# 基于改进相空间重构原理的支持向量机月径流模拟

胡昌军
(云南省麻栗坡县水务局, 云南 麻栗坡 663600)

**摘 要**: 基于交叉验证支持向量机(CV-SVM)原理及方法,构建以相空间重构理论与支持向量机相结合的径流时间序列模拟模型。针对相空间重构中确定延迟时间 $\tau$ 和嵌入维数 $m$ 的方法众多,且各方法确定结果不尽相同等缺点,本文采用试凑的方法,在延迟时间 $\tau$ 和嵌入维数 $m$ 取值范围为2 ~ 10内依次构建81个基于相空间重构理论的CV-SVM月径流模拟模型,以南利河董湖站月径流模拟为例进行分析,确定最佳延迟时间 $\tau$ 和嵌入维数 $m$,并与自相关函数法等相关延迟时间 $\tau$ 和嵌入维数 $m$ 的确定方法确定结果进行比较,旨在探寻延迟时间 $\tau$ 和嵌入维数 $m$ 对月径流模拟精度的影响及其规律,为基于时间序列的水文模拟及预测预报提供方法和参考。

**关键词**: 相空间重构; 支持向量机; 交叉验证; 径流模拟
**中图分类号**:P334.92 **文献标识码**: A **文章编号**: 1672-643X(2013)04-0210-07

# Monthly runoff simulation of support vector machine based on principle of improvement and phase space reconstruction

**HU Changjun**
(*Malipo County Water Authority of Yunnan Province*, *Malipo* 663000, *China*)

**Abstract**: Based on the CV-SVM principle and methods, the paper constructed the runoff time series model by combining phase space reconstruction theory with support vector machine. According to the many methods of phase space reconstruction to determine the delay time $\tau$ and embedding dimension $m$, and the shortcomings of determining different results by every method, this paper adopted the method of trial and error, the delay time $\tau$ and embedding dimension $m$ value range from 2 to 10 in order to construct 81 monthly runoff simulation model based on the theory of phase space reconstruction CV-SVM. Taking monthly runoff simulation in Nanli River Donghu Lake station for example, it determined the optimal delay time $\tau$ and embedding dimension $m$ and compared the results with that determined by autocorrelation function. The aim is to explore the influence of delay time $\tau$ and embedding dimension $m$ of monthly runoff simulation on accuracy and its regularity, and provide method and reference for the hydrological simulation and forecast of time series.

**Key words**: phase space reconstruction; support vector machine; cross validation; runoff simulation

## 1 概 述

水文系统受气候和人类活动影响,呈现出非常复杂的行为特征,借助诸如人工智能、模糊识别、知识工程等方法建立模型,可以处理多指标系统的综合识别问题。人工神经网络(Artifical Neural Network,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,是这类智能算法中运用最为广泛的算法之一,而BP网络(Back-Propagation Network,BP)是ANN最为常用的神经网络模型之一,广泛运用于径流预测[1-5]。然而,BP神经网络存在着学习收敛速度慢、易陷入局部极值以及网络结构难以确定等固有缺陷,难于解决因输入向量(嵌入维数)变化的实际问题。近年来,随着混沌理论和应用技术研究的不断深入,基于混沌时间序列的建模和预测已成为混沌信息处理研究领域中的热点,并成功应用于径流模拟及预测预报中[6-8],其相空间重构是影响径流模拟及预测成败的关键,而延迟时间 $\tau$ 和嵌入维数 $m$ 的选取对相空间重构具有十分重要意义。关于时延 $\tau$ 与嵌入维 $m$ 的选取,目前主要有两种观点:一种观点认为两者是互不相关的,

**收稿日期**:2013-01-11; **修回日期**:2013-02-23
**作者简介**:胡昌军(1979-),男,云南镇雄人,工程师,主要从事水利工程勘测设计与建设管理工作。

即 $\tau$ 和 $m$ 的选取是独立进行的,如求时延的自相关函数法[9]、互信息法[10],求嵌入维的 G-P 算法[11]或假最近邻法[12]等。另一种观点认为两者是相关的,即 $\tau$ 和 $m$ 的选取是互相依赖的,如嵌入窗法[13]、C-C 方法[14],可同时计算出时延和时间窗口。实际应用中,对于同一时间序列,上述各种方法的确定结果往往不尽相同,这对延迟时间 $\tau$ 和嵌入维数 $m$ 的选取带来困难。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[15-16],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,有效解决了传统 BP 神经网络存在着学习收敛速度慢、易陷入局部极值、网络结构难以确定等缺点,更为重要的是,SVM 构造回归函数的复杂程度取决于支持向量的个数,与特征空间的维数无关,能有效解决可能导致的"维数灾"问题。针对上述问题,本文采用试凑的方法,在延迟时间 $\tau$ 和嵌入维数 $m$ 取值为 2~10 范围内依次构建 81 个基于相空间重构理的 CV-SVM 月径流模拟模型,以南利河董湖站月径流时间序列为例进行分析,以确定最佳延迟时间 $\tau$ 和嵌入维数 $m$,并与自相关函数法等相关延迟时间 $\tau$ 和嵌入维数 $m$ 的确定方法确定结果进行比较,旨在探寻延迟时间 $\tau$ 和嵌入维数 $m$ 对月径流模拟精度的影响及其规律,为基于时间序列的水文模拟及预测预报提供方法和参考。

