

基于多嵌入维数的时用水量 LSSVM 组合预测

陈丽琳

(浙江大学 电气工程学院, 浙江 杭州 310027)

摘要:为解决供水系统调度所需混沌时用水量高精度预测等问题,将最小二乘支持向量机(LSSVM)组合预测模型应用到城市时用水量预测中。在分析不同嵌入维数和预测方法对模型预测精度影响程度的基础上,提出了基于多嵌入维数的 LSSVM 组合预测模型。采用互信息法和 G-P 方法求取多个嵌入维数,并建立了不同相空间模型,通过 LSSVM 算法对上述多个预测模型进行了组合预测,既综合了各不同嵌入维数各预测方法下的信息,又对单一模型下的预测偏差进行了融合,以有效地提高预测精度;最后在某地进行了时用水量序列的仿真实验。研究结果表明,该模型预测精度平均误差小于 2%,明显优于各单一模型的预测结果,证实了该组合模型的有效性和实用性。

关键词:时用水量预测;多嵌入维数;相空间重构;最小二乘支持向量机

中图分类号:O241.2;TP29

文献标志码:A

文章编号:1001-4551(2012)07-0869-04

Combined prediction of urban hourly water consumption using LSSVM based on multi-dimension embedding phase space

CHEN Li-lin

(College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China)

Abstract: In order to improve the accuracy of chaotic hour consumption prediction in urban water supply system, a combined forecasting model for urban hourly water consumption using least squares support vector machine(LSSVM) was investigated. After the analysis of different effects on chaotic system forecast accuracy by various forecast methods and parameters of phase space reconstruction, a combined forecasting model for urban hourly water consumption using LSSVM based on multi-dimension embedding phase space was established. The different embedding dimensions were estimated by combining mutual information method and G-P algorithm. Combined forecasting models were solved by LSSVM which can take advantage of all information in all dimension embeddings and forecast methods. The predictive bias under the single model was merged. In this way, the forecast accuracy was improved. The simulation results of hourly water consumption forecast in a place shows that the forecast error is below 2% and better than other forecasting results in single model. This proves the effectiveness and practicality of the approach.

Key words: urban hourly water consumption prediction; multi-dimension embedding; phase space reconstruction; least squares support vector machine(LSSVM)

0 引 言

目前,对城市时用水量时间序列的变化包含周期性、趋势性和随机扰动性成分的认识,已为从事时用水量预测研究的同行们所普遍认同^[1-3]。但近年来的一些研究表明,同时受社会经济因素和气候条件因素影响的城市用水量时间序列中存在混沌成分。而对于混

沌时间序列而言,即使对数据匹配得很好,传统的 AR、MA、ARMA 等模型有时也无法做出准确的预测,未来趋势会在性质上与原有时间序列趋势发生根本不同的变化。对于混沌时间序列的预测研究只能从混沌序列本身的特性入手。

相空间重构是混沌序列分析和混沌预测的基础,其中嵌入维数的选取尤为关键。不同的嵌入维数下,

收稿日期:2011-12-02

作者简介:陈丽琳(1985-),男,浙江丽水人,主要从事供水优化调度方面的研究. E-mail:chenlilin320@yahoo.com.cn

系统重构到不同的相空间中,则系统的吸引子及时序特性就会改变,其最终的预测结果也必然不同^[4]。同时,不同的混沌预测方法利用系统信息的方式迥异,从不同角度对时用水量做出预测,其结果也不尽相同。为避免有用信息的流失,应充分利用各预测方法的特点^[5]。

本研究分析探讨不同嵌入维数的影响,综合利用不同维数和不同方法,采用最小二乘支持向量机(LSSVM)进行组合预测,可有效利用系统包含的信息,减小单一模型可能出现预测误差偏大的问题。

1 基于多嵌入维数的时用水量 LSSVM 组合预测建模

1.1 混沌系统相空间重构

根据 Takens 定理,只要满足维数 $m \geq 2d + 1$,总可以在拓扑不变的意义上找到一个 m 维的嵌入相空间。对于混沌时间序列: $x_1, x_2, \dots, x_{n-1}, x_n$, 如果能够选取适当的延迟时间 t 和嵌入维数 m , 则相空间重构为:

$$X_m(i) = [x(i), x(i+t), x(i+2t), \dots, x(i+(m-1)t)] \quad (1)$$

式中: $i = 1, 2, \dots, n - (m-1)t$ 。

在相空间重构中,嵌入维数 m 是十分重要的一个参数,同时恰当的选取也是困难的^[6-7]。G-P 算法凭借混沌系统的一个测量数据序列就可以得到吸引子维数的信息,而虚假邻点法则是利用吸引子的几何特性,两种方法求解思路差异甚大,但在估计嵌入维数上都得到了广泛的应用。本研究采用 G-P 算法和虚假邻点法,分别从不同角度计算嵌入维数^[8]。

