

DOI: 10.3969/j.issn.1004-6933.2013.05.005

# 支持向量机在水资源类综合评价中的应用

## ——以全国 31 个省级行政区水资源合理性配置为例

崔东文

(云南省文山州水务局, 云南 文山 663000)

**摘要:** 简要分析水资源类综合评价中存在的问题, 针对各种人工神经网络(ANN)应用于水资源类综合评价中存在的缺点, 采取基于支持向量机(SVM)基本原理和交叉验证(CV)方法, 提出一种通用的 CV-SVM 水资源类综合评价模型, 并构建 GA-BP、传统 BP 及 RBF 评价模型作为对比模型。首先, 基于水资源配置基本内涵, 构建水资源合理配置评价指标体系和分级标准, 采用随机内插的方法在各评价分级标准阈值间生成训练样本和检验样本, 选用决定系数  $R^2$  等 6 个统计学指标用于评价模型性能, 在达到预期评价精度后将模型应用于水资源类综合评价中, 以解决此类综合评价中由于“欠拟合”与“过拟合”降低网络泛化能力的问题。结果表明, CV-SVM 模型性能和评价精度优于 GA-BP、RBF 和 BP 网络模型, 该模型应用于水资源类综合评价是合理可行和有效的。最后应用 CV-SVM 评价模型, 以全国 31 个省级行政区水资源合理性配置为例进行模型验证。

**关键词:** 水资源; 综合评价; 支持向量机; 神经网络; 交叉验证

中图分类号: P331      文献标志码: A      文章编号: 1004-6933(2013)05-0020-08

### Support vector machine for comprehensive evaluation of water resources: Application to reasonable allocation of water resources in 31 provincial-level administrative regions in China

CUI Dongwen

(Wenshan Water Conservancy Bureau of Yunnan Province, Wenshan 663000, China)

**Abstract:** A brief analysis of the existing problems in comprehensive evaluation of water resources was carried out. To overcome the shortcomings of a variety of artificial neural networks in comprehensive evaluation of water resources, a universal CV-SVM water resources comprehensive evaluation model was developed based on the fundamental principle of the support vector machine (SVM) and the cross-validation (CV) method, and GA-BP, traditional BP, and RBF evaluation models were built as comparison models. The random interpolation method was used to generate the evaluation classification threshold between training and testing samples, and six statistical indicators, including the coefficient of determination  $R^2$ , were used to evaluate the model's performance, in accordance with relevant evaluation index systems and grading standards. When the desired evaluation accuracy was achieved, the model was applied to comprehensive evaluation of water resources, in order to solve the problems of under-fitting and over-fitting, which can reduce the network generalization ability. The results show that the CV-SVM model has better performance and higher evaluation accuracy than GA-BP, RBF, and BP network models, and it is feasible and effective to apply the CV-SVM model to comprehensive evaluation of water resources. The CV-SVM model was validated through its application to reasonable allocation of water resources in 31 provincial-level administrative regions in China.

**Key words:** water resources; comprehensive evaluation; support vector machine; neural network; cross-validation

# 1 研究背景

水资源类综合评价问题可分为水资源开发利用评价、合理性配置评价、承载能力评价、水旱灾害评价、水生态环境评价、水资源管理有效性评价以及水资源紧缺度、水安全、用水水平、水资源再生能力等的评价<sup>[1]</sup>,其评价过程具有多指标、高维、非线性特征,评价结果对水资源规划、管理及保护具有重要意义。目前,此类综合评价常用的方法有模糊集理论方法<sup>[2-3]</sup>、人工神经网络方法<sup>[4-7]</sup>、灰色系统理论法<sup>[8]</sup>、数理统计方法等等,但由于水资源类综合评价是一个复杂巨系统,涉及水资源条件、经济、环境和水资源管理等多个方面,评价指标众多,对评价指标的筛选与权重的赋值普遍采用层次分析、Delphi 等方法<sup>[1]</sup>,存在明显的主观臆断成分。水资源类综合评价并不适宜建立常规数学模型,只能借助诸如人工智能、模糊识别、知识工程等方法建立模型,以处理多指标系统的综合评价问题<sup>[9]</sup>。模拟智能方法目前已成为建立和评价这类复杂系统最为有效的途径之一,而人工神经网络(artificial neural network,ANN)则是这类智能算法中运用最为广泛的算法之一。ANN 具有较强的非线性映射能力、鲁棒性、容错性和自适应、自组织、自学习等许多特性,适宜解决高维、非线性系统问题,是此类算法中运用最为广泛的算法之一,将 ANN 运用于水资源类综合评价中,存在两个关键性问题:①统一完善的评价指标体系和分级标准。目前具有统一完善的评价指标体系和分级标准的水资源类综合评价相对较少,如水质综合评价、湖库营养状态评价等,其余多无较为统一完善和普遍认同的评价指标和分级标准,需要参考相关文献进行构建<sup>[1]</sup>;②有效解决神经网络由于“欠拟合”与“过拟合”所导致的网络泛化能力降低的问题<sup>[10]</sup>。由于网络在训练过程中常出现的两种状态:“欠拟合”和“过拟合”,二者均会使网络的泛化能力下降。网络“欠拟合”表示模型无法完全探测到复杂数据集中的信号,使得训练达不到预期的评价或预测精度。网络“过拟合”表示模型会将信号连同噪声一起进行拟合,使得模型“记住”了训练样本的信息而使网络泛化能力降低,得不到理想的评价或预测精度<sup>[10-11]</sup>,目前,理论上尚无较好解决网络训练过程中由于“欠拟合”和“过拟合”导致泛化能力降低的方法。ANN 中常见的几种网络或多或少地存在缺陷或不足,如:BP 网络(back-propagation network,BP)存在着学习收敛速度慢、易陷入局部极值等不足,为克服 BP 算法中的不足,学者们提出基于附加动量、自适应调整参数、弹性方