# 2　基于改进相空间重构原理的 SVM 模拟模型

## 2.1　改进的混沌时间序列的相空间重构

相空间重构的目的在于在高维相空间中恢复混沌吸引子。由于吸引子的内在行为具有不规则性及混沌吸引子具有复杂的几何结构,一般来说,不同的混沌实测数据应建立不同的混沌模型。混沌模拟是在相空间中进行的,其原理就是在相空间中找到一个非线性模型去逼近系统动态特性,实现一定时期内的预测或模拟[17]。

设时间序列 $x(t),t=1,2,\cdots,N$,嵌入维为 $m$,时间延迟为 $\tau$,则重构相空间为:

$$y(t)=\{x(t),x(t+\tau),x(t+2\tau),\cdots,x[t+m\tau]\}\quad t=1,2,\cdots,M \tag{1}$$

式中:$M=N-m\tau$ 为相空间中的相点数。根据 Takens 定理[5],选择合适的嵌入维和时间延迟重构相空间,则重构相空间中的轨迹与原系统是动力学等价的。那么存在一个光滑映射 $F:R^m\rightarrow R^m$,给出相空间轨迹的表达式为:

$$y(t+\eta)=F[y(t)]\quad t=1,2,\cdots N \tag{2}$$

式中:$\eta$ 为模拟步长,这里是对月径流进行模拟,因此 $\eta=0$。

## 2.2　支持向量机模拟原理

利用 SVM 解决回归拟合问题时,Vapnik 等人在 SVM 分类的基础上引入了 $\varepsilon$ 不敏感损失函数,得到回归支持向量机(support vector machine for regrenssion,SVR),在回归拟合中取得了很好的性能和效果。SVR 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。从 SVM 分类判别函数的形式上看,它类似于一个 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。SVM 神经网络结构如图 1 所示。

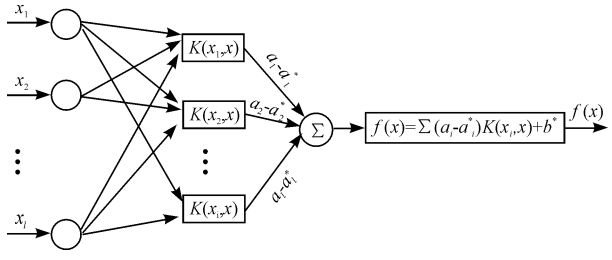


图 1　SVM 神经网络示意图

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

$$f(x)=w\Phi(x)+b \tag{3}$$

式中:$\Phi(x)$ 为非线性映射函数。

定义 $\varepsilon$ 线性不敏感损失函数为:

$$L(f(x),y,\varepsilon)=\begin{cases}0, & |y-f(x)|\leqslant\varepsilon\\ |f-f(x)|-\varepsilon, & |y-f(x)|>\varepsilon\end{cases} \tag{4}$$

式中:$f(x)$ 为回归函数返回的模拟值;$y$ 为对应真实值。

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

$$\begin{cases}\min \dfrac{1}{2}\| w \|^2 + C\sum\limits_{i=1}^{l}(\xi_i + \xi_i^*) \\ s.t. \begin{cases} y_i - w\Phi(x_i) - b \leqslant \varepsilon + \xi_i \\ -y_i + w\Phi(x_i) + b \leqslant \varepsilon + \xi_i, \\ \quad i = 1,2,\cdots,l \\ \xi_i \geqslant 0,\ \xi_i^* \geqslant 0 \end{cases} \end{cases} \tag{5}$$

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 Largrange 函数，并转换成对偶形式：

$$\begin{cases}\max\limits_{a,a^*}\Big[-\dfrac{1}{2}\sum\limits_{i=1}^{l}\sum\limits_{i=j}^{l}(a_i - a_i^*)(a_j - a_j^*)K(x_i,x_j) - \\ \qquad \sum\limits_{i=1}^{l}(a_i + a_i^*)\varepsilon + \sum\limits_{i=1}^{l}(a_i - a_i^*)y_i\Big] \\ s.t. \begin{cases} \sum\limits_{i=1}^{l}(a_i - a_i^*) = 0 \\ 0 \leqslant a_i \leqslant C \\ 0 \leqslant a_i^* \leqslant C \end{cases} \end{cases} \tag{6}$$

