

Evolving Support Vector Regression by Co-Evolution Algorithm for Runoff Forecasting Modeling Research and Application

Jiansheng Wu¹, Yongsheng Xie¹, Long Jin²

¹Department of Mathematical and Computer Sciences, Guangxi Science & Technology Normal University, Liuzhou Guangxi

²Guangxi Climate Center, Nanning Guangxi

Email: wjsh2002168@163.com, jinlong01@163.com

Received: Feb. 29th, 2020; accepted: Mar. 13th, 2020; published: Mar. 20th, 2020

Abstract

Aiming at the difficulty of selecting the appropriate kernel function type, kernel function parameters, penalty coefficients of support vectors, and insensitive loss parameters in practical applications of support vector model, in this paper, co-evolution algorithm is employed to simultaneously evolve selection the data features, the kernel type of support vector regression, the parameters of the kernel function and the model parameters, namely HPSOSA-SVR by hybrid the strengths of Particle Swarm Optimization (PSO) and Simulated Annealing (SA). In order to select the optimal support vector regression kernel function type, the optimal matching kernel function parameter and the support vector model parameter for runoff modelling, the co-evolution algorithm focuses on combining the advantages of PSO (fast calculation and easy mechanism) and SA (ability to jump away from local optimum solutions and converge to the global optimum solution) by combined the metropolis processes of SA into the movement mechanism and parallel processing of PSO. The HPSOSA algorithm has the capability of both fast calculation and searching in the direction of the global optimum solution, helping PSO jump out of local optima, avoiding into the local optimal solution early and leads to a good solution. Compared with HPSOSA, PSO and pure SVR, results show that the HPSOSA-SVR model can successfully identify the optimal type of kernel function and all the optimal values of the parameters of SVR with the lowest prediction error values in runoff forecasting, can significantly improve the runoff forecasting accuracy. Experimental results reveal that the predictions using the proposed approach are consistently better than those obtained using the other methods presented in this study in terms of the same measurements.

Keywords

Particle Swarm Optimization, Simulated Annealing Algorithm, Support Vector Regression, Kernel Function, Runoff Modeling

基于协同进化算法的支持向量回归径流建模研究与应用

吴建生¹, 谢永盛¹, 金龙²

¹广西科技师范学院数学与计算机科学学院, 广西 柳州

²广西气象局, 广西 南宁

Email: wjsh2002168@163.com, jinlong01@163.com

收稿日期: 2020年2月29日; 录用日期: 2020年3月13日; 发布日期: 2020年3月20日

摘要

针对支持向量回归模型在实际应用中难以选择建模数据的特征、核函数类型、最优核函数参数以及支持向量的惩罚系数和不敏感损失函数的参数问题, 本文提出利用粒子群算法混合模拟退火算法构造协同进化算法, 自适应选择建模数据的特征、协同进化选择支持向量回归的核函数类型, 匹配最优核函数的参数和模型参数, 以此建立协同进化的支持向量回归柳江径流预测模型。本文通过混合模拟退火算法 Metropolis接受准则和粒子群算法随机接受准则, 结合粒子群算法的并行计算和模拟退火算法的干扰机制, 减少在局部搜索点附近的震动偏离构造协同进化算法, 自适应筛选支持向量回归模型建模数据的特征、协同进化选择核函数类型, 最优匹配核函数参数和支持向量模型参数, 建立柳江径流建模, 实验结果表明, 该方法能够快速有效地选择建模数据的特征、核函数类型、最优匹配核函数参数和支持向量模型参数, 其泛化性能明显提升, 拟合效果更好。该方法用于柳江径流预测, 具有良好的建模效果和更高的预测精度。

关键词

粒子群优化, 模拟退火算法, 支持向量回归, 核函数, 径流模型

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

准确的径流预测建模是水文领域中一项具有十分挑战性的任务, 也是减灾防灾的重要研究课题, 精确可操作的径流预报模型对于水资源规划及综合利用, 水利枢纽运行管理等重大决策问题的提供基本决策依据, 对国民经济健康发展具有十分重要的意义。在水文科学研究中, 虽然基于水文过程的径流物理模拟有了很大进展, 但是由于降水径流过程的具有时空分布不均匀性和随机性、而且动力学模型的边界条件受地质和人类活动的因素影响太多[1] [2] [3], 自然界中降水形成有效径流受降水、渗透、蒸发等气候因素影响很大, 具有高度非线性、时变性、不确定性和强耦合等特性, 对参数极端敏感等问题, 使得径流过程的物理过程模型往往难以客观描述和构造, 即使该模型可以构造, 但是数学求解极为困难

[4] [5] [6]。而近年来随着计算机硬件和软件技术的发展,许多机器学习建模方法,如人工神经网络(Artificial Neural Network, ANN)、支持向量机(Support Vector Machine, SVM)方法迅速应用径流建模,它们不需要详细描述径流物理过程,由于它具有很强的处理非线性问题的能力[7] [8],比一般的方法具有更好的预测能力,它为水文的径流建模分析提供了新技术、新方法[9] [10]。

20世纪90年代由Vapnik等人提出的统计学习理论,在此基础上发展起来的支持向量机是一种新型机器学习方法。它是基于结构风险最小化原则,集成了最大间隔超平面、核技术、凸二次规划、稀疏解和松弛变量等多项技术[11] [12],在解决小样本、非线性及高维空间学习时显示了其卓越的优越性,被成功应用到函数拟合逼近,文本和手写识别等领域[13] [14] [15] [16]。SVM在实际应用建模过程中,但是对于建模数据的特征选择、支持向量的核函数类型、核函数的参数以及支持向量的惩罚系数和不敏感损失函数参数没有理论指导,实际应用中采用交叉验证方法确定参数。这种方法虽然简单,但是过多地依赖设计者的经验,缺乏理论基础,得到的结果不稳定,这些问题影响SVM模型在实际应用中推广[17] [18] [19] [20]。

本文利用粒子群-模拟退火算法构造协同进化计算技术,自适应选择建模数据的特征、协同进化支持向量机的核函数类型和匹配相关最优参数,以此建立广西柳江径流预测模型(Support Machine Regression with Hybrid Particle Swarm Optimization and Simulated Annealing, HPSOSA-SVR)。本文通过对广西柳江2008年,2009年和2010年汛期径流建立预测模型,计算结果表面该方法能够针对水文自身季节变化情况,选择建模数据的特征信息,协同进化最优核函数类型、匹配最优核函数和模型的参数,而且精度高、稳定性好,为径流建模预测分析提供了一种可靠、有效的途径和实际应用中可以操作的方法。

2. 混合进化支持向量回归建模

2.1. 支持向量机

支持向量机(Support Vector Machine, SVM)是在Vapnik等人提出的小样本统计学习理论基础发展而来的机器学习方法,其算法是基于VC维数理论和结构风险最小化准则。在处理非线性问题时,通过非线性核函数将输入向量映射到高维线性特征空间,在这个空间构造样本最优超平面,以此将非线性问题就转化为高维空间中的线性问题,在模型的复杂性和泛化能力之间寻求最佳折中,并且能有效地克服维数灾和有效提高泛化能力[17]。SVM按照使用的类别可以分为支持向量分类(Support Vector Classification, SVC)和支持向量回归(Support Vector Regression, SVR),其中SVR算法用于时间序列的预测、非线性建模与预测、优化控制等方面。

