

# 基于模拟退火算法支持向量机在枯水期 月径流预测中的应用

魏 胜

(云南省水文水资源局文山分局, 云南 文山 663000)

**摘 要:** 鉴于支持向量机(SVM)最佳算法参数难以确定的不足,利用模拟退火算法(SA)搜索SVM学习参数,提出SA-SVM预测模型,并与基于遗传算法(GA)搜索SVM学习参数的GA-SVM模型作对比,以云南省龙潭站枯水期1-3月月径流预测为例进行实例研究,利用实例前43年和后10年资料对模型进行训练和预测。结果表明:SA-SVM模型对实例后10年枯水期1-3月月均径流预测的平均相对误差绝对值分别为3.11%、4.93%和6.75%,精度优于GA-SVM模型,表明SA-SVM模型具有较高的预测精度和泛化能力。SA算法通过赋予搜索过程一种时变且最终趋于零的概率突跳性,有效避免了算法陷入局部极值并最终趋于全局最优。

**关键词:** 径流预测; 模拟退火算法; 支持向量机; 参数优化; 枯水期

中图分类号:P333.3

文献标识码:A

文章编号:1672-643X(2015)02-0135-04

## Application of support vector machine to prediction monthly runoff in dry season based on simulation annealing algorithm

WEI Sheng

(Wenshan Branch Bureau, Yunnan Province Hydrology Water Resources Bureau, Wenshan 663000, China)

**Abstract:** Aimed at the shortage that the best algorithm parameters of support vector machine (SVM) deficiency is difficult to determine, the paper used the simulation annealing algorithm (SA) to search SVM learning parameters and put forward SA-SVM prediction model. Based on comparison between genetic algorithm (GA) and GA-SVM model in searching SVM learning parameters, it took the runoff prediction in dry season of January to march at LongTan station in Yunnan Province as an example. It trained and forecasted the model by using data of the example before 43 years and after 10 years. The results show that the absolute values of average relative errors of prediction runoff in 3 months of dry season by SA-SVM model after 10 years of instance were 3.11%, 4.93% and 6.75%, the accuracy is better than that by GA-SVM model, which showed that SA-SVM model has higher prediction accuracy and generalization ability. The sudden jump time-varying and final trend to zero probability by SA algorithm through giving the search process can effectively avoid the algorithm falls into local extreme and tends to the global optimal.

**Key words:** runoff forecast; simulated annealing algorithm; support vector machine; parameter optimization; dry season

## 1 研究背景

枯水期径流预报模型及方法研究是水文研究工作中的热点和难点,提高枯水期径流预测预报精度对于合理配置水资源、制定水量分配方案以及实行最严格水资源管理制度具有重要意义。由于枯水期

径流受多种因素的影响和制约,表现出复杂、随机、多维等特性,传统回归分析、数理统计等方法往往难以达到理想的预测效果。支持向量机(Support Vector Machine, SVM)是继人工神经网络(Artificial Neural Network, ANN)之后机器学习领域新的研究热点,用于解决模式分类和非线性映射等问题。SVM

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,具有“理论全局最优”、“结构简单”、“克服维数灾问题”和“泛化能力好”等优点。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、高维和局部极值等问题<sup>[1-3]</sup>。

在实际应用中,SVM 惩罚因子  $C$ 、核函数参数  $g$  和不敏感系数  $\varepsilon$  3 个学习参数的选取对模型的预测精度有着关键性影响<sup>[4]</sup>,传统试凑等方法选取 SVM 学习参数不但繁琐、低效和耗时,而且难以获得理想的预测效果。模拟退火算法 (Simulated Annealing, SA) 是 Metropolis 等人基于 Monte - Carlo 迭代求解策略的一种随机寻优算法,其思想是基于物理中固体物质的退火过程与组合优化问题之间的相似性,通过赋予搜索过程一种时变且最终趋于零的概率突跳性,可有效避免陷入局部极值并最终趋于全局最优<sup>[5-6]</sup>。

本文鉴于 SVM 最佳学习参数难以确定的不足,利用 SA 算法搜索 SVM 最佳学习参数,提出 SA - SVM 预测模型,并构建 GA - SVM 模型作对比,以云南省龙潭站枯水期 1 - 3 月月径流预测为例进行实例验证。验证结果表明 SA - SVM 模型具有较好的预测精度和泛化能力。

## 2 SA - SVM 预测模型

### 2.1 模拟退火算法

模拟退火算法 (SA) 来源于固体退火原理,将固体加温至充分高,再让其冷却,加温时,固体内部粒子随温升变为无序状,内能增大,而冷却时粒子渐趋有序,在每个温度都达到平衡态,最后在常温时达到基态,内能减为最小。根据 Metropolis 准则,将固体退火模拟用于组合优化问题时,其内能模拟为目标函数,温度演化成控制参数,即得到解组合优化问题的模拟退火算法。算法通过赋予搜索过程一种时变且最终趋于零的概率突跳性,从而可有效避免陷入局部极值并最终趋于全局最优,在解决旅行商问题、背包问题、排序问题等典型优化问题中得到应用<sup>[5-7]</sup>。