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

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：

$$w^* = \sum_{i=1}^{l}(a_i - a_l^*)\Phi(x_i) \tag{7}$$

$$b^* = \frac{1}{N_{nsv}}\Big\{\sum_{0<a_i<C}\Big[y_i - \sum_{x_i\in SV}(a_i - a_i^*)K(x_i,x_j) - \varepsilon\Big] + \sum_{0<a_i<C}\Big[y_i - \sum_{x_i\in SV}(a_i - a_i^*)K(x_i,x_j) + \varepsilon\Big]\Big\} \tag{8}$$

式中：$N_{nsv}$ 为支持向量机个数。其回归函数为：

$$\begin{aligned} f(x) &= w^*\Phi(x) + b^* \\ &= \sum_{i=1}^{l}(a_i - a_i^*)\Phi(x_i)\Phi(x) + b^* \\ &= \sum_{i=1}^{l}(a_i - a_i^*)K(x_i,x) + b^* \end{aligned} \tag{9}$$

式中：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。

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 $(K(x,x_i) = x^{\mathrm{T}}x_i)$、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 $(K(x,x_i) = \tanh(\gamma x^{\mathrm{T}}x_i + r)$。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。普遍认为，基于径向机函数的 SVM 模型有着较好的模拟效果，本文选择径向基核函数为 SVM 的核函数。

### 2.3 基于相空间重构的 SVM 模拟方法

基于相空间重构的 SVM 预测模型是将原始时间序列数据进行相空间重构得到一个高维的特征空间，获得预测模型所需要的输入向量和输出向量。设观测到的时间序列为 $\{x(t_i)i = 1,2,\cdots,N\}$，分别计算时延 $\tau$ 与嵌入维 $m$，对该时间序列进行相空间重构，根据 Takens 定理[6,17] 有：

$$X_{i+T} = f(x) \qquad i = 1,2,\cdots,M \tag{10}$$

式中：$X_i$ 为相空间的第 $i$ 个相点；$T$ 为前向预测步长；$M$ 为 $m$ 维相空间中的相点数。

一般情况，取 $T = 1$，即前向预测一步，则输入、输出向量 $X$ 和 $Y$ 分别为：

$$X = \begin{bmatrix} x(t_1),x(t_1+\tau), & \cdots, & x[t_1+(m-1)]\tau \\ x(t_2),x(t_2+\tau), & \cdots, & x[t_2+(m-1)]\tau \\ \vdots & & \vdots \\ x(t_{M'}),x(t_{M'}+\tau), & \cdots, & x[t_{M'}+(m-1)]\tau \end{bmatrix}$$

$$Y = \begin{bmatrix} x(t_2+m\tau) \\ x(t_3+m\tau) \\ \vdots \\ x(t_{M'}+m\tau) \end{bmatrix} \tag{11}$$

式中：$M'$ 为满足 $t_{M'} + mt = t_N$ 的整数，显然 $M < M$。

在重构相空间后，即可对 SVM 进行训练，得到 $t$ 时刻 SVM 的进一步预测模型为：

$$\hat{x}_{t+1} = \sum_{i=1}^{M'}(a_i - a_i^*)K(x_i,x) + b^* \tag{12}$$

为表述方便，令 $x_{i_{t+1}} = x(t_i)$，则对于相空间的第 $t+1$ 点，有

$$x_{i(t+1)} = (\hat{x}_{t+1},\cdots,x(t-(m-2)\tau)) \tag{13}$$

由(7) 式可得到对 $t+2$ 点的预测为：

$$\hat{x}_{t+2} = \sum_{i=1}^{M'}(a_i - a_i^*)K(x_i,x_{i_{t+1}}) + b^* \tag{14}$$

依此类推，第 $p$ 步的 SVM 预测模型为：

$$\hat{x}_{t+p} = \sum_{i=1}^{M'}(a_i - a_i^*)K(x_i,x_{i\ t+p+1}) + b^* \tag{15}$$

由于本文着重研究同期模拟，即 $T = 0$，故模型的输入、输出向量 $X$ 和 $Y$ 分别为：

$$X = \begin{bmatrix} x(t_1),x(t_1+\tau), & \cdots, & x(t_1+m\tau) \\ x(t_2),x(t_2+\tau), & \cdots, & x(t_2+m\tau) \\ \vdots & & \vdots \\ x(t_{M'}),x(t_{M'}+\tau), & \cdots, & x(t_{M'}+m\tau) \end{bmatrix}$$

$$Y = \begin{bmatrix} x(t_1 + m\tau) \\ x(t_2 + m\tau) \\ \vdots \\ x(t_{M'} + m\tau) \end{bmatrix} \tag{16}$$

