

基于分形和支持向量机矿井涌水量的预测

黄存捍¹ 冯涛² 王卫军² 刘辉²

(1. 中南大学 资源与安全工程学院, 湖南 长沙 410083; 2. 湖南科技大学 能源与安全工程学院, 湖南 湘潭 411201)

摘要:针对矿井涌水量预测问题,提出一种新的非线性预测方法。首先利用分形理论对矿井涌水量的时间序列进行相空间重构,应用自相关系数法确定最小嵌入维数,并以最小嵌入维数作为支持向量机的输入节点,根据支持向量机原理建立矿井涌水量的预测模型。将河南鹤壁四矿 1982—1997 年的矿井涌水量作为时间序列的训练样本,在 Matlab 环境下,利用所建立的预测模型预测不同嵌入维数时 2000 和 2001 年的矿井涌水量。结果表明:与其他维数相比,当嵌入维数为 4 时,井筒涌水量的预测值误差最小,预测精度最高。为检验该方法预测的可靠性,分别将不同维数下井筒、巷道和工作面涌水量 1988—2001 年的预测值与观测值进行对比,发现预测值与观测值较一致。

关键词:分形;支持向量机;矿井涌水量;相空间重构

中图分类号:TD742.1 文献标志码:A

Mine water inrush prediction based on fractal and support vector machines

HUANG Cun-han¹, FENG Tao², WANG Wei-jun², LIU Hui²

(1. School of Resources and Safety Engineering, Central South University, Changsha 410083, China; 2. School of Mining and Safety Engineering, Hunan University of Science and Technology, Xiangtan 411201, China)

Abstract: A new nonlinear method was proposed for predicting the mine water inrush. According to the fractal theory, the phase space of time series obtained from the mine water inrush was reconstructed. And the minimum embedding dimension was determined by autocorrelation coefficient, then the minimum embedding dimension was used for the input node of the support vector machines. The prediction model of time series was established based on the support vector machines. The mine water inrush of the 4th Mine in Hebi in the years from 1982 to 1997 were taken as training samples of time series, under the Matlab environment, the mine water inrush in 2000 and 2001 years were forecasted with the established model in different dimensions. The results show that when the embedding dimension is four, the errors of predicted values of shaft water inrush are the minimum, and their precisions are the highest. For testing the prediction reliability of this method, in different dimensions the prediction values were compared with the observed ones of three kinds of water inrush from 1988 to 2001 respectively, it shows that they coincide with each other better.

Key words: fractal; support vector machines; mine water inrush; state space reconstruction

近年来煤矿突水事故频发,造成重大人员伤亡和财产损失,如 2006-05-18 左云新井煤矿发生特大透水事故,造成 58 人死亡,直接经济损失 5 312 万元;2007-03-10 抚顺老虎台煤矿发生特大透水事故,22 人死亡,7 人下落不明;2008-07-21 广西那

读特大透水事故,36 人死亡,直接经济损失达 989.8 万元;2009 年鸡西金利煤矿、湖南常宁市三角塘镇煤矿等均发生特大透水事故。因此,为预防和减少突水事故的发生,开展突水预测预报工作显得尤为重要。其中,对矿井涌水量进行准确的预测是一项必要的研

究措施。不少研究者采用 BP 神经网络、混沌时间序列等方法对矿井涌水量进行了大量预测工作^[1-2],但预测结果都不太理想,例如文献[2]对 18 个样本进行预测,其中 6 个样本误差超过了 10%,最大误差达到了 32.4%,明显不符合精度要求。所以寻求新的高精度的预测方法势在必行。而支持向量机(SVM)作为一种新的非线性预测方法,它不依赖于环境因子,只需要利用有限样本(实际观测值)就能有效预测出系统的未来特性^[3-4]。近年来 SVM 理论和方法广泛应用于径流、电力、经济等领域的预测研究,但是对矿井涌水量的预测鲜有报道。基于以上认识,利用 SVM 方法进行矿井涌水量的预测研究具有重要的理论意义和应用价值。

