基于多元变量组合的回归支持向量机集成模型及其应用

崔 东 文
（云南省文山州水务局，云南 文山 663000）

摘要：为进一步提高径流预测的精度和泛化能力，提出基于多元变量组合的回归支持向量机（SVR）集成年径流预测模型，以云南省龙潭站年均径流预测为例进行实例研究。首先，以实例1—10月月均流量作为预测因子，采用相关分析法确定预测因子与年均径流量的相关系数，按照相关系数大小顺序依次选取预测因子，构建2维输入变量～10维输入变量的9种SVR模型对实例后12年的年均径流量进行预测。最后，采用简单平均（SA）和加权平均（WA）两种集成方法对具有较高预测精度的7种SVR模型的预测结果进行综合集成。结果表明：①SVR模型的预测精度随着输入变量维数的增加明显提高。②SA-SVR和WA-SVR模型对实例后12年年均径流量预测的平均相对误差绝对值分别为1.73%和1.79%，最大相对误差绝对值分别为6.34%和6.47%，精度和泛化能力均优于各SVR模型。相对而言，由于采用多个SVR模型进行集成，SA-SVR模型预测效果略优于WA-SVR模型。

关 键 词：径流预测；集成模型；回归支持向量机（SVR）；简单平均法；加权平均法
中图分类号：P338 **文献标志码**：A **文章编号**：1009-640X(2014)02-0066-08

提高径流预测精度对于水文预测预报具有重要意义。由于河川径流受多种因素的影响和制约，其预测常反映出复杂、随机、多维等特点，目前径流预测主要是将成因分析、统计分析、模糊分析以及灰色系统理论等方法及理论引入径流中长期预测，在实际应用中取得了一定成效，但也存在不足，使其在应用中受到制约。譬如，难以分析其内部物理机制（影响因子相互关系）以及与所表观的水文现象（径流）之间的关系等；径流影响因子个数的选择以及预测精度不高等，带有明显的主观性，较适用于具有确定性趋势的预测问题；对于其他变化趋势，则拟合灰度较大，导致精度难以提高等[1]。人工神经网络（Artifical Neural Network，ANN）是一种模仿动物神经网络行为特征，进行分布式并行信息处理的数学模型。ANN依靠系统的复杂程度，通过调整内部大量节点之间相互连接的关系，从而达到处理信息的目的，其常见的BP网络（Back-Propagation Network，BP）、Elman网络、RBF网络（Radial Basis Function Neural Network，RBF）和GRNN网络（Generalized Regression Neural Network，GRNN）等均广泛运用于径流预测预报中[2-6]。然而，由于传统ANN算法是基于渐近理论，仅在样本容量趋向于无穷大时其经验风险才趋近于实际风险，但在实际应用中样本容量离无穷大相去甚远，导致ANN外推能力差、收敛速度慢以及存在局部极值等问题，且ANN还存在结构及参数选择的困难[7-8]。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，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等问题[9-10]。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，其具有的四大优势决定了它在机器学习领域有着举足轻重的地位：一是SVM以最小结构风险代替传统ANN经验风险，求

收稿日期：2013-06-15
作者简介：崔东文（1978-），男，云南玉溪人，高级工程师，主要从事水资源水环境研究及水资源保护等工作。
E-mail：cdwgr@163.com

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 ANN 算法中难于克服的局部极值缺陷；二是 SVM 拓扑结构由支持向量决定，弥补了 ANN 结构难以确定的不足；三是 SVM 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；四是 SVM 方法求解的是基于结构风险最小化原则，由经验风险和置信区间共同决定支持向量的实际风险，因此泛化能力要优于传统 ANN。尤其在解决小样本容量时，很大程度上解决了传统 BP 等网络在模型选择、过学习、高维和局部极值等方面的问题，在模式识别和回归预测中有着广泛的应用[11-12]。神经网络集成(Neural Network Ensemble，NNE)是 Hansen 和 Salamon 于 1990 年提出的集成方法，利用有限个神经网络对同一问题进行学习，将个体神经网络的输出以某种综合方法进行集成，集成输出的结果由各个体神经网络的输出共同决定[7,13]。NNE 的实现主要由“选择模型中的个体网络”和“集成输出方法”两部分组成，目前仍没有较好方法或理论指导如何获取具有较好性能的个体网络，常用的集成方法主要有简单平均(Simple Average，SA)和加权平均(Weighted Average，WA)等。

