

DOI:10.11918/j.issn.0367-6234.201704114

多变量时滞计算的最大独立互相关算法

耿 越, 马会超, 马 悦, 王振翀

(中国矿业大学(北京) 机电与信息工程学院, 北京 100083)

摘 要:为解决相空间重构中多变量时滞参数难以同时选择的问题,提出一种基于最大独立互相关的时滞计算方法. 将响应变量序列分段处理;对各段曲面拟合并将观测变量序列代入拟合函数;迭代运算至互相关最小,得到最优时滞. 对最大独立互相关算法与遗传神经网络、互信息法、极大联合熵法进行对比实验,并引入联合递归图与共有近邻比值法作为评价方法,结果表明:最大独立互相关算法克服了传统方法的不足. 选取某矿井下进风巷、上隅角、工作面 and 回风巷 4 个位置瓦斯浓度的真实数据进行四变量最优时滞选择实验并与互信息法对比,最大独立互相关算法的时滞计算结果为 16-3-10-11,共有近邻比为 0.58,联合递归密度为 0.34%,优于传统方法. 提出算法能够应用于实际多变量分析,具有一定实用价值.

关键词:最大独立互相关;延迟时间确定;联合递归密度;共有近邻比值;多变量瓦斯浓度

中图分类号: O415.5; TD712; TP391.4 **文献标志码:** A **文章编号:** 0367-6234(2018)01-0186-05

A calculation method for multivariate time-delay selection with the maximal independent cross-correlation algorithm

GENG Yue, MA Huichao, MA Yue, WANG Zhenchong

(School of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing 100083, China)

Abstract: To solve the problem of multivariate time-delay synchronous selection on phase space reconstruction (PSR), a maximal independent cross-correlation (MICC) algorithm was proposed to select multivariate phase space reconstruction time-delays. Firstly, the response variate sequence was segmented, and then, the segment surfaces were fitted and the observational sequences were substituted into the fitting function. At last, the optimal time-delay was computed iteratively when the cross-correlation was minimal. The simulation results of the Lorenz system were used to compare the binary and ternary time-delay selections of MICC algorithm with genetic neural networks, maximal entropy algorithm and mutual information algorithm. Joint recurrence plot (JRP) and mutual nearest neighbor ratio (MNNR) were applied to evaluate the selections, and MICC algorithm was superior to the others. Besides, the coal mine gas concentrations of four crucial undermine locations were chosen to be the one real coal mine system coupling variates. The contrast experiments between MICC and mutual information had been done and the time-delays computed using MICC were 16-3-10-11 respectively, meanwhile the MNNR and JRP densities were 0.58 and 0.34%. The results showed that the MICC algorithm had obvious advantages in selecting optimal time-delays of multivariate and could be applied on multivariate analysis in practical issues.

Keywords: maximal independent cross-correlation; time delay determination; joint recurrence density; mutual nearest neighbor ratio; multivariable gas concentrations

作为混沌理论的重要课题,混沌相空间重构技术被广泛应用于多个领域^[1-5]. 当重构嵌入维数大于系统自由度时,重构轨迹与原系统高维流行空间微分同胚,可以恢复复杂系统的动力学特征,因此相空间重构技术有着十分重要的应用价值^[6-8]. 但是该技术存在多变量时滞参数的计算与选择这一重要问题. 选择时滞过小,重构空间严重压缩扭曲;选择过大,则各变量坐标独立,很难提供系统吸引子的稳定投影^[9]. 所以对多变量最优时滞计算方法的研究

具有重要理论意义.

在传统方法中,互信息法在多变量问题的应用上受到限制^[1,10]. 陶慧等^[2]使用遗传神经网络计算煤矿冲击地压的最佳重构参数;张文专等^[3]将基于差分进化支持向量机与混沌理论相结合,根据模型预测效果进行时滞选择;但受样本数量影响,易出现过拟合问题. 张春涛等^[4]和张淑清等^[5]从信息熵的角度出发进行混沌时滞计算,利用了混沌系统的耗散性,物理意义明显,可解释性强,但计算复杂、误差较大,不易实现. Zhang 等^[11]将最小微分熵原则与神经网络模型组合,选择结果更为精准,但是计算复杂和过于依赖模型的根本问题没有得到解决. 上述

收稿日期: 2017-04-23
作者简介: 耿 越 (1989—), 男, 博士研究生;
王振翀 (1958—), 男, 教授, 博士生导师.
通信作者: 耿 越, 13011197054@163.com

方法难以推广用于多变量时滞的多参数同时选择. Palit 等^[9,12-13]针对互信息法的局限性提出了自相关法和改进高维互信息法,但高维互信息只能近似估计,误差较大;为此进一步研究了高维自相关算法,但可视化的定性评价方法仅适用于低维空间.

