

# 最小二乘支持向量机在太湖流域水质评价中的应用

陈曦<sup>1</sup>, 仇蕾<sup>1,2</sup>, 黄泽元<sup>1</sup>

(1. 河海大学商学院, 江苏 南京 211100;

2. 河海大学 水文水资源与水利工程科学国家重点实验室, 江苏 南京 210098)

**摘要:**为客观、准确地评价水质状况, 从而为水污染防治和水资源合理开发利用提供科学指导, 根据最小二乘支持向量机( LS-SVM) 的基本原理, 引入其分类算法构建太湖流域的水质评价模型, 以太湖流域 5 个重点断面为研究对象, 通过对已知训练样本进行学习训练, 对测试样本的水质等级进行评价, 并将其结果与 BP 神经网络、判别分析法相比较。结果表明, LS-SVM 在太湖流域水质评价方面有着更出色的效果, 可为太湖流域水资源管理提供新的参考方法。

**关键词:**最小二乘支持向量机; 太湖流域; 水质评价

**中图分类号:** X824, TP181 **文献标志码:** A **文章编号:** 1674-3075(2013)06-0016-06

太湖流域主要位于我国长江三角洲腹地, 自古以来即为“鱼米之乡”, 是我国最大的综合工业基地之一, 也是我国沿海主要对外开放地区(贾更华, 2013)。然而, 由于工业排污、生活垃圾等诸多原因, 城市湖泊目前都已处于重富营养或异常营养状态, 绝大部分大中型湖泊均已具备发生富营养化的条件或处于富营养化状态, 导致水量、水质和水生态系统发生巨大变化, 太湖流域的水生态环境出现了一定程度的恶化(李军等, 2006)。2012 年太湖流域 12 个水质监测点位中, 嘉兴王江泾多达 9 周为 V 类水质, 占全年监测周数的 33.33%, 多达 12 周为劣 V 类水质, 占全年监测周数的 44.44%, 太湖流域水体环境的改善迫在眉睫。

近年来, 许多学者从不同的角度对太湖流域水质评价做了研究, 常见的方法有多元统计分析(余辉等, 2010)、模糊综合评价(郑刚等, 2008)、灰色关联分析(安乐生等, 2012)、人工神经网络(吕恒等, 2006)等。支持向量机(Support Vector Machines, SVM)是一种非常有力的机器学习技术, 能够较好地解决小样本、非线性、高维数以及神经网络方法所面临的 结构选择、局部极小点等实际问题(张学工, 2000), 在诸多领域得到了广泛而成功的应用(Keun

Lee et al, 2011; 宋召青等, 2008; Fei-Long Chen et al, 2010)。最小二乘支持向量机(Least Squares Support Vector Machines, LS-SVM)是标准 SVM 的变形算法, 比标准的 SVM 有着更快的求解速度, 所需的计算资源较少(阎威武和邵惠鹤, 2003; S Deng et al, 2010)。本文根据 LS-SVM 模型的基本原理, 以太湖流域为例, 通过 LS-SVM 对已知样本的学习训练, 评价未知样本的水质等级, 并与 BP 神经网络、判别分析方法做了测试对比, 结果表明, LS-SVM 在太湖流域水质等级评价方面有着更可靠的结果。

## 1 最小二乘支持向量机原理

SVM 最早由 Vapnik 于 20 世纪 90 年代在统计学习理论(Statistical Learning Theory, STL)的基础上提出, 是从分类问题中线性可分情况下的最优分类面发展而来的, 其基本思想是通过某种特定的非线性映射将输入向量映射到高维特征空间中, 然后在此高维空间构建最优分类超平面, 将不同类别的样本无错误地分开且离超平面最近的向量到最优超平面的距离最大(张学工, 2000)。SVM 的基本思想可以用图 1 来表示, 图中  $\Delta$  和  $\blacktriangle$  各表示 1 类,  $H_1$ 、 $H_2$  分别为通过 2 类样本中离  $H$  最近的点且平行于  $H$  的直线( $H$  到  $H_1$ 、 $H_2$  的距离相等), 其中,  $H_1$ 、 $H_2$  为支持向量,  $H$  即为最优分割超平面。

