

doi:10.11835/j.issn.1674-4764.2016.S2.028

演化追踪法优化相空间的 SVM 供水量预测模型

赵雪凝, 杜坤, 周明, 任刚红, 李诚

(昆明理工大学 建筑工程学院, 昆明 650500)

摘要:相空间重构的支持向量机预测模型应用十分广泛,在城市供水量预测方面也占据着重要地位,传统的预测模型倾向于将重构的相空间整体带入,这样可能存在引入无效相点从而影响预测精度的问题,基于此将演化追踪法引入相空间重构的预测模型对有效相点进行筛选,优化预测模型的训练样本,达到提高预测精度目的。利用 MATLAB 编程软件将演化追踪法用于城市供水量的预测,预测结果的平均绝对误差由 0.52% 降低到了 0.29%,证明了演化追踪法的可利用性与有效性。

关键词:重构相空间; SVM; 演化追踪法; 供水量预测

中图分类号: TU991.31 **文献标志码:** A **文章编号:** 1674-4764(2016)S2-0147-04

Improvement of SVM regression forecast water supply model based on phase space reconstruction

Zhao Xuening, Du Kun, Zhou Ming, Ren Ganghong, Li Cheng

(Faculty of Civil Engineering and Mechanics, Kunming University of Science and Technology, Kunming 650500, P. R. China)

Abstract: Nowadays, SVM prediction model based on phase space reconstruction is widely used, and it also plays an important role in urban water supply prediction. The traditional prediction model tends to bring the reconstructed phase space into the whole, which may lead to ineffective introduction of SVM. Phase prediction method is used to improve the accuracy of prediction. Based on this, the evolutionary tracing method is introduced into the prediction model of phase space reconstruction to filter the effective points and to optimize the training samples of the prediction model. The evolutionary tracing method is used to forecast the urban water supply quantity by using MATLAB programming software. The average absolute error of forecasting result is reduced from 0.52% to 0.29%, which proves the availability and effectiveness of evolutionary tracing method.

Key words: reconstructed phase space; SVM; evolutionary tracing method; water supply prediction

如今水资源极度短缺,要解决这一问题,需要从节水和调度两方面入手,因此,城市供水优化调度显得尤为重要,而城市供水量的准确预测是供水优化调度的重要基础工作具有重要的研究意义。支持向量机(support vector machines, SVM)是建立在统

计学习理论基础上的数据挖掘方法^[1],它在解决小样本、非线性和结构风险最小化问题中有其特有的优势^[2],能成功处理回归问题(时间序列分析)^[3],并可推广于

预测领域的研究^[4-5]。根据混沌理论^[6],城市供

收稿日期: 2016-10-29

基金项目: 昆明理工大学 2016 年学生课外学术科技创新基金(2015YB025);国家自然科学基金(51608242);云南省人才培养计划项目(14118943)

作者简介: 赵雪凝(1992-),女,硕士生,主要从事市政工程研究,(E-mail)860655976@qq.com。

杜坤(通信作者),男,博士,(E-mail)250977426@qq.com。

水量数据是一组非线性混沌时间序列,相空间重构的目的就在于重构高维相空间,从中恢复出体现混沌系统规律性的混沌吸引子,从而获取更多的隐藏信息。对于将支持向量机与相空间重构相结合的预测许多学者对其进行了大量研究。郑永康^[7]使用相空间重构理论结合 SVM 方法进行电力系统短期负荷的预测,在不考虑气候因素的情况下仍获得了较高的预测精度;郑永康等^[8]提出了基于 SVM 的小波 SVM 与相空间重构相结合的预测电力系统短期负荷,预测效果好且证明小波核优于常用的径向基核;薛向红等^[9]提出了一种基于 SVM 与相空间重构的预测公交客流量的方法,通过实例证明该方法可靠有效;高伟等^[10]利用 SVM 与相空间重构结合方法与径向基函数神经网络法对浅海混响数据进行预测,结果表明 SVM 具有更好的预测精度;刘翼等^[11]利用基于重构相空间的 SVM 方法对中长期径流进行预测,通过实例分析该方法具有良好的推广泛化能力。由此可见,基于相空间重构的 SVM 模型对于非线性混沌时间序列的预测具有良好的应用前景。

