过核函数来处理数据的，具有很强的非线性数据处理能力；同时它引入软边缘，具有很好的泛化能力. 因此支持向量机方法可用于构建基于数据驱动的海水富营养化评价模型. 极限学习机（ELM）方法近来受到广泛的关注，缘于其模型结构简单、训练速度快、泛化性能较好，且具有较好的实时性，适合用来对未来一段时间的水质状况进行实时综合预测. 选取极限学习机的预测结果作为评价模型的输入变量，通过模型集成实现海水富营养化综合预警的目的.

影响海水富营养化评价模型预警结果的因素主要来自于预测误差的不确定性[4]. 海水富营养化预警模型的影响因素不仅众多，而且用来表征这一状况的实验数据也相对较少，另外，集成模型

收稿日期： 2015-10-28；　**修回日期：** 2016-01-10.
基金项目： 国家自然科学基金资助项目(61273068)；上海市科学技术委员会国际学术交流合作资助项目(15220721800)；上海市教育委员会科研创新资助项目(13YZ084).
作者简介： 张　颖*(1968-)，男，博士，副教授，E-mail：yingzhang@shmtu.edu.cn.

误差累积情况对预警的影响程度也不确定.对集成模型灵敏程度进行分析的主要目的在于获得该模型预测误差结果对预警结果的影响程度大小,来为集成预警模型的可靠性提供科学依据.参数灵敏度分析已在大气科学、经济学和生态学等领域得到了广泛的应用[5].其在该类模型中的作用不断增强,并且已成为模型行为分析和参数优化的重要方法和分析生态系统对环境影响的重要手段之一[6];分析方法也从比较单一的局部灵敏度分析进而发展到全局灵敏度分析等多种方法[7].本文以海水富营养化集成预警模型为研究对象,对集成模型中预测误差对预警结果的影响进行参数灵敏度分析,为集成模型的验证及后续研究奠定基础.

1 ELMR-SVMR 基本原理

1.1 极限学习机回归(ELMR)

极限学习机方法是建立在单隐含层前馈神经网络 SLFN 上的一种较为重要的算法,较强的逼近能力是前馈神经网络的优良特性,但是它的输入权值和隐含层神经元阈值的设置并不容易;除此之外,它的学习方法在处理非线性数据时会使得收敛速度变慢,进而影响模型的训练.极限学习机的优势在于通过隐含层的激活函数为无限可微函数来随机选择和调整节点的参数,进而将 SLFN 从非线性系统变为线性系统,使得传统神经网络训练时间较长和参数难以确定等问题得到很好的解决.

本算法中采用极限学习机回归是为了找到预测曲线的拟合误差最小函数,也即找到自变量 x 和因变量 y 的函数关系式.

设有 n 个变量 m 组样本,输入层第 i 个变量与隐含层各节点间的连接权值 $\boldsymbol{w}_i=(w_{i1}\quad w_{i2}\quad \cdots\quad w_{im})^{\mathrm{T}}$;设隐含层神经元的阈值 $\boldsymbol{b}=(b_1\quad b_2\quad \cdots\quad b_a)^{\mathrm{T}}$,隐含层节点与输出变量的连接权值 $\boldsymbol{\beta}_k=(\beta_1\quad \beta_2\quad \cdots\quad \beta_a)^{\mathrm{T}}$,隐含层神经元节点的激活函数为 $g(\boldsymbol{x})$.则具有 a 个隐含层节点的 ELMR 模型的决策函数可以用下式表示:

$$\sum_{k=1}^{a}\boldsymbol{\beta}_k g_k(x_i)=\sum_{k=1}^{a}\boldsymbol{\beta}_k g(w_k\cdot x_i+b_k)=\boldsymbol{y} \tag{1}$$

此式还可以表示为 $\boldsymbol{H\beta}=\boldsymbol{T}'$. $\boldsymbol{H}$ 为隐含层的输出矩阵,具体形式如下式所示:

$$\boldsymbol{H}(w_1\quad \cdots\quad w_a\quad b_1\quad \cdots\quad b_a\quad x_1\quad \cdots\quad x_n)=\begin{pmatrix} g(w_1\cdot x_1+b_1) & \cdots & g(w_a\cdot x_1+b_a)\\ \vdots & & \vdots\\ g(w_1\cdot x_n+b_1) & \cdots & g(w_a\cdot x_n+b_a)\end{pmatrix}_{n\times a} \tag{2}$$

Huang 等提出的 SLFN 定理表明了 ELMR 模型中的输入权重 $\boldsymbol{w}$ 和隐含层神经元的阈值 $\boldsymbol{b}$ 可以为任意值,其相应的数学分析如下所示:

假设将随机输入变量输入权重 $\boldsymbol{w}$ 和隐含层节点阈值 $\boldsymbol{b}$ 用向量 $\boldsymbol{e}$ 表示,$\boldsymbol{H}$ 是 $\boldsymbol{e}$ 的函数.输入变量为 n 维的子空间,对应向量 $\boldsymbol{e}$ 属于 $n-1$ 维的子空间,且存在向量 $\boldsymbol{\beta}$ 与这个子空间正交,如下式所示:

$$(\boldsymbol{\beta},(\boldsymbol{e}(b_i)-\boldsymbol{e}(c)))=0 \tag{3}$$

其中 $w_k\cdot x_i\neq w_k\cdot x_{i'}$,$c=w_k\cdot x_i$,由式(3)展开,可得出下式:

$$\begin{aligned}&\Rightarrow\beta_1\cdot g(b_i+c_1)+\beta_2\cdot g(b_i+c_2)+\cdots+\\&\quad\beta_n\cdot g(b_i+c_n)-y=0\\&\Rightarrow g(b_i+c_n)=\sum_{p=1}^{n-1}(\gamma_p\cdot g(b_i+c_p))+y/\beta_n;\\&\gamma_p=\beta_p/\beta_n,p=1,2,\cdots,n-1\end{aligned} \tag{4}$$

当 $g(\boldsymbol{x})$ 无限可微时,对上式方程求一阶导数,如下式所示:

$$\begin{aligned}&\Rightarrow g^{(l)}(a_i+b_n)=-\sum_{p=1}^{n-1}(\gamma_p\cdot g^{(l)}(a_i+b_p));\\&\quad l=1,2,\cdots,n,n+1,\cdots\end{aligned} \tag{5}$$

式(4)共有 $n-1$ 个系数:$\gamma_1,\gamma_2,\cdots,\gamma_{n-1}$,式(5)因激励函数无限可微可以派生出更多方程.因此,向量 $\boldsymbol{e}$ 不属于任意一个小于 $n-1$ 维的向量空间,也就是说 $\boldsymbol{H}$ 满秩且存在唯一解.这也证明了 $\boldsymbol{H}$ 是存在且唯一的,它与 $\boldsymbol{w}$ 和 $\boldsymbol{b}$ 的取值都无关,从而求得 Moore-Penrose 广义逆 $\boldsymbol{H}^{\dagger}$,得到预测模型的决策函数.

当激活函数 $g(\boldsymbol{x})$ 无限可微时,ELMR 的参数(输入连接权值和隐含层节点阈值)可以随机选择,无须全部调整,在训练过程中虽然不会改变大小,但是需要在训练之前先设置好隐含层节点个数和激活函数. $\boldsymbol{H}^{\dagger}$ 为 $\boldsymbol{H}$ 的广义逆矩阵.通过解线性方程组的最小二乘解就可以得出隐含层节点与输出节点的连接权值 $\boldsymbol{\beta}$,$\boldsymbol{\beta}$ 的值是唯一的,因此具有比较好的泛化性能,并且误差降低到最小.

$$\min_{\boldsymbol{\beta}}\|\boldsymbol{H\beta}-\boldsymbol{T}'\|,\ \hat{\boldsymbol{\beta}}=\boldsymbol{H}^{\dagger}\boldsymbol{T}' \tag{6}$$

其中 ELMR 模型需要考虑的是隐含层参数的确定问题,由上述极限学习机原理的分析可知,当模型的隐含层神经元个数 K 小于样本个数时,对于任意的输入权值和阈值,依然存在一个 K 值,使得训练误差无限逼近于一个任意大于零的 ε.因此,当样本数过多时,选取少于样本数的隐含层神经元个数即可.又因为隐含层神经元节点个数对预测结果的影响并不灵敏,因此可以推断,具有相同复杂度和样本个数的数据所需设置的隐含层个

数基本相同. 所以，可以通过 800 组数据进行样本训练来提前确定隐含层的个数.

1.2　支持向量机回归(SVMR)

SVMR 算法的主要思想是通过采用适当的核函数，把低维空间的非线性样本数据映射到高维空间，实现非线性数据的线性输出[8-9]. 另外，SVMR 算法在学习过程中主要采用折中考虑的方法(结构风险最小化原理)，缩小置信区间(结构复杂度造成的风险)和降低样本的经验风险(训练误差). SVMR 算法通过学习过程来找到一个函数 $f(\boldsymbol{x})$ 的值可以近似为 $y(\boldsymbol{x})$，也就是对独立同分布的数据进行处理以达到线性回归的目的. ε 表示可接受的误差值，即 $l_\varepsilon = |y(\boldsymbol{x}) - f(\boldsymbol{x})|_\varepsilon = \max(0, |y(\boldsymbol{x}) - f(\boldsymbol{x})|_\varepsilon - \varepsilon)$.

