

基于相空间重构的支持向量机方法在径流中长期预报中应用

刘 冀, 王本德*, 袁晶瑄, 周惠成

(大连理工大学 土木水利学院, 辽宁 大连 116024)

摘要: 水文中长期预报对于水资源规划管理、水库及水电站调度具有十分重要的意义. 针对常规混沌预测方法的局限性, 提出基于相空间重构的支持向量机(SVM)预报方法. 该方法首先对径流时间序列进行混沌辨识, 然后对其进行相空间重构, 采用基于结构风险最小化的SVM进行径流预报. 对于SVM的参数优选问题, 以径向基核函数作为核函数, 采用混沌变尺度优化方法进行参数寻优. 实例表明, 该方法优于SVM和人工神经网络(ANN)预报方法, 且具有良好的泛化推广能力.

关键词: 径流中长期预报; 相空间重构; 支持向量机; 混沌优化; 人工神经网络
中图分类号: P338 **文献标志码:** A

0 引言

成功的中长期水文预报对防汛抗旱、水资源规划管理以及水库等水利工程的综合利用有着十分重要的意义. 径流时间序列具有高度的非线性特征, 因此依据传统的线性方法难以描述其变化规律, 而混沌理论的兴起为此领域的研究开辟了一条崭新的道路. 就水文预报领域而言, 国内外已有一些相关的研究成果^[1~5]. 现有的混沌预测方法主要包括全域法、局域法、最大 Lyapunov 指数预测法、ANN 方法及在此基础上的各种改进算法等, 预测效果多以 ANN 方法较为理想. 但 ANN 结构的确定没有统一的准则, 容易出现过拟合现象, 模型泛化能力不佳等问题限制了其在径流预报领域中的应用. 支持向量机(support vector machine, SVM)是针对结构风险最小化原理提出的, 与 ANN 相比, 具有小样本学习、泛化能力强等特点, 能有效地避免过学习、局部最小点及“维数灾”等问题^[6]. 为此, 本文将 SVM 引入混沌预测方法进行径流中长期预报, 并与 ANN 方法进行比较. 针对 SVM 最优参数的辨识问题, 本文采用混沌优化算法进行参数寻优, 此方法是一种全局搜索算法, 且无需考虑模型的复杂度和变量维数.

1 混沌时间序列相空间重构方法

系统任一分量的演化是由与之相互作用的其他分量所决定的, 因此, 这些相关分量的信息就隐含在任一分量的发展过程中, 这样就可以从某一分量的一批时间序列数据中提取和恢复出系统原来的规律^[7]. 就一个实际的动态系统而言, 一个普遍存在的问题就是缺少关于系统内所有变量的信息, 相空间重构(phase-space reconstruction, PSR)方法为描述系统的动态特征提供了一条途径, 它将一维时间序列通过重构嵌入到多维相空间中, 按照 Takens 定理可以在拓扑等价意义下恢复吸引子的动力学特性. 对于时间序列 $x_i (i = 1, 2, \dots, N)$, 通过选定合适的嵌入维数(m)和时间延迟(τ), 便可重构相空间:

$$y_j = (x_j, x_{j+\tau}, x_{j+2\tau}, \dots, x_{j+(m-1)\tau}); \\ j = 1, 2, \dots, N - (m-1)\tau \quad (1)$$

式中: N 为时间序列长度. 通过重构相空间后便可采用局域法、ANN 等方法进行一步或多步预测. 对于 τ 的选取通常有自相关函数法、复相关函数法、互信息法等, 互信息法因包含了时间序列的非线性特征且计算结果明显优于自相关法, 因此本文采用互信息法计算 τ . 对于嵌入维数 m 的计算方法有 G-P 算法、伪最邻近点法、关联积分法

和 Cao 方法^[8]. Cao 方法不包括除延迟时间 τ 以外的任何主观参数,而且不过于依赖数据点的个数,能够有效地区分数据是随机的还是确定的,故本文采用 Cao 方法确定 m . 现将此方法做一简要介绍.

