

研究论文

DOI: 10.11949/j.issn.0438-1157.20161300

改进 CCM 算法检测外部扰动下系统变量间的时滞和因果关系

罗磊, 程非凡, 邱彤, 赵劲松
(清华大学化学工程系, 北京 100084)

摘要: 在化工过程中, 可以通过分析变量间的相互作用和时滞关系, 推理故障的传播路径和网络, 指出故障的根本原因。这对提高过程安全性, 增强经济效益具有重要意义, 是研究的热点和难点。常见的互相关函数和传递熵等方法, 由于只适用于线性或弱非线性系统, 或计算量较大, 往往无法准确地获得变量间的时滞信息和作用强度, 在实际应用中存在不足。近年来, 在生态领域研究中提出的交叉收敛映射(CCM)算法, 被认为是一种适用于非线性耦合过程因果分析的方法, 可适用于耦合变量间时滞关系的检测。但对于带有外部扰动的化工过程, CCM 无法根据随时受到扰动的过程数据构造出稳定的嵌入流形, 导致了时滞和因果分析失败。而基于 CCM 进行改进的扰动过滤交叉映射(DFCM)方法, 通过分析外部扰动对系统的影响, 预先筛选扰动数据, 再将过滤后的数据代入交叉映射的计算中。算例表明, 这种处理能有效地避免扰动下嵌入流形不稳定的问题, 适用于处于扰动下的化工过程, 并得到良好的时滞和因果关系分析效果。

关键词: CCM; 因果分析; 时滞检测; 算法; 过程系统; 安全

中图分类号: TP 277

文献标志码: A

文章编号: 0438—1157 (2016) 12—5122—09

An improved convergent cross mapping algorithm for causality identification and time delay analysis between systemic variables under external disturbance

LUO Lei, CHENG Feifan, QIU Tong, ZHAO Jinsong

(Department of Chemical Engineering, Tsinghua University, Beijing 100084, China)

Abstract: In chemical processes, fault propagation pathways and root cause identification could be discovered though analysis of interactions and time delay relationships among different process variables. Because of its importance in improving process safety and operation profit, fault discovery has been a popular and challenging research topic. Common methods such as correlation and entropy transfer functions, which usually cannot get accurate time delay and interaction strength between variables by limited applicability for linear and weak nonlinear systems or high computation demand, have experienced many disadvantages in actual application. Recently, a new convergent cross mapping (CCM) algorithm in ecology has been considered suitable for causality analysis and time delay identification for nonlinear coupling process variables. However, CCM fails to find application in externally disturbed chemical processes because it cannot establish stable embedded flow from process data. An improved CCM, disturbance filtered cross mapping method (DFCM), overcame many challenges of creating stable embedded flow by analyzing external disturbance, filtering disturbed process data, and applying filtrated data to CCM calculation. Case studies showed good results of time delays and causality analysis, thus DFCM could be applied to chemical processes under external disturbance.

2016-09-18 收到初稿, 2016-09-25 收到修改稿。

联系人: 邱彤。第一作者: 罗磊 (1992—), 男, 硕士研究生。

基金项目: 国家自然科学基金项目 (U1462206)。

Received date: 2016-09-18.

Corresponding author: Prof. QIU Tong, qiutong@tsinghua.edu.cn

Foundation item: supported by the National Natural Science Foundation of China (U1462206).

Key words: CCM; causality analysis; time-delay detection; algorithm; process system; safety