上述优化问题可以最终归为解下面二次规划问题：

$$\begin{aligned} &\min\left(\frac{1}{2}\|\boldsymbol{w}\|^2 + C\left(\sum_{i=1}^{n}\xi_i + \xi_i^*\right)\right) \\ &\text{s.t.}\ \ y_i - \langle \boldsymbol{w}, \boldsymbol{\Phi}(\boldsymbol{x}_i)\rangle - b \leqslant \varepsilon + \xi_i \\ &\qquad \langle \boldsymbol{w}, \boldsymbol{\Phi}(\boldsymbol{x}_i)\rangle + b - y_i \leqslant \varepsilon + \xi_i^* \\ &\qquad \xi_i, \xi_i^* \geqslant 0 \end{aligned} \tag{7}$$

式中：$\langle \boldsymbol{w}, \boldsymbol{\Phi}(\boldsymbol{x}_i)\rangle$ 表示输入权重与核函数的向量积，$k(\boldsymbol{x}, \boldsymbol{x}_i) = \boldsymbol{\Phi}(\boldsymbol{x}) \cdot \boldsymbol{\Phi}(\boldsymbol{x}_i)$，常数 $C(C>0)$ 为惩罚因子，用来平衡经验风险和置信范围的比例. 目标函数如下式所示：

$$\begin{aligned} \max(w(\lambda, \lambda^*)) = &-\frac{1}{2}\sum_{i,j=1}^{n}(\lambda_i - \lambda_i^*)(\lambda_j - \lambda_j^*) \cdot \\ &k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i=1}^{n}(\lambda_i - \lambda_i^*)y_i - \\ &\sum_{i=1}^{n}(\lambda_i - \lambda_i^*)\varepsilon \end{aligned} \tag{8}$$

求得系数 λ_i、λ_i^*，得到构造的回归函数为式(9)，常数项为式(10).

$$f(\boldsymbol{x}) = \sum_{i=1}^{n}(\lambda_i - \lambda_i^*)k(\boldsymbol{x}, \boldsymbol{x}_i) + b \tag{9}$$

$$b = y_i - \sum_{i,j=1}^{n}(\lambda_i - \lambda_i^*)k(\boldsymbol{x}_i, \boldsymbol{x}_j) - \varepsilon \tag{10}$$

其中采用的核函数为径向基函数(RBF). 如式(11)所示，$\boldsymbol{x}_i$ 为核中心，σ 主要用于控制函数的径向作用范围，即函数的宽度参数.

$$k(\boldsymbol{x}, \boldsymbol{x}_i) = \exp(-\|\boldsymbol{x} - \boldsymbol{x}_i\|^2/(2\sigma)^2) \tag{11}$$

惩罚系数 C 和核参数 σ 是 SVMR(核函数为 RBF)的两个最为重要的参数[10-11]. 其中，惩罚系数 C 为对错误分类的惩罚指数($C>0$)，C 不宜过大或过小，太大会导致回归准确率在训练过程中很高而在测试过程中很低，太小则回归准确率不可靠，模型无效；核参数 $\sigma(\sigma>0)$ 对特征空间的分割结果有较大的影响，对回归结果也很重要，核参数的值同样不宜过大或过小，太大会导致过拟合，太小会导致欠拟合.

SVMR 模型采用混沌布谷鸟算法对 SVM 的惩罚系数 C 和核参数 σ 进行寻优，提高预警性能. 优化初期先利用混沌动力学进行全局遍历性“粗搜索”，然后退出混沌进入下一阶段“细搜索”优化，这样能够有效地避免陷入局部极小点.

2　变量选取与评价标准

2.1　海水富营养化的定义

海洋富营养化是指“海水中的营养物质过度增加，并导致生态系统有机质增多、低氧形成、藻华爆发等一些异常改变的过程”. 现代富营养化模型[12]的关键理化因子主要以营养盐负荷信号为主，综合考虑浮游植物生物量等直接响应和营养盐循环、食物网络结构等间接响应，进一步发展了早期富营养化概念模型，能够更好地反映海水富营养化的特点. 现代时空富营养化概念模型如图 1 所示.