定义

$$a(i,m)=\frac{\|y_i(m+1)-y_{n(i,m)}(m+1)\|}{\|y_i(m)-y_{n(i,m)}(m)\|};$$
$$i=1,2,\cdots,N-m\tau \tag{2}$$
$$\|y_k(d)-y_l(d)\|=\max_{0\leq j\leq d-1}|x_{k+j\tau}-x_{l+j\tau}| \tag{3}$$

式中: $\|\cdot\|$ 为欧氏距离; $y_i(m+1)$ 为嵌入维数为 $m+1$ 的相空间中第 i 个向量; $y_{n(i,m)}$ 为在 m 维相空间中与 $y_i(m)$ 在 $\|\cdot\|$ 意义下的最邻近相点.

定义

$$E(m)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}a(i,m) \tag{4}$$
$$E_1(m)=E(m+1)/E(m) \tag{5}$$

$E(m)$ 独立于嵌入维数 m 和延迟 τ . 若时间序列来源于吸引子,当 m 大于某一个值 m_0 时, $E_1(m)$ 便达到饱和,便可选取 m_0+1 作为最小嵌入维数.

令

$$E^*(m)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}|x_{i+m\tau}-x_{n(i,m)+m\tau}| \tag{6}$$

定义

$$E_2(m)=E^*(m+1)/E^*(m) \tag{7}$$

对于随机时间序列, $E_2(m)$ 恒为 1,而对于确定性的时间序列,它将依 m 而变化,即存在不为 1 的值.

2 回归 SVM^[9] 及其参数辨识方法

2.1 回归 SVM

回归 SVM 遵循结构风险最小化(structural risk minimization, SRM)原理,而非传统的经验风险最小,求解的是一个二次寻优问题,从理论上分析得到的将是全局最优解;算法通过非线性变换转换到高维特征空间,在高维空间中构造线性判别函数来实现原空间中的非线性判别函数,从而巧妙地解决了维数问题,其算法复杂度与样本维数无关. 对于给定的样本数据集 $X=\{(x_i,y_i)|i=1,2,\cdots,k\}$ (x_i 为输入, y_i 为输出),用非线性

映射 ϕ 把数据集 X 映射到高维特征空间,在此空间中进行线性回归,回归函数形式为 $f(x)=\langle w,\phi(x_i)\rangle+b$. 根据 SRM 原理要寻求最优回归超平面使损失函数 Q 最小化,即

$$\min Q=\frac{1}{2}\|w\|^2+c\sum_{i=1}^N(\xi_i+\xi_i^*)$$
$$\text{s.t.}\begin{cases} y_i-\langle w,\phi(x_i)\rangle-b\leq\epsilon+\xi_i \\ \langle w,\phi(x_i)\rangle+b-y_i\leq\epsilon+\xi_i^* \\ \xi_i,\xi_i^*\geq 0 \end{cases} \tag{8}$$

式中: w 为权向量; c 为调节训练误差和模型复杂度之间的折中正则化常数; ξ_i,ξ_i^* 为松弛变量; ϵ 为不灵敏参数,用于控制回归逼近误差宽度以及支持向量的个数和泛化能力,其值越小,精度越高,支持向量数越多,但泛化能力减弱; b 为偏置.

式(8) 可通过求解其对偶问题获得解,即

$$\max J=-\frac{1}{2}\sum_{i,j=1}^N(\alpha_i-\alpha_i^*)(\alpha_j^*-\alpha_j)\times$$
$$(\phi(x_i),\phi(x_j))+\sum_{i=1}^N\alpha_i^*(y_i-\epsilon)-$$
$$\sum_{i=1}^N\alpha_i(y_i+\epsilon) \tag{9}$$
$$\text{s.t.}\begin{cases} \sum_{i=1}^N\alpha_i^*=\sum_{i=1}^N\alpha_i \\ 0\leq\alpha_i\leq c \\ 0\leq\alpha_i^*\leq c \end{cases}$$

式中: α_i,α_i^* 为 Lagrange 乘子. 求解式(9) 可得 $w=\sum_{i=1}^N(\alpha_i-\alpha_i^*)\phi(x_i)$. 系数 $(\alpha_i-\alpha_i^*)$ 不为 0 所对应的样本 (x_i,y_i) 即为支持向量, b 可由任一支持向量代入求得. 由此可得回归函数为

$$f(x)=\sum_{i=1}^N(\alpha_i-\alpha_i^*)\langle\phi(x_i),\phi(x)\rangle+b \tag{10}$$

定义高维特征变换空间的内积运算为 SVM 的核函数: $K(x_i,x)=\langle\phi(x_i),\phi(x)\rangle$,常用的核函数有线性核函数、多项式核函数、径向基核函数等.