SA 算法步骤简述如下<sup>[7-10]</sup>:

Step1 初始化。设定退火初始温度  $T_0$ , 温度冷却系数  $q$ , 终止温度  $T_{end}$ 。令  $T = T_0$ , 确定每个  $T$  时的迭代次数,即 Metropolis 链长  $L$ 。

Step2 任取初始解  $a_0$ , 计算初始解  $a_0$  对应的目

标函数值  $f(a_0)$ 。

Step3 模型扰动。随机产生新解  $a'_0$ , 并计算新目标函数值  $f(a'_0)$ 。

Step4 依据 Metropolis 准则,若  $f'(a_0) > f(a_0)$ , 则以新个体替换旧个体;否则接受新个体,舍弃旧个体。接受新解的概率  $P$  为:

$$P = \begin{cases} 1, & f(a'_0) < f(a_0) \\ \exp\left[-\frac{f(a'_0) - f(a_0)}{T}\right], & f(a'_0) \geq f(a_0) \end{cases} \quad (1)$$

Step5 对当前温度  $T$  和  $i = 1, 2, \dots, L$ , 重复 Step3 ~ Step4。

Step6 若  $T_i < T_{end}$ , 则算法结束,返回全局最优解;否则按式(2)执行降温操作,并转至 Step3。

$$T_{i+1} = qT_i \quad (2)$$

### 2.2 支持向量机

支持向量机 (SVM) 回归是通过核函数将低维空间中非线性回归问题映射到高维特征空间,然后在这个特征空间中求解凸优化问题,实现某一非线性变换后的线性回归。其回归算法原理如下<sup>[1-2,11]</sup>:

设含有  $l$  个训练样本的集合为:  $\{(x_i, y_i), i = 1, 2, \dots, l\}$ , 其中,  $x_i (x_i \in R^d)$  是第  $i$  个训练样本的输入列向量,  $x_i = [x_i^1, x_i^2, \dots, x_i^d]^T$ ,  $y_i \in R$  为对应输出值。SVM 回归用核函数将数据映射到高维特征空间,再在高维特征空间中进行线性回归,依据结构风险化最小化原则,将其学习过程转换为凸优化问题,即:

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

式中:  $\Phi(x)$  为非线性映射函数;  $\varepsilon$  为线性不敏感损失函数;  $\xi_i, \xi_i^*$  为松弛变量;  $C$  为惩罚因子。

其回归方程最终表述为:

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) K(x_i, x_j) + b \quad (4)$$

式中:  $K(x_i, x_j) = \Phi(x_i) \Phi(x_j)$  为满足 Mercer 条件的核函数;  $a_i^*, a_i$  为二次规划中 Lagrange 乘子。

$$b = \frac{1}{N_{sv}} \left\{ \sum_{0 < a_i < C} [y_i - \sum_{x_j \in SV} (a_i - a_i^*) K(x_i, x_j) - \varepsilon] + \sum_{0 < a_i < C} [y_i - \sum_{x_j \in SV} (a_i - a_i^*) K(x_i, x_j) + \varepsilon] \right\} \quad (5)$$

式中:  $N_{\text{svm}}$  为支持向量机个数。

本文选择径向基核函数作为 SVM 核函数, 径向基核函数表达式为:

$$K(x, x_i) = \exp(-g \|x - x_i\|^2) \quad (6)$$

式中:  $g > 0$ 。

依据上述 SVM 算法原理, SVM 性能主要受惩罚因子  $C$ 、核函数参数  $g$  和不敏感系数  $\varepsilon$  的影响。惩罚因子  $C$  过小, 样本惩罚减小, 训练误差就增大;  $C$  过大, 学习精度提高, 泛化能力变差。核函数参数  $g$  越小, 拟合误差越小, 训练时间延长, 过小的  $g$  会导致模型过拟合而降低泛化能力; 不敏感系数  $\varepsilon$  用于控制模型的预测能力,  $\varepsilon$  增加, 则支持向量个数减少, 模型由于欠拟合而导致预测精度降低;  $\varepsilon$  过小, 不但训练时间增加, 且易导致模型过拟合而使推广能力下降。

### 2.3 SA-SVM 预测模型

定义目标函数:

$$\begin{cases} \min f(C, g, \varepsilon) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ \text{s. t. } C \in [C_{\min}, C_{\max}], g \in [g_{\min}, g_{\max}], \\ \varepsilon \in [\varepsilon_{\min}, \varepsilon_{\max}] \end{cases} \quad (7)$$