1.2 LSSVM 基本原理

最小二乘支持向量机(LSSVM)是对标准支持向量机(SVM)的一种改进,通过将 SVM 中的不等式约束变为等式约束,并将误差平方和损失函数作为训练集的经验,把解二次规划问题转化为求解线性方程组问题,可大大提高求解问题的速度和收敛精度,且具有较好的推广性^[9]。

给定训练样本集: $\{(x_i, y_i)\}_{i=1}^N$, 其中, N 为训练样本数, $x_i \in R^n$ 为输入向量, $y_i \in R$ 为相应的输出向量。根据 Suykens 提出的 LSSVM 理论,在权空间中的函数估计问题可以描述为:

$$\begin{cases} \min J(w, x) = \frac{1}{2} w^T w + c \sum_{i=1}^n x_i^2 \\ \text{s. t. } y_i = wd(x_i) + b + x_i \end{cases} \quad (2)$$

式中: $d(\cdot)$ —输入空间到高维特征空间的非线性映射, w —权矢量, x_i —不敏感损失函数的松弛因子, b —偏移量, c —正则化参数。

为了求解式(2),引入拉格朗日函数:

$$L(w, b, x, \alpha) = J(w, x) - \sum_{i=1}^n \alpha_i \{w^T j(x_i) + b + x_i - y_i\} \quad (3)$$

式中: α_i —拉格朗日乘子。

通过式(3)对 w 、 b 、 x_i 、 α_i 求偏导,并使其等于零,可消除 w 、 x , 构造合适的核函数 $K(x_i, x_j)$ 。优化问题的解析解为:

$$\begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K(x_1, x_1) + 1/c & \cdots & K(x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_n, x_1) & \cdots & K(x_n, x_n) + 1/c \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (4)$$

由式(4)可得 LSSVM 的函数估计模型:

$$y = \sum_{i=1}^n [\alpha_i K(x, x_i)] + b \quad (5)$$

1.3 基于多嵌入维数的时用水量 LSSVM 组合预测模型

组合预测方法建立在信息最大化利用基础上,它融合了不同预测算法的优点,减小了单一预测模型所特有的不确定性,把局部不完整信息和相关信息加以综合和互补,获得研究对象的一致性认识,可以有效地降低预测风险,得到一个更高精度的预测结果^[10]。

鉴于相空间嵌入维数和预测方法的差异都会导致时用水量预测结果的不同,为了提高预测模型的准确性和可靠性,本研究利用不同维数下加权一阶局域法、BP 神经网络和 LSSVM 3 种模型作为组合预测的单项,然后采用 LSSVM 将不同维数下这 3 个单项的预测结果进行非线性组合,作为最终的预测结果,即:

$$y = f(f_1, f_2, \dots, f_i) \quad (6)$$

式中: y —最终的组合预测结果; $f_1, f_2, \dots, f_i$ —不同嵌入维数下不同预测方法预测函数; $f(\cdot)$ —由 LSSVM 来实现的不同维数下不同预测方法的组合预测函数。

根据上述 LSSVM 原理,本研究将参加组合预测的各单项预测结果作为 LSSVM 的输入向量,选取适当的核函数及参数,然后用一定数量的样本来训练 LSSVM,使不同的输入向量得到相应的输出量值,经过不断学习及测试,在输入向量和实际值之间建立一种非线性映射关系 $f(\cdot)$ 。在达到较高精度后,该 LSSVM 即可作为组合预测的模型,将样本向量 X_i 及待预测的向量 X 代入该 LSSVM,就能得到预测结果。

2 实验与结果

笔者针对某地 2008 年 5 月 1 日~2009 年 5 月 31 日的时用水量数据进行研究,利用 G-P 算法和虚假邻点法求出的嵌入维数分别为 10 和 8。笔者在所求嵌入维数间进行多维多模型预测,即在 m 为 8~10 时,分别建立基于相空间重构的单一算法模型。实验时本研究将历史数据分为两组,第 1 组以前 11 个月时用水量数据作为预测的原始序列,第 2 组是最后一个月数值,用来验证模型预测的有效性。