LS-SVM 是标准 SVM 的一种变形与扩展, 求解速度更快, 所需的计算资源较少, 基本原理(J A K Suykens, 1999; 顾燕萍等, 2010)如下:

假设训练样本集  $S = \{(x_k, y_k) | k = 1, 2, \dots, n\}$ ,  $x_k \in R^n$ ,  $y_k \in R$ , 其中,  $x_k$  是输入数据,  $y_k$  是输出数据。

收稿日期: 2013-06-02

基金项目: 国家社会科学基金重点项目(10AJY005)。

通讯作者: 仇蕾, 1978 年生, 女, 江苏阜宁人, 博士, 副教授, 硕士生导师, 主要从事管理科学理论与方法、水资源与水环境系统管理、资源-环境-经济复合系统预警等方面的研究。

作者简介: 陈曦, 1989 年生, 女, 江苏镇江人, 硕士研究生, 主要从事环境经济及政策研究。E-mail: chenxxi89@126.com

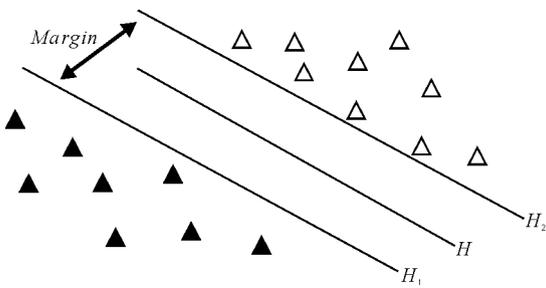


图1 支持向量机原理示意

Fig.1 Schematic diagram of SVM

在权  $w$  空间中最小二乘支持向量机分类问题可以描述为:

$$\min_{w,b,c} J(w, \varepsilon_k) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^N \varepsilon_k^2 \quad (1)$$

$$s. t. y_k [W^T \varphi(x_k) + b] = 1 - \varepsilon_k, k=1, 2, \dots, n$$

式(1)中,  $\varphi(x_k): R^n \rightarrow R^m$  是某种非线性映射,  $w \in R^m$  为权向量,  $\varepsilon_k \in R$  为误差变量,  $b$  是常值偏差, 损失函数  $J$  是误差和规则化量之和,  $\gamma$  是可调正则化参数, 控制对超出样本的惩罚。映射  $\varphi(x_k)$  的目的是从原始空间中抽取特征, 将原始空间中的 1 个样本映射为高维空间中的 1 个向量, 解决原始空间中线性不可分的问题。为解决式(1)所述优化问题, 构造 Lagrange 函数:

$$L(w, b, \varepsilon_k, a_k) = J(w, \varepsilon_k) - \sum_{k=1}^n a_k [w^T \varphi(x_k) + b + \varepsilon_k - y_k] \quad (2)$$

其中  $a_k \in R$  为 Lagrange 乘子。根据 KKT (Karush-Kuhn-Tucker) 条件, 可得到矩阵方程:

$$\begin{pmatrix} 0 & -Y^T \\ Y & \Omega + \frac{1}{\gamma} I \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (3)$$

式(3)中,  $Y$  为样本输出,  $I$  为单位矩阵,  $a = (a_1, a_2, \dots, a_n)$ ,  $\Omega_{kj} = K(x_k, x_j) = \varphi(x_k)^T \varphi(x_j)$ ,  $k, j = 1, 2, 3 \dots, n$ 。根据 Mercer 条件, 存在映射  $\varphi$  和核函数  $K(x_k, x_j)$  使得:

$$K(x_k, x_j) = \varphi(x_k)^T \varphi(x_j) \quad (4)$$

则可得 LS-SVM 的最优分类决策函数为:

$$y(x) = \text{sgn} \left\{ \sum_{k,j=1}^n a_k^* K(x_k, x_j) + b^* \right\} \quad (5)$$