对于给定的样本数据集 $\{(x_i, y_i) | i = 1, 2, \dots, N\}$,其中 $x_i \in R^n$ 是输入数据, $y_i \in R$ 是输出数据,SVR方法是首先利用非线性映射 $\phi(\cdot)$ 将样本空间从原低维空间映射到高维特征空间,在高维特征空间进行线性回归估计输入和输出非线性关系,回归函数:

$$f(x) = \omega^T \phi(x) + b \quad (1)$$

其中 ω 是权重向量, b 是偏置量,将非线性估计函数转化为高维特征空间中的线性估计函数,利用结构风险最小化原理,引入不敏感损失函数的参数 ε 和松弛变量 ξ, ξ^* ,支持向量回归问题就转化为以下目标函数的最小化问题:

$$\begin{cases} \min(\omega, \xi, \xi^*) = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^N (\xi + \xi^*) \\ s.t. \quad y_i - \omega^T \phi(x_i) - b \leq \varepsilon + \xi \\ \quad \quad -y_i + \omega^T \phi(x_i) + b \leq \varepsilon + \xi^* \\ \quad \quad \xi, \xi^* \geq 0, i = 1, 2, \dots, N \end{cases} \quad (2)$$

其中 C 为惩罚参数, 对其进行拉格朗日乘法求解二次规划得到回归函数:

$$f(x) = \sum_{i=1}^N (\beta_i^* - \beta_i) K(x_i, x) + b \quad (3)$$

其中 $(\beta_i^* - \beta_i) \neq 0$ 为训练样本支持向量, $K(x_i, x)$ 为核函数, 根据 Mercer 条件, 存在映射函数 $\varphi(x)$ 和核函数 $K(x_i, x)$, 使得满足:

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

在实际应用中, 常见的核函数有以下四种类型:

$$K_1(x_i, x_j) = x_i x_j^T \quad (5)$$

$$K_2(x_i, x_j) = (t + c x_i x_j^T)^d \quad (6)$$

$$K_3(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (7)$$

$$K_4(x_i, x_j) = \tanh[\alpha(x_i x_j^T + \beta)] \quad (8)$$

以上公式(5)是线性核函数, 在线性核函数中没有参数; 公式(6)是多项式核函数, 其中的参数 c 是斜率, d 是多项式维数; 公式(7)是高斯核函数, 其中的参数 σ 在 SVR 性能中起着主要作用, 如果估计过大核函数将呈线性, 高维投影失去其非线性功率, 如果估计过小, 核函数就缺乏正则化; 公式(8)是 Sigmoid 型核函数, 其中 α 是幅度调整参数, β 是控制映射阈值位移参数。

SVR 的性能主要取决于三个因素: 1) 核函数的类型及其核函数参数; 2) SVR 模型中惩罚系数 C 和不敏感损失函数的参数 ε ; 3) 建模数据集的特征信息。首先核函数类型的选择是决定的 SVR 模型的主要因素, 核函数是反映数据的先验知识, 决定训练数据高维特征空间情况, 不恰当的核函数会使得训练样本在高维空间的分布变得比输入空间的分布更复杂, 核函数的构造对于特定的核方法固然重要, 但当核函数构造完毕(核函数的类型固定)后, 其次核函数中的参数最优匹配同等重要。研究表明, 针对同一个核函数, 选择不同的核函数的参数, SVR 的泛化性能会相差很大[16], 这主要是因为不同的参数对应的高维特征空间的结构具有差异性, 而特征空间的性质直接决定着核方法的性能。再次 SVR 模型的惩罚系数和不敏感损失函数参数也是两个重要参数, 其中惩罚系数控制着最小训练误差和最小化模型复杂度之间的平衡, 目的是在总体误差与学习机模型复杂性之间进行折衷, 即在确定的特征空间中调节经验风险和学习机置信范围的比例, 以使得学习机器的泛化能力最好, 而不敏感损失函数参数是度量训练点上误差的代价, 在模型的光滑性和泛化性之间起着平衡作用[17] [18] [19] [20]。最后建模数据最优特征信息选择对 SVR 模型性能有一定的影响, 由于建模因子数据在收集中包含有噪声, 维数高, 而且数据是多数据源或异构数据集, 如果不对原始建模因子数据做特征提取处理, 将会影响 SVR 的性能, 增加模型的复杂度影响模型泛化性能[21] [22] [23] [24]。

由于核函数的类型及其核函数参数, 支持向量回归模型中惩罚系数 C 和不敏感损失函数的参数 ε , 以及建模数据的特征信息之间它们相互依赖, 相互影响, 在实际建模应有中需要同时考虑混合优化, 可以选择最为适用的模型, 但是目前的研究文献主要是针对特征信息提取, 或者核函数的类型及其核参数进行优化, 或者针对惩罚系数 C 和不敏感损失函数的参数 ε 进行优化, 如 Huang 和 Wang 在核函数类型为高斯核函数前提下, 提出利用遗传算法同时优化特征子集和 SVR 模型参数[25]极大提高分类准确率, 高斯核函数参数采用交叉验证方法确定, 论文没有对其它类型核函数选择优化; Oliveira 和 Braga 提出利用粒子群算法优化选择训练数据特征和高斯核函数的参数, 并以此建立软件工程量需求预测模型, 取得

较高得准确率,但是该论文没有对 SVR 模型的参数进行优化[26]; Li Xiaofeng 和 Wu Shijing 等人利用粒子群算法优化支持向量机的惩罚系数 C 和高斯核函数的 σ 建立高压断路器机械故障诊断,取得很好的效果,但是没有选择特征函数和优化其他参数[27]。Yang Ailing 和 Li Weide 等利用自相关函数提取输入信息特征,并利用灰狼算法和交叉验证优化最小二乘支持向量回归模型参数,建立短期电力负荷预测模型,实验室结果显示相对传统时间序列模型该方法能提高预测精度[28],但是对特征选择和模型参数没有优化。为此,本文利用粒子群-模拟退火算法构造混合进化算法技术,优化选择最佳的建模数据的特征,协同进化支持向量机的核函数类型以及参数,相应匹配最优模型参数,开展协同进化支持向量机径流预测建模研究,达到充分挖掘建模数据的有效特征信息,对不同的数据属性自适应选择不同的核函数类型,核函数的参数和模型参数。

2.2. 粒子群优化和模拟退火算法

粒子群优化(Particle Swarm Optimization, PSO)是由 Kennedy 和 Eberhart 在 1995 年提出的一种基于群体智能方法的进化计算技术,其思想主要来源于自然界中鸟群等生物种群觅食行为的仿真,它是通过个体之间的信息交换,互动协作来搜寻全局最优解,其算法简单、易实现同时又有深刻的群体智能背景,因此,在短短十几年间,获得了很快的发展,出现大量的研究成果,并在函数优化、神经网络训练、模式分类、模糊系统控制以及其他工程领域得到了广泛的应用[29]。但它也有缺陷,如容易早熟,易陷入局部极值点,在进化后期粒子种群容易失去多样性,优化精度较差。因此,许多研究者提出了大量的改进算法,这些改进算法在一定程度上改善了基本粒子群算法的性能[30],粒子群算法具体运算流程如图 1 所示。

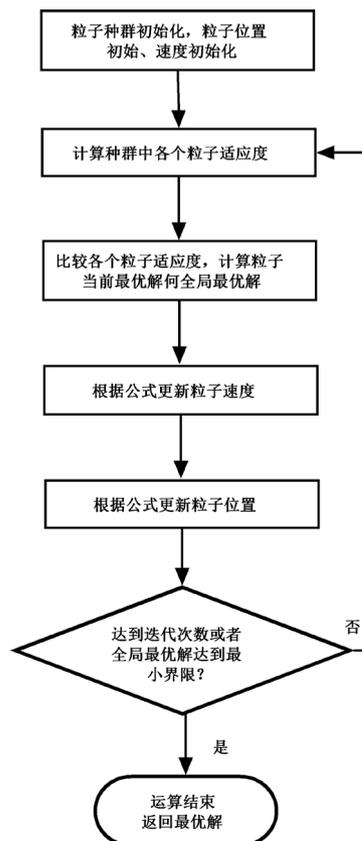


Figure 1. Flow chart of PSO
图 1. 粒子群优化算法流程图

模拟退火算法(Simulated Annealing, SA)最早的思想是由 N. Metropolis 等人于 1953 年提出。1983 年 S. Kirkpatrick 等成功地退火思想引入到组合优化领域[31]，它是基于 Monte-Carlo 迭代求解策略的一种随机寻优算法，其出发点是基于物理中固体物质的退火过程与一般组合优化问题之间的相似性。在迭代过程中不像粒子群算法只是用更好的值来代替原来的位置，在 SA 算法中伴随温度参数的不断下降，是以概率接受新状态在解空间中随机寻找目标函数的全局最优解，即在局部最优解能概率性地跳出并最终趋于全局最优[32] [33] [34]，模拟退火算法运算流程如图 2 所示。