本文提出基于最大独立互相关算法进行多变量时滞的同步选择,并推广到井下瓦斯多变量时滞计算,解决多变量瓦斯预测参数选择困难的问题^[14]. 以联合递归图^[15-17]和共有近邻比值法^[18]评价结果的优劣. 通过实验将本文算法与上述算法比较,并利用井下多变量瓦斯数据对算法进行真实数据验证.

1 最大独立多变量互相关算法

相空间重构实质上是对原始系统运动轨迹的还原过程,而响应变量与驱动变量的重构曲线是扭曲的系统轨迹. 当时滞参数最优时,相空间得到完美重构,此时的曲线扭曲程度最小,变量间的相关性也最小. 根据 Palit 等^[13]的研究,相关系数仅能度量二元序列的线性关系,难以解决实际问题多变量非线性问题. 因此,本文将相关系数改进为

$$R_{X,Y}(\tau) = \frac{\sum_{t=1}^{N-(m-1)\tau} \{Y(t) - \overline{Y(t)}\} \{f(X(t)) - \overline{f(X(t))}\}}{\sqrt{\sum_{t=1}^N \{Y(t) - \overline{Y(t)}\}^2} \sqrt{\sum_{t=1}^{N-(m-1)\tau} \{f(X(t)) - \overline{f(X(t))}\}^2}}.$$

式中: $R()$ 为互相关, m 为嵌入维, τ 为时滞参数, $X = \{x(t)\}_{t=1}^N$ 和 $Y = \{y(t)\}_{t=1}^N$ 为两序列, N 为序列长度, $f()$ 是从 $x(t)$ 中对 $y(t)$ 进行估计的函数, $\overline{Y(t)}$ 和 $\overline{f(X(t))}$ 分别为 $Y(t)$ 和 $f(X(t))$ 对应的平均值.

为将互相关系数用于解决多变量非线性问题,改进如下:

$$R(\tau_1, \dots, \tau_p) = \frac{\sum_{p,t=1}^{N-(m-1)\tau} \{Q(t) - \overline{Q(t)}\} \{W - \overline{W}\}}{\sqrt{\sum_{p,t=1}^N \{Q(t) - \overline{Q(t)}\}^2} \sqrt{\sum_{p,t=1}^{N-(m-1)\tau_{\max}} \{W - \overline{W}\}^2}}.$$

式中: p 为变量数, Q 为含有 p 个变量的多元序列, W 为对应的估值序列.

使用最小二乘曲面拟合作为优化方法,将 p 个变量拟合成一个曲面,变量间的相关性通过局部线性距离估计.

2 延迟时间的选择评价

本文以递归图和共有近邻比值法对实验结果进行评价. 递归图是观测变量的时间序列在不同时刻访

问同一区域二元图像^[14],以黑色表示递归状态白色表示非递归状态,图 1 为 Lorenz 系统 x 变量的递归图.