基于上述原因及回归支持向量机(Support Vector Regression，SVR)特性，笔者以云南省盘龙河龙潭站年径流预测为例，采用相关分析法确定预测因子与年径流的相关系数，按照相关系数大小顺序依次选取预测因子，构建 2 维输入变量～10 维输入变量的 9 种 SVR(SVR-2～SVR-10)年径流预测模型对龙潭站后 12 年的年径流进行预测，并基于 NNE 方法，采用 SA 和 WA 两种集成方法对具有较好预测精度的若干 SVR 模型的预测结果进行综合集成，集成结果表明 SA-SVR 和 WA-SVR 模型具有较好的预测精度和泛化能力。

1　回归支持向量机(SVR)及集成预测模型

1.1　回归支持向量机(SVR)

将 SVM 用于逼近函数的方法称为 SVR。SVR 是基于 VC 维概念和结构风险最小化原则，根据有限的样本信息在模型的复杂性(训练样本学习精度)和泛化能力之间寻求一种折中，以期获得最好的推广能力。在解决非线性回归问题时，SVR 是将样本通过核函数将低维空间中非线性回归问题映射到高维空间，并在高维空间中求解最优回归函数，这样，在高维空间中的线性回归就对应低维空间中的非线性回归，即实现某一非线性变换后的线性回归，在计算难度未增加的情况下利用线性空间的方法解决非线性问题。SVR 实现回归预测步骤归纳如下[8,11-12,14-15]。

步骤 1：给定含有 l 个训练样本的集合 $\{(x_i,y_i),i=1,2,\cdots,l\}$，其中，$x_i(x_i\in R^d)$ 是第 i 个训练样本的输入列向量，$x_i=[x_i^1,x_i^2,\cdots,x_i^d]^{\mathrm{T}}$，$y_i\in R$ 为对应输出值。在高维特征中建立的线性回归函数为：

$$f(x)=w\Phi(x)+b \tag{1}$$

式中：$\Phi(x)$ 为非线性映射函数。

步骤 2：定义 ε 线性不敏感损失函数为

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

式中：$f(x)$ 为回归函数返回的预测值；y 为对应的真实值。

步骤 3：类似于 SVM 分类情况，引入松驰变量 ξ_i,ξ_i^*，并将上述寻找 w,b 的问题用数学语言描述出来，即：

$$\begin{cases}\min\dfrac{1}{2}\|w\|^2+C\sum\limits_{i=1}^{l}(\xi_i+\xi_i^*)\\ \text{s.t.}\begin{cases}y_i-w\Phi(x_i)-b\leqslant\varepsilon+\xi_i\\ -y_i+w\Phi(x_i)+b\leqslant\varepsilon+\xi_i^*,\quad i=1,2,\cdots,l\\ \xi_i\geqslant0,\quad\xi_i^*\geqslant0\end{cases}\end{cases} \tag{3}$$

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

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

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

式中:$K(x_i,x_j)=\Phi(x_i)\Phi(x_j)$ 为核函数。

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

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

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

式中:N_{nsv}为支持向量机个数。

步骤 5:将 w^*,b^* 代入式(1)得到回归函数为:

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

其中,只要部分参数 $(a_i-a_i^*)$ 不为 0,其对应的样本 x_i 即为问题中的支持向量。

常用的核函数主要类型有线性核函数($K(x,x_i)-x^{\mathrm{T}}x_i$)、多项式核函数($K(x,x_i)=(\gamma x^{\mathrm{T}}x_i+r)^p,\gamma>0$)、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($K(x,x_i)=\tanh(\gamma x^{\mathrm{T}}x_i+r)$) 。

1.2 集成预测模型

基于上述 SVR 原理,SA-SVR 和 WA-SVR 模型年径流预测步骤可归纳如下:

步骤 1:计算年径流影响因子的相关系数。利用 SPSS 软件计算龙潭站 1—10 月月均流量与年径流相关系数,并按照相关系数大小进行排列。

步骤 2:构建各 SVR 年径流预测模型。基于 Matlab 软件环境和 libsvm 工具箱,选取相关系数最大的前 2 列月均流量作为 SVR 模型的输入变量,构建 2 维的 SVR 年径流预测模型,表示为 SVR-2;选取相关系数最大的前 3 列月均流量作为 SVR 模型的输入变量,构建 3 维的 SVR 年径流预测模型,表示为 SVR-3;依次类推,构建 10 维的 SVR 年径流预测模型,表示为 SVR-10。

步骤 3:训练及调试 SVR-2 ~ SVR-10 预测模型。确定训练样本及检验样本,并对样本进行归一化处理,利用训练样本对 SVR-2 ~ SVR-10 模型进行训练及调试,率定各模型的相关参数,并利用检验样本对 SVR-2 ~ SVR-10 模型进行预测精度及泛化能力检验。

步骤 4:SA-SVR 集成。利用平均相对误差和最大相对误差对 SVR-2 ~ SVR-10 模型的预测精度及泛化能力进行分析评价,并选取若干较优模型的预测结果按下式进行 SA 集成:

$$\bar{x}=\frac{1}{k}\sum_{i=1}^{k}\bar{x}_i \tag{8}$$

式中:$\bar{x}_i(i=1,2,\cdots,k)$ 为第 i 个模型的预测值。

步骤 5:WA-SVR 集成。WA 集成是按照 SVR 模型的预测效果优劣给出不同的权重,然后求加权平均值,以加权平均值作为集成模型的预测值。同步骤 4,选取若干最优模型的预测结果按下式进行 WA 集成:

$$\bar{x}=\sum_{i=1}^{k}w_i\bar{x}_i \tag{9}$$

式中:w_i 为第 i 个模型的权重,$\sum_{i=1}^{k}w_i=1$,$w_i\geqslant 0$。根据 SVR 模型预测值的平均相对误差确定权重 w_i,并给予预测相对误差绝对值较小的模型更大的权重,计算公式如下:

$$w_i = \frac{1/|e_i|}{\sum_{i=1}^{k} 1/|e_i|} \tag{10}$$

式中：e_i 为第 i 个模型预测相对误差的绝对值。

步骤 6：对 SA-SVR 和 WA-SVR 模型预测结果进行分析。若 SA-SVR 和 WA-SVR 模型预测值达不到期望的精度和泛化能力要求，则返回步骤 3 对各 SVR 个体模型进行调试和检验，直至 SA-SVR 和 WA-SVR 模型的预测值满足预期的精度要求。

2　实例应用

2.1　数据来源与分析

龙潭寨水文站位于盘龙河上游，建于 1951 年 4 月，属国家基本站，观测项目有水位、流量、降水、蒸发、泥沙和水温，系列均为 1951 年 4 月至 2005 年 12 月。控制流域面积 3 128 km^2，流域多年平均流量 24.6 m^3/s，降水量 956.8 mm。本文以龙潭站 1952—2005 年共 54 年的实测资料为例进行分析。利用 SPSS 软件分析年均径流与 1—10 月月均流量的相关性，可得年均径流与 1—10 月月均流量的相关系数分别为 0.455**，0.441**，0.333*，0.486**，0.497**，0.519**，0.616**，0.822**，0.782** 和 0.798**（“**”表示在 0.01 水平（双侧）上显著相关；“*”表示在 0.05 水平（双侧）上显著相关）。