其中  $a_k^*$  和  $b^*$  由式(3)求解出,  $K(x_i, x_j)$  为核函数。常用的核函数有多项式核函数、径向基(RBF)核函数和 Sigmoid 核函数(许桂梅和黄圣国, 2010)。本文选择了应用范围最广泛的径向基(RBF)核函数:

$$K(x_k, x_j) = \exp \left( -\frac{\|x_k - x_j\|^2}{2\sigma^2} \right) \quad (6)$$

## 2 基于 LS-SVM 的太湖流域水质评价

### 2.1 模型构建

为利用 LS-SVM 分类算法进行太湖流域水质评价, 首先要确定影响太湖流域水质的主要因素, 然后选择样本数据进行 LS-SVM 模型的学习训练, 最后利用训练好的模型进行预测, 得出未知水质等级样本的评价结果。

根据国家地表水环境质量标准(GB3838-2002)的规定, 依据地表水水域环境功能和保护目标, 按功能高低依次划分为 5 类(I、II、III、IV、V)。不同功能类别分别执行相应类别的标准值, 水域功能类别高的标准值严于水域功能类别低的标准值。考虑到数据的可获得性及有效性, 最终选取国家环保部数据中心公布的 pH 值、溶解氧(DO)、高锰酸钾盐指数(COD<sub>Mn</sub>)、氨氮(NH<sub>3</sub>-N) 4 个评价指标数据进行太湖流域水质评价研究。国家地表水环境质量标准(GB3838-2002)中以上 4 个评价指标的标准限值见表 1。

表 1 4 个评价指标的标准限值

Tab.1 Standard limit of 4 assessment criteria

指 标	分 类				
	I 类	II 类	III 类	IV 类	V 类
pH	6~9	6~9	6~9	6~9	6~9
DO/mg · L <sup>-1</sup> ≥	7.5	6	5	3	2
COD <sub>Mn</sub> /mg · L <sup>-1</sup> ≤	15	15	20	30	40
NH <sub>3</sub> -N/mg · L <sup>-1</sup> ≤	0.15	0.5	1.0	1.5	2.0

本文以我国环保部 2010 年第 1 周~2013 年第 10 周公布的太湖流域浙江湖州新塘港、浙江嘉兴王江泾、江苏苏州西山、江苏无锡沙渚、江苏宜兴兰山嘴 5 个重点断面水质周报数据为研究对象, 原始数据来源于国家环保部数据中心(<http://datacenter.mep.gov.cn/>)。每个监测点位均有 138 组数据, 包括表 1 中 4 个评价指标及对应的水质等级。其中, 每个监测点位均随机抽取 80 组数据作为 LS-SVM 模型的训练样本以完成 LS-SVM 模型的学习训练和确定最佳模型; 剩余 58 组数据作为 LS-SVM 模型的测试样本(假设水质等级未知), 利用训练好的 LS-SVM 模型根据输入测试样本指标数据对其水质等级进行自动分类评价, 然后将其结果与实际水质等级进行对比, 并与 BP 神经网络、判别分析法评价结果进行比较, 以测试 LS-SVM 模型的水质评价效果。根据 LS-SVM 基本原理, 基于 LS-SVM 模型的太湖流域水质评价流程如图 2。