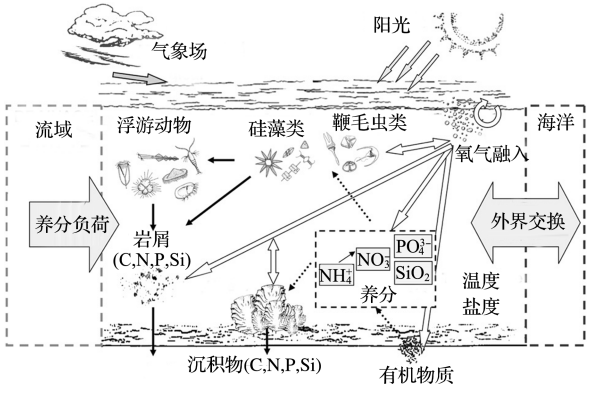

图 1　现代时空富营养化概念模型

Fig. 1　The concept model of modern space-time eutrophication

虽然导致沿海海洋富营养化过程的因素是复杂的，但是，是否需要将所有的因素都进行评估是值得探究的问题. 根据海水富营养化的定义，考虑以下两方面：(1)选取合适的海洋环境监测参数，参数应具有以下特点：易于监测，且保存有长期监测数据；能够进行空间规模的富营养化描述；被大量研究证实能够有效建立预测模型的参量. (2)以生物群落结构信息量化富营养化的变化趋势. 这

种方法有助于理解具有物种多样性的生态系统功能及相关的影响参量.

2.2 样本的来源及选取

本文所采用的样本数据均来自于杭州湾东部某近岸海域布设的监测站点(每隔 1 h 连续采样).将采集到的 800 组数据作为实验研究的样本,每组数据依次分别包括叶绿素-a 浓度、硝酸盐浓度、光照强度(包括浅层水深和深层水深)、温度、含盐量、溶解氧、酸碱度、传导率、NO_3^- 浓度、浑浊度 11 个因子.本文采用主成分分析法(PCA)对海水富营养化因子进行统计学分析.根据下式分别计算主成分贡献率 η、累计贡献率 η_t:

$$\eta = \lambda_i \Big/ \sum_{i=1}^{m} \lambda_i;\ \eta_t = \sum_{k=1}^{i} \lambda_k \Big/ \sum_{i=1}^{m} \lambda_i \qquad (12)$$

贡献率的大小与包含原始主成分信息的多少成正比.通常情况下,取累计贡献率达 85%以上的 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 所对应的前 $k(k \leqslant m)$ 个主成分作为主要的影响因子.主成分筛选结果如表 1 所示.

表 1 主成分分析表

Tab. 1 Analysis table of principal components

主成分	特征值	贡献率/%	累计贡献率/%
1	3.504 60	31.860	31.860
2	1.786 30	16.239	48.099
3	1.302 70	11.842	59.941
4	0.992 29	9.020	68.961
5	0.910 77	8.279	77.240
6	0.806 38	7.331	84.571
7	0.605 61	5.506	90.077
8	0.495 14	4.501	94.578
9	0.394 04	3.582	98.160
10	0.202 02	1.837	99.997
11	0.000 20	0.003	100.000

从表 1 分析看出,排前 6 的主成分的累计贡献率达到 84.571%,基本上可以保留原来 11 个变量所表征的信息,提取这 6 个主成分作为系统主要的输入变量.

如图 2 所示,计算出各个影响因子在前 6 个主成分中所对应的权重值,计算公式为 $F_j = e_{j1}x_1 + e_{j2}x_2 + \cdots + e_{jm}x_m$.

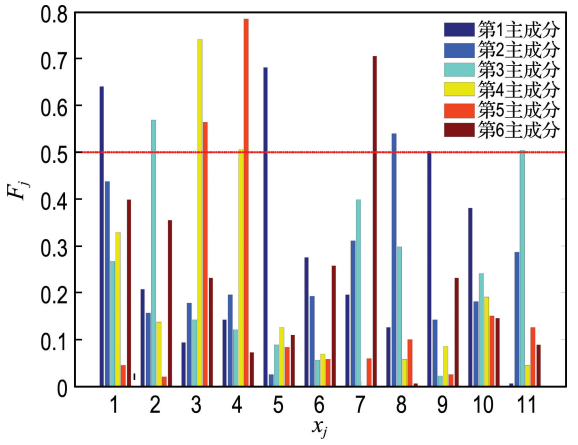

图 2 各影响因子在前 6 个主成分中所占的权重

Fig. 2 The weights of the influence factors on the front 6 principal components

结合表 1 和图 2 可得水体富营养化影响因子的前 6 个主成分,从图 2 中可以看出温度、浑浊度、酸碱度、溶解氧、硝酸盐浓度、叶绿素-a 浓度、光照强度(深、浅)为主要变量,但光照强度与其他变量均有关联,为了避免变量冗余,不将光照强度纳入评价参量中.

2.3 综合水质评价标准

近海水域富营养化评价等级划分如表 2 所示.表中各参数评价标准的确定主要参考了国家海水水质标准(GB 3097—1997)、海水水域富营养化各指标评价阈值以及中国近海海域水化学要素质量评价指数[13].