可见，年均径流与 1—10 月月均流量均呈显著正相关。本文选取与年均径流在 0.05 水平（双侧）上显著相关的 1—10 月月均流量作为影响因子预测年均径流，并以 1952—1993 年 42 年的实测资料作为训练样本，1994—2005 年 12 年的实测资料作为检验样本。

2.2　年径流预测的实现

2.2.1　*数据归一化处理*　利用下式对径流序列进行归一化处理：

$$\hat{x} = (x - x_{\min})/(x_{\max} - x_{\min}) \tag{11}$$

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 0～1 范围之内，有利于网络训练。

2.2.2　*SVR 模型的构建及网络训练*　按照上述方法从 1—10 月月均流量中按相关系数大小顺序选取不同维数的输入向量，构建 SVR-2～SVR-10 年径流预测模型，并基于 Matlab 环境和 libsvm 工具箱，创建 SVR-2～SVR-10 模型对龙潭站 54 年的实测资料进行拟合和预测分析。

由于在核函数选定的条件下，SVR 模型中的惩罚因子 c 和不敏感系数 ε 对模型的预测精度有着关键性影响。惩罚因子 c 决定着由训练样本产生的经验风险对模型性能的影响，即经验风险随着 c 值的增加而增加，减小而减小，当 c 值无穷大时，SVR 结构风险趋于经验风险；当 c 值趋于零时，由于 SVR 模型无法获取训练样本信息，模型失去了解决具体问题的能力。不敏感系数 ε 用于控制支持向量的个数，平衡模型的复杂程度与模型对训练样本维数的依赖程度。在实际应用中，若 ε 值过小，可能导致模型“过拟合”，并且增加训练时间；值过大，则可能导致模型“欠拟合”。参考多个文献[8,11-12,14-15]后选择径向基函数为 SVR 的核函数，设置惩罚因子 c 和核函数参数 g 的搜索空间为 2^{-2}～2^{6}，K 取值 2～5，g 和 c 的步进大小均取 0.1～0.5，不敏感系数 ε 均取 0.001～0.1（其他参数采用默认值），利用交叉验证法（Cross Validation，CV）确定模型中的惩罚因子 c 和核函数参数 g。经过反复调试，在下述参数设置情况下，SVR-2～SVR-10 模型具有较好的预测效果（除此之外的其他参数采用系统默认值）。

SVR-2～SVR-10 模型的最佳参数设置见表 1，拟合及预测效果见表 2 及图 1。由表 2 及图 1 可以看出：①SVR 模型的预测精度和泛化能力随着输入向量维数的增加具有明显的提高趋势，其中 SVR-8～SVR-10 模型对龙潭站后 12 年年径流预测的平均相对误差绝对值均小于 3%，最大相对误差绝对值均在 10% 以内，具有较好的预测精度和泛化能力；②由于 SVR-2、SVR-3 模型输入向量维数较低，训练样本中信息量不足，导致其拟合及预测效果不够理想。

表 1 SVR-2 ~ SVR-10 模型的最佳相关参数

Tab. 1 Optimal parameters of SVR-2 ~ SVR-10 models

模 型	SVR-2	SVR-3	SVR-4	SVR-5	SVR-6
输入向量	8 月/10 月	8—10 月	7—10 月	6—10 月	5—10 月
惩罚因子 c	0.933 0	0.307 8	0.329 9	16	1.414 2
核函数参数 g	64	4	2.143 6	0.353 6	0.353 6
不敏感系数 ε	0.1	0.1	0.01	0.01	0.01
模 型	SVR-7	SVR-8	SVR-9	SVR-10	
输入向量	4—10 月	1 月/4—10 月	2—10 月	1—10 月	
惩罚因子 c	1.4142	8	4	4	
核函数参数 g	0.25	0.25	0.25	0.25	
不敏感系数 ε	0.001	0.001	0.001	0.01	