# 3　实例应用

## 3.1　基本资料

董湖水文站位于南利河下游,设立于 1958 年 12 月,属国家基本站,观测项目有水位、流量、蒸发,系列均为1959 年1 月 -2005 年12 月。董湖站控制流域面积2 369 $km^2$,多年平均流量40.3 $m^3/s$,降水量1 191 mm。本文研究董湖水文站从 1959 年 1 月 -2005 年 12 月共47 年历月实测径流量,将资料按月展开,得到 564 个数据的月径流时间序列。

## 3.2　相空间滞时和嵌入维数的确定

本文采用自相关函数法、互信息法和 C - C 方法求时间延迟 $\tau$;采用 G - P 算法、CAO 算法和 C - C 方法求嵌入维数 $m$。计算结果见表 1。

**表 1　董湖水文站月径流相空间重构参数计算结果**

| 相空间重构参数 | 计算方法 | 结果 | 相空间重构参数 | 计算方法 | 结果 |
|---|---|---|---|---|---|
| 时间延迟 | 自相关函数法 | 2 | 嵌入维数 | G - P 法 | 4 |
| | 互信息法 | 10 | | CAO 法 | 4 |
| | C - C 法 | 4 | | C - C 法 | 3 |

## 3.3　径流时间序列演变混沌特性识别

目前关于混沌时间序列识别方法主要有关联维数法、Kolmogorov 熵法、Lyapunov 指数法、功率谱方法、庞卡莱截面法等众多方法,但尚无可以在混沌系统和随机系统之间做出准确判断的统一方法[17-18]。本文用 Lyapunov 指数法来判断董湖水文站月径流时间序列演变的混沌特性。确定 Lyapunov 指数的方法较多,主要有 Wolf 方法、Jacobian 法、p - 范数法和小数据量法等。本文利用小数据量法求得董湖站月径流序列的最大 Lyapunov 指数以 2 为底时等于 0.4898,以 e 为底时等于 0.3396,大于零,根据若最大 Lyapunov 指数大于零,则系统一定存在混沌特性的判别方法,董湖水文站月径流序列的演变具有混沌特性。因此可以用上述基于相空间重构理论的 CV - SVM 等模型对月径流时间序列进行模拟。

## 3.4　径流预测的实现

3.4.1　*数据处理*　数据处理的方法很多,本文采用以下方法将各指标数据无量纲化到[0.1,0.9]之间,有利于网络训练。公式如下:

$$\hat{x} = 0.1 + (0.9 - 0.1)[(x - x_{\min})/(x_{\max} - x_{\min})] \tag{17}$$

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

3.4.2　*训练及测试样本设计*　由上述可知,将董湖水文站 1959 年 1 月 -2005 年 12 月的实测资料按月展开,得到 564 个月径流时间序列数据,由于延迟时间 $\tau$ 和嵌入维数 $m$ 的取值范围均为 2 ~ 10,故对董湖水文站径流时间序列进行相空间重构,得到输入、输出样本的范围为 464 ~ 560。为能更客观地评价模型的泛化能力,本文选取前 400 个样本作为训练样本,余下的样本作为测试样本。

3.4.3　*SVM 设计*　SVM 用于处理模式分类或非线性映射问题时,在选定径向机核函数条件下,模型中的惩罚因子 $c$ 和核函数参数 $g$ 的选取对模型的预测精度有着关键性影响[16]。由于惩罚因子和核函数参数的选取目前尚无理论上的指导原则,最优参数的选取多凭经验、实验对比等进行搜寻,极大地制约了 SVM 模型精度[19-20]。目前普遍采用交叉验证法(Cross Validation,CV)来搜寻合理的参数。CV 是用来验证分类器性能一种统计分析方法,基本思想是把在某种意义下将原始数据(dataset)进行分组,一部分作为训练集(train set),另一部分作为验证集(validation set),首先用训练集对分类器进行训练,再利用验证集来测试训练得到的模型(model),以此来作为评价分类器的性能指标。CV 方法可以有效的避免模型"过学习"以及"欠学习"现象的发生。本文基于 MATLAB 环境,构建 81 个基于相空间重构理论的 CV - SVM 月径流模型模型,对董湖水文站月径流进行模拟。