引 言

一直以来针对大型复杂系统的因果关系识别和时滞检测都较为困难,尤其是当系统具有非线性特征时。目前有人从不同角度出发提出了若干方法用于解决过程系统中变量间的因果关系识别和时滞分析。最简单的是互相关函数法 (CCF) [1],这一方法通过计算在不同的时滞假设点上的互相关函数值,选取其绝对值的最大值对应的时滞作为时滞检测的结果。尽管该方法计算简单,也适用于某些实际过程,但它对非线性系统适应性较差。另一种被广泛使用的格兰杰因果检验,则往往因为无法满足待检测变量间相互独立的条件,不能适用于耦合系统的分析。Schreiber 等 [2-10] 基于信息熵的概念提出了传递熵 (transfer entropy) 的方法,这一方法也被广泛地应用于化工、大气科学、神经科学等领域。但由于其基于概率计算的理论,对概率密度函数分布的计算必须足够精确以确保结果的准确性,这就带来了较大的计算量。另外,有时传递熵并不能准确地计算出变量间的时滞关系,所以无法获得类似于符号有向图 (signed directed graph) [11-17] 的结果以清晰地总结和分析结论。总之,上述的方法或者基于概率理论,计算量较大;或者只能适用于线性

或者是弱非线性系统,难以应用于对非线性系统的分析。

近年来提出的交叉收敛映射算法 (convergent cross mapping, CCM) [18],利用非线性耦合系统中变量间微分同胚映射的特点,将原有的因果关系识别转化为对嵌入流形相互预测效果的对比。Ye 等 [19] 的工作表明 CCM 算法适用于某些自振系统中变量因果关系的检测和时滞分析。但是,化工过程中处于常见的外部扰动时,Ye 等的方法也许不再适用。本文对 CCM 算法进行改进,提出 DFCM 方法,解决了该问题。

1 CCM 算法

CCM 算法由 Sugihara 等 [18] 在 2012 年提出,相比于互相关函数和传递熵,它能更清楚地把握耦合系统中变量间的相互作用,得到更准确的结果。根据 Takens 嵌入定理,在一动态系统中,相互关联两变量其对应的影子流形 M_x 和 M_y 都与原系统的流形 M 微分同胚。这里的影子流形 M_x 和 M_y 都是通过对原始数据 x 和 y 进行时序嵌入重构得到,其元素分别用 x_t 和 y_t 表示。这样,流形 M_x 和 M_y 之间便形成了连续单射的联系。对于 M_x 上选定的点 $x_{t,1}, x_{t,2}, \dots, x_{t,k}$,在 M_y 上也一定能找到对应的 $y_{t,1}, y_{t,2}, \dots, y_{t,k}$,若