许多学者的侧重点在于相空间重构或 SVM 算法的研究,对于预测数据的预处理方面研究较少,大部分研究偏向于将相空间重构后的所有样本带入模型进行预测,这样不仅增加了计算用时,而且容易引进过多无效预测数据进而影响最终预测精度。基于此,本文提出了利用演化追踪法筛选相空间中的有效相点,带入模型进行预测,研究其对于经过相空间重构的 SVM 预测模型预测精度的影响,并结合实例进行了分析验证。

1 相空间重构

1.1 相空间重构理论

由于实际系统的一维时间序列所反映出的有效信息是有限的,导致了从一维时间序列难以认识到系统的复杂性和全面性,因此相空间重构的目的在于重构高维相空间,从中恢复出体现混沌系统规律性的混沌吸引子,从而获取更多的隐藏信息^[12-13]。

对于单变量时间序列, Packard 等^[14]人提出了坐标延迟重构法:设有单变量时间序列 $\{x(i), i=1, 2, \dots, N\}$, N 是序列总长度,以延迟时间 τ 和嵌入维数 m 进行相空间重构,相点可以表示为

$$X(i) = [x(i), x(i+\tau), \dots, x(i+(m-1)\tau)]$$

$$i = 1, 2, \dots, M \quad (1)$$

$$M = N - (m-1)\tau \quad (2)$$

根据 Tankens^[15]的嵌入理论, d 维吸引子能够嵌入到 $m \geq 2d+1$ 维相空间时,可重现出原吸引子的几何特征,由此来研究系统的演化规律。由上可知,相空间的重构需要确定两个重要的参数,即延迟时间 τ 和嵌入维数 m 。目前,对于延迟时间的确定方法较多,如自相关函数法、互信息法、平均位移法等;同时确定嵌入维数的方法也很多,如饱和关联维数法、临近点维数法、虚假临近点法等。

2 基于相空间重构的支持向量机预测模型

2.1 支持向量机原理

支持向量机是建立在统计学理论基础上的的一种新的通用学习机^[16]。其回归预测原理总的来说就是通过某种事先选择的非线性映射将输入空间映射到一个高维特征空间中,在这个特征空间中的平面上构造最优分类超平面^[3](即进行线性回归)。线性回归函数为

$$f(x, \omega) = \omega \cdot \Phi(x) + b \quad (3)$$

式中: ω 为权向量; b 为阈值。所求的回归函数 $f(x)$ 是使式(4)所示目标函数最小

$$\min\left(\frac{1}{2} \|\omega\|^2 + c \cdot R_{emp}\right) \quad (4)$$

式中: ω 为权重; c 为误差惩罚因子, R_{emp} 为训练误差。

根据统计学习理论,基于的 SVM 混沌时间序列的拟合函数为

$$y(x) = \sum_{k=1}^N a_k k(x_k, x) + b \quad (5)$$

式中: a_k 为支持向量; b 为偏差; $k(x_k, x)$ 为核函数,这里采用最常用的径向基核函数(RBF)

$$k(x_k, x_1) = \exp(-\|x_k - x_l\|^2 / 2\sigma^2) \quad (6)$$

2.2 基于相空间重构的 SVM 模型预测步骤

1) 根据模型假定选取自变量和因变量; 2) 对训练集和测试集进行归一化预处理,将原始数据规整到 $[0, 1]$ 范围内; 3) 对原始数据进行相空间重构; 4) 将相空间重构的数据带入 SVM 模型进行预测,输出预测结果。

3 演化追踪法

所谓预测即是根据已有相点的运动趋势来推断

目标相点的运动趋势,因此演化趋势相近的相点的选择直接影响最终预测结果的准确与否^[17]。目前,多数相点的选取方法是根据当前各相点与预测中心点 X_M 的距离来确定有效相点^[18],当它们的欧式距离小于某阈值就认为是有效的相点,并不考虑相点的演化与运动趋势,但事实是,如图 1 所示。