表 2 近海水域富营养化评价等级划分

Tab. 2 Hierarchy of offshore seawater eutrophication

评价类别		物理参数		化学参数			叶绿素-a 浓度/($\mu g \cdot L^{-1}$)	综合水质
警级	等级	温差/℃	浑浊度/ntu	pH	溶解氧浓度/($mg \cdot L^{-1}$)	硝酸盐浓度/($mg \cdot L^{-1}$)		
无警	好	(0,1]	(0,5]	(8.0,8.3]	(6,8]	(0,0.3]	(0,1]	(0,1]
	较好	(1,2]	(5,10]	(7.7,8.0]∩(8.3,8.5]	(5,6]	(0.3,0.5]	(1,3]	(1,2]
弱警	中	(2,3]	(10,15]		(4,5]	(0.5,0.7]	(3,5]	(2,3]
	差	(3,4]	(15,20]	(6.8,7.7]∩(8.5,8.8]	(3,4]	(0.7,1.2]	(5,7]	(3,4]
有警	很差	(4,5]	(20,25]		(0,3]	(1.2,3.2]	(7,10]	(4,5]

3　预警模型的灵敏度分析

通过前面的分析可以看出海水富营养化的影响因子众多，模型输入数据的随机误差及模型本身受到各种干扰等引起的输出误差的不确定性等都可能造成集成模型评价结果的不确定性，并可能使得预警结果失效. 验证模型的可靠性是模型分析的必要环节，灵敏度的大小反映了所建模型的风险程度，是模型可靠性的一种衡量标准，通过分析预测误差对集成模型的灵敏度，可实现对预警模型的可靠性验证.

某个因素的变化会对一些关键指标产生不同的影响，敏感性分析就是对这类不确定性的影响程度进行定量分析的技术[14]. 敏感系数度量是敏感性分析的一种重要方法，主要通过相关变量的改变对关键指标的影响程度进行分析，得出关键指标受这些因素变化影响的规律. 敏感系数＝关键指标浮动百分比/相关变量浮动百分比. 敏感系数越大，表明该因素对关键指标的影响越大，即对输出结果的影响越大；反之，敏感系数越小，则该因素对关键指标或输出结果的影响就越小.

参数灵敏度是通过分析扰动输入变量并利用统计学方法定量或定性地评估预测误差对仿真结果的影响，灵敏度＝状态变量变化百分比/误差变化百分比. 状态变量(y)对第 j 个变量误差值(x_j)的灵敏度如下式所示：

$$S=(\Delta y/y)\Big/\left(\sum_{j=1}^{6}\Delta x_j\Big/x_j\right) \tag{13}$$

式中：Δy 表示相应状态变量的变化量，Δx_j 表示第 j 个变量的误差值变化量. 本文的灵敏度等级划分标准为 $|S|\leqslant 0.1$：该参数不灵敏；$0.1<|S|\leqslant 0.5$：该参数较灵敏；$0.5<|S|\leqslant 1$：该参数灵敏；$1<|S|$：该参数非常灵敏. 该模型的灵敏度分析表征了预警模型的风险程度，误差灵敏度越大，说明此误差对模型的影响程度越大，即模型的风险性较大.

根据式(13)中介绍的灵敏度等级计算方法可得到 ELMR-SVMR 模型的水质等级预测的灵敏度为 0.1，处于对输出误差不确定性的不太灵敏状态，说明该模型的预测误差对集成模型的灵敏度处于不太灵敏状态，预警模型可靠性具有一定保证，说明该集成预警模型具有适用性.

4　仿真实验与分析

基于 ELMR-SVMR 集成模型的海水富营养化综合预警建模分为两部分. 首先，根据评价标准建立 SVMR 综合水质评价标准模型；其次，将 ELMR 的单因子预测结果作为标准模型的输入变量进行综合水质预警. 关键环节的结果分析如下所述.

4.1　海水富营养化预测与评价结果的比较

根据等级判定标准和样本各指标值，采用等间隔均匀分布方式，内插水质指标标准数据，生成训练样本，经过网络反复训练，得出海水富营养化评价标准模型[15]. 建模说明如下：

(1)训练样本生成. 本文内插水质指标共生成 350 组数据，每组数据有 6 个输入指标(温度、浑浊度、酸碱度、溶解氧、硝酸盐浓度、叶绿素-a 浓度)，输出为综合水质指标. 选取前 300 组数据样本作为训练集.

(2)参数优化. 对 SVMR 参数进行初始化，设定 C、σ 的初始解空间均为(0，10)，设定鸟窝的个数为 20，即将解空间(0，10)分成符合高斯分布的 20 个区域，为后续的寻优做准备. 寻优过程中，最大迭代次数设为 50，淘汰率设为 0.25.