法、拟牛顿法、共轭梯度法以及 Levenberg-Marquardt 等的改进算法<sup>[12-14]</sup>,但在实际应用中仍不能完全克服 BP 算法固有的缺陷。针对传统 BP 算法的固有缺陷,目前普遍采用增加隐含层数或运用遗传算法(genetic algorithm,GA)对 BP 网络权值及阈值进行优化,以期提高网络的预测精度。然而增加隐含层数或运用 GA 优化网络权值及阈值面临着复杂的隐含层、相关参数、编码等的选取困难以及早熟收敛问题<sup>[10]</sup>,且均不能较大地改善网络泛化能力。Elman 网络是一种典型的动态神经网络,网络结构及算法比之 BP 有着较大改善,但由于 Elman 网络是在 BP 网络的基础上改进而来,也采用 BP 算法进行权值修正,因此同样存在着学习速度较慢,易陷入局部极小值等缺点。RBF 网络(radial basis function neural network,RBF)虽然有着良好的非线性逼近能力和精度,但 RBF 径向基函数和隐含层神经元中心的选取是制约其精度提高的主要因素。PNN 网络(probabilistic neural networks,PNN)主要用于模式识别,在用于预测或识别时,由于其采用的平滑参数,使得预测精度或识别能力不高。GRNN 网络(generalized regression neural network,GRNN)具有良好的泛化性能和逼近能力,但用于预测分析时,由于 GRNN 平滑因子难以确定,对网络的预测性能影响较大。以上网络均在一定程度上影响了 ANN 在水资源类综合评价中的精度和泛化能力。支持向量机(support vector machine,SVM)是 20 世纪 90 年代中后期发展起来的基于统计学习理论构建的典型神经网络<sup>[15-16]</sup>,它由 Vapnik 首先提出,是一种通用的前馈神经网络,用于解决模式分类和非线性映射问题。SVM 具有严谨的数学基础,通过统计学习中的 VC 维(vapnik-chervonenkis dimension)理论和寻求结构风险最小化原理来提高泛化能力,已成为继 ANN 之后机器学习领域新的研究热点,其三大优势决定了它在机器学习领域有着举足轻重的地位:①SVM 以最小结构风险代替传统 BP 等网络的经验风险,求解的是一个二次寻优问题,理论上得到全局最优,解决了传统 BP 算法中难于克服的局部极值缺陷;②SVM 拓扑结构由支持向量决定,弥补了 BP 等网络结构难以确定的不足;③SVM 决策函数由少数的支持向量确定,计算的复杂程度取决于支持向量的数目,而不是样本空间的维数,避免了“维数灾”问题。

针对上述问题及原因,笔者参考相关文献<sup>[1]</sup>,以全国 31 个省级行政区水资源合理性配置为例,基于 SVM 基本原理和交叉验证(cross validation,CV)方法,提出一种通用的 CV-SVM 水资源类综合评价

方法,采用随机内插的方法在各评价分级标准阈值间生成训练样本和检验样本,选用决定系数  $R^2$ ,以及平均绝对误差  $e_{MAE}$ 、均方根绝对误差  $e_{RMSE}$ 、平均相对误差  $e_{MRE}$ 、均方根相对误差  $e_{RMAPE}$  以及最大相对误差  $e_{maxRE}$  6个统计学指标,对 SVM、GA-BP、RBF 以及 BP 网络模型进行性能评价,以控制网络训练过程中的“欠拟合”和“过拟合”问题,在达到预期的评价精度后将模型运用于水资源类综合评价中,并认为,在训练样本、检验样本的  $R^2$  足够大,  $e_{MAE}$ 、 $e_{RMSE}$ 、 $e_{MRE}$ 、 $e_{RMAPE}$  和  $e_{maxRE}$  足够小,且满足训练样本、检验样本的各统计学指标相近时,认为模型具有较好的泛化能力,此时可将该模型应用于该类水资源综合评价。

## 2 支持向量机(SVM)

SVM 应用于回归时,其基本思想不再是寻找最优分类面将样本分开,而是寻找一个最优超平面,使得所有训练样本离该最优超平面距离最短,这个超平面可看作拟合好的曲线。SVM 类似于一个 3 层前馈神经网络,其隐层节点数对应于输入样本与一个支持向量机的内积核函数,输出节点数对应于隐层输出的线性组合<sup>[15-16]</sup>。SVM 神经网络结构参见文献[17]。

为不失一般性,设含有  $l$  个训练样本的集合  $\{(x_i, y_i), i=1, 2, \dots, l\}$ , 其中,  $\mathbf{x}(x_i \in R^d)$  是第  $i$  个训练样本的输入列向量,  $\mathbf{x} = [x_i^1, x_i^2, \dots, x_i^d]^T$ ,  $y_i \in R$  为对应输出值。在高维特征中建立的线性回归函数为

$$f(x) = \mathbf{w}^T \Phi(x) + b \quad (1)$$

式中:  $\Phi(x)$  为非线性映射函数;  $\mathbf{w}$  为超平面的权值向量;  $b$  为偏置项。

定义  $\varepsilon$  线性不敏感损失函数为

$$L(f(x), y, \varepsilon) = \begin{cases} 0 & |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & |y - f(x)| > \varepsilon \end{cases} \quad (2)$$