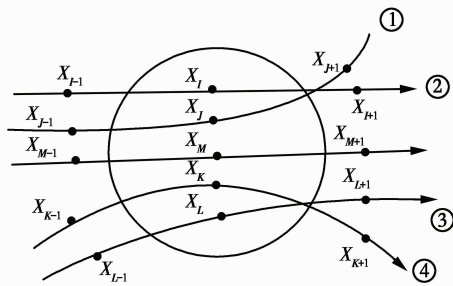


图 1 预测中心点及其邻近点的演化示意图

X_I, X_J, X_K, X_L 均为符合与中心点 X_M 欧式距离要求的相点, $X_{i-1} (i=I, J, M, K, L)$ 分别为各相点上一步的历史相点, $X_{i+1} (i=I, J, M, K, L)$ 分别是各相点下一步的演化相点, ①②③④是相点可能的运动演化轨迹。相点 X_I 与中心相点 X_M 具有近似相同的演化运动轨迹, 而相点 X_L 的历史相点 X_{L-1} 虽与 X_{M-1} 距离较大, 但其演化相点呈现趋近于 X_{M+1} 的趋势, 因此, 说明 X_I 与 X_L 为有效相点; 相反的相点 X_J 与 X_M 的历史相点虽均趋近, 但其演化相点都呈现远离 X_{M+1} 的趋势, 相点 X_K 历史相点与演化相点均远离中心相点, 对目标相点的演化轨迹没有参照意义, 所以二者可以认定为无效相点。如将重构后的相空间中全部相点带入到预测模型中去, 就会存在将无效相点用于预测的可能, 从而降低预测精度, 因此找出有效相点剔除无效相点是模型准确预测的重要基础, 而演化追踪法就是这一基础的实现手段。

4 实例分析

选用某市 2006 年 1 月 1 日至 2006 年 1 月 31 日时间段的供水数据用于经相空间重构的 SVM 模型预测实例。该时间段的供水数据可以看作是一组混沌的时间序列, 采用互信息法确定最佳延迟时间 $\tau=2$, 采用关联函数法确定嵌入维数 $m=4$ 。根据延迟时间和嵌入维数进行相空间重构, 得样本数 $K=25$ 个。

利用演化追踪法回溯 2 步历史选取有效相点: 根据欧式距离筛选离预测中心点 X_M 最近的 K 个初

始相邻相点 $X_i^1, (i=1, 2, \dots, K)$, 看是否与中心相点在演化过程中“一直邻近”, 如果出现不邻近情况, 则认为是伪邻点, 予以剔除; 再根据欧氏距离寻找离 X_M 的上一步演化点 X_{M-1} 最近的 K 个初始邻近相点 $X_i^2, (i=1, 2, \dots, K)$; 接着再根据欧氏距离寻找离 X_M 的上 2 步演化点 X_{M-2} 最近的 K 个初始邻近相点 $X_i^3, (i=1, 2, \dots, K)$; 如果 $X_i^1, (i=1, 2, \dots, K)$ 的上一步演化点在点集 $\{X_i^2, (i=1, 2, \dots, K)\}$ 中, 而且, $X_i^2, (i=1, 2, \dots, K)$ 的上两步演化点在点集 $\{X_i^3, (i=1, 2, \dots, K)\}$ 中, 则点 $X_i^1, (i=1, 2, \dots, K)$ 是中心点 X_M 的有效相点, 否则是无效相点, 同理依次判断剩余相点。经计算最终相空间重构后的有效相点 $K=19$ 个, 比原相空间相点数少 6 个。