表 2 SVR-2 ~ SVR-10 模型对龙潭站年径流拟合及预测相对误差

Tab. 2 Annual runoff fittings of Longtan station given by SVR-2 ~ SVR-10 models and their relative forecast errors %

	模 型	SVR-2	SVR-3	SVR-4	SVR-5	SVR-6
训练样本	平均相对误差绝对值	10.50	8.11	4.71	2.24	2.81
	最大相对误差绝对值	25.83	25.71	18.50	9.53	9.98
检验样本	平均相对误差绝对值	5.42	5.31	4.16	3.22	3.36
	最大相对误差绝对值	12.01	15.72	8.57	10.16	9.55
	模 型	SVR-7	SVR-8	SVR-9	SVR-10	
训练样本	平均相对误差绝对值	2.04	0.82	1.04	1.64	
	最大相对误差绝对值	8.10	8.57	8.12	9.26	
检验样本	平均相对误差绝对值	3.37	2.77	2.70	2.68	
	最大相对误差绝对值	7.05	7.27	7.02	7.89	

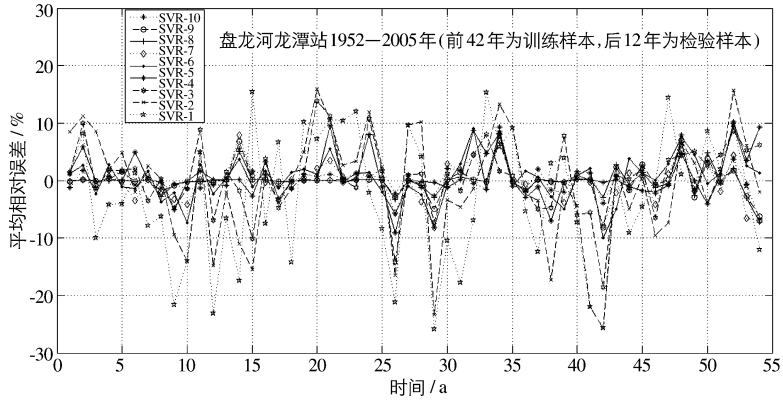


图 1 SVR-2 ~ SVR-10 模型对龙潭站年径流拟合及预测相对误差

Fig. 1 Annual runoff fittings of Longtan station given by SVR-2 ~ SVR-10 models and their relative errors of forecast values

本文基于预测精度及泛化能力考虑,选取具有较好预测效果的 SVR-4 ~ SVR-10 模型作为 SA-SVR 和 WA-SVR 模型集成的个体模型。

2.3 WA-SVR 模型权重的确定及预测结果分析

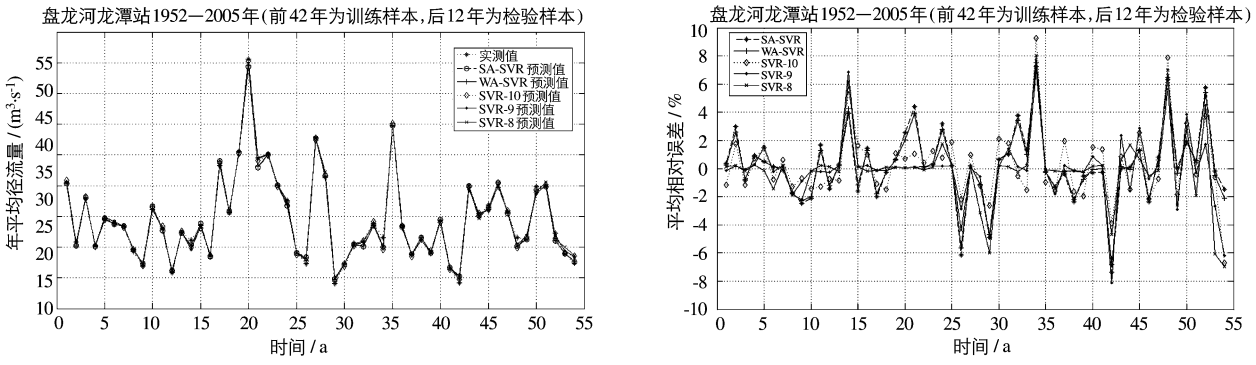
利用 SVR-4 ~ SVR-10 模型对龙潭站后 12 年年径流预测的平均相对误差绝对值确定各自权重,结果分别为 0.1068,0.1380,0.1323,0.1319,0.1605,0.1646 和 0.1659。