2.2 SVM 参数辨识方法

采用 SVM 进行回归分析时需要确定的参数有 c 和 ϵ 以及所选取的核函数参数. 以往的参数辨识算法结构通常较复杂,算法参数的取值需经多次试算确定,主观性较大. 而混沌变尺度优化算法^[10]简单易行,编程方便,利用混沌的遍历性和随机性,在解空间内不断缩小优化变量的搜索空间直至找到全局最优解,文献^[11]证明混沌优化算法的

寻优效率优于模拟退火算法等随机算法。

3 计算流程

采用本文方法进行径流预报的步骤如下：

- (1) 采用互信息法计算时间延迟 τ ，采用 Cao 方法计算嵌入维数 m ；
- (2) 以 m 和 τ 对径流时间序列进行相空间重构；
- (3) 以每个相点作为 SVM 的输入，以预报点的实际值为理想输出从而组成训练样本，采用混沌变尺度优化算法进行 SVM 最优参数辨识；
- (4) 以优选的参数值作为 SVM 的参数取值，采用相空间中除待预报点所在相点(y_f)以外的所有相点($y_i, i \neq f$)对 SVM 进行训练；
- (5) 以 y_f 的前一个相点 y_{f-1} 作为 SVM 的输入，采用训练好的 SVM 进行计算便得到预报值。

4 应用实例

采用文献[12]中的某水库 30 a(1974~2003 年)的月径流实测资料为研究对象,该水库多年平均月径流量为 1 230 m³/s,变差系数 0.717. 选用 1974~1998 年的月径流数据作为训练样本,剩余 5 a 数据作为检验样本. 首先进行混沌识别,采用互信息法计算延迟时间,当 $\tau=3$ 时,曲线第一次达到极小点,见图 1. 采用 Cao 方法计算嵌入维数,当 $m=11$ 时曲线基本保持不变,见图 2,因此采用 $m+1=12$ 作为嵌入维数. 同时计算出 E_2 不恒为 1,表明月径流时间序列来自于确定性系统,而非随机性的.

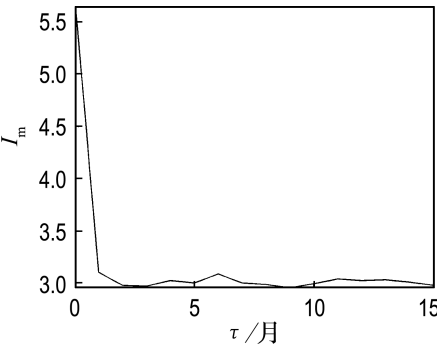


图 1 互信息法确定延迟时间

Fig. 1 Delay time determination using mutual information method

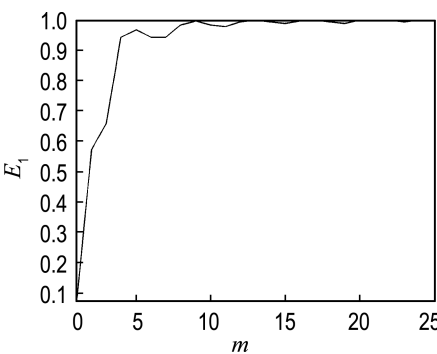


图 2 Cao 方法确定嵌入维数

Fig. 2 Embedding dimension determination using Cao's method

采用相空间重构方式(以下简称为方式 1),取 $m=12, \tau=3$ 对月流量时间序列进行重构. 以重构后的每个相点作为 SVM 的输入,以月径流实测值为期望输出,进行 SVM 的参数识别. 首先对数据进行零均值和归一化处理,将时间序列转化为(0,1),预报时再将其转化到解区间. SVM 的核函数选取径向基核函数,其形式如下:

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{11}$$

此核函数仅有一个参数 σ ,因此,所需优选的参数有(c, ϵ, σ). 采用混沌变尺度优化方法进行参数辨识,目标函数取预报值与实际值的均方误差最小,优化后的参数值(c, ϵ, σ)分别为(6.347, 0.003, 1.326). 为便于方法比较,同时采用没有进行相空间重构的方式(以下简称为方式 2)对月流量序列进行预报,即相当于取 $m=12, \tau=1$,在此方式下的 SVM 优化参数值(c, ϵ, σ)为(6.356, 0.001 6, 1.298). 针对两种预报方式,分别采用 SVM 对训练和检验样本进行预报,同时采用 ANN 方法进行预报,采用三层网络结构,为避免 ANN 过度训练,采用贝叶斯正则化方法训练网络,通过试算确定当隐层节点数为 4 时训练精度相对较高,因此选取各层神经元数分别为(12, 4, 1). 取均方根误差和相关系数作为预报结果的评价指标,为便于比较,将两种方式的预报结果一并列如表 1.