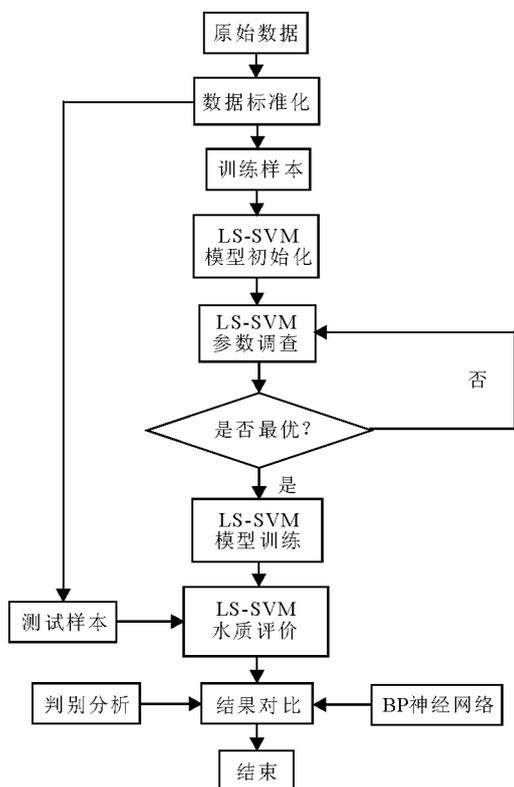


图2 基于LS-SVM的太湖流域水质评价流程

Fig.2 Evaluation procedure of Lake Tai Basin based on LS-SVM

## 2.2 数据标准化

在抽选训练样本和测试样本之前,为了消除指标量纲所带来的不可公度性影响,达到较好的评价效果,需首先对原始数据进行标准化处理,使指标值均标准化为区间 $[0,1]$ 内的无量纲数值,且数值越大越优、越小越劣。对于太湖流域水质评价中138个研究样本、4个指标,原始评价可记为:

$$X = (x_{k1}, x_{k2}, x_{k3}, x_{k4})$$

$$= \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ \dots & \dots & \dots & \dots \\ x_{138,1} & x_{138,2} & x_{138,3} & x_{138,4} \end{pmatrix}, k=1, 2, \dots, 138 \quad (8)$$

式(8)中,pH值、DO、 $\text{COD}_{\text{Mn}}$ 、 $\text{NH}_3\text{-N}$  4个评价指标的138个样本值分别记为 $x_{k1}, x_{k2}, x_{k3}, x_{k4}$  ( $k=1, 2, \dots, 138$ )。经过标准化处理后,记标准评价矩阵为:

$$Z = (z_{k1}, z_{k2}, z_{k3}, z_{k4})$$

$$= \begin{pmatrix} z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & z_{23} & z_{24} \\ \dots & \dots & \dots & \dots \\ z_{138,1} & z_{138,2} & z_{138,3} & z_{138,4} \end{pmatrix}, k=1, 2, \dots, 138 \quad (9)$$

根据国家地表水环境质量标准(GB 3838-

2002), I~V类水质的pH值在6~9,pH值超过此范围的均列为劣V类水质,DO为正向型指标(值越大水质越好), $\text{COD}_{\text{Mn}}$ 和 $\text{NH}_3\text{-N}$ 均为负向型指标(值越大水质越劣)。因此,假设最佳pH值为 $x_{k1}^o$ (本文设 $x_{k1}^o=7$ ),则pH值采用式(10)、DO采取式(11)、 $\text{COD}_{\text{Mn}}$ 和 $\text{NH}_3\text{-N}$ 采取式(12)所示方式进行标准化处理。

$$z_{k1} = \begin{cases} 0 & \text{若 } x_{k1} \leq 6 \text{ 或 } x_{k1} \geq 9 \\ 1 - \frac{x_{k1}^o - x_{k1}}{x_{k1}^o - 6} & \text{若 } 6 < x_{k1} < x_{k1}^o \\ 1 & \text{若 } x_{k1} = x_{k1}^o \\ 1 - \frac{x_{k1} - x_{k1}^o}{9 - x_{k1}^o} & \text{若 } x_{k1}^o < x_{k1} < 9 \end{cases}, (1 \leq k \leq 138) \quad (10)$$

$$z_{k2} = \frac{x_{k2} - \min x_{k2}}{\max x_{k2} - \min x_{k2}}, (1 \leq k \leq 138) \quad (11)$$

$$z_{kj} = \frac{\max x_{kj} - x_{kj}}{\max x_{kj} - \min x_{kj}}, (1 \leq k \leq 138, j=3, 4) \quad (12)$$