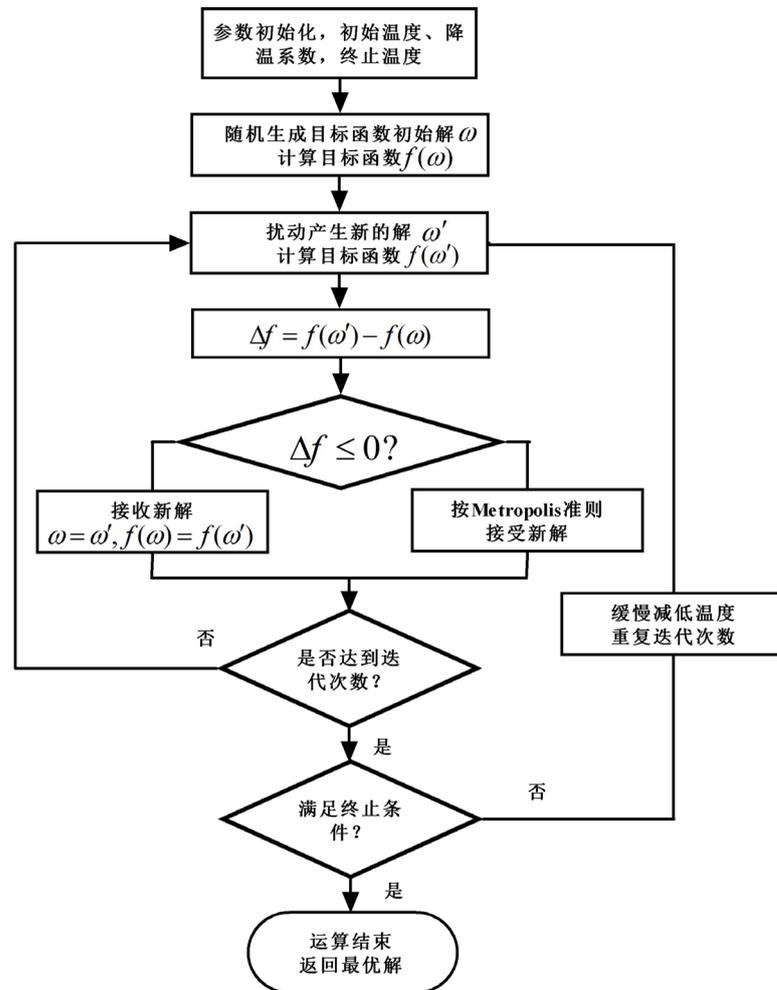


Figure 2. Flow chart of SA
图 2. 模拟退火算法流程图

2.3. 粒子群 - 模拟退火算法协同进化支持向量回归建模

对于给定样本 $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N), x_i \in R^n, y_i \in R\}$ ，其中 n 为影响径流变量因子个数， N 是样本个数，为此把选取的建模数据分成训练数据集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_{N_1}, y_{N_1}), x_i \in R^n, y_i \in R\}$ 和测试数据 $\{(x_{N_1+1}, y_{N_1+1}), (x_{T+2}, y_{T+2}), \dots, (x_N, y_N), x_i \in R^n, y_i \in R\}$ ，为此定义适应度函数为：

$$f(T) = \frac{1}{N_1} \sum_{i=1}^{N_1} (y_i - \hat{y}_i)^2 \tag{9}$$

本文是利用粒子群 - 模拟退火算法构造协同进化算法优化选择训练数据集的特征, 核函数类型及其参数, 支持向量参数, 其数学模型为最大最小化离散随机动态规划问题:

$$\left\{ \begin{array}{l} \max f(T) = \frac{1}{N_1} \sum_{i=1}^{N_1} (y_i - \hat{y}_i)^2 \\ \text{s.t. } \hat{y}_i = \sum_{i=1}^{N_1} (\beta_i^* - \beta_i) K(c_i x_i, x) + b \\ \left\{ \begin{array}{l} \min(\omega, \xi, \xi^*) = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^N (\xi + \xi^*) \\ \text{s.t. } y_i - \omega^T \varphi(x_i) - b \leq \varepsilon + \xi \\ \quad - y_i + \omega^T \varphi(x_i) + b \leq \varepsilon + \xi^* \\ \quad \xi, \xi^* \geq 0, i = 1, 2, \dots, N \\ c_i \in \{0, 1\}, i = 1, 2, \dots, N_1 \\ K_j \in \{K_1, K_2, K_3, K_4\} \end{array} \right. \end{array} \right. \quad (10)$$

由于混合模型需要进化选择变量和核函数类型, 优化对应核函数参数和模型参数, 为此本文编码采用 0-1 布尔代数和浮点数混合编码方法, 其中 0-1 编码主要控制变量选择和核函数类型, 浮点数编码主要控制核函数参数和支持向量参数大小, 控制变量选择编码 0 表示该变量没有选中, 1 表示该变量被选中; 控制核函数选择 00 表示线性核函数、01 表示多项式核函数、10 表示高斯核函数、11 表示 Sigmoid 核函数; 所有 0-1 编码形成整数编码串, 浮点数 R_1 和 R_2 对应核函数参数编码, R_3 和 R_4 则对应 SVR 模型参数编码, 所有浮点数形成浮点数编码串, 具体构成和表示见表 1。

Table 1. Relevant parameter coding components involved in the SVR model
表 1. SVR 模型所涉及的相关参数编码构成

控制变量的选择				控制核函数的选择				核函数参数		模型的参数	
X^1	X^2	...	X^N	K_1	K_2	K_3	K_4	P_1	P_2	C	ε
0	1	...	1	00	01	10	11	R_1	R_2	R_3	R_4

粒子群 - 模拟退火混合进化 SVR 的具体实现步骤如下:

1) 参数初始化, 随机生成 n 行 m_1 列布尔代数矩阵和 n 行 m_2 列浮点数矩阵, n 表示种群个数, 本文 $n = 40$, 粒子群种群个数 40 个, m_1 是影响径流变量建模因子个数和四个核函数决定, 为此 $m_1 = N + 4$, m_2 是核函数参数个数和模型参数决定, 故此 $m_2 = 4$ 。布尔代数是控制矩阵, 控制变量选择和核函数类型, 浮点数矩阵取上的 $[0,10]$ 均匀分布随机数, 它控制核函数的参数和 SVR 模型参数的大小。

2) 计算适应度, 输入训练样本并标准化, 通过解码每行布尔代数矩阵和浮点数矩阵, 即可获得参与计算建模样本因子、核函数类型、核函数参数和模型参数, 输入相应选择建模样本的特征, 通过(10)式计算每个粒子的适应度, 比较种群个体的适应度获得最好位置 $p_{best}(t)$, 以及群体中全局的最好位置 $p_{gbest}(t)$ 。

3) 粒子速度更新方程为:

$$v_i^{t+1} = \omega v_i^t + c_1 r_1 (p_{best}(t) - s_i^t) + c_2 r_2 (p_{gbest}(t) - s_i^t) \quad (11)$$

$$\omega = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{iter_{max}} \cdot iter \quad (12)$$

其中 $\omega_{max}, \omega_{min}$ 分别是惯性权重的最大值和最小值, $iter, iter_{max}$ 分别是当前迭代次数和最大的迭代次数, c_1, c_2 是学习因子, r_1, r_2 是随机常数。