式中:  $f(x)$  为回归函数返回的预测值;  $y$  为对应的真实值;  $\varepsilon$  为不敏感损失函数所定义的误差。

类似于 SVM 分类情况,引入松弛变量  $\xi_i, \xi_i^*$ , 并将上述寻找  $\mathbf{w}$ 、 $b$  的问题用数学语言描述出来,即

$$\begin{cases} \min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s. t.} \begin{cases} y_i - \mathbf{w}^T \Phi(x_i) - b \leq \varepsilon + \xi_i \\ -y_i + \mathbf{w}^T \Phi(x_i) + b \leq \varepsilon + \xi_i^*, i=1, 2, \dots, l \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \end{cases} \quad (3)$$

式中:  $C$  为惩罚因子,  $C$  越大表示对训练误差大于  $\varepsilon$  的样本惩罚越大;  $\varepsilon$  规定了回归函数的误差要求,  $\varepsilon$

越小,表示回归函数的误差越小。求解式(3)时,同时引入 Lagrange 函数,并转换成对偶形式:

$$\begin{cases} \max \left[ -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (a_i - a_i^*) (a_j - a_j^*) K(x_i, x_j) - \sum_{i=1}^l (a_i + a_i^*) \varepsilon + \sum_{i=1}^l (a_i - a_i^*) y_i \right] \\ \text{s. t.} \begin{cases} \sum_{i=1}^l (a_i - a_i^*) = 0 \\ 0 \leq a_i \leq C \\ 0 \leq a_i^* \leq C \end{cases} \end{cases} \quad (4)$$

式中:  $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$  为核函数;  $a_i, a_i^*$  为拉格朗日乘子。

设求解式(4)得到的最优解为  $a_i = [a_1, a_2, \dots, a_l]$ ,  $a_i^* = [a_1^*, a_2^*, \dots, a_l^*]$ , 则有

$$\mathbf{w}^* = \sum_{i=1}^l (a_i - a_i^*) \Phi(x_i) \quad (5)$$

$$b^* = \frac{1}{N_{\text{nsv}}} \left\{ \sum_{0 < a_i < C} [y_i - \sum_{x_j \in \text{SV}} (a_i - a_i^*) K(x_i, x_j) - \varepsilon] + \sum_{0 < a_i < C} [y_i - \sum_{x_j \in \text{SV}} (a_i - a_i^*) K(x_i, x_j) + \varepsilon] \right\} \quad (6)$$

式中,  $N_{\text{nsv}}$  为支持向量机个数。

将  $\mathbf{w}^*$ 、 $b^*$  代入式(1)得到回归函数为

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) K(x_i, x) + b^* \quad (7)$$

其中,只要部分参数  $(a_i - a_i^*) \neq 0$ , 其对应的样本  $\mathbf{x}_i$  即为问题中的支持向量。

研究表明,只要满足 Mercer 条件的对称函数即可作为 SVM 核函数,常用的核函数主要类型有线性核函数 ( $K(x, x_i) = \mathbf{x}^T x_i$ )、多项式核函数 ( $K(x, x_i) = (\gamma \mathbf{x}^T x_i + r)^p, \gamma > 0$ )、径向基核函数 ( $K(x, x_i) = \exp(-\gamma \|x - x_i\|^2), \gamma > 0$ ) 和两层感知核函数 ( $K(x, x_i) = \tanh(\gamma \mathbf{x}^T x_i + r)$ )。核函数可以看成是实际问题的特征提取过程,核函数的合理选取有助于提高模型精度。SVM 模型选择主要有两个步骤,一是核函数的选择,二是 SVM 本身的两个参数以及所选取的核函数所对应的参数<sup>[18-19]</sup>。本文选择径向基核函数为 SVM 的核函数。

## 3 水资源合理性配置综合评价

### 3.1 评价指标

水资源配置是指在流域或特定区域范围内,通过工程与非工程措施,利用系统分析方法、决策理论和计算机技术,统一调配水资源,协调各区域、各部门之间的利益与矛盾,使有限的、不同形式的水资源在各用水户之间进行科学分配,以提高区域整体的

用水效率,保障区域水资源的可持续开发利用,促进区域经济社会的可持续发展<sup>[1,7,20-21]</sup>。水资源配置的目的是在抑制需求与增加供给之间找到平衡点。水资源合理性配置综合评价是在水资源现状配置的基础上,遵循公平、高效、合理的准则,从经济、社会、生态、效率及水资源开发利用5个方面综合研究水资源在区域内不同属性和不同作用的生产、生活、生态用水的满足程度及分配合理性,并对在现状配置和规划配置条件下产生的综合效益进行评价,由此判断现状水资源配置的合理性与否及未来规划水平年配置结果合理性变化趋势<sup>[1]</sup>。

对水资源合理性配置进行综合评价,首先必须确定一套评价指标<sup>[1]</sup>,只有确定了恰当的评判标准,才能科学地评判各区域水资源配置成果的合理性,开展水资源配置成果的评比和择优。笔者综合考虑反映水资源合理配置的社会、经济、生态环境、效率和开发利用5个层面,依据相关文献<sup>[1,22]</sup>,选取缺水率等13个指标作为评价因子,对区域水资源配置合理性进行综合评价。合理性配置评价指标体系及分级标准见表1~2。