利用 SA 和 WA 两种集成方法将上述具有较好预测效果的 SVR-4 ~ SVR-10 模型进行综合集成,构建 SA-SVR 和 WA-SVR 模型对龙潭站后 12 年年径流进行预测,预测结果与 SVR-8 ~ SVR-10 个体模型进行比较,结果见表 3 及图 2。

表 3　龙潭站 1994—2005 年 SA-SVR 和 WA-SVR 模型径流预测结果及比较

Tab. 3　Runoff forecast results given by SA-SVR and WA-SVR models from 1994 to 2005 and their comparison ($m^3 \cdot s^{-1}$)

年　份	实测值	SA-SVR		WA-SVR		SVR10		SVR9		SVR8	
		预测值	相对误差/ %	预测值	相对误差/ %	预测值	相对误差/ %	预测值	相对误差/ %	预测值	相对误差/ %
1994	30.0	30.01	−0.01	29.96	0.15	29.74	0.86	29.29	2.37	29.80	2.74
1995	25.1	25.08	0.08	25.12	−0.06	25.47	1.49	25.50	1.61	24.68	0.76
1996	26.8	26.46	1.28	26.44	1.33	26.12	2.53	26.04	2.85	26.61	1.38
1997	29.8	30.50	−2.37	30.45	−2.16	30.45	2.17	30.02	0.72	29.96	1.19
1998	25.7	25.49	0.81	25.54	0.62	25.89	0.75	25.65	0.21	25.73	0.56
1999	21.6	20.23	6.34	20.20	6.47	19.90	7.89	20.08	7.02	20.37	6.71
2000	21.3	21.29	0.04	21.37	−0.32	21.69	1.82	21.92	2.89	21.85	2.54
2001	29.9	29.35	1.84	29.31	1.98	28.93	3.23	28.74	3.87	29.01	4.81
2002	30.0	29.82	0.59	29.88	0.41	30.13	0.43	30.09	0.30	30.57	0.08
2003	22.3	21.02	5.76	21.09	5.41	21.46	3.75	21.92	1.72	21.29	2.12
2004	18.9	18.94	−0.19	18.99	−0.48	19.01	0.61	19.40	2.65	20.05	3.03
2005	17.4	17.66	−1.49	17.77	−2.11	18.57	6.70	18.47	6.18	18.61	7.27
平均相对误差绝对值/ %		1.73		1.79		2.68		2.70		2.77	
最大相对误差绝对值/ %		6.34		6.47		7.89		7.02		7.27	



(a)　年径流拟合及预测值　　(b)　年径流拟合及预测比较

图 2　集成模型与 SVR−8 ~ SVR−10 模型年径流拟合及预测比较

Fig. 2　Comparison between annual runoff fitting and forecast values given by integrated model and SVR−8 ~ SVR−10 models