4) 粒子个体位置更新分为二进制离散粒子群和实数粒子群的更新, 二进制离散粒子群位置更新方程为:

$$s_i^{t+1} = \begin{cases} 0, & r \geq \frac{1}{1 + \exp(-v_i^{t+1})} \\ 1, & r < \frac{1}{1 + \exp(-v_i^{t+1})} \end{cases} \quad (13)$$

其中, r 为 $[0,1]$ 均匀分布的随机数。实数粒子群位置更新方程为:

$$s_i^{t+1} = s_i^t + v_i^{t+1} \quad (14)$$

5) 计算新的位置和旧的位置适应度之差 $\Delta f = f(s_i^{t+1}) - f(s_i^t)$, 若适应度变大, 则新的位置被接受, 作为下一代的粒子, 若适应度变小, 对其进行模拟退火 Metropolis 抽样准则接收进入下一代。在 $\Delta f < 0$ 时, 在 $[0,1]$ 均匀分布生成的随机数 R , 在当前温度 T_i 下, 如果 $R > e^{(\Delta f/T_i)}$, 则接收到下一代 s_i^{t+1} , 否则, 在当前温度 T_i 下, 对其粒子速度高斯扰动, 产生新的速度, 并以此更新位置, 使其达到抽样稳定准则。

6) 反复进行 2)~5), 直到适应度满足要求或者达到总的进化代数(总的进化代数)。

7) 把进化后的最后一代适应度最高个体解码, 得到相应 SVR 模型建模样本的特征、核函数的类型、核函数的参数和模型参数, 输入测试样本进行预测, HPSOSA-SVR 算法的流程框架如图 3 所示:

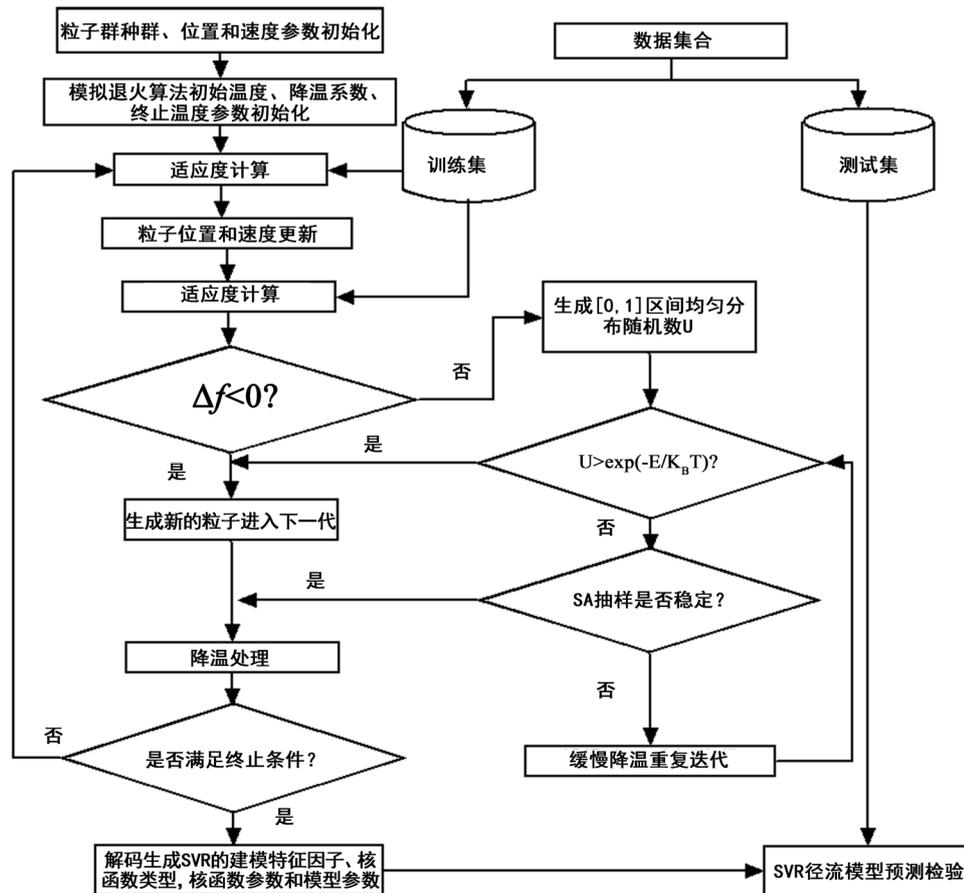


Figure 3. Flow chart of the HPSOSA-SVR algorithm
图 3. HPSOSA-SVR 算法的流程图

3. 应用实例及其结果分析

3.1. 建模数据

广西柳江属珠江水系,干流全长 773.3 公里,流域面积 57,173 平方公里,跨黔、桂、湘三省 30 个县、市,天然落差 1306 米,平均比降 1.68‰,年均流量 1865 立方米/秒,每年 4~9 月为汛期。从 1988 年至今的近 20 年间,大洪水的发生较为频繁,灾害也很严重,发生超过 100 年一遇的特大洪水 1 次(“1996.7”),20 至 50 年一遇的大洪水 3 次(“1988.8”、“1994.6”和“2004.7”洪水),10 至 20 年一遇的较大洪水 1 次(“2000.6”洪水),建立精度较高的柳江汛期径流预测模型对防灾减灾具有重要意义。

为此本文建模数据选取广西柳江老桥口的每天 12 时的径流数据进行应用实例分析,总的数据为 2006 年 1 月 1 日到 2010 年 12 月 31 日共 1826 个数据,分别建立 2008 年,2009 年和 2010 年 4~9 月汛期径流预测,依据建模需要把数据分为训练数据和测试数据,训练数据用于建立径流拟合模型,测试数据用于检验模型预测效果,具体如表 2 所示。

Table 2. Establish model training data and test model test data

表 2. 建立模型的训练数据和检验模型的检测数据

训练数据(拟合数据)	建模样本个数	测试数据(预测数据)	测试样本个数
2006 年 1 月 1 日~2008 年 3 月 31 日	821	2008 年 4 月 1 日~9 月 30 日	183
2007 年 1 月 1 日~2009 年 3 月 31 日	821	2009 年 4 月 1 日~9 月 30 日	183
2008 年 1 月 1 日~2010 年 3 月 31 日	821	2010 年 4 月 1 日~9 月 30 日	183

3.2. 建模因子

径流模型选择自变量对于预报模型非常重要。首先本文采用对区域径流有显著影响的降水、水汽、动力和热力等四类因子作为候选因子,从中国气象局 96 小时 T213 数值产品的 17 种常规气象要素和物理要素,数据覆盖柳江径流,范围从 150 N 到 300 N,经度从 1000 E 到 1200 E,分辨率为 $10^\circ \times 10^\circ$,共有 336 个网格点因子,经过初步筛选获得 45 个候选因子,对候选因子通过协同进化选择特征变量作为建模因子建立柳江径流模型。

3.3. 模型参数和性能评价

为考察 HPSOSA-SVR 模型的效果,分别建立支持向量回归(SVR)模型、单纯粒子群进化支持向量回归模型(PSO-SVR),SVR 模型采用高斯核函数,其所有参数确定采用交叉验证确定,PSO-SVR 是利用粒子群优化选择训练数据集的特征,核函数类型及其参数建立径流模型,其中两个进化模型参数为粒子群个数 40,进化迭代次数 200,模拟退火算法初始温度 1000,降温系数 0.9。模型分别对 821 个样本拟合和对 183 个样本预测,比较结果用来检验预测模型的效果。设和分别表示实际值和神经网络集成输出,为样本个数,根据数理统计的基本原理,采用 4 个指标来检验模型的拟合和预测效果。引入以下 4 种统计指标:

- 1) 均方根误差(Root Mean Square Error, RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (15)$$