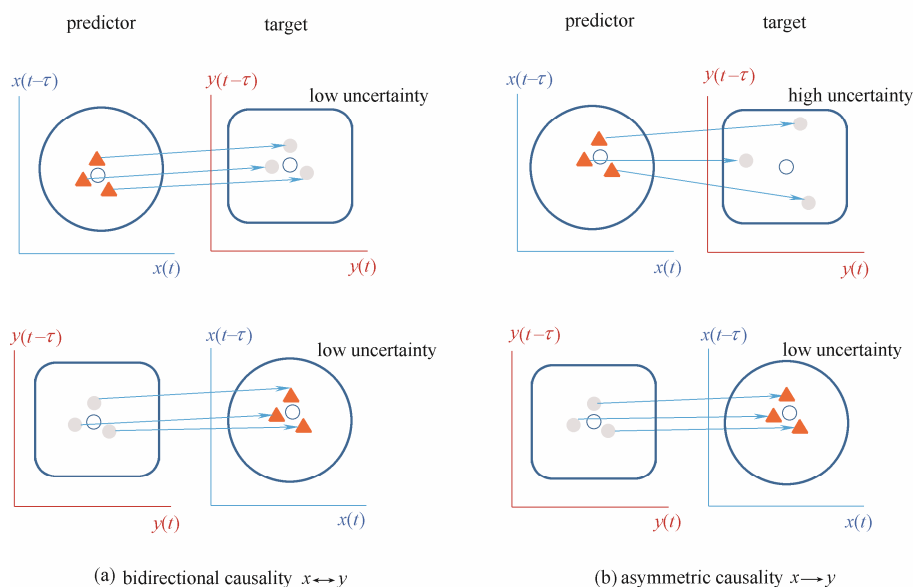


图 1 CCM 原理

Fig.1 Principle of CCM

所有的 $x_{t,i} (i=1, \dots, k)$ 均收敛于特定的 $x_{t,0}$, 那么相应地, $y_{t,i} (i=1, \dots, k)$ 也会收敛于 $x_{t,0}$ 在 M_y 上所对应的 $y_{t,0}$ 。对于 M_x 上每个点 $x_{t,i}$, CCM 都寻找其相邻点及其在流形 M_y 上的对应点, 并对 $y_{t,i}$ 进行预测, 记作 $\hat{y}_{t,i}$ 。这样变量 x 对 y 的预测能力 $\rho_{x \rightarrow y}$ 即是预测值 $\{\hat{y}_{t,i}\}$ 和真实值 $\{y_{t,i}\}$ 之间的相关系数。随着样本总数 L 的增加, 预测值序列 $\{\hat{y}_{t,i}\}$ 逐渐收敛, $\rho_{x \rightarrow y}$ 也将会收敛到一定值, 这样便可使用变量 x 预测 y , 反之亦然。然而, 如果变量 x 和 y 之间不存在耦合关系, 由于信息传递的不对称将会导致作为结果的某一变量对另一作为原因的变量预测效果的下降。图 1 解释了这一原理。

对于 n 维动态系统

$$dX/dt = f(X, U) \quad (1)$$

其中 X 和 U 分别对应变量和参数向量, $X = [x_1, x_2, \dots, x_n]^T$, $U = [u_1, u_2, \dots, u_n]^T$ 。对于 X 中的两变量 x 和 y , 其对应的影子流形构造如下

$$M_{x,t} = [x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(E-1)\tau}] \quad (2)$$

$$M_{y,t} = [y_t, y_{t-\tau}, y_{t-2\tau}, \dots, y_{t-(E-1)\tau}] \quad (3)$$

E 是嵌入维数, τ 是延迟时间, 这样根据前面所述, 便有

$$\lim_{M_{x,i} \rightarrow M_{x,k_0}} \|M_{y,i} - M_{y,k_0}\| = 0 \quad (4)$$

若在 M_x 上找到一系列最接近 M_{x,k_0} 的点 $\{M_{x,k_i}\} = \{M_{x,k_1}, M_{x,k_2}, \dots, M_{x,k_{E+1}}\}$, 那么 M_{y,k_0} 的值便可由 $\{M_{y,k_i}\}$ 进行估计

$$d(M_{x,k_i}, M_{x,k_0}) = \exp\left(-\frac{\|M_{x,k_i} - M_{x,k_0}\|}{\|M_{x,k_1} - M_{x,k_0}\|}\right) \quad (5)$$

$$\hat{M}_{y,k_0} \Big|_{M_x} = \sum_{i=1}^{E+1} \frac{d(M_{x,k_i}, M_{x,k_0})}{\sum_{j=1}^{E+1} d(M_{x,k_j}, M_{x,k_0})} M_{y,k_i} \quad (6)$$

如果样本数 L 足够大以保证 $\{M_{x,k_i}\}$ 足够接近 M_{y,k_0} , 那么最终 M_x 预测 M_y 的能力 $\rho_{x \rightarrow y}$ 将会收敛到 0 和 1 之间。

$$\lim_{L \rightarrow +\infty} \text{cov}(M_y, \hat{M}_y \Big|_{M_x}) = \rho_{x \rightarrow y} \quad (7)$$

Ye 等^[19]将 CCM 的应用拓展到对变量间时滞关系的确定中。根据他的想法, 如果变量 y 对 x 的作用存在时滞 Δt , 即 $x(t) = f(y(t - \Delta t))$, 那么使用 CCM 将 x 对 y 进行估计, 最强的作用应出现在时滞为 $+\Delta t$ 的位置, 这里正号表示 x 落后于 y 。图 2 对 Ye 等的思想进行了简要的解释。