(3)对 800 组数据在不同隐含层神经元个数的情况下进行样本训练，通过比较来确定极限学习机隐含层的个数，训练结果如图 3 所示.

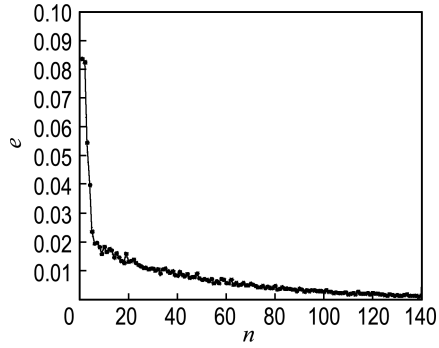

图 3　隐含层神经元个数对 ELMR 性能的影响

Fig. 3　The influence on the ELMR performance for the number of the hidden layer neurons

图 3 显示，在 ELMR 模型的隐含层神经元个数取 120～140 时，误差变化曲线趋于平坦，趋近于 0. 故 ELMR 模型的隐含层神经元节点设置为 140 个.

(4)海水富营养化的综合水质评价. 选取 800 组采集样本中的 300 组数据作为已建好的综合水质评价模型的输入样本，通过模型评价，得到采集区域的海水富营养化的综合水质等级.

BP 神经网络及 T-S 模糊神经网络也常被用来构建综合水质等级评价模型，ELMR-SVMR 与这两种模型针对综合水质等级评价的对比结果如表 3 所示.

表 3 综合水质等级评价结果对比

Tab. 3 Comparison of comprehensive seawater quality level prediction

模型	预测误差	预测误差百分比	平均绝对误差	均方误差	泛化误差
ELMR-SVMR	0.035 0	0.001 1	0.003 3	0.3×10^{-4}	−4.212 3
T-S	0.540 0	0.003 5	0.010 8	1.9×10^{-4}	−3.534 3
BP	5.909 7	0.033 6	0.118 2	0.017 5	−1.235 8

从表 3 中容易看出，3 种模型都拥有一定的泛化能力和预测精度，表明所建的模型是有效的. 通过 3 种模型的数据对比可以发现，T-S 模糊神经网络模型和 ELMR-SVMR 模型在各种指标上都要比 BP 神经网络模型好很多. 另外，T-S 模糊神经网络模型的各项误差指标值都比 ELMR-SVMR 模型大，说明 ELMR-SVMR 模型在预测精度和泛化能力方面均比 T-S 模糊神经网络模型要好.

为了验证 ELMR-SVMR 模型在水质等级预测方面的性能，在相同的实验条件下，分别对具有较好评价效果的 T-S 模糊神经网络模型、ELMR-SVMR 模型进行海水等级预测结果对比，仿真结果分别如图 4 和 5 所示.