3.4.4　*性能评价*　依据水文情报预报规范[21]等,选取平均相对误差 $e_{MRE}$、最大相对误差 $e_{\max RE}$、决定系数 $DC$、合格率 $QR$、平均绝对误差 $e_{MAE}$、均方根绝对误差 $e_{RMSE}$ 和均方根相对误差 $e_{RMAPE}$7 个统计学指标作为模型的评价指标。决定系数 $DC$ 范围在[0,1]内,愈接近 1,表明模型的性能越好;合格率 $QR$ 是以相对误差小于 20% 为合格标准,但考虑到模型模拟精度较高,故选用相对误差小于 10% 的测试样本为合格样本;其他评价指标越小,表明模型的性能越好。

评价指标计算公式如下:

$$e_{MRE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \tag{18}$$

$$e_{\max RE} = \max_{1\leqslant i\leqslant n} \frac{|\hat{y}_i - y_i|}{y_i} \tag{19}$$

$$DC = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{(\sum_{i=1}^{n} y_i - \bar{y})^2} \tag{20}$$

$$QR = \frac{k}{n} \times 100\% \tag{21}$$

$$e_{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{22}$$

$$e_{RMSE} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{23}$$

$$e_{RMAPE} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left[\frac{|\hat{y}_i - y_i|}{y_i}\right]^2}, \quad i = 1,2,\cdots,n \tag{24}$$

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；$\bar{y}$ 为实测值的均值；$k$ 为相对误差小于10% 的样本数；$n$ 为测试样本数，数量为 64 ~ 160 不等。

### 3.5 预测结果及分析

利用上述训练好的基于相空间重构理论的 CV - SVM 月径流模拟模型对董湖水测试样本进行模拟，模拟结果见表 2 ~ 8。

表 2 董湖站 SVM 月径流模拟模型平均相对误差 $e_{MRE}$ 效果评价表 %

| 时间延时 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.5748 | 0.5405 | 0.6753 | 0.7551 | 0.9476 | 1.2563 | 1.2347 | 1.5663 | 2.0760 |
| 3 | 0.3509 | 0.5424 | 0.6527 | 0.7460 | 1.0162 | 1.4176 | 2.0907 | 1.7317 | 2.1802 |
| 4 | 0.2589 | 0.2906 | 0.3965 | 0.6854 | 0.7851 | 0.9968 | 1.2114 | 1.5422 | 1.5899 |
| 5 | 0.3261 | 0.4415 | 0.7076 | 0.8185 | 0.9687 | 1.3086 | 1.4360 | 1.6382 | 1.8655 |
| 6 | 0.2302 | 0.4278 | 0.7546 | 0.8352 | 1.0162 | 1.1154 | 0.9573 | 0.9485 | 1.0040 |
| 7 | 0.3379 | 0.3707 | 0.5695 | 0.7059 | 0.8737 | 1.1528 | 1.2000 | 1.2974 | 1.4688 |
| 8 | 0.2804 | 0.3719 | 0.5452 | 0.5795 | 0.7517 | 0.7932 | 0.8299 | 1.0373 | 1.1824 |
| 9 | 0.2865 | 0.5324 | 0.6406 | 0.5800 | 0.9846 | 0.8962 | 1.0701 | 1.1450 | 1.1419 |
| 10 | 0.3764 | 0.4228 | 0.5628 | 0.6976 | 0.8314 | 0.8532 | 1.0578 | 1.0568 | 1.5674 |

注：填充部分平均相对误差≤0.5%。

表 3 董湖站 SVM 月径流模拟模型最大相对误差 $e_{maxRE}$ 效果评价表 %

| 时间延时 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 14.3371 | 16.8240 | 14.6578 | 12.3397 | 13.7809 | 21.4039 | 17.5902 | 21.3579 | 26.7563 |
| 3 | 8.8517 | 10.8863 | 13.0823 | 16.5604 | 23.7672 | 32.5980 | 82.5890 | 62.8242 | 64.7760 |
| 4 | 2.8413 | 5.7599 | 6.3591 | 17.8353 | 29.8049 | 30.1064 | 25.1297 | 28.3070 | 23.8808 |
| 5 | 7.1838 | 8.6761 | 31.0607 | 23.2092 | 27.4832 | 30.0243 | 32.3809 | 36.0799 | 38.9200 |
| 6 | 7.6650 | 4.8476 | 21.2703 | 16.6011 | 21.8017 | 22.3381 | 24.1035 | 25.7803 | 19.7727 |
| 7 | 4.2211 | 5.1288 | 12.7201 | 12.3195 | 15.0956 | 19.8458 | 19.9289 | 18.9334 | 20.8037 |
| 8 | 6.5986 | 8.3617 | 12.4309 | 15.6071 | 16.5084 | 17.5287 | 14.7566 | 25.3617 | 30.1588 |
| 9 | 8.9121 | 19.6104 | 22.2119 | 16.2257 | 40.2169 | 37.3607 | 34.9063 | 39.0488 | 45.2110 |
| 10 | 2.7589 | 7.0353 | 17.7115 | 18.6107 | 16.2653 | 18.2874 | 30.4496 | 37.8181 | 58.5225 |