表2 将水资源合理性配置指标评价标准量化分

表1 区域水资源合理性配置评价指标

分层	社会合理性			经济合理性		生态环境合理性		效率合理性			开发合理性		
指标	缺水率/%	10年用水变化均衡率/%	用水结构系数	用水弹性系数	用水GDP/(元·m <sup>-3</sup> )	河道内生态用水比例/%	水功能区达标率/%	农业灌溉水利用系数	工业用水重复利用率/%	管网漏损率/%	水资源开发利用率/%	人均用水量/m <sup>3</sup>	地下水利用指数
公式及含义	区域缺水总量/区域需求总量	10年区域用水增长率/同期全国用水增长率	10年内生活、工业、农业和生态用水结构变化系数	用(需)水量增长率/同期GDP增长率	地区生产总值/总用水量	河道内生态用水量/水资源总量	评价期功能区内达标的评价河长/总评价功能区内评价河长	田间用水量/取水口取水量	工业重复用水量/工业用水量	漏损水量/水源取水总量	实际供水量/当地年均水资源量	用水量/总人口	地下水实际开采量/地下水可开采量
表征	反映区域水资源需求未满足程度	反映区域用水变化的相对程度	反映生活、工业、农业和生态用水情况	反映用水量对经济增长的弹性影响	反映用水效益和效率的一个指标	反映河道内生态用水状况	反映区域水功能区达标状况	反映灌溉水利用程度	反映工业用水重复利用程度	反映管网漏损状况	反映区域水资源开发利用的潜力	反映区域人均总体用水水平	反映区域地下水开发利用状况

表2 水资源合理性配置评价指标分级标准

标准等级	缺水率/%	10年用水变化均衡率/%	用水结构系数	用水弹性系数	用水GDP/(元·m <sup>-3</sup> )	河道内生态用水比例/%	水功能区达标率/%	农业灌溉水利用系数	工业用水重复利用率/%	管网漏损率/%	水资源开发利用率/%	人均用水量/m <sup>3</sup>	地下水利用指数	期望输出值
1级	0.5	100	0.0	0.00	50	50	100	0.70	75	10	10	550	1.00	1
2级	0.8	±10	0.1	0.05	45	48	90	0.60	70	12	15	520	1.05	2
3级	1.2	±15	0.2	0.10	40	46	80	0.52	65	14	20	480	1.10	3
4级	2.0	±20	0.3	0.15	35	44	70	0.44	60	16	22	450	1.15	4
5级	5.0	±25	0.4	0.20	30	42	60	0.42	55	18	24	420	1.20	5
6级	8.0	±30	0.5	0.30	25	40	55	0.40	50	20	26	350	1.25	6
7级	12.0	±35	0.6	0.40	20	38	50	0.38	48	21	28	320	1.30	7
8级	15.0	±40	0.7	0.50	15	36	45	0.36	46	22	30	300	1.35	8
9级	18.0	±45	0.8	0.60	10	34	40	0.34	44	24	32	280	1.40	9
10级	22.0	±50	0.9	0.70	5	32	35	0.32	42	26	34	260	1.45	10
11级	25.0	±55	1.0	0.80	1	30	30	0.30	40	28	40	250	1.50	11

注:文献[1]中水资源合理性配置评价指标标准分为水资源丰水区、短缺区和平衡区。本文基于可比性原则,采用水资源平衡区评价标准。

级,每个指标分为11个等级,从1到11级别依次递减,即1级表示为水资源配置合理性最优,11级为最差,其余相互对应<sup>[1,22]</sup>,并以1~11作为标准等级1级~11级的期望输出。

### 3.2 综合评价的实现

#### 3.2.1 指标数据标准化处理

为了消除表2中各评价指标不同量纲对评价结果的影响,需对评价指标数据进行标准化处理。由于用水GDP、河道内生态用水比例、水功能区达标率、农业灌溉水利用系数、工业用水重复利用率和人均用水量6个指标值均为其值越大则合理性越优,而其他指标的取值正好相反,即相应的值越小则合理性越优,因此,需对评价指标作一致性处理,对于标准值越大其合理性越优型指标按式(8)进行归一化处理:

$$\hat{x} = 1 - (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (8)$$

对于标准值越小其合理性越优型指标按式(9)进行归一化处理:

$$\hat{x} = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (9)$$

式中: $\hat{x}$ 为经过标准化处理的数据, $x$ 为原始数据, $x_{\max}$ 和 $x_{\min}$ 分别为数据序列中的最大数和最小数。

经过标准化处理后,数据处于 $[0 \sim 1]$ 范围之内,有利于网络训练。

### 3.2.2 训练及检验样本设计

为不失一般性,采用随机内插的方法在各评价分级标准阈值间生成30个样本,随机选取20个样本作为训练样本,10个作为检验样本,以此计算共随机内插得到330个样,其中220个作为训练样本,110个作为检验样本。为使各评价指标具有相同的权重以及能够尽可能涵盖实际各评价指标范围,依据表2,规定以各评价指标上限的10倍和下限的0.1倍(极大与极小值)作为限值对各指标进行标准化处理。

### 3.2.3 SVM等模型设计

本文基于MATLAB环境,构建CV-SVM、传统BP、GA-BP及RBF水资源合理性配置综合评价模型。以表2中各评价因子作为输入向量,即输入层神经元个数为13个;以各等级对应的期望输出作为输出向量,即输出层的神经元数为1个,构建13输入1输出的评价模型。本例经反复调试,在下述参数搜索范围及设置条件下,CV-SVM、传统BP、GA-BP及RBF模型具有较好的评价性能。

CV-SVM模型:惩罚因子 $C$ 和核函数参数 $g$ 的搜索空间均设置为 $2^{-2} \sim 2^{10}$ , $K$ 取值5, $g$ 和 $C$ 的步进大小均取0.5,参考文献[17-18],利用CV法确定惩罚因子 $C$ 和核函数参数 $g$ 分别为22.6274、64(其他参数采用默认值)时CV-SVM模型性能达到最优。