- 2) 平均绝对百分比误差(Mean Absolute Percentage Error, MAPE):

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t|} \quad (16)$$

3) 纳什效率系数(Nash-Sutcliffe Efficiency coefficient, NSE):

$$SE = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (17)$$

式(15)~(17)中, y_t, \hat{y}_t 分别代表径流实际值和模型的输出值, n 为样本个数, 在纳什效率系数的基础上, 提出洪水峰值模型预测性能评价指标:

4) 峰值纳什效率系数

$$NSE_{peak} = 1 - \frac{\sum_{i=1}^p (y_{pi} - \hat{y}_{pi})^2}{\sum_{i=1}^p (y_{pi} - \bar{y}_{pi})^2} \quad (18)$$

式(18)中、 y_{pi}, \hat{y}_{pi} 分别表示洪水峰值实际值和预测洪水峰值输出值, \bar{y}_{pi} 表示洪水峰值的平均值, 本文把径流大于 4000 (m^3/s)看作是洪峰值, p 为洪水峰值期间样本个数。性能评价指标(1)和(2)主要是从数值上评价模型的性能, 其值越小说明模型的预测性能越好; 性能指标(3)主要是看模型能否对径流的趋势做出正确的判断, 其值越大就说明模型的可信度高, 能够准确跟踪径流趋势; 性能指标(4)是洪水峰值趋势的判断指标, 对于径流模型在实际应用中, 最关心的就是能否较为准确预测灾害性洪水, 即洪水峰值达到时刻和对洪水峰值趋势跟踪。

3.4. 结果分析

图 4 是 HPSOSA 和 PSO 分别进化 SVR 过程中, 适应度随迭代次数变化过程。从图 4 可以看出, 在 200 次迭代次数过程中, HPSOSA-SVR 前 60 次中出现振荡情况, 之后随着迭代次数增加适应度逐渐趋于变小而且稳定, 而 PSO-SVR 在进化过程前 140 次还在反复振荡, 直到迭代 200 次还没有趋于稳定状态。从图 4 迭代次数和适应度变化过程可以看出, HPSOSA-SVR 随着训练次数的增加, 适应度逐渐稳定, 避免过早收敛的问题并避免局部最优。

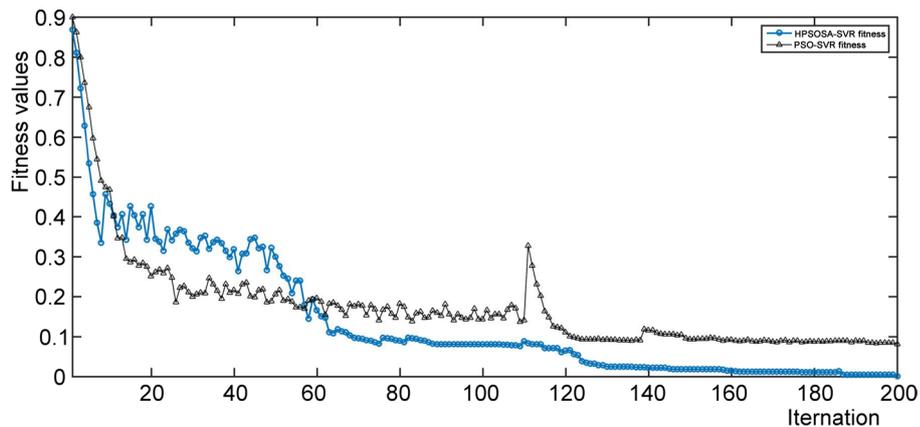


Figure 4. Fitness values of PSO-SVR and HPSOSA-SVR process with the iterations
图 4. HPSOSA-SVR 和 PSO-SVR 进化迭代过程中适应度变化过程

Table 3. Fitting and prediction results of three models during the 2008 flood season
表 3. 三种模型对 2008 年汛期的拟合和预测结果

拟合(2006 年 1 月 1 日~2008 年 3 月 31 日)				
模型	RMSE	MAPE	NSE	NSE_{peak}
SVR	167.8633	15.7583%	0.9866	0.8754

Continued

PSO-SVR	154.0910	12.2492%	0.9887	0.8921
HPSOSA-SVR	74.2182	4.0399%	0.9974	0.9526
预测(2008年4月1日~2008年9月30日)				
模型	<i>RMSE</i>	<i>MAPE</i>	<i>NSE</i>	<i>NSE_{peak}</i>
SVR	1211.1081	49.6047%	0.6586	0.4339
PSO-SVR	1012.6213	30.0892%	0.5449	0.3757
HPSOSA-SVR	340.3019	9.3433%	0.9747	0.9447

注：拟合数据中径流大于 4000 的峰值为 47 个，预测数据中径流大于 4000 的峰值是 24 个。

表 3 是三种模型对柳江径流对 2006 年 1 月 1 日~2008 年 3 月 31 日期间的 821 建模样本拟合和对 2008 年 4 月 1 日~9 月 30 日共 183 个样本预测的各种性能统计表，图 5 是三种模型对 2008 年 4 月 1 日~9 月 30 日期间径流预测效果图，图 6 是对 2008 年 4 月 1 日~9 月 30 日期间径流大于 4000 (m³/s)预测效果。

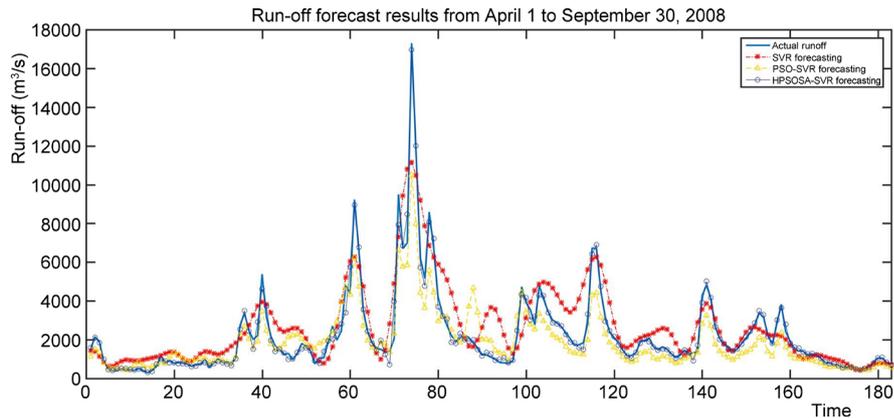


Figure 5. Prediction results of runoff from April 1 to September 30, 2008

图 5. 不同模型对 2008 年 4 月 1 日~9 月 30 日径流预测结果

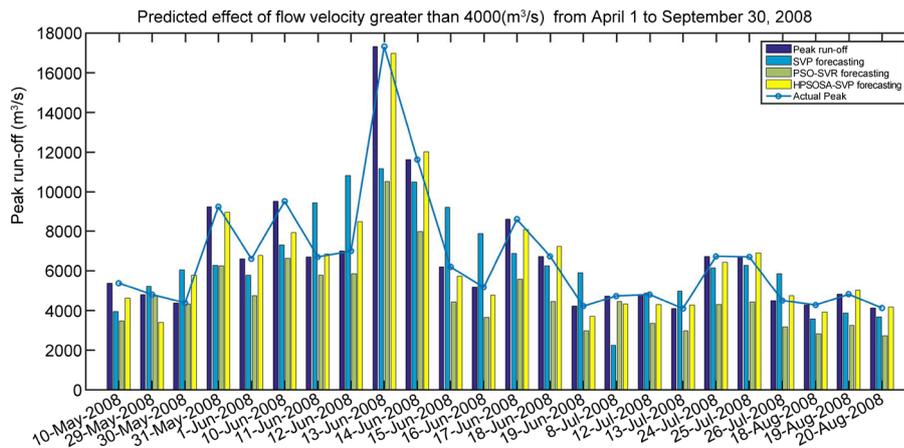


Figure 6. Prediction results on runoff greater than 4000 (m³/s) from April 1 to September 30, 2008