注：填充部分平均相对误差≤10%。

表 4 董湖站 SVM 月径流模拟模型决定系数 DC 效果评价表

| 时间延时 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.9994 | 0.9992 | 0.9995 | 0.9995 | 0.9995 | 0.9985 | 0.9992 | 0.9989 | 0.9987 |
| 3 | 0.9999 | 0.9999 | 0.9999 | 0.9993 | 0.9986 | 0.9990 | 0.9978 | 0.9979 | 0.9981 |
| 4 | 1.0000 | 1.0000 | 0.9996 | 0.9995 | 0.9994 | 0.9984 | 0.9985 | 0.9985 | 0.9978 |
| 5 | 1.0000 | 0.9999 | 0.9998 | 0.9997 | 0.9997 | 0.9996 | 0.9994 | 0.9993 | 0.9991 |
| 6 | 1.0000 | 0.9995 | 0.9994 | 0.9990 | 0.9990 | 0.9984 | 0.9987 | 0.9987 | 0.9985 |
| 7 | 1.0000 | 0.9999 | 0.9999 | 0.9998 | 0.9995 | 0.9993 | 0.9959 | 0.9960 | 0.9944 |
| 8 | 1.0000 | 1.0000 | 0.9996 | 0.9995 | 0.9995 | 0.9991 | 0.9991 | 0.9990 | 0.9955 |
| 9 | 1.0000 | 0.9999 | 0.9999 | 0.9996 | 0.9993 | 0.9992 | 0.9989 | 0.9979 | 0.9974 |
| 10 | 0.9999 | 0.9998 | 0.9998 | 0.9995 | 0.9995 | 0.9995 | 0.9988 | 0.9988 | 0.9983 |

注：填充部分相对误差≥0.9998。

表 5　董湖站 SVM 月径流模拟模型合格率 QR 效果评价表　　%

| 时间延时 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 99.4 | 99.4 | 98.7 | 99.4 | 99.3 | 97.3 | 98.6 | 96.6 | 92.4 |
| 3 | 100.0 | 99.4 | 98.7 | 98.0 | 96.6 | 94.4 | 94.3 | 94.9 | 94.0 |
| 4 | 100.0 | 100.0 | 100.0 | 98.6 | 97.9 | 97.8 | 96.2 | 93.0 | 92.7 |
| 5 | 100.0 | 100.0 | 98.6 | 97.1 | 97.0 | 95.3 | 95.2 | 94.1 | 94.7 |
| 6 | 100.0 | 100.0 | 98.6 | 98.5 | 97.7 | 97.5 | 98.3 | 98.2 | 97.1 |
| 7 | 100.0 | 100.0 | 98.5 | 97.7 | 96.7 | 95.7 | 94.4 | 94.1 | 91.5 |
| 8 | 100.0 | 100.0 | 99.2 | 99.2 | 98.3 | 97.2 | 96.0 | 96.7 | 91.7 |
| 9 | 100.0 | 99.3 | 98.4 | 98.3 | 95.5 | 96.0 | 93.5 | 90.4 | 91.9 |
| 10 | 100.0 | 100.0 | 98.4 | 97.4 | 98.1 | 94.7 | 91.7 | 94.6 | 87.5 |

注:填充部分相对误差≥0.99%。

表 6　董湖站 SVM 月径流模拟模型平均绝对误差 $e_{MAE}$ 效果评价表　　%

| 时间延时 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.2295 | 0.2527 | 0.2460 | 0.2641 | 0.3210 | 0.6083 | 0.6250 | 0.7212 | 0.8384 |
| 3 | 0.1289 | 0.1447 | 0.1939 | 0.3160 | 0.4366 | 0.4729 | 0.7076 | 0.7945 | 0.8939 |
| 4 | 0.0954 | 0.1032 | 0.2082 | 0.2876 | 0.3198 | 0.5676 | 0.6347 | 0.7180 | 0.8622 |
| 5 | 0.1029 | 0.1853 | 0.2473 | 0.2918 | 0.3415 | 0.4119 | 0.5897 | 0.6628 | 0.8253 |
| 6 | 0.0590 | 0.2106 | 0.2929 | 0.4498 | 0.5092 | 0.6678 | 0.6400 | 0.6258 | 0.7062 |
| 7 | 0.1010 | 0.1474 | 0.2153 | 0.2841 | 0.4388 | 0.5757 | 0.8342 | 0.8301 | 1.0320 |
| 8 | 0.0936 | 0.1328 | 0.3169 | 0.2823 | 0.3742 | 0.5005 | 0.5176 | 0.6161 | 0.8953 |
| 9 | 0.0995 | 0.1667 | 0.2013 | 0.2933 | 0.4522 | 0.4630 | 0.5987 | 0.7609 | 0.8866 |
| 10 | 0.1546 | 0.2022 | 0.2292 | 0.3127 | 0.3842 | 0.4124 | 0.6003 | 0.7128 | 0.9705 |