由表 1 可见,采用方式 1 对训练样本和检验样本进行预报,无论采用 SVM 或 ANN 进行预报,均方根误差和相关系数均优于方式 2 的相应结果. 此外,对高流量(本文取大于径流序列均值

表 1 SVM 与 ANN 流量预报结果比较

Tab. 1 Comparison of streamflow prediction using SVM and ANN

预报方法	训练				检验			
	均方根误差/	均方根误差*/	相关系数	相关系数*	均方根误差/	均方根误差*/	相关系数	相关系数*
	($\text{m}^3 \cdot \text{s}^{-1}$)	($\text{m}^3 \cdot \text{s}^{-1}$)			($\text{m}^3 \cdot \text{s}^{-1}$)	($\text{m}^3 \cdot \text{s}^{-1}$)		
SVM	352.05	343.28	0.919	0.922	353.05	344.58	0.942	0.946
ANN	356.59	354.64	0.913	0.916	360.54	338.58	0.933	0.944

注:表中*项表示采用相空间重构方式(方式1)的计算结果

的流量为高流量)的预报精度进行分析,当采用 SVM 进行预报时,方式 1 在训练及检验期的均方根误差分别为 516.5 和 $497.2 \text{ m}^3 \cdot \text{s}^{-1}$,方式 2 的相应值为 526.6 和 $510.4 \text{ m}^3 \cdot \text{s}^{-1}$,表明采用方式 1 进行预报优于方式 2,从而验证了本方法的有效性和可行性.究其原因为通过对径流时间序列进行相空间重构,在高维空间中恢复出径流动

态系统原来的轨迹,而此轨迹经过一定时间的变化后,最终会做出有规律的运动,从而便于预报;其次,无论是对训练还是检验样本的预报结果(方式 1 中检验样本的均方根除外),SVM 方法均优于 ANN 方法.基于相空间重构的 SVM 的月径流预报结果见图 3、4(其中 Q 为流量, Q_a 为实际流量, Q_p 为 SVM 预报流量).

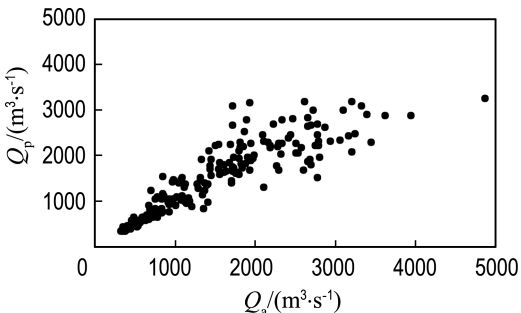
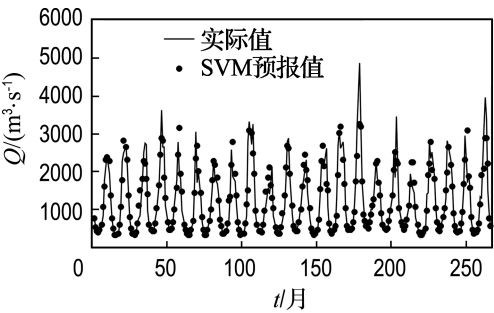


图 3 训练期 SVM 流量预报结果

Fig. 3 Streamflow hydrograph prediction of SVM (calibration period)

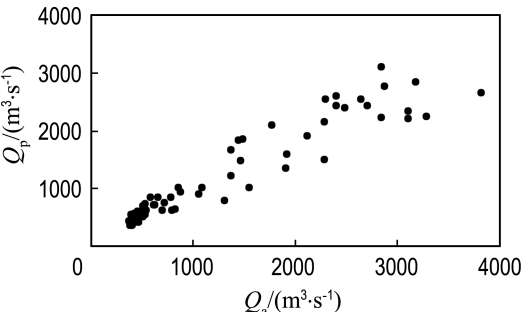
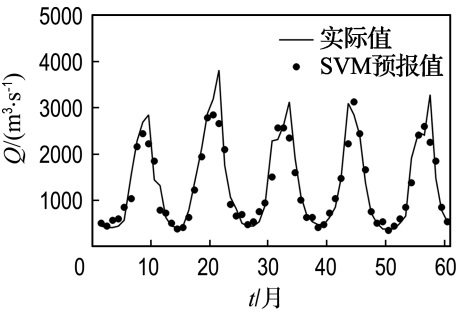