式中:  $y_i$  为第  $i$  个样本实测值;  $\hat{y}_i$  为第  $i$  个样本模拟值, 可由(4)计算得到。SVM 模型参数优化的思想就是通过迭代算法搜寻一组参数  $(C, g, \varepsilon)$ , 通过 SVM 学习使适应度函数式(7)达到最小。

利用 SA 算法对参数  $(C, g, \varepsilon)$  进行寻优。具体算法步骤描述如下:

表1 上年度1月-次年3月月均流量与次年1-3月月相关系数

月份	1	2	3	4	5	6	7	8	9	10	11	12	次年1月	次年2月
次年1月	0.263	0.288*	0.311*	0.515**	0.205	0.213	0.347*	0.676**	0.561**	0.688**	0.694**	0.914**		
次年2月	0.208	0.239	0.234	0.461**	0.239	0.217	0.288*	0.608**	0.545**	0.634**	0.583**	0.864**	0.832**	
次年3月	0.222	0.248	0.193	0.366**	0.137	0.110	0.177	0.545**	0.436**	0.498**	0.492**	0.737**	0.800**	0.842**

注: “\*\*”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。

从表1可以看出, 该站上年度月径流与次年度枯水期1-3月月径流相关系数在 0.110~0.914 之间。本文选取 0.01 水平(双侧)上显著相关的上年度4月、8-12月月均径流作为影响因子预测次年枯水期1月月均径流量; 选取上年度4月、8-12月及次年1月月均径流作为影响因子预测次年枯水期2月月均径流量; 选取上年度4月、8-12月及次年1-2月月均径流作为影响因子预测次年枯水期3月月均径流量。并以 1952-1994 年实测月月均径流作为训练样本拟合 1953-1995 年枯水期月月均径流量, 1995-2004 年实测月月均径流作为预测样本预测

Step1 初始化参数。选取 SVM 模型训练样本和预测样本, 进行样本归一化处理, 设定惩罚因子  $C$ 、核函数参数  $g$  和不敏感系数  $\varepsilon$  的搜寻空间; 设定 SA 算法初始温度  $T_0$ , 温度冷却系数  $q$ , 终止温度  $T_{\text{end}}$  及 Metropolis 链长  $L$ 。确定式(7)为优化目标适应度函数。

Step2 退火操作。按上述 SA 算法进行全局极值寻优。

Step3 预测。获得最优解, 即  $(C, g, \varepsilon)$  值。将  $(C, g, \varepsilon)$ , 作为 SVM 的最佳学习参数对预测样本进行预测。

## 3 应用实例

### 3.1 数据来源与分析

云南省龙潭寨水文站设立于 1951 年 4 月, 位于盘龙河干流上游, 控制径流面积 3 128 km<sup>2</sup>, 观测有水位、流量、降雨等项目, 是国家重要水文站和中央报汛站。盘龙河属红河流域泸江水系, 发源于红河州蒙自县三道湾, 自西北向东南流经砚山、丘北、文山、马关、西畴、麻栗坡六县, 于天保船头注入越南, 交泸江汇入红河, 为非闭合的亏水流域。中越国界以上流域面积 6 497 km<sup>2</sup>, 河长 247 km, 落差 1 803 m, 平均坡降 7.24‰, 主要支流有岔河、德厚河、马过河、暮底河、布都河、白石岩暗河、九股水暗河等。本文以该站 1952-2005 年 54 a 的实测资料为例进行实例研究, 利用 SPSS 软件分析龙潭寨水文站 1952-2005 年上年度月径流与次年度枯水期 1-3 月月径流相关关系, 见下表 1。

1996-2005 年枯水期月月均径流量。并利用式(8)对该站各径流序列进行归一化处理:

$$\hat{x} = \frac{2(x - x_{\min})}{x_{\max} - x_{\min}} - 1 \quad (8)$$

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。

### 3.2 模型构建及参数设置

(1)模型构建。本文基于 MatlabR2011b 软件环境, 利用 libsvm 工具箱、模拟退火算法工具箱及谢菲尔德遗传算法工具箱编程构建枯水期 1-3 月月均径流的 SA-SVM、GA-SVM 预测模型, 并选取平

均相对误差绝对值  $MRE$  和最大相对误差绝对值  $\max RE$  对各模型的预测效果进行评价:

$$MRE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \times 100 \quad (9)$$

$$\max RE = \max_{1 \leq i \leq n} \frac{|\hat{y}_i - y_i|}{y_i} \times 100 \quad (10)$$

式中:  $\hat{y}_i$  为第  $i$  个样本预测值;  $y_i$  为第  $i$  个样本实测值;  $i = 1, 2, \dots, n$ ;  $n$  为预测样本数。