从表 3 及图 2 可见：①SA-SVR 和 WA-SVR 模型预测龙潭站后 12 年年径流的平均相对误差绝对值分别为 1.73% 和 1.79%，最大相对误差绝对值分别为 6.34% 和 6.47%，预测精度和泛化能力均优于各 SVR 个体模型，表明本文提出的集成模型和集成方法用于径流预测是合理可行的。集成模型具有预测精度高、泛化能力强以及稳健性能好等优点。②普遍认为，加权平均集成模型通过给予预测误差较小的个体模型更大的权重能够得到比简单平均集成模型更好的预测精度和泛化能力，但从本实例预测结果来看，由于采用多个(7 个)个体模型进行集成，SA-SVR 模型的预测效果要优于 WA-SVR 模型。③在系列长达 54 年的年径流拟合及预测中，相对误差均在−10% ~10% 之间，表明 SA−SVR、WA−SVR 以及 SVR−8 ~ SR−10 模型均具有较好的拟合及预测效果，其中 SA-SVR 和 WA-SVR 模型相对误差均在−8% ~8% 之间，具有更高的预测精度、泛化能力及稳健性能。

3　结　语

基于 SVR 原理和 NNE 基本思想，利用多元变量择优组合的方式构建具有不同维数的 SVR 年径流预测

模型,采用 SA 和 WA 两种集成方法构建 SA-SVR 和 WA-SVR 模型对龙潭站后 12 年的年径流进行预测。预测结果表明,SA-SVR 和 WA-SVR 模型具有较高的预测精度和泛化能力。本文在以下两方面可为相关预测研究提供参考:一是仅基于 SVR 原理及算法,提出利用不同的输入维数构建由多个 SVR 个体网络集成的年径流预测模型和方法;二是在实际应用中,决定模型预测精度和泛化能力的关键因素是问题本身的复杂程度,对于不同的预测问题,很难片面地认为某一模型或算法的优劣,只有尝试不同模型或算法,反复进行调试,以期获得理想的预测效果。

参 考 文 献:

[1] 杨旭, 栾继虹, 冯国章. 中长期水文预报研究评述与展望[J]. 西北农业大学学报, 2000, 28(6): 203-207. (YANG Xu, LUAN Ji-hong, FENG Guo-zhang. Discussion and prospect on mid-to-long-term hydrological forecasting[J]. Acta Univ. Agric. Boreali-occidentalis, 2000, 28(6): 203-207. (in Chinese))

[2] 崔东文. 多隐层 BP 神经网络模型在径流预测中的应用[J]. 水文, 2013, 33(1): 68-73. (CUI Dong-wen. Multi-layers BP neural network model in runoff prediction[J]. Journal of China Hydrology, 2013, 33(1): 68-73. (in Chinese))

[3] 邓霞, 董晓华, 薄会娟. 基于 BP 网络的河道径流预报方法与应用[J]. 人民长江, 2010, 41(2): 56-59. (DENG Xia, DONG Xiao-hua, BO Hui-juan. River runoff forecasting methods based on BP network and its applications[J]. Yangtze River, 2010, 41(2): 56-59. (in Chinese))

[4] 杨新华, 马建立, 苏军希, 等. 基于 Elman 网络的黄河源区径流评估模型[J]. 人民长江, 2007, 38(8): 134-135, 174. (YANG Xin-hua, MA Jian-li, SU Jun-xi, et al. Runoff assessment model for the source area of the Yellow River based on Elman neural network[J]. Yangtze River, 2007, 38(8): 134-135, 174. (in Chinese))

[5] 刘荻, 周振民. RBF 神经网络在径流预报中的应用[J]. 华北水利水电学院学报, 2007, 28(2): 12-14. (LIU Di, ZHOU Zhen-min. Application of RBF neural network in runoff prediction[J]. Journal of North China Institute of Water Conservancy and Hydroelectric Power, 2007, 28(2): 12-14. (in Chinese))

[6] 陈仁升, 康尔泗, 张济世. 应用 GRNN 神经网络模型计算西北干旱区内陆河流域出山径流[J]. 水科学进展, 2002, 13(1): 333-338. (CHEN Ren-sheng, KANG Er-si, ZHANG Ji-shi. Application of the generalized regression neural network to simulating runoff from the mountainous watersheds of inland river basins in the arid area of northwest China[J]. Advances in Water Science, 2002, 13(1): 333-338. (in Chinese))