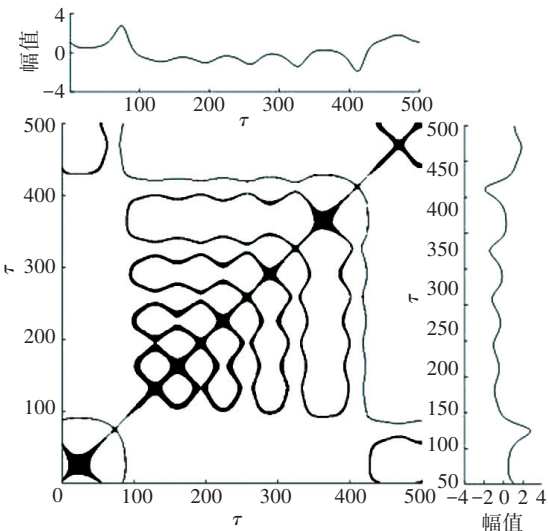


图 1 Lorenz 系统 x 变量递归图

Fig.1 The recurrence graph of Lorenz system x variable
递归图的初衷是将高维动力系统的动态演变转化为一个 N 阶方阵并进行可视化展示^[15],但可视化方法局限于三维空间展示,无法用于高维系统. 本文使用联合递归密度来度量高维空间的扭曲程度,联合递归密度^[16]如下式:

$$JR_{i,j}^M(e^{x(1)}, \dots, e^{x(p)}) = \prod_{k=1}^p R_{i,j}^{m_k}(e^{x(k)}),$$
$$M = \sum_{k=1}^p m_k, \quad i, j = 1, \dots, N.$$

式中: $JR()$ 为联合递归密度函数, M 为所有变量嵌入维之和, e 为自然对数函数的底数.

为了对延迟时间选择进行定量评价,引入共有近邻比值法^[17],度量相空间重构前后各变量轨迹共有近邻点的变化,比值越大则说明重构后共有近邻点减少的越多,相空间的扭曲程度越小重构效果越好,利用这一规律进行最优时滞选择的评价.

综上所述,本文提出的以联合递归密度和共有近邻比为评价方法的最大独立多变量互相关算法总结如下:

- 1) 初始化,设定嵌入维数、延迟时间和最大迭代次数,系统观测序列进行相空间延时重构,得到重构结果 $Y_n^{m_1}$ 和 $X_n^{m_2}$;
- 2) 对 $Y_n^{m_1}$ 进行分段处理,得到 $p + 1$ 段子序列 $\{Y_{1,n}, \dots, Y_{p+1,n}\}_{n=1}^{150-500}$;
- 3) 对 $p + 1$ 段子序列进行曲面拟合,得到函数近似估计;
- 4) 将重构后的系统驱动变量代入拟合函数,计算得到估计值 W ,进行迭代运算;
- 5) 迭代结束,最大独立互相关对应的坐标值即

为最优时滞.

3 实验验证与分析

为了验证本文算法相对其他算法的优越性,利用标准 Lorenz 混沌系统的仿真数据将本文算法与文献[2]的遗传神经网络预测法、文献[4]的极大联合熵法和传统的互信息法 3 种方法的计算结果对比,并采用递归图法和共有近邻比值法对计算结果进行评判. Lorenz 系统的微分方程为

$$\frac{dx}{dt} = s(y - x),$$
$$\frac{dy}{dt} = x(r - z) - y, \text{ , } s = 10, r = 28, b = 8/3.$$
$$\frac{dz}{dt} = xy - bz.$$

采用四阶 Runge-Kutta 法求解方程,设定相关参数分别与陶慧^[2]、张春涛^[4]一致,进行互信息和本文算法实验. 陶慧设定步长为 0.04, $x_0 = 15.34$, $y_0 = 13.68$, $z_0 = 37.91$, $m_1 = m_2 = 3$, 对 x 、 y 各采样

6 000点去除前5 500点的暂态过程;张春涛设定步长为 0.02, $x_0 = 8$, $y_0 = 9$, $z_0 = 2$, $m_1 = m_2 = m_3 = 2$, 对 x 、 y 各采样8 000点去除前5 000点的暂态过程. 直接采用文献[2,4]中的结果分别为 $\tau_1 = 3, \tau_2 = 4$ 和 $\tau_1 = 9, \tau_2 = 9, \tau_3 = 9$, 二变量和三变量的互信息法和本文算法结果如表 1 与图 2~5 (时滞为量纲一变量).