图 6. 2008 年 4 月 1 日~9 月 30 日径流大于 4000 (m³/s)预测结果

表 4 是三种模型对柳江径流 2007 年 1 月 1 日~2009 年 3 月 31 日期间的 821 训练样本拟合和对 2009 年 4 月 1 日~9 月 30 日期间共 183 个样本预测的误差性能统计表，图 7 是三种模型对 2009 年 4 月 1 日~9 月 30 日期间径流预测效果图，图 8 是对 2009 年 4 月 1 日~9 月 30 日期间径流大于 4000 (m³/s)预测效果。

Table 4. Fitting and prediction results of three models during the 2009 flood season
表 4. 三种模型对 2009 年汛期的拟合和预测结果

拟合(2007 年 1 月 1 日~2009 年 3 月 31 日)				
模型	RMSE	MAPE	NSE	NSE_{peak}
SVR	235.7658	13.0707%	0.9782	0.7954
PSO-SVR	196.1539	12.3797%	0.9849	0.8521
HPSOSA-SVR	67.8030	2.8373%	0.9982	0.9635
预测(2009 年 4 月 1 日~2009 年 9 月 30 日)				
模型	RMSE	MAPE	NSE	NSE_{peak}
SVR	1417.7070	53.6823%	0.7864	0.5481
PSO-SVR	1371.9913	46.2099%	0.7994	0.6799
HPSOSA-SVR	333.2365	6.0931%	0.9882	0.9679

注：拟合数据中径流大于 4000 的峰值为 49 个，预测数据中径流大于 4000 的峰值是 19 个。

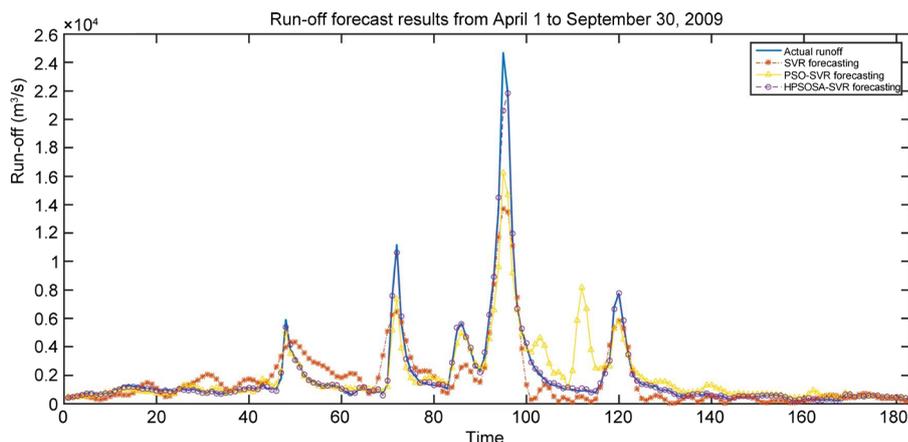


Figure 7. Prediction results of runoff from April 1 to September 30, 2009
图 7. 不同模型对 2009 年 4 月 1 日~9 月 30 日径流预测结果

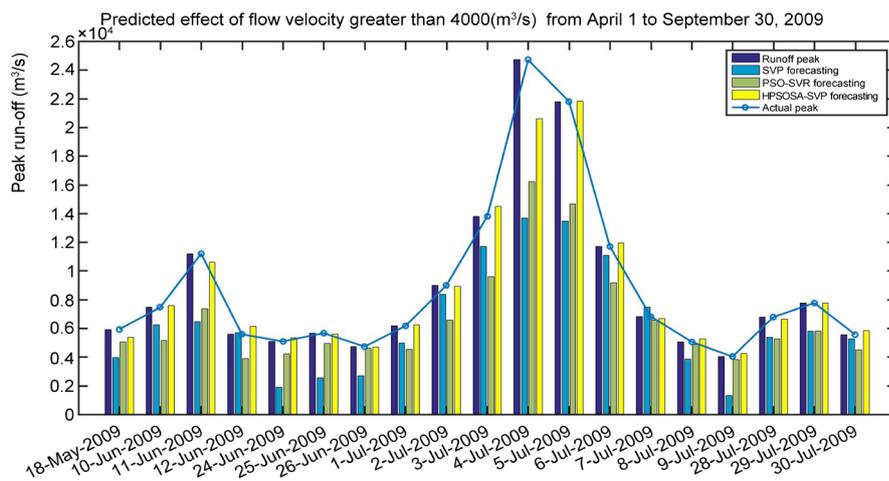


Figure 8. Prediction results on runoff greater than 4000 (m³/s) from April 1 to September 30, 2009
图 8. 2009 年 4 月 1 日~9 月 30 日径流大于 4000 (m³/s)预测结果

表 5 是两种模型对柳江径流在 2008 年 1 月 1 日~2010 年 3 月 31 日期间的 821 训练样本拟合和对 2010 年 4 月 1 日~9 月 30 日共 183 个样本预测的各种误差统计, 图 9 是三种模型对 2010 年 4 月 1 日~9 月 30 日径流预测效果, 图 10 是 2010 年 4 月 1 日~9 月 30 日径流大于 4000 (m³/s)预测效果。

Table 5. Fitting and prediction results of three models during the 2010 flood season

表 5. 三种模型对 2010 年汛期的拟合和预测结果

拟合(2008年1月1日~2010年3月31日)				
模型	RMSE	MAPE	NSE	NSE _{peak}
SVR	244.2786	14.0535%	0.9054	0.7863
PSO-SVR	195.9623	13.5166%	0.9723	0.8941
HPSOSA-SVR	78.1758	4.8994%	0.9845	0.9053
预测(2010年4月1日~2010年9月30日)				
模型	RMSE	MAPE	NSE	NSE _{peak}
SVR	1180.4190	111.8562%	0.7244	0.4377
PSO-SVR	862.7260	43.9980%	0.8517	0.2068
HPSOSA-SVR	192.9120	19.1050%	0.9926	0.9863

注: 拟合数据中径流大于 4000 的峰值为 48 个, 预测数据中径流大于 4000 的峰值是 21 个。

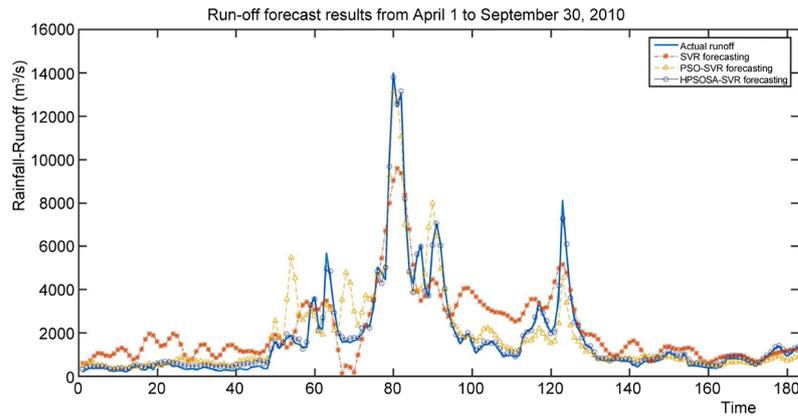


Figure 9. Prediction results of runoff from April 1 to September 30, 2010

图 9. 不同模型对 2010 年 4 月 1 日~9 月 30 日径流预测结果

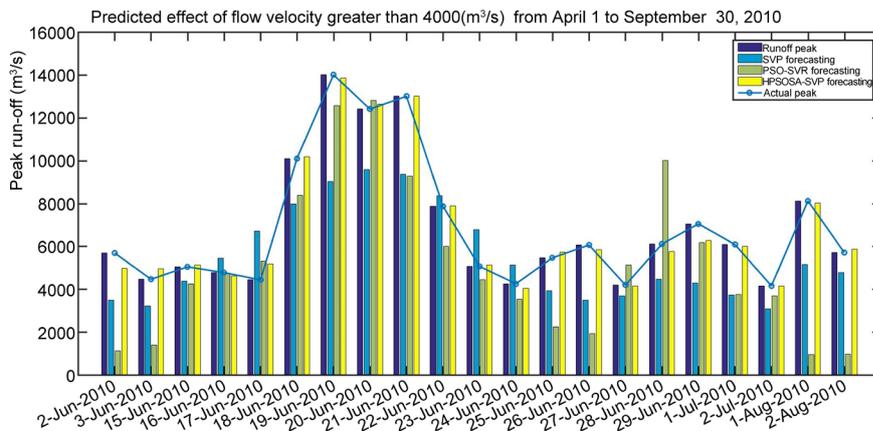


Figure 10. Prediction results on runoff greater than 4000 (m³/s) from April 1 to September 30, 2010