(2) 参数设置。为了便于模型间性能的比较, SA-SVM、GA-SVM 模型相同参数设置如下: 最大迭代次数  $L = 200$ , 惩罚因子  $C$ 、核函数参数  $g$  的搜索空间均设为  $0.001 \sim 1000$ , 不敏感系数  $\varepsilon$  搜索空间设为  $0.001 \sim 0.1$ , 交叉验证参数  $V = 2$ 。其中 SA-SVM 模型初始温度  $T_0 = 500$ , 终止温度  $T_{end} = 0.001$ , 温度冷却系数  $q = 0.95$ 。GA-SVM 模型种群规模  $N = 30$ , 交叉概率  $P_c = 0.7$ , 变异概率  $P_m = 0.05$ 。

### 3.3 算法寻优及枯水期月径流预测评价

利用 SA、GA 算法(进化过程以 1 月数据拟合为例)及 SA-SVM、GA-SVM 模型对龙潭站 1996-2005 枯水期月均径流进行拟合及预测, 结果见表 2。

表 2 龙潭站枯水期 1-3 月径流预测结果

年份	SA-SVM 模型			GA-SVM 模型		
	相对误差/%			相对误差/%		
	1月	2月	3月	1月	2月	3月
1996	-4.22	-0.18	1.03	-4.35	-1.56	-8.76
1997	0.44	5.14	12.17	-1.36	5.46	-0.08
1998	-1.32	-7.58	7.43	5.07	-7.51	11.19
1999	4.37	-4.32	5.40	-4.70	-5.28	-6.68
2000	1.90	9.77	13.52	0.80	9.22	-2.03
2001	3.07	5.81	-4.76	3.35	4.79	-16.88
2002	9.11	7.85	14.11	11.25	7.92	1.44
2003	-1.13	-7.19	1.99	4.32	-5.06	-5.33
2004	-2.21	0.29	-7.10	-8.18	1.07	-16.59
2005	-3.34	-1.16	-0.05	-4.99	-10.24	-6.37
$MRE$	3.11	4.93	6.75	4.84	5.81	7.54
$\max RE$	9.11	9.77	14.11	11.25	10.24	16.88
$\min f$	0.0328	0.0261	0.0549	0.0584	0.331	0.0588

(1) 从表 2 可以看出, SA-SVM 模型对龙潭站 1996-2005 年 1-3 月月均径流预测的  $MRE$  分别为 3.11%、4.93% 和 6.75%, 预测精度分别比 GA-SVM 模型提高了 35.7%、15.1% 和 10.5%, 表明 SA-SVM 模型具有较好的预测精度和泛化能力。

(2) 从目标适应度函数值  $\min f$  来看, SA 算法的  $\min f$  均小于 GA 算法, 表明 SA 算法具有较好的全局寻优能力。

## 4 结 语

利用 SA 算法搜寻 SVM 模型最优惩罚因子  $C$ 、核函数参数  $g$  和不敏感系数  $\varepsilon$ , 提出 SA-SVM 预测模型, 以云南省龙潭站枯水期 1-3 月月均径流预测为例进行研究, 结果表明:

(1) SA 算法具有较快的收敛速度和全局寻优能力, 利用 SA 算法搜索 SVM 学习参数, 实现了 SVM 模型惩罚因子  $C$ 、核函数参数  $g$  和不敏感系数  $\varepsilon$  的自动确定, 提高了 SVM 模型的预测精度和泛化能力。

(2) 从实例 1996-2005 年枯水期月径流预测结果来看, SA-SVM 模型预测效果优于 GA-SVM 模型。本文 SA-SVM 模型及方法可为枯水期月径流预测预报提供一种新的途径和方法。

## 参考文献:

- [1] 崔东文. 支持向量机在湖库营养状态识别中的应用[J]. 水资源保护, 2013, 29(4): 26-30.
- [2] 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): 20-27.
- [3] 崔东文. 基于多元变量组合的回归支持向量机集成模型及其应用[J]. 水利水运工程学报, 2014(2): 66-73.
- [4] 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.
- [5] 田雨波. 混合神经网络技术[M]. 北京: 科学出版社, 2009.
- [6] 马良, 朱刚, 宁爱兵. 蚁群优化算法[M]. 北京: 科学出版社, 2010.
- [7] 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.
- [8] 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.
- [9] 贾伟娜, 刘顺兰. 模拟退火遗传算法在 DOA 估计技术中的应用[J]. 计算机工程与应用, 2014, 50(12): 266-270.
- [10] 李政伟, 谭国俊. 改进的退火遗传优化策略应用研究[J]. 计算机工程与应用, 2010, 46(4): 245-248.
- [11] 刘伟, 王建平, 刘长虹, 等. 基于粒子群寻优的支持向量机番茄红素含量预测[J]. 农业机械学报, 2012, 43(4): 143-147+155.