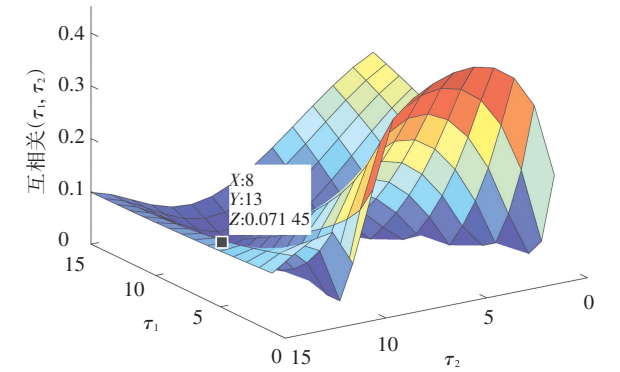


图 2 本文算法的二变量计算结果

Fig.2 The MICC calculation result for two-variate time-delays

表 1 本文算法与 3 种算法对比

Tab.1 Comparison of the paper's algorithm with 3 algorithms

方法	变量数	τ_1	τ_2	τ_3	$R(\tau_1, \cdots, \tau_p)$	共有近邻比	递归图密度/%
遗传神经网络法	2	3	4			1.00	6.18
互信息法	2	13	16			1.07	7.97
本文算法	2	13	8		8.93×10^{-6}	1.08	8.71
极大联合熵方法	3	9	9	9		0.88	3.89
互信息法	3	17	16	15		0.68	3.15
本文算法	3	15	30	5	4.21×10^{-3}	0.94	4.56

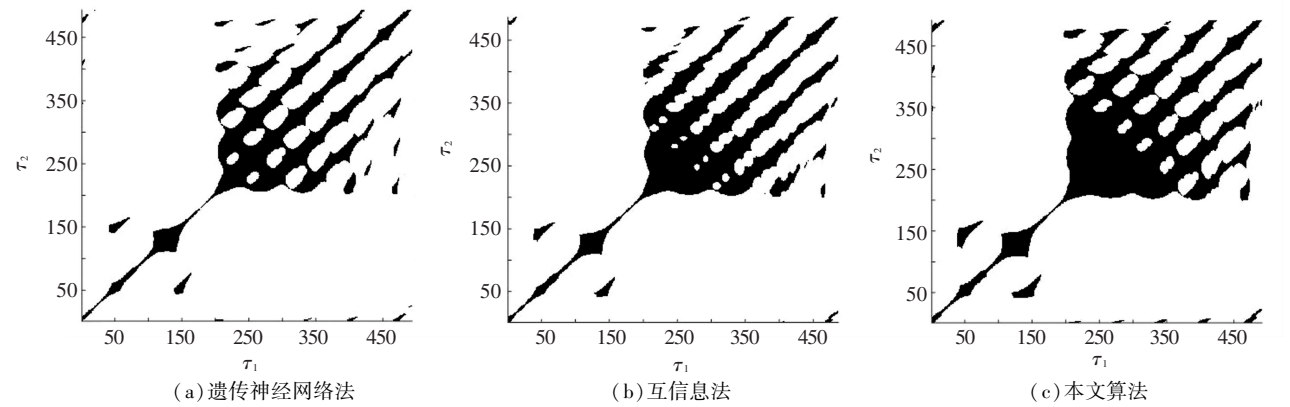


图 3 二变量时遗传神经网络法、互信息法和本文算法局部联合递归图

Fig.3 The local joint recurrence graph of genetic neural network,mutual information algorithm and MICC algorithm under the two-variable situation

对于二变量问题,本文算法选择结果共有近邻比最大(1.08),表明相空间重构后 Lorenz 系统的二元变量轨迹联合空间轨迹展开程度最大,所以共有近邻点数减少最多,并且联合递归密度最大,说明变量内部的相似度最高,能够更好地恢复原 Lorenz 系统的动态特性,即选择结果最优;其次是互信息法分

别确定的结果,再次是遗传神经网络法,文献[2]中实验结果证明的遗传神经网络法优于互信息法,是因为该方法过于依靠神经网络的非线性拟合能力,并且实验只选取了 Lorenz 系统的 500 个数据,神经网络精度依赖于带有标签信息的大数据量,迭代次数多、计算量大、普适性差,而对于相空间重构的时

滞参数选择问题很难获取带有标记的大数据量训练样本,人工标注成本较高,只能对少量样本使用. 而本文算法对数据量没有具体要求,且能够对多变量同时确定延迟时间,优于互信息法. 对于三变量问题,本文算法的共有近邻点变化最大,三变量联合递归状态呈现最多,相空间重构效果最好,选择最优,其次是极大联合熵法,再次互信息法. 极大联合熵法克服了只能进行二元选择的局限性,但是多变量熵值的计算复杂,很难推广到更多变量的情况,而本文算法避免了该问题. 但是在二元和三元变量小数据量情况下,算法的优越性没有得到最大体现,所以实验结果中的评价参数与其他算法相差较小.