传统BP模型:由于BP网络各隐层神经元的选取,目前并没有统一的计算方法,本文采用目前较为普遍的Kolmogorv定理<sup>[12-13]</sup>确定隐层单元数,最终确定BP模型结构为13-17-1,隐含层和输出层传递函数分别采用tansig和purelin,训练函数采用traingdx,学习速率 $lr$ 为0.05,设定期望误差为0.01,最大训练轮回为2000次。

GA-BP模型:基于公平原则,GA-BP模型的网络结构、隐含层和输出层传递函数、期望误差、最大训练轮回数等均与上述传统BP模型相同,在此条件下运用GA来优化BP神经网络的初始权值和阈值。GA参数设置为:种群规模为50,进化次数为100次,交叉概率为0.1,变异概率为0.05。

RBF模型:RBF神经网络人为调节的参数少,只有1个阈值,程序采取循环训练算法,最终确定RBF神经网络的SPREAD、期望误差和轮回次数分别为1.2、0.01时和1000。

### 3.2.4 模型性能评价

选用决定系数 $R^2$ ,平均绝对误差 $e_{MAE}$ 、均方根绝

对误差 $e_{RMSE}$ 、平均相对误差 $e_{MRE}$ 、均方根相对误差 $e_{RMAPE}$ 以及最大相对误差 $e_{maxRE}$ 6个统计学指标对SVM、GA-BP、RBF以及BP网络模型进行性能评价,以控制网络训练过程中的“欠拟合”和“过拟合”问题。其中, $R^2$ 范围在 $[0,1]$ 内,其值愈接近1,表明模型的性能越好;其他评价指标越小,表明模型的性能越好。

$$R^2 = \frac{\left( l \sum_{i=1}^l \hat{y}_i y_i - \sum_{i=1}^l \hat{y}_i \sum_{i=1}^l y_i \right)^2}{\left( l \sum_{i=1}^l \hat{y}_i^2 - \left( \sum_{i=1}^l \hat{y}_i \right)^2 \right) \left( \sum_{i=1}^l y_i^2 - \left( \sum_{i=1}^l y_i \right)^2 \right)} \quad (10)$$

$$e_{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (11)$$

$$e_{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (12)$$

$$e_{MRE} = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (13)$$

$$e_{RMAPE} = \frac{1}{n} \sqrt{\sum_{i=1}^n \left[ \frac{|\hat{y}_i - y_i|}{y_i} \right]^2} \quad (14)$$

$$e_{maxRE} = \max_{1 \leq i \leq n} \frac{|\hat{y}_i - y_i|}{y_i} \quad (15)$$

式中: $\hat{y}_i$ 为第 $i$ 个样本模拟值; $y_i$ 为第 $i$ 个样本实测值; $i=1,2,\dots,n$ , $n$ 为模拟样本数。

由于训练样本和检验样本是采用随机内插生成,因此模型每次运行结果均不一样,采用某次随机连续运行5次的平均值作为CV-SVM、传统BP、GA-BP及RBF评价模型的对比试验结果,以此来评价网络性能的优劣,对比试验结果见表3。

表3 CV-SVM、传统BP、GA-BP及RBF评价模型对比试验结果

评价模型	样本类型	各种模型随机5次运行平均值					
		$R^2$	$e_{MAE}$	$e_{RMSE}$	$e_{MRE}$	$e_{RMAPE}$	$e_{maxRE}$
CV-SVM	训练	0.9993	0.0640	0.0059	0.0158	0.0019	15.4799
	检验	0.9991	0.0741	0.0090	0.0211	0.0033	15.7392
传统BP	训练	0.9967	0.1405	0.0123	0.0503	0.0065	53.3819
	检验	0.9964	0.1502	0.0182	0.0538	0.0096	46.9204
GA-BP	训练	0.9991	0.0785	0.0066	0.0211	0.0023	17.5218
	检验	0.9989	0.0885	0.0104	0.0248	0.0040	20.2254
RBF	训练	0.9990	0.0789	0.0066	0.0223	0.0025	20.9526
	检验	0.9988	0.0892	0.0106	0.0275	0.0048	24.7815

从各模型的决定系数 $R^2$ 、平均绝对误差 $e_{MAE}$ 、均方根绝对误差 $e_{RMSE}$ 、平均相对误差 $e_{MRE}$ 、均方根相对误差 $e_{RMAPE}$ 以及最大相对误差 $e_{maxRE}$ 6个统计学指标来看,无论是训练样本还是检验样本,模型的评价精度及性能优劣排序依次是:CV-SVM,GA-BP,RBF,传统BP。由此可以看出,CV-SVM评价模型具