经过计算发现, Ye 等的方法在对自振系统进行计算时能取得良好的效果, 但是, 如果系统变量振荡来源于外部输入的扰动, Ye 等的方法并没有考虑

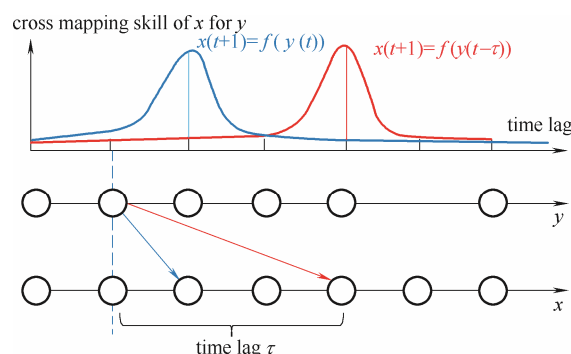


图 2 Ye 等^[19]的时滞判断原理

Fig.2 Schematic diagram of Ye *et al.*^[19] time-delay detection method

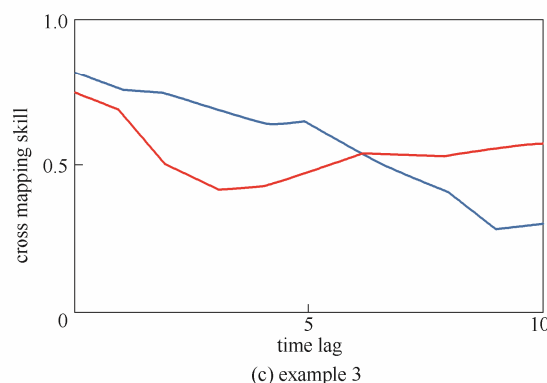
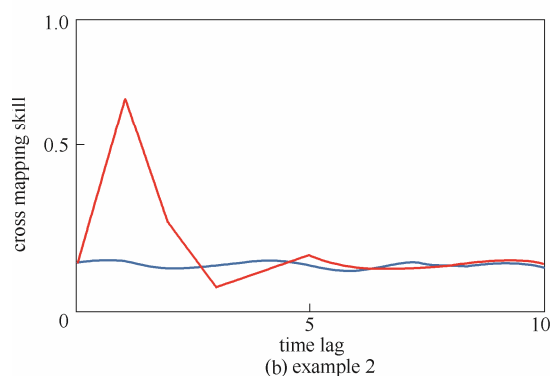
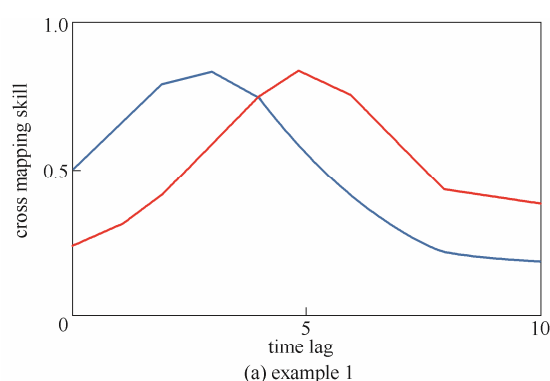


图 3 Ye 等^[19]的方法对 3 个不同算例的计算结果

Fig.3 Results of Ye *et al.*^[19] method applied to 3 examples

blue line— x maps y ; red line— y maps x

到这部分变化对 CCM 计算结果的影响, 计算结果并不理想。图 3 列出了将 Ye 等的方法应用于 3 个不同的非线性时滞系统的计算结果, 例 1 是一自振系统, 例 2 和例 3 中都加入了参数 u 的扰动, 正是该扰动导致本应稳定的系统出现了波动。从最后的结果可以看出, Ye 等的方法能准确地计算出例 1 中变量之间的时滞, 但却在例 2 和例 3 中无法得到准确的结果。