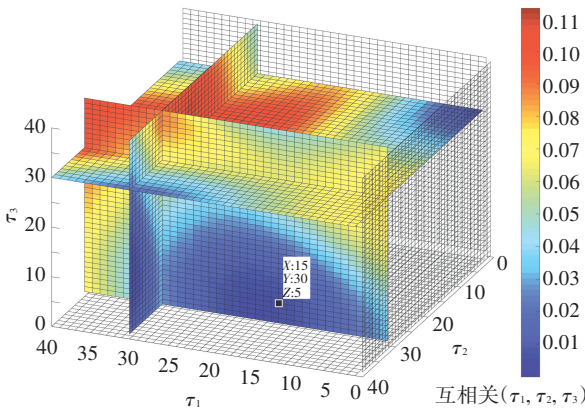


图 4 本文算法的三变量计算结果

Fig.4 The MICC calculation result for three-variate time-delays

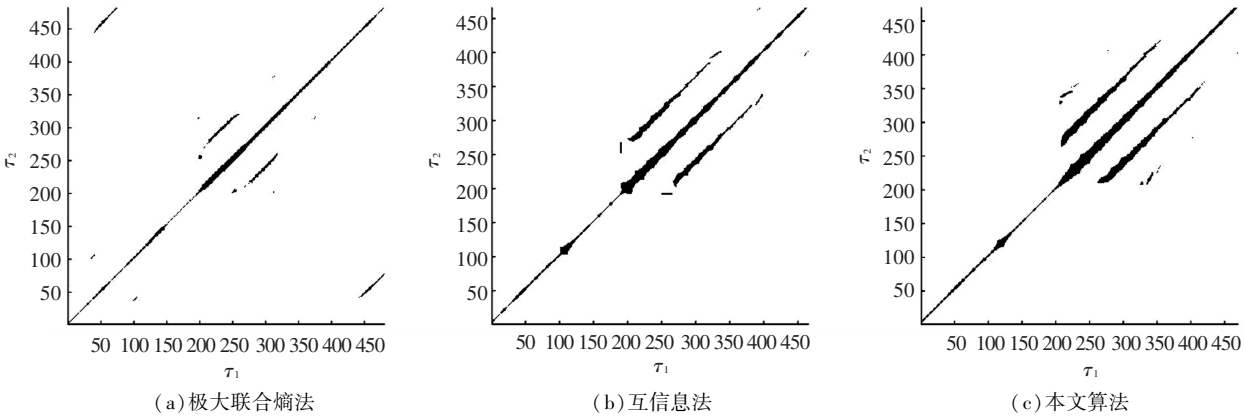


图 5 三变量时极大联合熵法、互信息法和本文算法的局部联合递归图

Fig.5 The local joint recurrence graph of maximal entropy algorithm, mutual information algorithm and MICC algorithm under the three-variable situation

为了说明本文算法的实际应用价值,对某煤矿井下巷道的瓦斯检测数据进行相空间重构,通过真实数据对算法进行验证. 将整个回采工作面视为一个局部动态复杂系统,设工作面瓦斯为该复杂系统的响应变量,进风巷、上隅角和回风巷为系统的驱动

变量,选择 2015 年 1 月 1 日至 1 月 2 日 303 回采工作面的每分钟历史检测瓦斯浓度数据,受篇幅所限在这里不具体列出数据. 设定重构嵌入维为 3,对巷道中 4 个位置的瓦斯浓度数据进行归一化处理,如表 2 所示.