## 2.3 模型结果及对比

在完成数据标准化处理后,分别输入太湖流域5个监测点位的训练样本数据和测试样本数据即可完成LS-SVM模型的学习训练以及对测试样本水质等级的评价。太湖流域水质评价问题属于多分类问题,借助LS-SVMlab v1.8程序包完成LS-SVM模型的初始化和参数优化,以及完成了多分类问题。本文借助该程序包,在Matlab环境中完成了LS-SVM模型训练和对测试样本未知水质等级的评价。

惩罚参数 $\gamma$ 和核参数 $\sigma^2$ 是LS-SVM模型的关键参数,参数的选择对模型性能有很大影响,其优化是LS-SVM模型的难点,本文采用网格搜索(Grid Search)方法(李琳和张晓龙,2006)进行交叉验证(Cross Validation)完成。对于浙江湖州新塘港、浙江嘉兴王江泾、江苏苏州西山、江苏无锡沙渚、江苏宜兴兰山嘴5个点位的训练样本数据,各监测点位的LS-SVM模型参数如表2,训练情况如表3。从表3所示各个监测点位训练样本的训练情况看,LS-SVM模型对各个监测点位训练样本的水质评价均达到了较好的效果。为了检验LS-SVM模型的泛化能力,还需利用测试样本指标数据。假设各测试样本水质等级未知,利用训练好的LS-SVM模型根据输入测试样本的指标数据自动地完成其水质等级评价,然后将评价结果与实际水质等级进行对比,并通过与BP神经网络、判别分析法进行性能比较,测试

其评价能力。其中, BP神经网络采用单层隐含层网络结构,记 $n_i$ 为输入节点个数, $n_0$ 为输出节点个数, $a$ 为1~10的整数,则隐含层最佳节点 $n$ 个数参照经验公式(13)确定(张丰德,2009);判别分析法借助统计分析软件 SPSS 19.0 完成。

$$n = \sqrt{n_i + n_0} + a \quad (13)$$

LS-SVM 模型对各监测点位各类测试样本的水质等级评价情况见表 4, LS-SVM 模型与 BP 神经网络及判别分析法的评价效果对比情况见表 5。

综合表 3~表 5 可知,基于 LS-SVM 的太湖流域水质评价模型不仅可以对已知水质等级的训练样本

达到较好的训练识别效果,还有出色的泛化能力,即可对测试样本(假设水质等级未知)取得较高的分类正确率,并且相比于 BP 神经网络方法及判别分析方法也有着更高的分类正确率。

表 2 各监测点位 LS-SVM 模型参数寻优结果

**Tab.2 LS-SVM optimization results of every monitoring point**

模型参数	监测点位				
	湖州 新塘港	嘉兴 王江泾	苏州 西山	无锡 沙渚	宜兴 兰山嘴
惩罚参数	1.599	20.688	50 964.398	0.176	2.508
核参数	1.426	2.216	727.028	0.022	1.200

表 3 各监测点位训练样本的 LS-SVM 模型训练情况

Tab.3 Training of LS-SVM samples of every monitoring point

监测点位	项 目	分 类						
		整体	I 类	II 类	III 类	IV 类	V 类	劣 V 类
湖州新塘港	样本数/个	80	-	22	44	14	-	-
	分类正确率/%	97.50	-	95.45	97.73	100	-	-
嘉兴王江泾	样本数/个	80	-	-	2	26	29	23
	分类正确率/%	96.25	-	-	100.00	92	97	100
苏州西山	样本数/个	80	2	42	26	10	-	-
	分类正确率/%	92.50	0.00	98	88.46	100	-	-
无锡沙渚	样本数/个	80	-	57	22	-	-	1
	分类正确率/%	98.75	-	100	100.00	-	-	0.00
宜兴兰山嘴	样本数/个	80	-	14	32	27	5	2
	分类正确率/%	96.25	-	92.86	100.00	100.00	100	0.00

注:“-”表示该监测点位无此类样本,劣 V 类水质污染程度已超过 GB3838-2002 界定的 V 类。

Note:“-”means no sample for the monitoring point, class worsen than V was exceeded ClassV defined by GB3838-2002.