注:填充部分相对误差≤0.3。

表 7　董湖站 SVM 月径流模拟模型均方根绝对误差 $e_{RMSE}$ 效果评价表　　%

| 时间延时 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.0814 | 0.0984 | 0.0752 | 0.0777 | 0.0760 | 0.1388 | 0.0999 | 0.1164 | 0.1282 |
| 3 | 0.0319 | 0.0275 | 0.0369 | 0.0923 | 0.1311 | 0.1135 | 0.1724 | 0.1713 | 0.1615 |
| 4 | 0.0220 | 0.0202 | 0.0684 | 0.0827 | 0.0922 | 0.1490 | 0.1437 | 0.1493 | 0.1798 |
| 5 | 0.0230 | 0.0409 | 0.0536 | 0.0668 | 0.0638 | 0.0729 | 0.0961 | 0.1042 | 0.1262 |
| 6 | 0.0104 | 0.0787 | 0.0937 | 0.1157 | 0.1210 | 0.1595 | 0.1476 | 0.1504 | 0.1692 |
| 7 | 0.0167 | 0.0259 | 0.0396 | 0.0519 | 0.0865 | 0.1128 | 0.2662 | 0.2666 | 0.3206 |
| 8 | 0.0162 | 0.0231 | 0.0791 | 0.0850 | 0.0940 | 0.1278 | 0.1224 | 0.1368 | 0.2849 |
| 9 | 0.0212 | 0.0385 | 0.0418 | 0.0801 | 0.1075 | 0.1209 | 0.1430 | 0.1987 | 0.2347 |
| 10 | 0.0379 | 0.0554 | 0.0600 | 0.0891 | 0.0950 | 0.0925 | 0.1456 | 0.1596 | 0.2103 |

注:填充部分相对误差≤0.1。

表 8　董湖站 SVM 月径流模拟模型均方根相对误差 $e_{RMAPE}$ 效果评价表　　%

| 时间延时 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.0013 | 0.0014 | 0.0016 | 0.0016 | 0.0020 | 0.0027 | 0.0024 | 0.0032 | 0.0045 |
| 3 | 0.0008 | 0.0013 | 0.0016 | 0.0019 | 0.0029 | 0.0038 | 0.0075 | 0.0057 | 0.0064 |
| 4 | 0.0004 | 0.0006 | 0.0008 | 0.0019 | 0.0026 | 0.0030 | 0.0032 | 0.0042 | 0.0043 |
| 5 | 0.0008 | 0.0010 | 0.0025 | 0.0025 | 0.0030 | 0.0039 | 0.0043 | 0.0050 | 0.0055 |
| 6 | 0.0006 | 0.0009 | 0.0024 | 0.0023 | 0.0030 | 0.0034 | 0.0030 | 0.0033 | 0.0036 |
| 7 | 0.0006 | 0.0008 | 0.0016 | 0.0021 | 0.0026 | 0.0036 | 0.0038 | 0.0045 | 0.0051 |
| 8 | 0.0007 | 0.0010 | 0.0015 | 0.0021 | 0.0024 | 0.0028 | 0.0030 | 0.0043 | 0.0057 |
| 9 | 0.0008 | 0.0019 | 0.0024 | 0.0022 | 0.0048 | 0.0048 | 0.0056 | 0.0067 | 0.0081 |
| 10 | 0.0007 | 0.0010 | 0.0021 | 0.0028 | 0.0032 | 0.0037 | 0.0056 | 0.0069 | 0.0125 |

注:填充部分相对误差≤0.002。

分析表 1 ~ 表 8 可以得出以下结论:

(1)从表 2 ~ 表 8 可以看出,填充部分多集中在表格左侧,即嵌入维数 $m \leqslant 5$ 时模型具有较好的模拟精度和泛化能力,表明 CV - SVM 模型的模拟精度对嵌入维数较为敏感,即随着嵌入维数 $m$ 的增大,模型的模拟精度呈降低趋势;而对时间延迟 $\tau$ 敏感性较差,即随着时间延迟 $\tau$ 的增大,模型的模拟精度变化趋势不明显。