表 2 进风巷-上隅角-采面-回风巷瓦斯变量的延迟时间

Tab.2 The delay time of gas variables in air intake alley-upper corner-mining face-return airway

方法	τ_1	τ_2	τ_3	τ_4	$R(\tau_1, \tau_2, \tau_3, \tau_4)$	共有近邻比	递归图密度/%
本文算法	16	3	10	11	2.27×10^{-8}	0.58	0.34
互信息法	20	1	15	20		0.47	0.18

该巷道煤岩瓦斯动态系统受进风瓦斯、上隅角瓦斯、采面瓦斯和回风巷瓦斯共同影响,互相关最小为 2.27×10^{-8} ,4 个位置对应最优时滞的归一化结果为 16-3-10-11. 进风巷瓦斯延迟时间最大,其次为采面和回风巷瓦斯时滞,上隅角瓦斯的延迟时间最小且与其他 3 个变量相差较多. 由于巷道中隅角区域不易通风,且温度湿度都较高,容易积聚瓦斯且与进风巷、采面和回风巷瓦斯的交互流通较少. 而进

风巷-采面-回风巷 3 个位置通风顺畅,对应的 3 个瓦斯变量间相互作用、相互影响,所以四变量间最大独立时,这 3 个位置瓦斯变量的延迟时间较为接近. 0.58 的共有近邻比和 0.34% 的联合递归密度说明受 4 个变量作用的局部巷道煤岩瓦斯动态系统较为复杂,导致 4 个巷道位置对应的瓦斯变量重构后的高维空间轨迹依旧严重的压缩扭曲,从而造成四变量较少的联合递归状态.

4 结 论

1) 本文针对相空间重构中多变量时滞选择计算的问题,提出了最大独立互相关算法,并引入最大联合递归密度与最大共有近邻比的方法作为实验评价标准。

2) 通过仿真实验证明了本文算法在面对二变量与三变量的时滞计算问题时优于传统方法。对井下进风巷、上隅角、采面和回风巷的瓦斯多变量进行实验,结果优于传统方法,具有现实意义。