图 10. 2010 年 4 月 1 日~9 月 30 日径流大于 4000 (m³/s)预测结果

从表 3, 表 4 和表 5 的结果可以看出 HPSOSA-SVR 模型对训练样本拟合和测试样本预测的 RMSE 和 MAPE 和性能指标均小于 SVR、PSO-SVR 模型, 说明 HPSOSA-SVR 模型拟合值和真实值, 预测值和真实值之间的偏离程度小, 拟合精度和预测高于 SVR、PSO-SVR 模型; 同时可以看出 SVR-PSOSA 模型对训练样本和测试样本的 NSE 和 NSE_{peak} 性能指标均大于 SVR、PSO-SVR 模型, 说明 SVR-PSOSA 能够准确把握汛期径流的趋势, 这些显示出 SVR-PSOSA 模型相对 SVR 模型、PSO-SVR 具有很好的学习能力和预测能力。进一步从图 5~10 可以看出, PSO-SVR 性能优于 SVR 模型, 这进一步说明在实际应用中需要对 SVR 模型进一步优化, 提高 SVR 模型拟合和预测精度。

洪水峰值性能指标是预测灾害性洪水达到时刻和对洪水峰值趋势跟踪重要指标, 也是检验模型能否预测灾害性洪水以及在实际应用的重要指标。从表 3 的结果可以看出, 在对 2006 年 1 月至 2008 年 3 月建模样本拟合中 SVR 的峰值纳什效率系数是 0.8754, PSO-SVR 的峰值纳什效率系数是 0.8921, HPSOSA-SVR 的峰值纳什效率系数是 0.9526, 三者相差不大; 但是建立的模型对 2008 年 4 月~9 月样本预测时, SVR 的峰值纳什效率系数仅为 0.4539, PSO-SVR 的峰值纳什效率系数为 0.3757, HPSOSA-SVR 的预测峰值系数是 0.9447, 三者相差特别悬殊, SVR、PSO-SVR 模型基本不能在实际应用中预测洪水峰值, 而 HPSOSA-SVR 可以对洪水峰值做出精度很高的预测。同时从图 6 也可以看出, HPSOSA-SVR 模型能够对 2008 年汛期径流大于 4000 (m³/s) 峰值趋势很好的跟踪。同样 HPSOSA-SVR 模型在对 2009 年和 2010 年汛期峰值预测中, 对汛期洪水峰值趋势很好的跟踪能力, 这些结果均说明 HPSOSA-SVR 模型用于预测灾害性洪水可信度很高, 而且结果稳定。

表 6 给出 HPSOSA 协同进化支持向量回归得到核函数类型, 核函数的参数和支持向量回归的参数, 从表 6 可以看出, 三种不同数据的径流模型, 通过粒子群 - 模拟退火算法协同进化支持向量回归的核函数, 都得的核函数类型为 Sigmoid 型核函数, 以及核函数参数和模型参数, 表 6 的结果说明 Sigmoid 型核函数最为适宜用于建立柳江汛期径流 SVR 预测模型, 而且预测结果精度高, 稳定性好。

Table 6. Fitting and prediction results of three models during the 2010 flood season
表 6. 三种模型对 2010 年汛期的拟合和预测结果

模型	核函数参数类型	核函数参数		SVR 参数	
		α	β	C	ϵ
模型 1	Sigmoid 核函数	0.2504	-0.0346	26.20	0.0686
模型 2	Sigmoid 核函数	0.1627	-0.0627	25.39	0.0645
模型 3	Sigmoid 核函数	0.3253	-0.0265	18.60	0.1059

4. 结语

径流是水文系统中研究人员关注的重大课题, 它关系到水利工程建设和水资源的开发利用, 并且由于径流系统包含着系统空间和时间变化的非线性特性, 传统的分析方法很难对它们的演化规律和变化特征做出清晰的了解和准确的预测。本文利用粒子群算法和模拟退火算法构造协同进化算法, 同时选择建模数据的特征、协同选择核函数类型、优化核函数及其核函数的参数, 建立柳江径流预测模型, 实例计算结果得出以下结论:

1) HPSOSA-SVR 算法能避免局部最优解, 通过混合 SA 的接受准则和 PSO 随机接受原则, 减少了局部搜索端点附近的震动偏离; 根据新旧位置之间不同的适应度值, 运用规则决定是接受新位置, 还是重新计算另一个可信位置。这使得模型可以跳出局部最优解, 并减少局部搜索终点附近的震动偏离。这些特性提高了解决方案的质量和收敛速度。

2) HPSOSA-SVR 算法具有运算速度快和能在全局方向搜索最优解的优点,通过结合 PSO 运行机制、并行处理干扰机制以及 SA 解决搜索路径的特点,形成了一个良好的质量解决方案。

3) 根据本文的研究结果,我们可以得出以下结论: Sigmoid 曲线核函数的 SVR 模型可以作为实际径流预测应用中的一种有效的核函数,而且最为适宜汛期径流 SVR 预测模型,预测结果精度高,稳定性好。

4) 采用 Sigmoid 核函数时,支持向量回归实现的是一种前向神经网络,但是由于使用 SVM 方法,网络隐含层节点数、连接权值都是在设计的过程中进化确定,而且 SVR 的理论决定了它最终求得的是全局最优值而不是局部最小值,也保证了它对于未知样本的良好泛化能力而不会出现过学习现象,以此可以解释 Sigmoid 曲线核函数的 SVR 模型最为适宜径流预测模型。

基金项目

国家自然科学基金资助项目(41575051、41565005),广西科技厅科技基地和人才专项(No. AD16450003),广西科技厅创新项目(No. 2018AB14003),广西高校中青年教師基础能力提升项目(No. 2018KY0699、2018KY0700、2018KY0701),广西来宾市科学研究与技术开发项目(来科能 193305)。