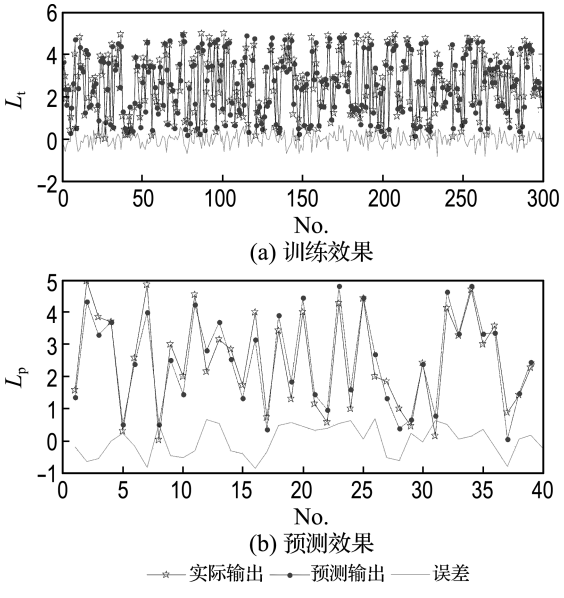

图 4 T-S 模糊神经网络训练与预测效果

Fig. 4 The effect of training and prediction of the T-S fuzzy neural networks

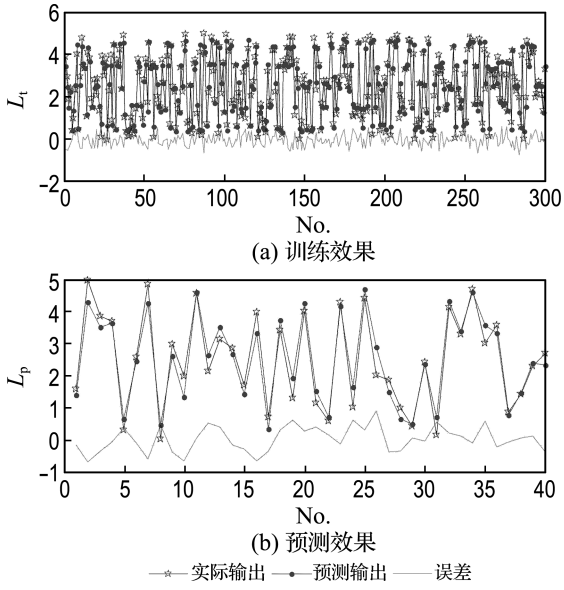

图 5 ELMR-SVMR 训练与预测效果

Fig. 5 The effect of training and prediction of ELMR-SVMR

从图 4、5 的水质等级预测及误差曲线可以看出，ELMR-SVMR 评价模型的总体预测误差要略小于 T-S 模糊神经网络模型，尤其是第 27 个样本之后，其误差波动比 T-S 模糊神经网络模型要小，说明 ELMR-SVMR 模型的海水等级预测的综合性能优于 T-S 模糊神经网络模型.

4.2 ELMR-SVMR 综合水质预警结果及分析

针对杭州湾东部某近海海域 2009 年第 4 季度的部分水质因子监测数据，基于 ELMR-SVMR 集成模型的海水水质综合预警仿真结果如图 6 所示.

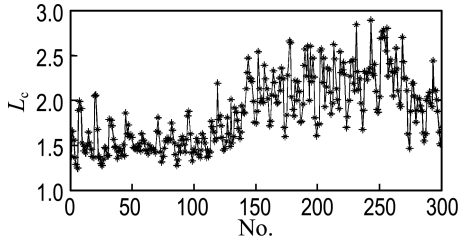

图 6 ELMR-SVMR 综合水质预警结果

Fig. 6 The result of comprehensive water quality forewarning by ELMR-SVMR

根据 2009 年中国环境状况公报，东海近岸海域为中度污染. 由水质评价结果可以看出，基于支持向量机评价模型的综合水质评价结果中，综合水质等级整体呈动态变化状态，数据变化较为齐整，能够较好地表现综合水质的动态变化趋势；从

水质等级上分析，水质处于由良好向中度污染过渡的状态趋势，基本与 2009 年中国环境状况公报显示的结果一致.

在本文中，通过一段时期内从杭州湾某监测海域获得的 800 组历史数据的分析及基于数据驱动的学习建模，获得一种具有水质预测及评价功能的集成预警模型. 在集成模型中，支持向量机回归用于构建海水水质评价模型，极限学习机回归用来对未来一段时间的水质状况进行综合预测，极限学习机回归的预测结果被作为评价模型的输入变量，最终实现海水富营养化综合预警的目的. 通过对集成模型灵敏程度的分析，可获得该模型预测误差结果对预警结果的影响程度，为集成预警模型的可靠性提供依据. 该模型能够对该区域内的海水水质状况做出基本正确的预测及评价，在趋势预测上具有一定的准确性，在该区域具有一定的适用性. 应该看到，不同地理环境下的海域特征还是具有一定差异性的，在起主要影响的海水理化因子的选择上会有所不同，一种环境下建立起来的模型不可能适应所有的地理环境. 但是，该类模型的创建机理还是值得借鉴的，其针对这类问题的模型构造方法具有一定的适用性.

5　结　语

本文首先对海水富营养化的时空概念模型进行分析，揭示了海水富营养化的成因及其演化过程，并从描述性分析和统计学分析这两方面对导致海水富营养化的多种因素进行筛选，选取出影响海水富营养化的主要因素，以准确地反映海水富营养化的整体状况. 针对海水环境中各水质理化因子状态变化的非线性和不确定性等特性，采用 SVMR 方法建立综合水质评价模型，通过布谷鸟算法对 SVMR 参数进行寻优. 将极限学习机预测结果作为该评价模型的输入变量进行海水富营养化的综合水质预警评价. 通过与其他方法的对比分析，验证了 ELMR-SVMR 集成模型的有效性.

参考文献：

[1] 俞志明，沈志良. 长江口水域富营养化[M]. 北京：科学出版社，2011.
YU Zhi-ming, SHEN Zhi-liang. **Eutrophication in Changjiang Estuary and Its Adjacent Waters** [M]. Beijing: Science Press, 2011. (in Chinese)

[2] LIU Zai-wen, WANG Xiao-yi, CUI Li-fen, *et al*. Research on water bloom prediction based on least squares support vector machine [C] // **2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009**. Piscataway: IEEE Computer Society, 2009: 764-768.