图 3 中具体算例如下。

例 1

$$x_{i+1}=x_i(3.78-3.78x_i-0.07y_{i-3})$$

$$y_{i+1}=y_i(3.77-3.77y_i-0.08x_{i-5})$$

例 2

$$x_{i+1}=0.3x_i+(1+0.3u_i-y_{i-25})x_i/5$$

$$y_{i+1}=0.9y_i+0.02(x_{i-10}-5.5)$$

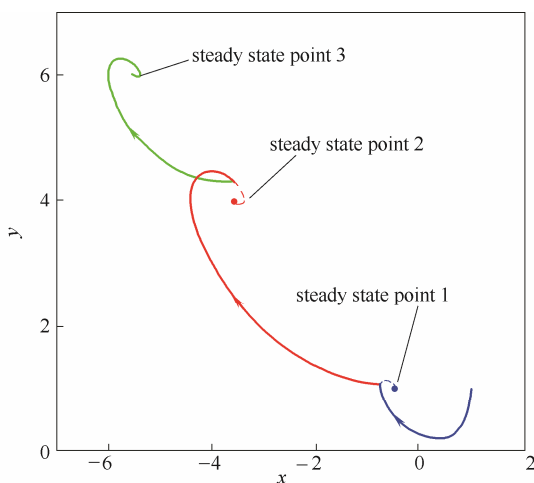
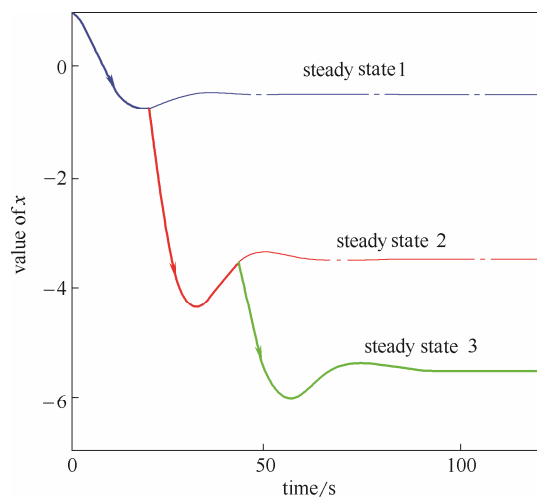


图 4 不同参数导致不同稳态解

Fig.4 Different parameters lead to different results

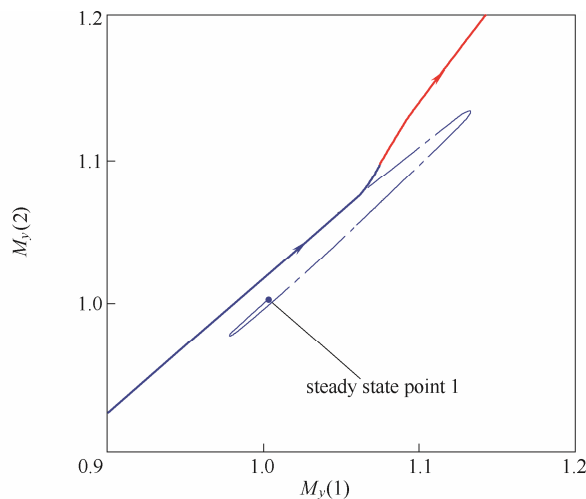
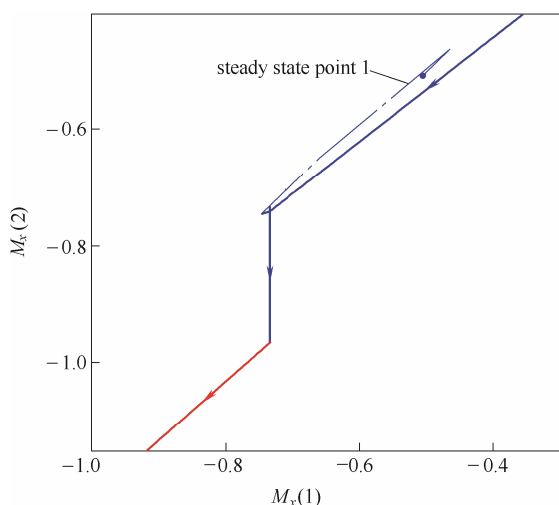