2009 年 5 月 1 日不同嵌入维数下相同预测模型的预测结果如图 1 所示。从图 1 可以看出,在 1~10 时刻, $m=8$ 时, BP 神经网络的预测结果明显优于另两组的预测结果,与实际值更为接近。而在 11~17, 18~24 这两个时间段, m 分别取为 9 和 10,预测效果更佳。

2009 年 5 月 1 日同一嵌入维数下不同预测模型的预测结果如图 2 所示。从图 2 可以看出,不同预测方法的预测结果差异甚大,任一方法的预测结果都不可能任意时刻优于其他方法。

综合分析图 1 和图 2,可以得出结论:嵌入维数和预测模型的选取对时用水量预测的结果都有着不可忽视的影响,如果只考虑单一嵌入维数下的单一预测模型,则很可能造成有效信息的流失,给预测带来风险。

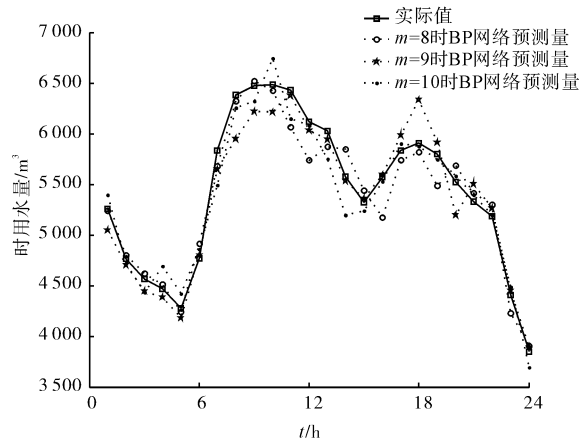


图 1 不同嵌入维数下相同预测模型预测结果与实际值比较

为了验证基于多嵌入维数的时用水量 LSSVM 组合预测模型较强的预测能力,本研究将其与传统 LSSVM 的时用水量模型及基于 BP 网络的时用水量预测模型进行对比,并采用平均相对误差 A_{MAPE} 和均方根相对误差。

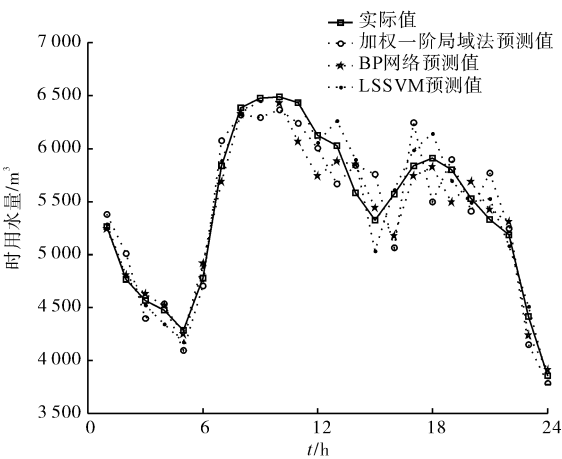


图 2 $m=8$ 时各单项预测模型预测结果与实际值比较

A_{MSE} 作为各种方法预测效果评价的依据,3 个模型预测结果与实际值的比较如图 3 所示,预测效果分析如表 1 所示。

表 1 预测效果分析

	加权一阶局域法 预测模型			BP 网络 预测模型			LSSVM 预测模型			基于多嵌 入维数的 LSSVM 组合预测
嵌入维数 m	8	9	10	8	9	10	8	9	10	
$A_{MAPE}/(\%)$	4.96	4.47	4.57	3.04	3.14	2.87	2.21	2.55	2.37	1.58
$A_{MSE}/(\%)$	5.54	5.12	5.05	3.51	3.63	3.33	2.67	2.83	2.77	1.84

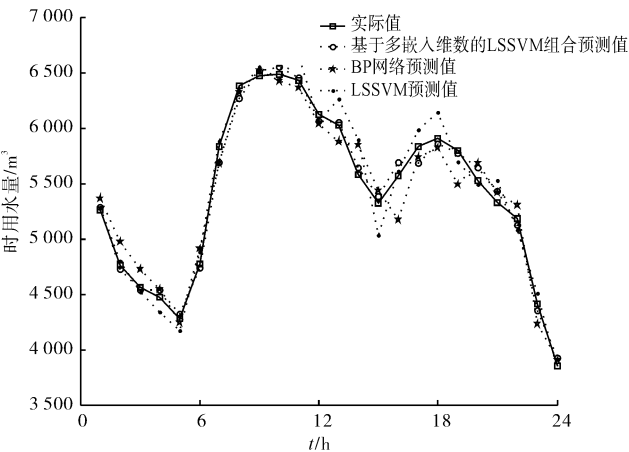


图 3 3 种模型预测结果与实际值的比较

由图 3 可以看出:各预测模型的预测结果基本上都反映出了实际时用水量的变化规律,尤其是以基于多嵌入维数的 LSSVM 组合预测模型,其预测结果与实际值的重合度最好;单一维数的单一预测模型在某些时刻误差很大,而基于多嵌入维数的 LSSVM 组合预测模型在这些时刻显示出其优良的稳定性,在一定程度上减小了预测风险。