图 4 检验期 SVM 流量预报结果

Fig. 4 Streamflow hydrograph prediction of SVM (validation period)

5 结 语

混沌理论将水文时间序列嵌入到重构的相空间中以揭示水文动力系统的复杂运动特征,为探索水文现象的复杂变化规律开辟了一条新途径.本文引入混沌理论对径流时间序列进行相空间重构,采用 SVM 对水文时间序列进行预报.针对

SVM 的参数辨识问题,采用混沌变尺度优化算法,该算法易于搜索到全局最优解,且较其他化优算法更简单实用.实例应用表明,采用基于相空间重构的 SVM 进行径流预报,较单独采用 SVM 或 ANN 方法的预报精度有所提高,且达到可利用程度,验证了本方法的有效性.进一步研究 SVM 参数的敏感性以及如何提高对于极值的预报精度,

具有重要的现实意义。

参考文献：

[1] SIVAKUMAR B. Chaos theory in hydrology: important issues and interpretations [J]. **Journal of Hydrology**, 2000, **227**(1):1-20

[2] SIVAKUMAR B, JAYAWARDENA A W, FERNANDO T M K G. River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches [J]. **Journal of Hydrology**, 2002, **265**(1-4):225-245

[3] SIVAKUMAR B. Forecasting monthly streamflow dynamics in the western United States: a nonlinear dynamical approach [J]. **Environmental Modeling & Software**, 2003, **18**(8-9):721-728

[4] KHAN S, GANGULY A R, SAIGAL S. Detection and predictive modeling of chaos in finite hydrological time series [J]. **Nonlinear Processes in Geophysics**, 2005(12):41-53

[5] 刘少华, 毛红梅. 一种水文时间序列预报的新方法 [J]. **武汉大学学报(工学版)**, 2002, **35**(4):52-55

[6] VAPNIK V N. An overview of statistical learning theory [J]. **IEEE Transactions on Neural Networks**, 1999, **10**(5):988-999

[7] 吕金虎, 陆君安, 陈士华. 混沌时间序列分析及其应用 [M]. 武汉: 武汉大学出版社, 2002

[8] CAO Liang-yue. Practical method for determining the minimum embedding dimension of a scalar time series [J]. **Physica D**, 1997, **110**(1-2):43-50

[9] SMOLA A J, SCHOLKOPF B. A tutorial on support vector regression [J]. **Statistics and Computing**, 2004, **14**(3):192-222

[10] 张 彤, 王宏伟, 王子才. 变尺度混沌优化方法及其应用 [J]. **控制与决策**, 1999, **14**(3):285-288

[11] 李 兵, 蒋慰孙. 混沌优化方法及其应用 [J]. **控制理论及应用**, 1997, **14**(4):613-615

[12] 林剑艺, 程春田. 支持向量机在中长期径流预报中的应用 [J]. **水利学报**, 2006, **37**(6):681-686

Application of support vector machine based on phase-space reconstruction to medium-term and long-term runoff forecast

LIU Ji, WANG Ben-de*, YUAN Jing-xuan, ZHOU Hui-cheng

(School of Civil and Hydraulic Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: Effective medium-term and long-term runoff forecast is of great significance to the successful water resources planning and management, reservoirs and hydropower stations operation. The combined method, called support vector machine (SVM) based on the phase-space reconstruction, is developed according to the limitation of the past chaotic forecasting methods. Firstly, the existence of chaos in runoff time series is determined. Secondly, phase-space reconstruction of the runoff series is conducted and SVM, based on a principle that aims at minimizing the structural risk, is employed for forecasting, and the radial basis kernel is used as kernel function. Mutative scale chaos optimization algorithm is employed to search optimal parameters of the SVM. Comparison of the proposed method and artificial neural network indicates that the former is superior to the latter both in the forecasting accuracy and generalization ability.

Key words: medium-term and long-term runoff forecast; phase-space reconstruction; support vector machine; chaos optimization; artificial neural network