[7] 田雨波. 混合神经网络技术[M]. 北京: 科学出版社, 2009. (TIAN Yu-bo. Hybrid neural network technology[M]. Beijing: Science Press, 2009. (in Chinese))

[8] 王雷. 支持向量机在汽轮机状态监测中的应用[M]. 北京: 北京师范大学出版社, 2012. (WANG-Lei. Applications of support vector machine on turbine state monitoring[M]. Beijing: Beijing Normal University Press, 2012. (in Chinese))

[9] 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. The nature of statistical learning theory[M]. Translated by ZHANG Xue-gong. Beijing: Tsinghua University Press, 2000. (in Chinese))

[10] 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. (TIAN Jing-wen, GAO Mei-juan. Research and application of artificial neural network algorithm[M]. Beijing: Beijing Institute of Technology Press, 2006. (in Chinese))

[11] 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 Dong-wen. Application of support vector machine to lake and reservoir trophic status recognition[J]. Water Resources Protection, 2013, 29(4): 26-30. (in Chinese))

[12] 崔东文. 支持向量机在水资源类综合评价中的应用研究—以全国 31 个省级行政区水资源合理性配置为例[J]. 水资源保护, 2013, 29(5): 61-68. (CUI Dong-wen. Application of support vector machines to comprehensive assessment of water resources class[J]. Water Resources Protection, 2013, 29(5): 61-68. (in Chinese))

[13] 张大斌, 张景广, 彭森. 基因表达式编程在组合预测建模中的应用[J]. 系统工程理论与实践, 2012, 32(3): 568-573. (ZHANG Da-bin, ZHANG Jing-guang, PENG Sen. Application of gene expression programming on combination forecasting modeling[J]. Systems Engineering-Theory & Practice, 2012, 32(3): 568-573. (in Chinese))

[14] 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.（SHI Feng，WANG Hui，YU Lei，et al. 30 cases study on MATLAB intelligent algorithm[M]. Beijing：Beihang University Press，2011.（in Chinese））

[15] 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.（MATLAB Chinese forum. MATLAB neural network analysis of 30 cases[M]. Beijing：Beihang University Press，2010.（in Chinese））

A regression support vector machine integrated model based on multivariate combinations and its application

CUI Dong-wen

（*Wenshan Water Bureau*，*Yunnan Province*，*Wenshan*　663000，*China*）

Abstract：In order to improve the accuracy of runoff forecast and generalization ability，a regression support vector machine（SVR）integrated annual runoff forecasting model is developed based on multivariate combinations，and the annual runoff forecasting of Longtan hydrologic station in Yunnan Province is taken as an example for the case studies. First，the average monthly discharge from January to October is taken as predictor，and correlation coefficients of the predictior and the average annual runoff are determined by a correlation analysis method. Then，the predictor is sequentially selected，according to the correlation coefficients from the maximum to the minimum，to develop nine SVR models for 2-D to 10-D input variables，and annual runoffs of next 12 years are forecasted by the models. At last，the simple average（SA）and the weighted average（WA）methods are applied to forecasting comprehensive integration for seven SVR models with high accuracy. The analysis research results show that：①the prediction accuracy of SVR model improves significantly with the increase of imput variables dimension；②for annual runoff predictions of next 12 years，the absolute average relative errors of SA-SVR and WA-SVR models are 1.73% and 1.79%，and the absolute maximum relative errors are 6.34% and 6.47%，which indicates that the accuracy and generalization ability of the SA-SVR and WA-SVR models are better than that of other SVR models，therefore SA-SVR model is better than WA-SVR model slightly.

Key words：runoff forecasting；integrated model；regression support vector machine（SVR）；simple average method；weighted average method