1 相空间重构

相空间重构是时间序列预测分析的第 1 步,即把时间序列所处的低维空间拓扑到高维空间,并保持时间序列所对应的原动力系统内在信息的不变性。根据 Takens 定理^[5],对于观测时间序列 $\{x_i\}$, $i=1,2,3,\dots,n$,选择嵌入维数 m 可得另一组列向量 $y_i = \{x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau}\}$, $i=1,2,\dots$,其中 τ 为延迟时间。则由观测序列和延序列所构成的 m 维状态空间即为重构后的相空间,它与原始的状态空间是微分同胚的。所以进行相空间重构的重点是确定非线性时间序列重构的最佳延迟时间 τ 和最小嵌入维数 m 。

1.1 最佳延迟时间的确定

最佳延迟时间 τ 常用自相关函数法来确定^[6],它提取了任意 2 个时间序列值之间的线性相关性。设 $R(\tau)$ 为序列时间的间隔为 τ 的自相关系数,则

$$R(\tau) = \frac{\sum_{i=1}^{n-\tau} (x_i - \bar{x})(x_{i+\tau} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

其中, \bar{x} 为时间序列 $\{x_i\}$ 中所有观测值的平均值。当自相关系数降低到初值的 $1 - 1/e$ 倍时,所得到的 τ 即为最佳时间延迟^[7]。

1.2 最小嵌入维数的确定

嵌入维数的确定方法主要有饱和关联维数法(G-P 法)^[8]、伪最近邻域法和真实矢量场法等,但是前两者在一定条件下存在假临近现象,而矢量场法在实际应用中计算量较大,Cao 方法则规避了上述缺点^[9],通过计算相空间中各点与在嵌入维数下最邻近点的欧式距离判定最小嵌入维数。

定义

$$a(i, m) = \frac{\|y_i(m+1) - y_{n(i, m)}(m+1)\|}{\|y_i(m) - y_{n(i, m)}(m)\|} \quad (2)$$

式中, $\|\cdot\|$ 为 ∞ -范数; $i=1,2,3,\dots,n-m\tau$; $y_i(m+1)$ 是嵌入维数为 $m+1$ 的重构相空间中第 i 个向量; $y_{n(i, m)}(m)$ 是距离 $y_i(m)$ 最近的向量; $1 \leq n(i, m) \leq n - d\tau$ 。

根据式(2)计算均值 $E(m)$,有

$$E(m) = \frac{1}{n - m\tau} \sum_{i=1}^{n-m\tau} a(i, m) \quad (3)$$

为研究此均值从 m 到 $m+1$ 的变化,定义

$$E_1(m) = \frac{E(m+1)}{E(m)} \quad (4)$$

据式(4)可以判定,当 $E_1(m)$ 趋于不变时的维数 m 即为最小嵌入维。

2 支持向量机非线性预测原理

支持向量机(SVM)是一种新的非线性回归预测方法,它是建立在 VC 维理论和结构风险最小原理基础之上,通过非线性变换将输入向量映射到高维特征空间,并构造最优决策函数,利用原空间的核函数取代高维特征空间中的点积运算,应用有限样本的学习训练,获得全局最优解^[10]。

SVM 估计函数为

$$f(X) = W^T \varphi(X) + b \quad (5)$$

利用 ε -不敏感损失函数将估计函数转化为优化问题:

$$\min W^T W + C \sum_{i=1}^n (\xi_i + \hat{\xi}_i) \quad (6)$$

约束条件为

$$\begin{cases} [W \varphi(X_i)] + b - y_i \leq \varepsilon + \xi_i \\ y_i - [W \varphi(X_i)] - b \leq \varepsilon + \hat{\xi}_i \\ \xi_i, \hat{\xi}_i \geq 0 \end{cases} \quad (7)$$