表 4 全国各省级行政区评价指标特征值

省级行政区	缺水率%	10 年用水变化均衡率/%	用水结构系数	用水弹性系数	用水 GDP/(元·m <sup>-3</sup> )	河道内生态用水比例/%	水功能区达标率/%	农业灌溉水利用系数	工业用水重复利用率/%	管网漏损率/%	水资源开发利用率/%	人均用水量/m <sup>3</sup>	地下水利用指数
北京	5.82	69.89	0.42	0.04	62.11	31.00	75.7	0.72	90	16	97.72	348.6	1.03
天津	4.52	30.76	0.77	0.03	61.79	33.53	29.4	0.6	84	18	51.86	283.52	1.11
河北	23.96	80.16	0.52	0.07	29.66	37.46	39.7	0.58	72	20	83.79	253.31	1.28
山西	3.44	65.57	0.73	0.03	22.49	36.55	45.5	0.56	83	19	62.07	224.18	0.78
内蒙古	8.31	179.81	0.76	0.35	8.09	43.01	20.3	0.38	50	18	31.97	737.97	0.39
辽宁	7.61	44.65	0.44	-0.10	34.75	46.70	54.6	0.46	67	20	39.31	322.18	0.75
吉林	2.28	80.10	0.53	0.18	14.29	48.67	36.0	0.40	63	27	31.95	462.19	0.35
黑龙江	2.96	250.00	0.90	0.63	12.98	50.00	17.0	0.43	45	22	30.82	669.75	0.31
上海	0	126.56	0.34	0.22	37.63	50.00	44.3	0.63	78	17	13.36	879.78	0.91
江苏	1.45	18.26	0.13	0.02	17.97	46.47	36.3	0.53	52	22	19.9	649.20	0.23
浙江	0.05	160.64	0.32	0.22	29.86	50.00	93.8	0.48	49	20	21.16	453.49	0.13
安徽	3.28	124.46	0.37	0.07	12.88	47.72	89.0	0.43	52	21	32.81	368.34	0.14
福建	0.34	228.00	0.28	0.02	20.22	50.00	99.6	0.40	47	26	16.42	580.12	0.18
江西	0.77	61.53	0.47	0.06	8.65	50.00	95.0	0.38	47	25	14.79	515.34	0.17
山东	9.40	73.85	0.48	-0.03	34.35	43.94	29.0	0.53	65	19	83.20	273.68	0.94
河南	10.53	173.65	0.52	0.22	23.12	43.64	68.8	0.50	63	22	55.41	233.49	0.83
湖北	0.51	139.96	0.36	0.14	13.62	50.00	78.8	0.43	55	26	30.30	502.47	0.09
湖南	0.54	95.97	0.34	0.11	10.96	50.00	83.1	0.43	60	25	19.94	492.69	0.18
广东	1.81	117.98	0.13	0.14	22.07	50.00	58.9	0.41	52	22	23.93	609.96	0.08
广西	1.00	16.03	0.24	0.01	6.84	50.00	81.2	0.36	51	19	15.83	637.04	0.05
海南	1.27	112.33	0.68	0.19	12.16	50.00	74.7	0.46	57	22	13.87	560.46	0.08
重庆	1.08	263.18	0.32	0.33	20.66	50.00	75.3	0.41	53	17	13.55	244.20	0.21
四川	2.18	153.06	0.26	0.20	14.35	50.00	88.0	0.38	58	21	10.69	311.59	0.16
贵州	2.39	67.27	0.26	0.12	11.29	50.00	86.0	0.40	55	28	8.28	241.40	0.25
云南	3.47	172.20	0.55	0.30	12.12	50.00	72.0	0.36	45	28	7.30	383.62	0.03
西藏	1.08	481.96	0.93	0.85	4.03	50.00	15.6	0.23	27	32	0.66	1152.39	0.01
陕西	6.88	46.04	0.37	0.05	18.05	42.34	63.7	0.52	57	20	21.86	253.34	0.52
甘肃	6.87	72.53	0.60	0.12	7.74	36.74	78.6	0.46	35	24	47.53	489.09	0.43
青海	4.37	136.69	0.67	0.16	6.91	50.00	87.8	0.36	38	22	6.06	740.10	0.06
宁夏	1.49	-0.67	0.77	-0.06	3.09	49.27	52.6	0.35	46	25	56.20	1557.58	0.35
新疆	4.62	19.81	0.83	0.05	2.94	48.97	88.7	0.38	43	19	55.84	2581.61	0.22

有较高的评价精度和较好的泛化能力,可以用于水资源类综合评价。本文基于模型的评价精度和泛化能力考虑,选取 CV-SVM 评价模型对全国 31 个省级行政区水资源合理性配置进行综合评价。

## 4 实例应用

### 4.1 资料来源

以全国 31 个省级行政区水资源合理性配置为例进行实例分析(评价数据来源参见文献[1],见表 4 所示)。

### 4.2 评价结果及分析

利用 CV-SVM 模型对全国 31 个省级行政区水资源合理性配置进行综合评价,并将评价等级标准中的分级临界值进行“模拟计算”,将“模拟计算”结果作为划分水资源合理性配置等级的依据,并采用“最合理”等适当语言对评价等级进行定性描述,结果见表 5~6。

分析表 5、表 6 可以得出以下结论:

表 5 水资源合理性配置分级临界值模拟结果 (随机 5 次平均)

定性评价	评价等级	临界值模拟结果	等级范围
最合理	1 级	1.7115	[0, 1.7115)
	2 级	2.4269	[1.7115, 2.4269)
	3 级	3.3795	[2.4269, 3.3795)
合理	4 级	4.3043	[3.3795, 4.3043)
	5 级	5.1969	[4.3043, 5.1969)
较合理	6 级	6.3692	[5.1969, 6.3692)
	7 级	7.3870	[6.3692, 7.3870)
不合理	8 级	8.3584	[7.3870, 8.3584)
	9 级	9.4054	[8.3584, 9.4054)
极不合理	10 级	10.5031	[9.4054, 10.5031)
	11 级	11.1489	[10.5031, 11.1489]

a. CV-SVM 模型对全国 31 个省级行政区水资源合理性配置综合评价等级为 3~8 级,处于最合理~不合理之间,在相同评价标准下能客观反映各省区水资源合理性配置状况,评价结果可为水资源规划、管理及保护提供参考。CV-SVM 模型从定性和定量两方面评价各省级行政区水资源合理性配置状

表 6 2000 年全国 31 个省级行政区水资源合理性配置综合评价结果 (随机 5 次平均)