参考文献

- [1] Kisi, O., Shiri, J. and Tombul, M. (2013) Modeling Rainfall-Runoff Process Using Soft Computing Techniques. *Computers & Geosciences*, **51**, 108-117. <https://doi.org/10.1016/j.cageo.2012.07.001>
- [2] Rezaie-Balf, M., Zahmatkesh, Z. and Kim, S. (2017) Soft Computing Techniques for Rainfall-Runoff Simulation: Local Non-Parametric Paradigm vs. Model Classification Methods. *Water Resources Management*, **31**, 3843-3865. <https://doi.org/10.1007/s11269-017-1711-9>
- [3] Kumar, D., Sarthi, P.P. and Ranjan, P. (2016) Rainfall-Runoff Modeling Using Computational Intelligence Techniques. *International Conference on Advances in Computing, Communications and Informatics*, Jaipur, 21-24 September 2016, 800-806. <https://doi.org/10.1109/ICACCI.2016.7732144>
- [4] Wu, J.S. and Xie, Y.S. (2019) Hybrid Support Vector Regression with Parallel Co-Evolution Algorithm Based on GA and PSO for Forecasting Monthly Rainfall. *Journal of Software Engineering and Applications*, **12**, 524-539. <https://doi.org/10.4236/jsea.2019.1212032>
- [5] Granata, F., Gargano, R. and De Marinis, G. (2016) Support Vector Regression for Rainfall-Runoff Modeling in Urban Drainage: A Comparison with The EPA's Storm Water Management Mode. *Water*, **8**, 69. <https://doi.org/10.3390/w8030069>
- [6] Kan, G., Li, J., Zhang, X., et al. (2017) A New Hybrid Data-Driven Model for Event-Based Rainfall-Runoff Simulation. *Neural Computing and Applications*, **28**, 2519-2534. <https://doi.org/10.1007/s00521-016-2200-4>
- [7] Talei, A., Chua, L. H. and Quek, C. (2010) A Novel Application of a Neuro-Fuzzy Computational Technique in Event-Based Rainfall-Runoff Modeling. *Expert Systems with Applications*, **37**, 7456-7468. <https://doi.org/10.1016/j.eswa.2010.04.015>
- [8] Wu, J.S., Jin, L. and Liu, M.Z. (2015) Evolving RBF Neural Networks for Rainfall Prediction Using Hybrid Particle Swarm Optimization and Genetic Algorithm. *Neurocomputing*, **148**, 136-142. <https://doi.org/10.1016/j.neucom.2012.10.043>
- [9] Wu, J.S. (2018) Co-Evolution Algorithm for Parameter Optimization of RBF Neural Networks for Rainfall-Runoff Forecasting. In: Huang, D.S., Bevilacqua, V., Premaratne, P. and Gupta, P., Eds., *Intelligent Computing Theories and Application. ICIC, Lecture Notes in Computer Science*, Springer, Cham, 195-206. https://doi.org/10.1007/978-3-319-95930-6_19
- [10] Song, X., Kong, F., Zhan, C., et al. (2012) Hybrid Optimization Rainfall-Runoff Simulation Based on Xinanjiang Model and Artificial Neural Network. *Journal of Hydrologic Engineering*, **17**, 1033-1041. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000548](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000548)
- [11] Vapnik, V. (2013) *The Nature of Statistical Learning Theory*. Springer Science & Business Media, Cham.
- [12] 邓乃杨, 田英杰. 数据挖掘中的新方法-支持向量机[M]. 北京: 科学出版社, 2004.
- [13] 朱波, 刘飞. 基于优化多核支持向量回归的制造过程均值偏移幅度估计[J]. 中国机械工程, 2014, 25(3): 630-635.
- [14] 汪洪桥, 孙富春, 蔡艳宁, 等. 多核学习方法[J]. 自动化学报, 2010, 36(8): 1037-1050.

- [15] Salat, R. and Salat, K. (2010) New Approach to Predicting Proconvulsant Activity with the Use of Support Vector Regression. *Computers in Biology and Medicine*, **42**, 575-581. <https://doi.org/10.1016/j.compbiomed.2012.02.001>
- [16] Kaytez, F., Taplamacioglu, M.C., Cam, E., et al. (2015) Forecasting Electricity Consumption: A Comparison of Regression Analysis, Neural Networks and Least Squares Support Vector Machines. *International Journal of Electrical Power & Energy Systems*, **67**, 431-438. <https://doi.org/10.1016/j.ijepes.2014.12.036>
- [17] 常群. 支持向量机的核方法及其模型选择[D]: [博士学位论文]. 哈尔滨: 哈尔滨工业大学, 2007.
- [18] 罗林开. 支持向量机的核选择[D]: [博士学位论文]. 厦门: 厦门大学, 2007.
- [19] 汪延华. 支持向量机模型选择研究[D]: [博士学位论文]. 北京: 北京交通大学, 2009.
- [20] 廖杰, 王文圣, 李跃清, 等. 支持向量机及其在径流预测中的应用[J]. *四川大学学报(工程科学版)*, 2006, 38(6): 24-29.
- [21] López, F.J.M., Puertas, S.M. and Arriaza, J.A.T. (2014) Training of Support Vector Machine with the Use of Multivariate Normalization. *Applied Soft Computing*, **24**, 1105-1111. <https://doi.org/10.1016/j.asoc.2014.08.020>
- [22] Güraksin, G.E., Haklı, H. and Uğuz, H. (2014) Support Vector Machines Classification Based on Particle Swarm Optimization for Bone Age Determination. *Applied Soft Computing*, **24**, 597-602. <https://doi.org/10.1016/j.asoc.2014.08.007>
- [23] Sheikhpour, R., Sarram, M.A. and Sheikhpour, R. (2016) Particle Swarm Optimization for Bandwidth Determination and Feature Selection of Kernel Density Estimation Based Classifiers in Diagnosis of Breast Cancer. *Applied Soft Computing*, **40**, 113-131. <https://doi.org/10.1016/j.asoc.2015.10.005>
- [24] Wu, J.S., Jin, L. and Liu, M. (2010) A Hybrid Support Vector Regression Approach for Rainfall Forecasting Using Particle Swarm Optimization and Projection Pursuit Technology. *International Journal of Computational Intelligence and Applications*, **9**, 87-104. <https://doi.org/10.1142/S1469026810002793>
- [25] Huang, C.L. and Wang, C.J. (2016) A GA-Based Feature Selection and Parameters Optimization for Support Vector Machines. *Expert Systems with Applications*, **31**, 231-240. <https://doi.org/10.1016/j.eswa.2005.09.024>
- [26] Oliveira, A.L.I., Braga, P.L., Lima, R.M.F., et al. (2010) GA-Based Method for Feature Selection and Parameters Optimization for Machine Learning Regression Applied to Software Effort Estimation. *Information and Software Technology*, **52**, 1155-1166. <https://doi.org/10.1016/j.infsof.2010.05.009>
- [27] Li, X., Wu, S., Li, X., et al. (2020) Particle Swarm Optimization-Support Vector Machine Model for Machinery Fault Diagnoses in High-Voltage Circuit Breakers. *Chinese Journal of Mechanical Engineering*, **33**, 6. <https://doi.org/10.1186/s10033-019-0428-5>
- [28] Yang, A., Li, W. and Yang, X. (2019) Short-Term Electricity Load Forecasting Based on Feature Selection and Least Squares Support Vector Machines. *Knowledge-Based Systems*, **163**, 159-173. <https://doi.org/10.1016/j.knosys.2018.08.027>
- [29] Poli, R., Kennedy, J. and Blackwell, T. (2007) Particle Swarm Optimization. *Swarm Intelligence*, **1**, 33-57. <https://doi.org/10.1007/s11721-007-0002-0>
- [30] Rini, D.P., Shamsuddin, S.M. and Yuhaniz, S.S. (2011) Particle Swarm Optimization: Technique, System and Challenges. *International Journal of Computer Applications*, **14**, 19-26. <https://doi.org/10.5120/1810-2331>
- [31] Dowsland, K.A. and Thompson, J. (2012) Simulated Annealing. In: *Handbook of Natural Computing*, Wiley, New York, 1623-1655. https://doi.org/10.1007/978-3-540-92910-9_49
- [32] Van, Laarhoven, P.J.M. and Aarts, E.H.L. (1987) Simulated Annealing. In: *Simulated Annealing: Theory and Applications*, Springer, Dordrecht, 7-15. https://doi.org/10.1007/978-94-015-7744-1_2
- [33] Mafarja, M.M. and Mirjalili, S. (2017) Hybrid Whale Optimization Algorithm with Simulated Annealing for Feature Selection. *Neurocomputing*, **260**, 302-312. <https://doi.org/10.1016/j.neucom.2017.04.053>
- [34] Assad, A. and Deep, K. (2018) A Hybrid Harmony Search and Simulated Annealing Algorithm for Continuous Optimization. *Information Sciences*, **450**, 246-266. <https://doi.org/10.1016/j.ins.2018.03.042>