图 5 二元系统的影子流形 M_x 和 M_y

Fig.5 Shadow manifolds M_x, M_y of dual system

$$u_i = \text{rand}(0,1), i=1, 2, 3, \dots$$

例 3

$$x_{i+1}=x_i+\Delta t[0.4(1+0.1u_i)-0.2x_i+0.3y_{i-80}]$$

$$y_{i+1}=y_i+\Delta t[0.2x_{i-50}-0.3y_i-0.4(1+0.1u_i)y_i]$$

$$u_i = \text{rand}(0,1), i=1, 2, 3, \dots$$

2 扰动过滤交叉映射 (DFCM) 法

尽管 CCM 最初被应用于生态学领域, 但其中微分同胚的特点也常见于化工过程数据的重构流形中, 因此 CCM 算法的思想也适用于化工过程, 但算例 2 和例 3 中参数 u 的频繁改变使得 M_x 和 M_y 不再稳定, 这导致 CCM 算法无法识别到正确的时滞并对变量间的相互作用作出计算。以一个二元系统为例, 如图 4 所示, 如果参数 u 被设定为不同值, 该系统变量 x 和 y 将会收敛到不同的

稳态值。同时,由于稳态点的不断变化,相轨迹也会出现相互交叉,这将会导致原 CCM 方法计算准确度的下降。

$$\frac{dx}{dt} = 0.2y - u \quad (8)$$

$$\frac{dy}{dt} = -0.2x - 0.2y + 0.1 \quad (9)$$

例如图 4 中红色和蓝色轨迹的交点对应于图 5 中曲线的突变,这说明,即使在原流形 M 或者 M_y 中发现两相邻点,由于相轨迹的交叉等原因也不能保证它们在 M_x 中也满足相邻条件。还有一种可能是它们对应于不同的参数 u ,只是恰好在图中相邻而已。这样,CCM 寻找到的相邻点在其之后的计算中是没有意义的。

根据以上对 CCM 算法的分析,对其进行改进,希望它能对处于外部扰动下的系统进行计算,从而提出扰动过滤交叉映射 (disturbance filtrated cross-mapping, DFCM) 法。该方法基于以下几点假设:

- (1) 系统对于每个 u 有且仅有一个稳态, u 的扰动不会导致多稳态现象的发生;
- (2) 稳态点的变化是连续的;

(3) 如果参数 u 和 u_0 接近,那么 u 对应的从相同初值出发的相轨迹也会收敛到 u_0 所对应的相轨迹上。

图 6 对该过程进行了阐述。图中从左至右参数 u 的变化逐渐增大,不同 u 下对应的影子流形与原流形的差别也逐渐增大。这样,通过限定扰动 u 的变化便可获得稳定的流形。

图 7 是 DFCM 计算的大致流程。在 DFCM 中,首先选定合适的嵌入维数 E 和时间延迟 τ ,将过程数据根据 u 进行过滤筛选,然后再将筛选后的数据代入交叉收敛映射的计算。