省级行政区	CV-SVM 输出结果	评价等级	定性评价	省级行政区	CV-SVM 输出结果	评价等级	定性评价	省级行政区	CV-SVM 输出结果	评价等级	定性评价
北京	7.1490	7 级	较合理	安徽	5.2154	6 级	较合理	四川	4.6477	5 级	合理
天津	6.5216	7 级	较合理	福建	5.4454	6 级	较合理	贵州	4.4794	5 级	合理
河北	7.9199	8 级	不合理	江西	3.9283	4 级	合理	云南	5.7460	6 级	较合理
山西	7.1493	7 级	较合理	山东	7.7432	8 级	不合理	西藏	7.3811	7 级	较合理
内蒙古	7.1728	7 级	较合理	河南	8.1732	8 级	不合理	陕西	4.3610	5 级	合理
辽宁	5.0416	5 级	合理	湖北	5.2408	6 级	较合理	甘肃	6.2830	6 级	较合理
吉林	5.6250	6 级	较合理	湖南	4.0110	4 级	合理	青海	4.6658	5 级	合理
黑龙江	7.6624	8 级	不合理	广东	4.2685	4 级	合理	宁夏	5.7768	6 级	较合理
上海	3.3365	3 级	最合理	广西	3.0263	3 级	最合理	新疆	5.3761	6 级	较合理
江苏	3.3664	3 级	最合理	海南	4.1056	4 级	合理				
浙江	4.5870	5 级	合理	重庆	6.4886	7 级	较合理				

态,从评价结果可以看出,本文所研究建立的 CV-SVM 评价模型和评价方法是合理可行的。

b. 从模拟结果及评价等级上看,各省级行政区水资源配置处于最合理(即 3 级)的有广西、上海和江苏;处于合理(即 4~5 级)的有江西、湖南、广东、海南、辽宁、四川、贵州陕西和青海;处于较合理(即 6~7 级)的有吉林、安徽、福建、湖北、云南、甘肃、宁夏、新疆、北京、天津、山西、内蒙古、重庆和西藏;处于不合理(即 8 级)的有河北、黑龙江、山东和河南。从全国分布来看,水资源配置处于最合理~合理之间的地区基本都分布在经济发达地区和水资源相对丰沛地区;处于较合理~不合理之间的地区基本都分布在华北、东北及部分西北和西南地区,此类地区主要在缺水率、用水 GDP、水功能达标率及水资源开发利用效率等方面表现较差,说明该类地区的水资源配置工作在体现社会公平性、经济合理性及生态协调性等方面存在诸多问题,配置结构有待调整和提高。

## 5 结 语

从评价标准和评价方法上简要分析了目前水资源类综合评价中存在的问题。针对目前 ANN 应用于水资源类综合评价中的缺点,基于 SVM 基本原理和特点,提出一种通用的 CV-SVM 水资源类综合评价模型,并构建 GA-BP、传统 BP 及 RBF 模型作为对比模型。采用随机内插的方法在各评价分级标准阈值间生成训练样本和检验样本,选用决定系数  $R^2$  等 6 个统计学指标对 SVM、GA-BP、RBF 以及 BP 网络模型进行性能评价,以控制网络训练过程中的“欠拟合”和“过拟合”问题。通过模型对比试验,证明了 CV-SVM 模型评价精度和评价性能均优于 GA-BP、RBF 和传统 BP,表明研究建立的 CV-SVM 评价模型具有较高的评价精度和较好的泛化能力,可以用于水资源类综合评价。最后以全国 31 个省级行政区水资源合理性配置为例进行实例分析,获得了较为理想的评价效果。当然,SVM 核函数、惩

罚因子和核函数参数的合理选取是制约 SVM 进一步提高精度的关键性因素,采用遗传算法、粒子群算法,以及针对遗传算法可能存在着早熟收敛、易陷入局部极值等缺点提出的基于改进的多种群遗传算法、量子遗传算法和免疫遗传算法<sup>[11,23]</sup>对 SVM 核参数和误差惩罚因子进行优化,使 SVM 的识别精度得到更进一步的提高。

## 参考文献:

- [1] 耿雷华,卞锦华,徐澎波,等. 水资源合理配置评价指标体系研究[M]. 北京:中国环境科学出版社,2008.
- [2] 张先起,梁川. 基于熵权的模糊物元模型在水质综合评价中的应用[J]. 水利学报,2005,36(9):1057-1061. (ZHANG Xianqi, LIANG Chuan. Application of fuzzy matter-element model based on coefficients of entropy in comprehensive evaluation of water quality[J]. Journal of Hydraulic Engineering, 2005, 36(9): 1057-1061. (in Chinese))
- [3] 陈守煜,柴春岭. 区域水资源可持续利用评价的模糊可变评价方法[J]. 水利水电科技进展,2007,25(5):1-5. (CHEN Shouyu, CHAI Chunling. Fuzzy variable assessment method for sustainable utilization of regional water resources [J]. Advances in Science and Technology of Water Resources,2007,25(5):1-5. (in Chinese))
- [4] 崔东文,郭荣. 基于 GRNN 模型的区域水资源可持续利用评价:以云南文山州为例[J]. 人民长江,2012,43(5):26-31. (CUI Dongwen, GUO Rong. Evaluation of sustainable utilization of regional water resources based on GRNN neural network model: case of Wenshan Prefecture of Yunnan Province[J]. Yangtze River,2012,43(5):26-31. (in Chinese))
- [5] 崔东文. 基于 BP 神经网络的文山州水资源承载能力评价分析[J]. 长江科学院院报,2012,29(5):9-15. (CUI Dongwen. Evaluation and analysis of water resources carrying capacity in Wenshan prefecture based on BP neural network [J]. Yangtze River Scientific Research Institute,2012,29(5):9-15. (in Chinese))
- [6] 崔东文. RBF 与 GRNN 神经网络模型在河流健康评价