(2) 可用于模型测试样本数随着时间延迟 $\tau$ 和嵌入维数 $m$ 的增大而急剧减少,减少数为 $\tau \times m$,而由于实测时间序列数据长度有限,因此在实际应用中,应尽可能选取既满足模拟精度,$\tau \times m$ 又小的时间延迟 $\tau$ 和嵌入维数 $m$。

(3) 本实例中,合适的嵌入维数 $m$ 取值范围为 2 ~ 5;而由于 CV - SVM 模型的模拟精度对时间延迟 $\tau$ 敏感性较差,考虑到实测时间序列数据长度有限等因素,合适的时间延迟 $\tau$ 取值范围为 3 ~ 5。本实例中,最佳 CV - SVM 模拟模型的时间延迟 $\tau$ = 4,嵌入维数 $m$ = 2,此时各模型性能评价指标均处于最优。

(4)比较表 1 可以发现,自相关函数法等几种方法均很难确定最佳 CV - SVM 模型的延迟时间 $\tau$ 和嵌入维数 $m$,不过与 C - C 算法确定的延迟时间 $\tau$ 和嵌入维数 $m$ 最为接近。

## 4 结 语

(1)基于改进的相空间重构理论构造月径流模拟模型的输入、输出向量,利用 SVM 收敛速度快、全局最优、避免"维数灾"等优点,依次构建 81 个 CV - SVM 月径流模拟模型,以董湖站月径流进行实例分析,结果表明,CV - SVM 模型的模拟精度随着嵌入维数 $m$ 的增大呈降低趋势,而随着时间延迟 $\tau$ 的增大其模拟精度变化趋势不明显。

(2)从表 2 ~ 8 可以看出,CV - SVM 模型用于月径流模拟有着较高的模拟精度和泛化能力,用于水文径流模拟是合理可行的。

**参考文献:**


[1] 谷晓平,王长耀,王 汶,等. 应用于水文预报的优化 BP 神经网络研究[J]. 生态环境,2004,13(4): 524 - 527.

[2] 庞 博,郭生练,熊立华,等. 改进的人工神经网络水文预报模型及应用[J]. 武汉大学学报(工学版),2007,40(1):33 - 36 + 41.

[3] 蓝永超,康尔泗,徐中民,等. BP 神经网络在径流长期预测中的应用[J]. 中国沙漠,2001,21(1):97 - 100.

[4] 尹 晔,梁 川. 改进的 BP 网络模型及其在日径流预测中的应用[J]. 云南水力发电,2005,21(3):14 - 17.

[5] 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.

[6] 于国荣,夏自强. 混沌时间序列支持向量机模型及其在径流预测应用[J]. 水科学进展,2008,19(1):116 - 122.

[7] 陈南祥,黄 强,曹连海,等. 径流序列的相空间重构神经网络预测模型[J]. 河海大学学报(自然科学版),2005,33(5):490 - 493.

[8] 战国隆,马孝义,刘继龙,等. 泾河月径流量的混沌特征识别及预测[J]. 水利水电科技进展,2010,30(1):7 - 9(增刊).

[9] Kantz H, Schreiber T. Nonlinear time series analysis[M]. Cambridge: Cambridge University Press, 1997.

[10] Fraser A M, Swinney H L. Independent coordinates for strange attractors from time series[J]. Physical Review A: 1986, 33: 1134 - 1140.

[11] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors[J]. Physica D: Nonliner Phenomena, 1983, 9: 189 - 208.

[12] Kennel M B, Brown R, Abarbanel H D I. Determining embedding dimension for phase - space reconstruction using a geometrical construction[J]. Physical Review A., 1992, 45: 3403 - 3411.

[13] Kugiurmtzis D. State space reconstruction parameters in the analysis of chaotic times series - the role of the time window length[J]. Physica D: Nonlinear Phenomena, 1996, 95: 13 - 28.

[14] Kim H S, Eykholt R, Salas J D. Nonlinear dynamics, delay times and embedding windows[J]. Physica D: Nonlinear Phenomena, 1999, 127: 48 - 60.

[15] Vladimir N Vapnik. 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,2006.7.

[17] 邵东国,刘丙军,阳书敏,等. 水资源繁杂系统理论[M]. 北京:科学出版社,2012.6.

[18] 吴学文. 考虑生态的多目标水电站水库混沌优化调度研究[M]. 北京:中国水利水电出版社,2012.12.

[19] 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,2010.4.

[20] 史 峰, 王 辉,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,2011.7.

[21] 中华人民共和国水利部. SL250 - 2000,水文情报预报规范[S]. 北京:中国水利水电出版社,2000.