3) 未来将研究多变量非线性混沌系统相空间重构时滞与嵌入维参数的耦合计算。

参考文献

- [1] 赵皓,高智勇,高建民,等.一种采用相空间重构的多源数据融合方法[J].西安交通大学学报,2016,50(8):84-89. DOI:10.7652/xjtub201608014.
ZHAO Hao, GAO Zhiyong, GAO Jianmin, et al. A fusion method of multisource data using phase space reconstruction[J]. Journal of Xi'an Jiaotong University, 2016, 50(8): 84-89. DOI:10.7652/xjtub201608014.
- [2] 陶慧,马小平,乔美英.基于多变量混沌时间序列的冲击地压预测[J].煤炭学报,2012,37(10):1624-1629. DOI:10.13225/j.cnki.jccs.2012.10.016.
TAO Hui, MA Xiaoping, QIAO Meiyong. Rock burst prediction on multivariate chaotic time series[J]. Journal of China coal society, 2012, 37(10): 1624-1629. DOI:10.13225/j.cnki.jccs.2012.10.016.
- [3] 张文专,龙文,焦建军.基于差分进化算法的混沌时间序列预测模型参数组合优化[J].物理学报,2012,61(22):126-132. DOI:10.7498/aps.61.220506.
ZHANG Wenzhuan, LONG Wen, JIAO Jianjun. Parameter determination based on composite evolutionary algorithm for reconstructing phase-space in chaos time series[J]. Acta Phys Sin, 2012, 61(22): 126-132. DOI: 10.7498/aps.61.220506.
- [4] 张春涛,马千里,彭宏,等.基于条件熵扩展的多变量混沌时间序列相空间重构[J].物理学报,2011,60(2):118-125.
ZHANG Chuntao, MA Qianli, PENG Hong, et al. Multivariate chaotic time series phase space reconstruction based on extending dimension by conditional entropy[J]. Acta Phys Sin, 2011, 60(2): 118-125.
- [5] 张淑清,师荣艳,张立国,等.混沌预测模型改进及在电力日负荷预测中的应用[J].仪器仪表学报,2016,37(1):208-214. DOI:10.3969/j.issn.0254-3087.
ZHANG Shuqing, SHI Rongyan, ZHANG Liguang, et al. Improvement of chaotic forecasting model and its application in power daily load forecasting[J]. Chinese Journal of Scientific Instrument, 2016, 37(1): 208-214. DOI:10.3969/j.issn.0254-3087.
- [6] TAKENS F. Detecting strange attractors in turbulence[J]. Dynamical Systems and Turbulence, 1981, 898: 366-381. DOI: 10.1007/bfb0091924.
- [7] MOHAMED G T. Handbook of research on hydroinformatics[M]. PA: Idea Group Inc (IDC), 2010: 199-228.
- [8] 吴虎胜,倪丽萍,张凤鸣,等.一种多变量时间序列的分形维数计算方法[J].控制与决策,2014,29(3):455-459. DOI:10.13195/j.kzyjc.2012.1681.
WU Husheng, NI Liping, ZHANG Fengming, et al. A method for calculation fractal dimension of multivariate time series[J]. Control and Decision, 2014, 29(3): 455-459. DOI: 10.13195/j.kzyjc.2012.1681.
- [9] PALIT S K, MUKHERJEE S, BHATTACHARYA D K. New types of nonlinear auto-correlations of bivariate data and their applications[J]. Applied Mathematics and Computation, 2012, 218(17): 8951-8967. DOI: 10.1016/j.amc.2012.02.055.
- [10] 韩敏,刘晓欣.基于互信息的分步式输入变量选择多元序列预测研究[J].自动化学报,2012,38(6):999-1006. DOI:10.3724/SP.J.1004.2012.00999.
HAN Min, LIU Xiaoxin. Stepwise input variable selection based on mutual information for multivariate forecasting[J]. 2012, 38(6): 999-1006. DOI: 10.3724/SP.J.1004.2012.00999.
- [11] ZHANG S, HU Y, BAO H, et al. Parameters determination method of phase-space reconstruction based on differential entropy ratio and RBF neural network[J]. Journal of Electronics (China), 2014, 31(1): 61-67. DOI: 10.1007/s11767-014-3125-7.
- [12] PALIT S K, MUKHERJEE S, BHATTACHARYA D K. A high dimensional delay selection for the reconstruction of proper phase space with cross auto-correlation[J]. Neurocomputing, 2013, 113(1): 49-57. DOI: 10.1016/j.neucom.2013.01.034.
- [13] PALIT S K, MUKHERJEE S, BANERJEE S, et al. Some time-delay finding measures and attractor reconstruction[M]//Applications of Chaos and Nonlinear Dynamics in Science and Engineering Understanding Complex Systems. Berlin Heidelberg: Springer International Publishing, 2015: 215-526.
- [14] 董丁稳,李树刚,常心坦,等.工作面多变量瓦斯体积分数时间序列预测模型[J].采矿与安全工程学报,2012,29(1):135-139. DOI: 10.3969/j.issn.1673-3363.
DONG Dingwen, LI Shugang, CHANG Xintan, et al. Prediction model of gas concentration around working face using multivariate time series[J]. Journal of Mining and Safety Engineering, 2012, 29(1): 135-139. DOI: 10.3969/j.issn.1673-3363.
- [15] 李新杰,胡铁松,董秀明.递归图法在径流时间序列非线性分析中的应用[J].武汉大学学报(工学版),2013,46(1):62-66+72.
LI Xinjie, HU Tiesong, DONG Xiuming. Application of recurrence plots to nonlinear analysis of runoff time series[J]. Engineering Journal of Wuhan University, 2013, 46(1): 62-66+72.
- [16] MARWAN N, ROMANO M C, THIEL M, et al. Recurrence plots for the analysis of complex systems[J]. Physics reports, 2007, 438(5): 237-329. DOI: 10.1016/j.physrep.2006.11.001.
- [17] MARWAN N, WEBBER C L Jr. Mathematical and computational foundations of recurrence quantifications[M]//Recurrence Quantification Analysis. Understanding Complex Systems. Berlin Heidelberg: Springer International Publishing, 2015: 3-43.
- [18] RULKOV N F, SUSHCHIK M M, TSIMRING L S, et al. Generalized synchronization of chaos in directionally coupled chaotic systems[J]. Physical Review E, 1995, 51(2): 980-994. DOI:10.1103/PhysRevE.51.980.

(编辑 杨 波)