表 4 LS-SVM 模型对各监测点位各类测试样本的水质等级评价情况

Tab.4 Grade evaluation of testing samples of every monitoring point based on LS-SVM

监测点位	项 目	分 类						
		整体	I 类	II 类	III 类	IV 类	V 类	劣 V 类
湖州新塘港	样本数/个	58	-	18	38	2	-	-
	分类正确率/%	86.21	-	83.33	86.84	100.00	-	-
嘉兴王江泾	样本数/个	58	-	-	2	28	15	13
	分类正确率/%	87.93	-	-	0.00	92.86	86.67	92.31
苏州西山	样本数/个	58	1	34	7	6	-	-
	分类正确率/%	84.48	0.00	85.29	88.24	83.33	-	-
无锡沙渚	样本数/个	58	-	46	12	-	-	-
	分类正确率/%	86.21	-	95.65	50.00	-	-	-
宜兴兰山嘴	样本数/个	58	-	13	28	14	3	0
	分类正确率/%	77.59	-	84.62	78.57	85.71	0.00	-

注:“-”表示该监测点位无此类样本,劣 V 类水质污染程度已超过 GB3838-2002 界定的 V 类。

Note:“-”means no sample for the monitoring point, class worsen than V was exceeded ClassV defined by GB3838-2002.

### 3 结束语

本文构建的太湖流域水质评价 LS-SVM 模型能自动评价未知样本的水质等级,且与 BP 神经网络、判别分析法等常用的分类评价方法相比,其评价结果与实际情况更为相符,可以为太湖流域水资源管理及其他环境质量评价工作提供一种新的科学手段

与决策参考。本文选取的样本在时间上有一定的跨度,由于不同季节的水质指标可能会有较大的波动,从而影响模型的评价效果,但即使是在这种情况下,基于 LS-SVM 的太湖流域水质评价模型依然能有着较高的评判正确率,且要优于 BP 神经网络等常用方法,更说明其可靠性;如果缩小时间跨度、减小不同时期水质波动带来的影响,LS-SVM 模型的分

表5 3种方法太湖流域水质等级评价效果比较

Tab.5 Comparison of evaluating results of 3 methods for water quality assessment in Lake Tai Basin

监测点位	分类正确率/%		
	LS-SVM	BP神经网络	判别分析
湖州新塘港	86.21	81.03	67.24
嘉兴王江泾	87.93	79.31	65.52
苏州西山	84.48	75.86	60.34
无锡沙渚	86.21	81.03	74.14
宜兴兰山嘴	77.59	68.97	67.24

评价正确率和可靠性将会更好。LS-SVM 模型的分类评价仍需做进一步的深入研究。例如,本文选择采用径向基(RBF)核函数,在不同的复杂水体环境下必定不能都适用,如何选择或构造适当的核函数是一个重要问题;此外,由于客观条件限制,本文只选取 pH 值、DO、COD<sub>Mn</sub>、NH<sub>3</sub>-N 4 个评价指标,而从不同角度反映水质信息的水环境评价指标非常多,如果指标增加到十几个甚至更多,模型的评价效果需作进一步研究。