- 中的应用;以文山州区域中小河流健康评价为例[J]. 中国农村水利水电,2012(3):56-61. (CUI Dongwen. RBF and GRNN neural network model in the evaluation of river health[J]. China Rural Water and Hydropower,2012(3):56-61. (in Chinese))
- [7] 崔东文,郭荣. 基于概率神经网络的文山州水资源配置合理性评价分析[J]. 长江科学院院报,2012,29(10):63-67. (CUI Dongwen, GUO Rong. evaluation of rational water allocation based on probabilistic neural network; case study of Wenshan prefecture [J]. Yangtze River Scientific Research Institute, 2012, 29 (10): 63-67. (in Chinese))
- [8] 谷红梅,郭文献,徐建新,等. 区域水资源开发利用程度的灰色关联分析评价[J]. 人民黄河,2006,28(1):47-51. (GU Hongmen, GUO Wenxian, XU Jianxin, et al. Analysis and evaluation of water resources development and utilization degree of grey incidence area [J]. Yellow River,2006,28(1):47-51. (in Chinese))
- [9] 金菊良,王文圣,洪天求,等. 流域水安全智能评价方法的理论基础探讨[J]. 水利学报,2006,37(8):918-925. (JIN Juliang, WAN Wensheng, HONG Tainqiu, et al. Theoretical basis of intelligent evaluation methods of watershed water security [J]. Journal of Hydraulic Engineering,2006,37(8):918-925. (in Chinese))
- [10] 施彦,韩力群,廉小亲. 神经网络设计方法与实例分析[M]. 北京:北京邮电大学出版社,2009.
- [11] 史峰,王辉,郁磊,等. MATLAB 智能算法 30 个案例分析[M]. 北京:北京航空航天大学出版社,2011.
- [12] 张德丰. MATLAB 神经网络应用设计[M]. 北京:机械工业出版社,2009.
- [13] 张良均,曹晶,蒋世忠. 神经网络实用教程[M]. 北京:机械工业出版社,2008.
- [14] 傅荟璇,赵红. MATLAB 神经网络应用设计[M]. 北京:机械工业出版社,2009. 10.
- [15] VIADIMIR N V. 统计学习理论的本质[M]. 张学工,译. 北京:清华大学出版社,2000.
- [16] 田景文,高美娟. 人工神经网络算法研究及应用[M]. 北京:北京理工大学出版社,2006.
- [17] 崔东文. 支持向量机在湖库营养状态识别中的应用研究[J]. 水资源保护,2013,29(4):26-30. (CUI Dongwen. Application of support vector machine to lake and reservoir trophic status recognition [J]. Water Resources protection,2013,29(4):26-30. (in Chinese))
- [18] MATLAB 中文论坛. MATLAB 神经网络 30 个案例分析[M]. 北京:北京航空航天大学出版社,2010.
- [19] 白鹏,张喜斌,张斌,等. 支持向量机理论及工程应用实例[M]. 西安:西安电子科技大学出版社,2008.
- [20] 王顺久,张欣莉,倪长键,等. 水资源优化配置原理及方法[M]. 北京:中国水利水电出版社,2007.
- [21] 杜守建,崔振才. 区域水资源优化配置与利用[M]. 郑州:黄河水利出版社,2009.
- [22] 崔东文. 基于极限学习机的长江流域水资源开发利用综合评价[J]. 水利水电科技进展,2013,33(2):14-19. (CUI Dongwen. Comprehensive evaluation of water resources development and utilization in Yangtze River Basin based on extreme learning machine [J]. Advances in Science and Technology of Water Resources,2013,33(2):14-19. (in Chinese))
- [22] 雷德明,严新平. 多目标智能优化算法及其应用[M]. 北京:科学出版社,2009.

(收稿日期:2013-01-30 编辑:高渭文)

## 欢迎订阅《水资源保护》杂志

中国科技核心期刊 RCCSE 核心期刊

ISSN 1004-6933 CN 32-1356/TV

《水资源保护》是由河海大学和中国水利学会环境水利专业委员会主办的科技期刊。该杂志针对我国水资源短缺、用水效率不高、水污染严重等突出问题,探讨水资源保护工作中的基础研究、防治技术、宏观管理及水环境治理问题,重点关注水环境、水资源和大江大湖的环境生态问题和可持续发展。主要栏目有科学研究、应用技术、综述述评、管理研究等。

《水资源保护》是中国学术期刊综合评价数据库来源期刊、中国核心期刊(遴选)数据库统计源期刊,已被美国化学文摘(CA)数据库、美国《剑桥科学文摘》(CSA)数据库、波兰哥白尼索引(IC)数据库、中国期刊网、中国数字化期刊群、水信息网、北极星网、中华期刊网、中文科技期刊数据库(VIP)、中国科技论文与引文数据库(CSTPCD)等收录和引用,长期以来一直都是水利界和环保部门备受关注的重点期刊,2012年被教育部科技司授予“中国高校特色科技期刊”称号。

《水资源保护》主要读者对象是全国从事与水资源保护工作有关的工程技术人员、科研人员、管理人员以及大专院校师生,邮发代号:28-298,双月刊,96页,12元/期,全年共计72元,每逢单月20日出版。欲订购者,请向当地邮局订购。若无法从邮局订阅,亦可登陆本刊网站下载征订单。

编辑部地址:南京市西康路1号 河海大学《水资源保护》编辑部

邮政编码:210098 电话/传真:(025)83786642 E-mail:bh@hhu.edu.cn

http://kkb.hhu.edu.cn/web/indexbh.asp?\_d\_id=37