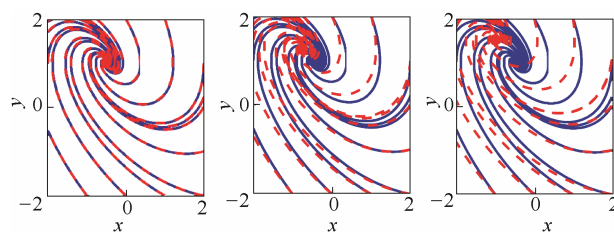


图 6 参数 u 的改变使影子流形不再稳定

Fig.6 Variation of u leads to unstable shadow manifold

blue—original manifold; red—new manifold

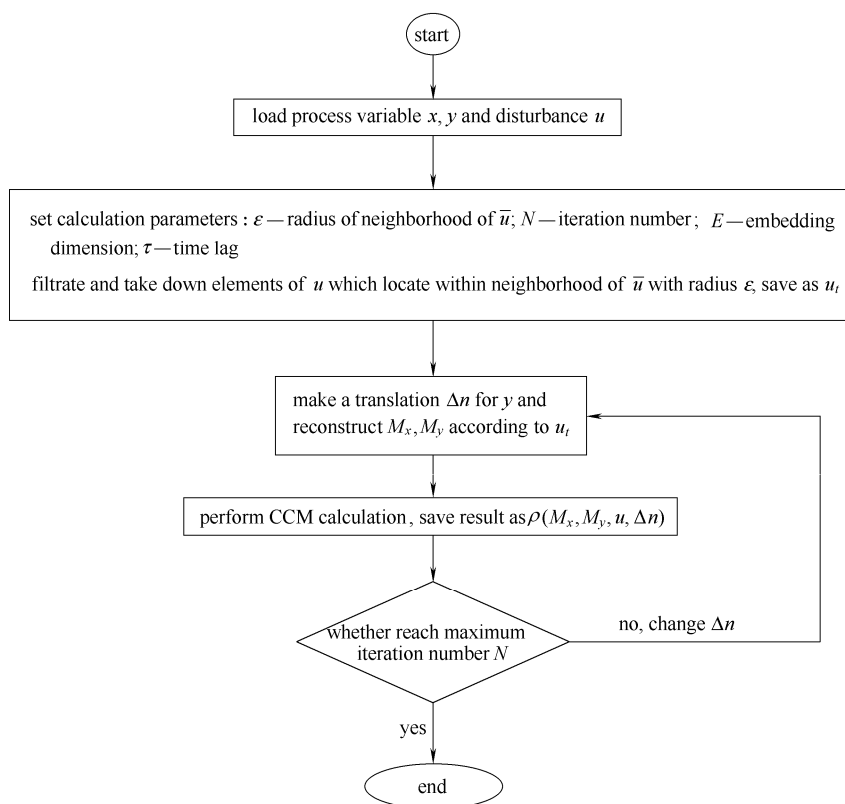


图 7 DFCM 计算流程

Fig.7 Calculation procedure of DFCM

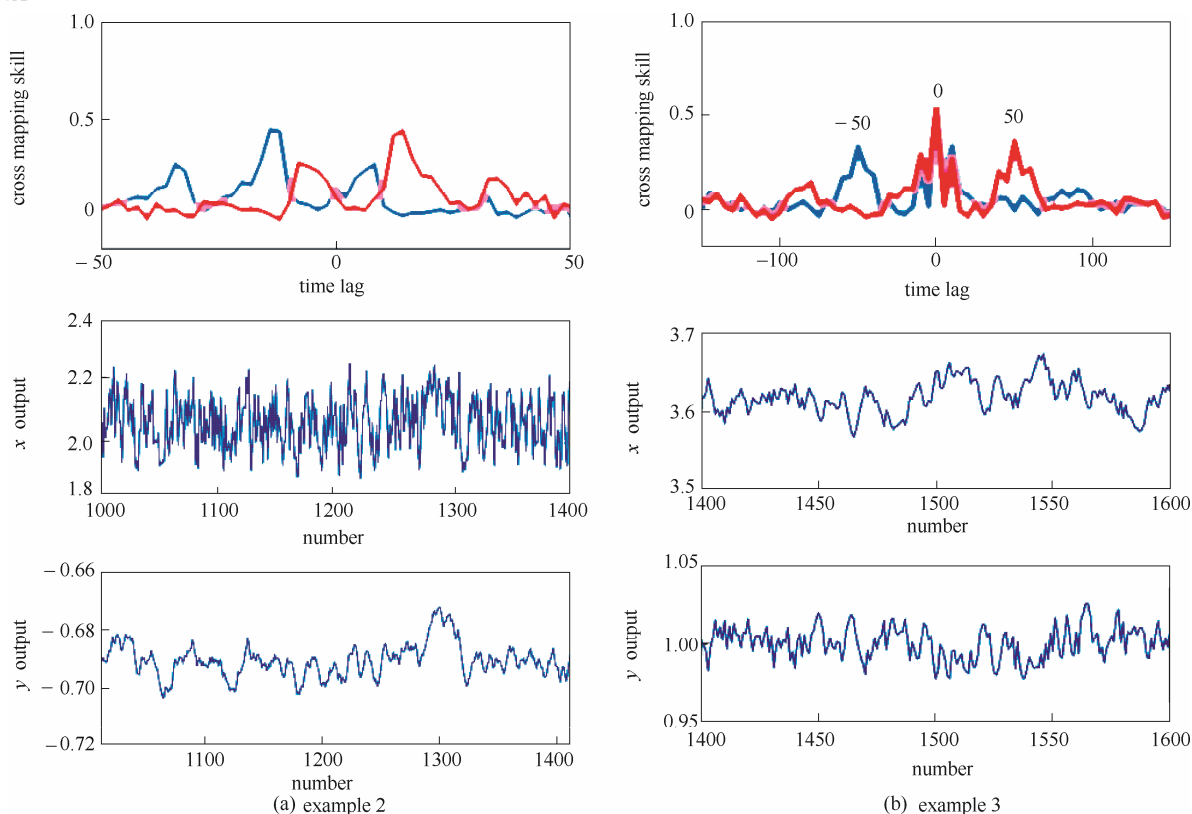


图 8 DFCM 对例 2 和例 3 的计算结果

Fig.8 Results of examples 2&3 calculated by DFCM

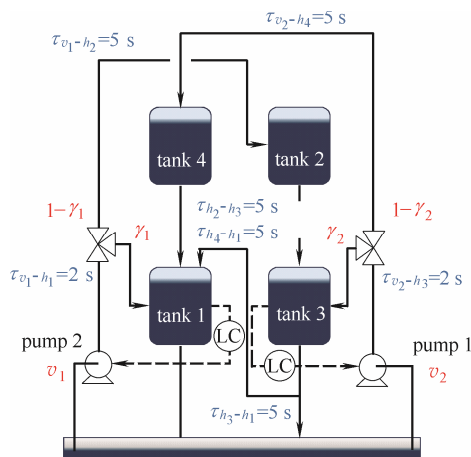