[3] Wu Q. Hybrid wavelet v-support vector machine and chaotic particle swarm optimization for regression estimation [J]. **Expert Systems with Applications**, 2011, **38**(12): 14624-14632.

[4] CHEN Fa-fa, TANG Bao-ping, SONG Tao, *et al*. Multi-fault diagnosis study on roller bearing based on multi-kernel support vector machine with chaotic particle swarm optimization [J]. **Measurement: Journal of the International Measurement Confederation**, 2014, **47**(1): 576-590.

[5] Pinto U, Maheshwari B, Shrestha S, *et al*. Modelling eutrophication and microbial risks in peri-urban river systems using discriminant function analysis [J]. **Water Research**, 2012, **46**(19): 6476-6488.

[6] 李新艳，杨丽标，晏维金. 长江输出溶解态无机磷的通量模型灵敏度分析及情景预测[J]. 湖泊科学，2011，**23**(2)：163-173.
LI Xin-yan, YANG Li-biao, YAN Wei-jin. Scenario prediction and sensitivity of modeling dissolved inorganic phosphorus export from the Yangtze River [J]. **Journal of Lake Science**, 2011, **23**(2): 163-173. (in Chinese)

[7] Eccles S R, Subbe C, Hancock D, *et al*. CREWS: Improving specificity whilst maintaining sensitivity of the national early warning score in patients with chronic hypoxaemia [J]. **Resuscitation**, 2014, **85**(1): 109-111.

[8] Rivas-Perea P, Cota-Ruiz J. An algorithm for training a large scale support vector machine for regression based on linear programming and decomposition methods [J]. **Pattern Recognition Letters**, 2013, **34**(4): 439-451.

[9] Patil S G, Mandal S, Hegde A V. Genetic algorithm based support vector machine regression in predicting wave transmission of horizontally interlaced multi-layer moored floating pipe breakwater [J]. **Advances in Engineering Software**,

2012, **45**(1):203-212.

[10] Wu Q, Law R, Wu E, *et al*. A hybrid-forecasting model reducing Gaussian noise based on the Gaussian support vector regression machine and chaotic particle swarm optimization [J]. **Information Sciences**, 2013, **238**:96-110.

[11] Ju Fuh-yuh, Hong Wei-chiang. Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting [J]. **Applied Mathematical Modelling**, 2013, **37**(23):9643-9651.

[12] Lovato T, Ciavatta S, Brigolin D, *et al*. Modelling dissolved oxygen and benthic algae dynamics in a coastal ecosystem by exploiting real-time monitoring data [J]. **Estuarine, Coastal and Shelf Science**, 2013, **119**:17-30.

[13] 暨卫东. 中国近海海洋环境质量现状与背景值研究[M]. 北京:海洋出版社, 2011.
JI Wei-dong. **China Offshore Marine Environment Quality Present Situation and the Background Research** [M]. Beijing: China Ocean Press, 2011. (in Chinese)

[14] Niero M, Pizzol M, Bruun H G, *et al*. Comparative life cycle assessment of wastewater treatment in Denmark including sensitivity and uncertainty analysis [J]. **Journal of Cleaner Production**, 2014, **68**:25-35.

[15] SHEN Ping-ping, LI Gang, HUANG Liang-min, *et al*. Spatiotemporal variability of phytoplankton assemblages in the pearl river estuary with special reference to the influence of turbidity and temperature [J]. **Continental Shelf Research**, 2011, **31**:1672-1681.

Research on forewarning model of seawater quality based on ELMR-SVMR

ZHANG Ying*, LI Mei, GAO Qian-qian, SHI Jia

(College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China)

Abstract: Forewarning of seawater quality has important significance for monitoring and protecting of marine environment. ELMR-SVMR (extreme learning machine regression-support vector machine regression) integrated model takes the ways of combining the prediction and evaluation to achieve the purpose of comprehensive forewarning for the state of seawater eutrophication. Therein SVMR method is used to establish the evaluation model, ELMR method is used to comprehensively predict the seawater quality state for a future time, and the result of the prediction by ELMR is used as the input variable of the evaluation model. The reliability of integrated model directly affects the effectiveness of the forewarning. The reliability of ELMR-SVMR model can be obtained by analyzing the sensitivity regarding the prediction error of ELMR to the evaluation model of SVMR. The prediction errors of all the parameters can be selected as the influence parameters of sensitivity of the evaluation model. The sensitivity evaluation of ELMR-SVMR model can be achieved by the sensitivity calculation. Compared with other methods by experiment and analysis, the effectiveness of the forewarning model of ELMR-SVMR is validated in this scope of sea area. It provides a pathway to establish the forewarning model of seawater quality effectively.

Key words: integrated forewarning model; extreme learning machine; support vector machine; seawater eutrophication; sensitivity analysis