在基于相空间重构的 SVM 预测模型中分别带入未经筛选的相空间时序和利用演化追踪法筛选过的相空间时序对城市日供水量进行预测, 预测结果如图 2、3 所示。

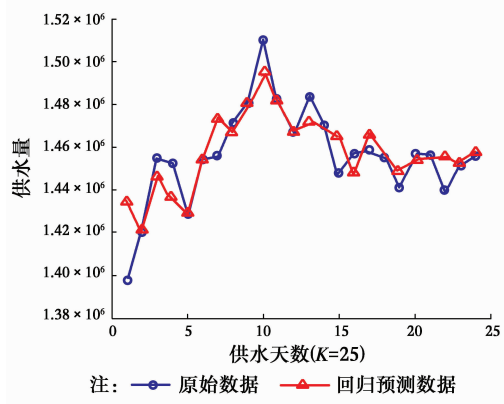


图 2 $K=25$ 原始数据与预测数据对比图

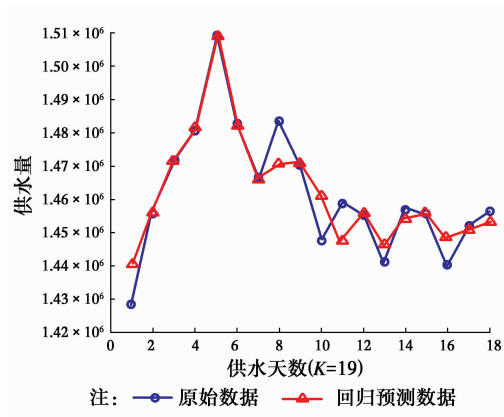


图 3 $K=19$ 原始数据与预测数据对比图

由图 2、3 对比可以看出 $K=19$ 的供水量预测值与原始数据的吻合程度更高, 因而预测效果也更好, 预测结果的平均绝对误差 (MAPE) 见表 1。

表 1 预测精度比较 %

训练样本数	各预测日相对误差 P_i			
$K=25$	0.11	-1.35	-0.14	-0.29
$K=19$	0.05	0.04	-0.03	-0.85
训练样本数	各预测日相对误差 MAPE			
$K=25$	-0.82	-0.18	-0.61	0.52
$K=19$	0.87	0.01	-0.02	0.29

可以看出 $K=19$ 的预测平均绝对误差比 $K=25$ 的要小,而且相同预测日的预测相对误差也都小于 $K=25$ 的相对误差,这说明有时 K 值的增加与最终预测结果的精度不是成正比关系的,反而会增加误差,影响预测的准确性;同时也说明时间序列在经过相空间重构后存在对于预测来说无效的相点,对准确预测起反作用,演化追踪法在这其中起到了很好的辨别剔除作用。

5 结论

介绍了支持向量机的原理及其在城市供水量预测方面的应用,同时提出对于支持向量机预测模型的数据处理方面的改进,即运用演化追踪法对相点进行选择,筛选出有效相点带入支持向量机预测模型对城市供水量进行预测。使用 Matlab 软件进行编程对实例数据进行验证,结论如下:

1) 支持向量机模型预测城市用水量是可行的,但将相空间重构后的相点全部带入可能引入无效相点使得预测精度有所降低;

2) 利用演化追踪法对相点进行筛选,之后带入预测模型,其预测结果的平均绝对误差由 0.52% 降低到 0.29%,证明了该方法的有效性。

参考文献:

[1] 汪海燕,黎建辉,杨风雷. 支持向量机理论及算法研究综述[J]. 计算机应用研究, 2014, 31(5):1281-1286.
 [2] 邵福波,董玉林,胡运红. 支持向量机理论发展与应用综述[J]. 泰山学院学报, 2013, 35(6):78-81.
 [3] 张博洋. 支持向量机理论与应用研究综述[J]. 无线互联科技, 2015(19):111-112.

[4] TONG S, CHANG E. Support vector machine active learning for image retrieval [C]//Acm International Conference on Multimedia, 2001.
 [5] NOBLE W S. What is a support vector machine? [J]. Nature Biotechnology, 2006, 24(12):1565-1567.
 [6] THIÉTART R A, FORGUES B. Chaos theory and organization[J]. Organization Science, 1995, 6(1): 19-31.
 [7] 郑永康. 相空间重构与支持向量机结合的短期负荷预测研究[D]. 成都:西南交通大学, 2008.
 [8] 郑永康,陈维荣,戴朝华. 小波支持向量机与相空间重构结合的短期负荷预测研究[J]. 电力系统保护与控制, 2008, 36(7):29-33.
 [9] 薛向红,薛小锋. 基于相空间重构—SVM 的公交客流量预测研究[J]. 江苏理工学院学报, 2010, 16(9): 24-28.
 [11] 刘冀,王本德,袁晶瑄,等. 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[J]. 大连理工大学学报, 2008, 48(4):591-595.
 [12] 吴彤. 非线性动力学混沌理论方法及其意义[J]. 清华大学学报(哲学社会科学版), 2000(3):72-79.
 [13] LEVY D. Chaos theory and strategy: Theory, application, and managerial implications[J]. Strategic Management Journal, 1994, 15(Sup2):167-178.
 [14] PACKARD N H, CRUTCHFIELD J P, FARMER J D, et al. Geometry from a time series [J]. Physical Review Letters, 1980, 45(9): 712-716.
 [15] TAKENS F. Detecting strange attractors in turbulence, in: Dynamical systems and turbulence, warwick 1980 [M]. Springer: Berlin Heidelberg, 1981.
 [16] LAPIN M, HEIN M, SCHIELE B. Learning using privileged information: SVM+ and weighted SVM[J]. Neural Networks, 2014, 53:95-108.
 [17] 高俊杰. 混沌时间序列预测研究及应用[D]. 上海:上海交通大学, 2013.
 [18] 白云. 时间序列特性驱动的供水量预测方法研究及应用[D]. 重庆:重庆大学, 2014.

(编辑 郭飞)