式中, C 为惩罚因子; $\xi_i, \hat{\xi}_i$ 表示松弛因子; b 为偏置量; $i=1,\dots,n$ 。

式(7)一般采用对偶理论转化成二次规划求解,则约束表达式的对偶式为

$$\begin{aligned} \min & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i^* - a_i)(a_j^* - a_j) < \varphi(X_i) \\ \varphi(X_j) & > \sum_{i=1}^n (a_i - a_i^*) \varepsilon - \sum_{i=1}^n (a_i - a_i^*) y_i \end{aligned} \quad (8)$$

约束条件为

$$\begin{cases} 0 \leq a_i, a_i^* \leq C \\ \sum_{i=1}^n (a_i - a_i^*) = 0 \end{cases} \quad (9)$$

通过二次规划算法可得 SVM 回归预测模型为

$$y(X) = \sum_{i=1}^n a_i K(X, X_i) + b \quad (10)$$

式中 $K(\cdot)$ 为支持向量机的核函数, 根据向量间的不同算法可将核函数分为 4 类^[11-12]: 线性核函数、多项式核函数、RBF 函数和 sigmoid 函数, 其中 RBF 是最常用的核函数, 可表示为

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

以上 a_i 、 a_i^* 、 b 等可根据 Karush-Kuhn-Tucker 定理求得。

3 基于分形和支持向量机回归的矿井涌水量预测分析

3.1 预测模型

将实测涌水量数据作为时间序列 $\{x_i\}$, 利用相空间重构理论确定最小嵌入维数 m , 将 m 作为支持向量机输入的特征节点来建立训练样本, 即

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-m} & x_{n-m+1} & \cdots & x_{n-1} \end{bmatrix}$$

$$Y = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_n \end{bmatrix} \quad (12)$$

训练样本经过学习后, 根据 SVM 回归预测函数 (10), 得到预测值^[13]为

$$x_{n+1} = \sum_{i=1}^{n-m} a_i \exp \left(-\frac{\|x_i - x_{n-m+1}\|^2}{\sigma^2} \right) + b \quad (13)$$

3.2 预测实例分析

3.2.1 涌水量时间序列重构

矿井涌水量的时间序列数据见表 1^[14], 共有 3 个时间数列, 根据上述预测原理分别利用 1982—1997 年的 16 个样本预测 2000—2001 年的涌水量, 并进行对比分析。

利用第 1 节的介绍确定时间序列的最佳延迟时间和最小嵌入维数。以井筒涌水量为例, 计算得最佳延迟时间 $\tau=2$, 相应的最小嵌入维数可由图 1 判断。从图 1 可以看出, 当 $m=4$ 时, $E_1(m)$ 在随 m 的演变中开始趋于稳定, 即时间序列的最小嵌入维数为 4。 $E_2(m)$ 用于判别一个时间序列是否具有分维特性, 虽然本例训练样本较少, 但是 $E_2(m)$ 仍然具有归一的

趋势, 即此时间序列是可以分维的。

表 1 鹤壁四矿历年矿井涌水量
Table 1 Historical well water inrush of the 4th Mine in Hebi m^3/h

年度	井筒涌水量	巷道涌水量	工作面涌水量
1982	465.3	758.0	2 162.7
1983	648.1	987.5	1 467.8
1984	1 251.3	687.4	1 815.3
1985	1 249.2	712.2	2 204.1
1986	1 187.1	814.2	1 991.5
1987	921.0	1 191.0	2 164.0
1988	76.5	1 990.0	2 256.1
1989	54.0	1 670.5	2 266.4
1990	108.0	2 348.0	2 296.4
1991	170.2	2 114.0	2 022.2
1992	213.0	1 972.0	1 965.3
1993	166.4	1 852.0	1 484.2
1994	179.9	1 890.0	1 516.0
1995	271.6	1 609.3	3 587.9
1996	514.0	1 417.9	8 880.1
1997	439.1	1 171.3	3 283.7
1998	403.7	600.1	2 686.9
1999	393.9	538.2	2 745.1
2000	437.8	578.8	3 877.0
2001	452.2	496.3	4 189.8