同时,由表 1 可以看出:本研究所提出的基于多嵌入维数的 LSSVM 组合预测模型预测精度大幅提高,其预测性能明显优于单一维数下加权一阶局域法预测模型、BP 网络预测模型和 LSSVM 预测模型,能达到令人满意的预期效果。

3 结束语

本研究在综合考虑不同嵌入维数、不同预测方法所包含信息的基础上,提出了基于多嵌入维数的时用水量 LSSVM 组合预测模型。通过对某地实际时用水量进行仿真,结果证明该模型的预测精度明显高于单一维数下各个单项预测方法,能快速跟踪时用水量变化的趋势,有效地减小预测风险,具有较好的实用价值。

另一方面,本研究的预测模型是建立在不同维数多种单项预测模型的基础之上的,接下来的研究工作中如能提高单项预测的精度,则组合预测模型的预测精度能得到进一步提高。

参考文献 (References):

- [1] CHEN Lei. Genetic least squares support vector machine approach to hourly water consumption prediction[J]. **Journal of Zhejiang University : Engineering Science**, 2011, 45 (6):1100-1103.
- [2] HERRERA M, TORGO L. Predictive models for forecasting hourly urban water demand[J]. **Journal of Hydrology**, 2010, 387 (2):141-150.
- [3] SARA N, MOHAMMAD K. Pressure management model for urban water distribution networks[J]. **Water Resources Management**, 2010, 24(3):437-458.
- [4] 王丽婕. 基于多嵌入维数的风力发电功率组合预测模型[J]. **控制与决策**, 2010, 25(4):577-580.
- [5] DENG W, HU Si-ji. Non-linear combined prediction model of medium and long-term civil vehicle population of China[J]. **Journal of Transportation Systems Engineering and Information Technology**, 2010, 10(6):103-108.
- [6] TANG Hai-yan, QI Wei-gui. Prediction of elevator traffic flow based on SVM and phase space reconstruction[J]. **Journal of Harbin Institute of Technology**, 2011, 18(3):111-114.
- [7] YOU Rong-yi, HANG Xiao-jing. Phase space reconstruction of chaotic dynamical system based on wavelet decomposition[J]. **Chinese Physics B**, 2011, 20(2):114-118.
- [8] WANG Jun-guo, ZHOU Jian-zhong, PENG Bing. A hybrid neural genetic method for load forecasting based on phase space reconstruction[J]. **Kybernetes**, 2010, 39(8):1291-1297.
- [9] ISMAIL S, SHABRI A, SAMSUDIN R. A hybrid model of self-organizing maps and least square support vector machine for time-series forecasting[J]. **Expert Systems with Applications**, 2011, 38(8):10574-10578.
- [10] 程 健, 白静宜. 基于混沌时间序列的煤矿瓦斯浓度短期预测[J]. **中国矿业大学学报**, 2008, 37(2):231-235.

[编辑:李 辉]

“透明”管理,洞察企业运营智慧

近日,国内管理软件及服务供应商鼎捷软件在上海发布了“智慧企业”的全年核心战略。鼎捷软件认为,智慧企业应该是一个前瞻性的,并具备迅速、灵活、准确解决问题能力的企业,而这样的企业具备透明、协同、敏捷、精准、前瞻、创新和绿色七大基本素质。其中,“透明”在七大要素中占据首位,是指企业的业务流程、制造现场、销售数据及决策信息等透明清晰,它是一家“智慧企业”最基本的要求与保障。

鼎捷认为,“透明”的智慧企业主要指三点:第一是公开,即对公司内部发生的每件事情,不论是水平还是垂直,都能够公开、及时地看到;第二是一致,即每个员工在企业中使用的资料是一致的、相同的,因此数据在时间上或是在转移过程中都不会有偏失,从而保证了提供给财务用于生产、结算的数据是生产线真正所需要的数据;第三是可监控,因为全部的资料数据都是透明一致的,那么将更容易进行追踪、衡量、管理等一连串的行动,从而使企业管理状况得到改善。当企业走向全球运营,“透明”还指跨地域的透明,这样企业管理者能及时监控到不同地域之下所发生的现象。