#### 参考文献

- 安乐生,赵全生,刘贯群,等. 2012. 代表性水质评价方法的比较研究[J]. 中国环境监测, 26(5): 47-50.
- 顾燕萍,赵文杰,吴占松. 2010. 最小二乘支持向量机的算法研究[J]. 清华大学学报:自然科学版, 50(7): 1063-1066,1071.
- 贾更华. 2013. 太湖流域综合治理的科技需求分析[J]. 人民黄河, 35(1): 66-68.
- 李军,刘丛强,肖化云,等. 2006. 太湖北部夏季浮游藻类多样性与水质评价[J]. 生态环境, 15(3): 453-456.
- 李琳,张晓龙. 2006. 基于RBF核的SVM学习算法的优化计算[J]. 计算机工程与应用, 42(29): 190-192.
- 吕恒,李新国,曹凯. 2006. 基于BP神经网络模型的太湖

- 悬浮物浓度遥感定量提取研究[J]. 武汉大学学报:信息科学版, 31(8): 683-686.
- 宋召青,崔和,胡云安. 2008. 支持向量机理论的研究与进展[J]. 海军航空工程学院学报, 23(2): 143-152.
- 许桂梅,黄圣国. 2010. 应用LS-SVM的飞机重着陆诊断[J]. 系统工程理论与实践, 30(4): 763-768.
- 阎威武,邵惠鹤. 2003. 支持向量机和最小二乘支持向量机的比较及应用研究[J]. 控制与决策, 18(3): 358-360.
- 余辉,燕姝雯,徐军. 2010. 太湖出入湖河流水质多元统计分析[J]. 长江流域资源与环境, 19(6): 696-702.
- 张学工. 2000. 关于统计学习理论与支持向量机[J]. 自动化学报, 26(1): 32-42.
- 张丰德. 2009. Matlab神经网络仿真与应用[M]. 北京:电子工业出版社.
- 郑刚,刘庄,张永春,等. 2008. 基于模糊综合评价的流域社会经济活动对太湖生态影响评价研究[J]. 环境工程学报, 2(12): 1705-1710.
- Fei-Long Chen, Feng-Chia Li. 2010. Combination of feature selection approaches with SVM in credit scoring[J]. Expert Systems with Applications, 37(7): 4902-4909.
- J A K Suykens, J Vandewalle. 1999. Least Squares Support Vector Machine Classifiers [J]. Neural Processing Letters, 9(3): 293-300.
- Keun Lee, Sohyung Cho, Shihab Asfour. 2011. Web-based algorithm for cylindricity evaluation using support vector machine learning [J]. Computers & Industrial Engineering, 60(2): 228-235.
- S Deng, Tsung-Han Yeh. 2010. Applying least squares support vector machines to the airframe wing-box structural design cost estimation [J]. Expert Systems with Applications, 37(12): 8417-8423.

(责任编辑 张俊友)

## Application of Least Square-Support Vector Machine in Water Quality Assessment of Taihu Lake Basin

CHEN Xi<sup>1</sup>, QIU Lei<sup>1,2</sup>, HUANG Ze-yuan<sup>1</sup>

(1. Business School of Hohai University, Nanjing 211100, P. R. China;  
2. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering,  
Hohai University & NHRI, Nanjing 210098, P. R. China)

**Abstract:** In order to achieve an objective and accurate assessment of water quality and provide a scientific guidance to water pollution prevention and rational utilization of water resources, a model based on LS-SVM was constructed to evaluate the Taihu Lake Basin's water quality. Five monitoring sections in the Taihu Lake Basin were taken as examples. The LS-SVM models were established with the samples which water quality grades were already known, then water quality of testing samples were evaluated by using the well trained LS-SVM models. Furthermore, the Back-Propagation Neural Network (BPNN) and Discriminant Analysis were also used with the same testing samples to testify the method's efficiency and accuracy. The results comparison of three methods showed that the LS-SVM method presented in this paper performed much better than BPNN and Discriminant Analysis Method in the water quality assessment of the Taihu Lake Basin. In conclusion, the LS-SVM might be a new reference method for the Taihu Lake Basin's water resources management.

**Keywords:** Least Square-Support Vector Machines; Taihu Lake Basin; water quality assessment