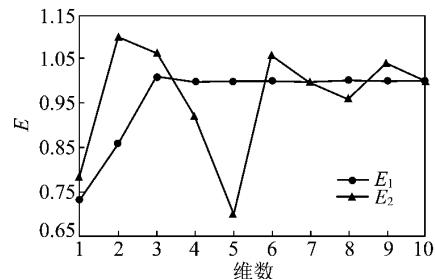


图 1 井筒涌水量最小嵌入维数

Fig. 1 The minimum embedding dimension of shaft water inrush

3.2.2 支持向量机预测结果分析

当 $m=4$ 时, 建立井筒涌水量的训练样本见表 2。

本文采用基于 Matlab 下 LIBSVM 工具^[15]进行支持向量机学习训练, 核函数选用 RBF 函数。首先对样本进行归一化验证, 并利用 Grid-Search 方法自动搜索最优参数: 惩罚参数、灵敏度及径向宽度等核函数参数。经过编程计算确定模型的最优参数为 $C=64$, $\sigma^2=0.125$, $\epsilon=0.00255$, 利用该模型对 2000—2001 年的井筒涌水量进行预测, 并分别建立 $m=3$ 、4、5、6 等不同维数下的预测模型, 其预测结果与真实结果的误差和相关系数见表 3。从表 3 可以看出, 当

$m=4$ 时,2000 和 2001 年井筒涌水量预测值和观测值的误差与其他节点时的预测结果相比,其误差均最小;且其预测相关系数最大,即表明其预测精度最高。

表 2 井筒涌水量支持向量机训练样本

Table 2 Training samples of SVM of shaft water inrush

目标值	特征值 1	特征值 2	特征值 3	特征值 4
1 187.1	465.30	648.10	1 251.30	1 249.20
921.0	648.10	1 251.30	1 249.20	1 187.10
76.5	1 251.30	1 249.20	1 187.10	921.00
54.0	1 249.20	1 187.10	921.00	76.50
108.0	1 187.10	921.00	76.50	54.00
170.2	921.00	76.50	54.00	108.00
213.0	76.50	54.00	108.00	170.20
166.4	54.00	108.00	170.20	213.00
179.9	108.00	170.20	213.00	166.40
271.6	170.20	213.00	166.40	179.90
514.0	213.00	166.40	179.90	271.60
439.1	166.40	179.90	271.60	514.00
403.7	179.90	271.60	514.00	439.10
393.9	271.60	514.00	439.10	403.70

表 3 井筒涌水量预测误差与相关系数

Table 3 Protecting errors and correlation coefficients of shaft water inrush

m	误差 / %		相关系数
	2000 年	2001 年	
3	-13.41	-7.92	0.743 0
4	-6.17	-2.59	0.963 5
5	-21.00	-26.00	0.382 0
6	7.49	-9.60	0.848 5

为研究该方法预测多年矿井涌水量的可靠性,分别将不同维数下井筒、巷道和工作面涌水量 3 组时间序列 1988—2001 年的预测值与观测值进行对比,结果分别如图 2 所示,图中曲线 r 代表涌水量的实际值,其他 4 条曲线为各嵌入维数下的预测值。

从图 2(a) 可以看出,相对于其他维数而言,当 $m=4$ 时,其预测结果与真实值吻合较好,说明引入最小嵌入维数对模型的预测效果是有利的,同时也证明了时间序列分维的正确性。从图 2 的预测值和真实值对比来看,除了个别预测结果偏离真实值曲线外,其他预测结果基本满足精度要求,准确地反应了涌水量的总体发展规律;这同时也说明,重构后的涌水量时间序列经过 SVM 训练后,既包含了原有系统的信息,又成功地预测了系统的未来发展趋势。

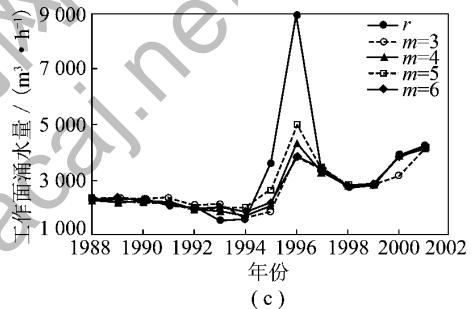
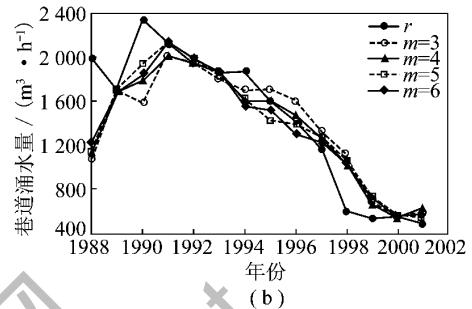
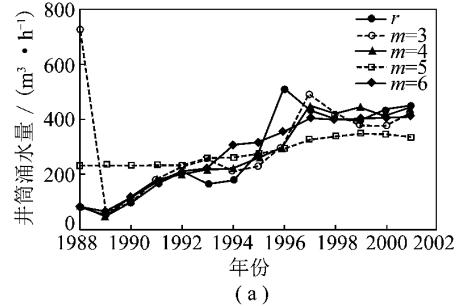


图 2 不同维数时涌水量预测结果

Fig. 2 Predicting results of water inrush under different dimensions

4 结 论

(1) 将时间序列最小嵌入维数作为支持向量机的输入节点,并据此建立矿井涌水量预测的训练样本,所得到的预测结果误差最小,精度最高。

(2) 应用该预测方法对多年矿井涌水量进行预测,得到的预测值和观测值较吻合,说明该方法既能保留实际矿井涌水量的历史信息,又能准确描述矿井涌水量的未来发展趋势,这为实际生产中预防矿井突水事故的发生提供了一个新思路。

(3) 利用有限样本进行回归预测分析是 SVM 方法的优势,但是也存在一些问题:样本中的个别突跳值对分析结果有一定的干扰(如图 2(c) 中第 1 个观测值与所有预测值相差均较大),如何更有效地对样本进行去噪化需要进一步的探索;另外,模型本身核参数的选取对回归预测结果也有影响,例如,就 RBF 核函数而言,选取不同的 C 和 σ^2 会得到不同的误差值,因此,为进一步提高预测精度需要继续深入研究。

参考文献:

- [1] 梁冰,李刚,王宗林,等. 基于神经网络的降雨充水矿井涌水量预测[J]. 中国地质灾害与防治学报, 2009, 21(1): 122-125.
Liang Bing, Li Gang, Wang Zonglin, et al. Prediction of water inflow of mine with rainfall yield based on BP artificial neural network[J]. The Chinese Journal of Geological Hazard and Control, 2009, 21(1): 122-125.
- [2] 陈玉华,杨永国,彭高辉,等. 矿井涌水量混沌时间序列分析与预测[J]. 煤田地质与勘探, 2008, 36(4): 34-36.
Chen Yuhua, Yang Yongguo, Peng Gaohui, et al. Chaotic time series analysis and prediction for mine water inflow[J]. Coal Geology and Exploration, 2008, 36(4): 34-36.
- [3] Vapnik V N, Golowich S E, Smöla A J. Support vector method for function approximation, regression estimation and signal processing [A]. In Advances in Neural Information Processing Systems 9 [C]. 1997: 281-287.
- [4] Smöla A J, Scholkopf B. A tutorial on support vector regression [J]. Statistics and Computing, 2004, 14(3): 199-222.
- [5] Takens F. Detecting strange attractor in turbulence [M]. Heidelberg: Springer Berlin, 1981: 368-381.
- [6] 骆振华. 时间序列分析引论 [M]. 厦门: 厦门大学出版社, 1987: 23-26.
Luo Zhenhua. The time series analysis introduction [M]. Xiamen: Xiamen University Press, 1987: 23-26.
- [7] Rosenstern M T, Collins J J, Deluca C J. A practical method for calculating largest Lyapunov exponents from small data sets [J]. Physica D, 1993, 65: 117-134.
- [8] Grassberger P, Procaccia L. Measuring the strangeness of strange attractors [J]. Physica D, 1983, 9: 189-208.
- [9] Cao Liangyue. Practical method for determining the minimum embedding dimension of a scalar time series [J]. Physica D, 1997, 110: 43-50.
- [10] 韩敏. 混沌时间序列预测理论与方法 [M]. 北京: 中国水利水电出版社, 2007: 207-210.
Han Min. Prediction theory and method of chaotic time series [M]. Beijing: China Water Power Press, 2007: 207-210.
- [11] Lahiri S K, Ghanta K C. Prediction of pressure drop of slurry flow in pipeline by hybrid support vector regression and genetic algorithm model [J]. Chinese Journal of Chemical Engineering, 2008, 16(6): 841-848.
- [12] 刘丁,任海鹏,孔志强. 基于径向基函数神经网络的未知模型混沌系统控制 [J]. 物理学报, 2003, 52(3): 531-535.
Liu Ding, Ren Haipeng, Kong Zhiqiang. Control of chaos solely based on RBF neural network without an analytical model [J]. Acta Physica Sinica, 2003, 52(3): 531-535.
- [13] 刘涵,刘丁,李琦,等. 基于支持向量机的混沌时间序列非线性预测 [J]. 系统工程理论与实践, 2005(9): 94-99.
Liu Han, Liu Ding, Li Qi, et al. Chaotic time series nonlinear prediction based on support vector machines [J]. System Engineering Theory and Practice, 2005(9): 94-96.
- [14] 鲁静,陈江峰,黄存捍,等. 鹤壁四矿矿井涌水量的分形预测 [J]. 矿业安全与环保, 2002, 29(4): 25-26.
Lu Jing, Chen Jiangfeng, Huang Cunhan, et al. Fractal prediction of mine water inrush in 4th mine of Hebi [J]. Mining Safety and Environmental Protection, 2002, 29(4): 25-26.
- [15] Chih-Chung Chang, Chih-Jen Lin. LIBSVM: a library for support vector machines [DB/OL]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html>, 2009-02-27.

2015 年全国煤矿将建立井下避险系统

2010-05-19, 国家安全生产监督管理总局副局长、国家煤矿安全监察局局长赵铁锤在山西长治表示, 我国将在全国煤矿建立完善监测监控、人员定位、紧急避险、压风自救、供水施救和通信联络等井下安全避险六大系统, 以提高煤矿安全保障能力。

赵铁锤表示, 井下安全避险“六大系统”的总体目标要求: 到 2015 年全国所有煤矿都建立“六大系统”, 并达到“设施完备、系统可靠、管理到位、运转有序”。

煤矿井下安全避险六大系统包括: 监测监控系统, 实现对煤矿井下瓦斯、CO 浓度、温度、风速的动态监控, 完善紧急情况下及时断电撤人制度; 人员定位系统, 准确掌握各个区域作业人员的情况; 救生舱、避难硐室等紧急避险系统, 实现井下灾害突发时的安全避险; 压风自救系统, 确保灾变时现场作业人员有充分的氧气供应; 供水施救系统, 在灾变后为井下作业人员提供清洁水源或必要的营养液; 通信联络系统, 实现井上、井下和各个作业地点通信畅通。