blue line— x maps y ; red line— y maps x 

图 9 四罐耦合系统

Fig.9 Schematic diagram of quadruple-tank model

3 算例

3.1 DFCM 对例 2 和例 3 的计算

图 8 显示了将 DFCM 用于之前的例 2 和例 3 后的计算结果, 对于例 2, DFCM 的计算表明最强的作用出现在变量 x 比 y 领先 $\Delta t=12$ 的位置, 这一结果与图中给出的两变量的输出结果相符合, 注意为了方便比较, 两输出的横坐标已按照 DFCM 的结

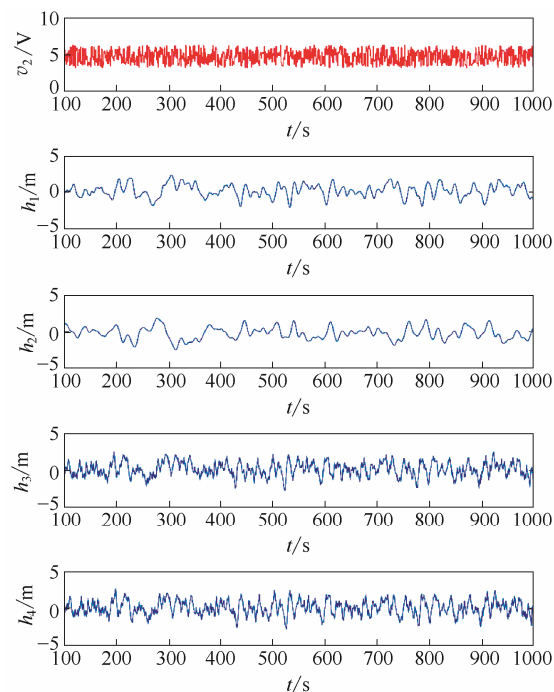


图 10 四罐耦合系统的输出

Fig.10 Outputs of quadruple-tank model

果进行了相应平移。对于例 3, DFCM 不仅在 x 比 y 领先 $\Delta t_1=50$ 和 $\Delta t_2=-80$ 发现了时滞点, 还发现两

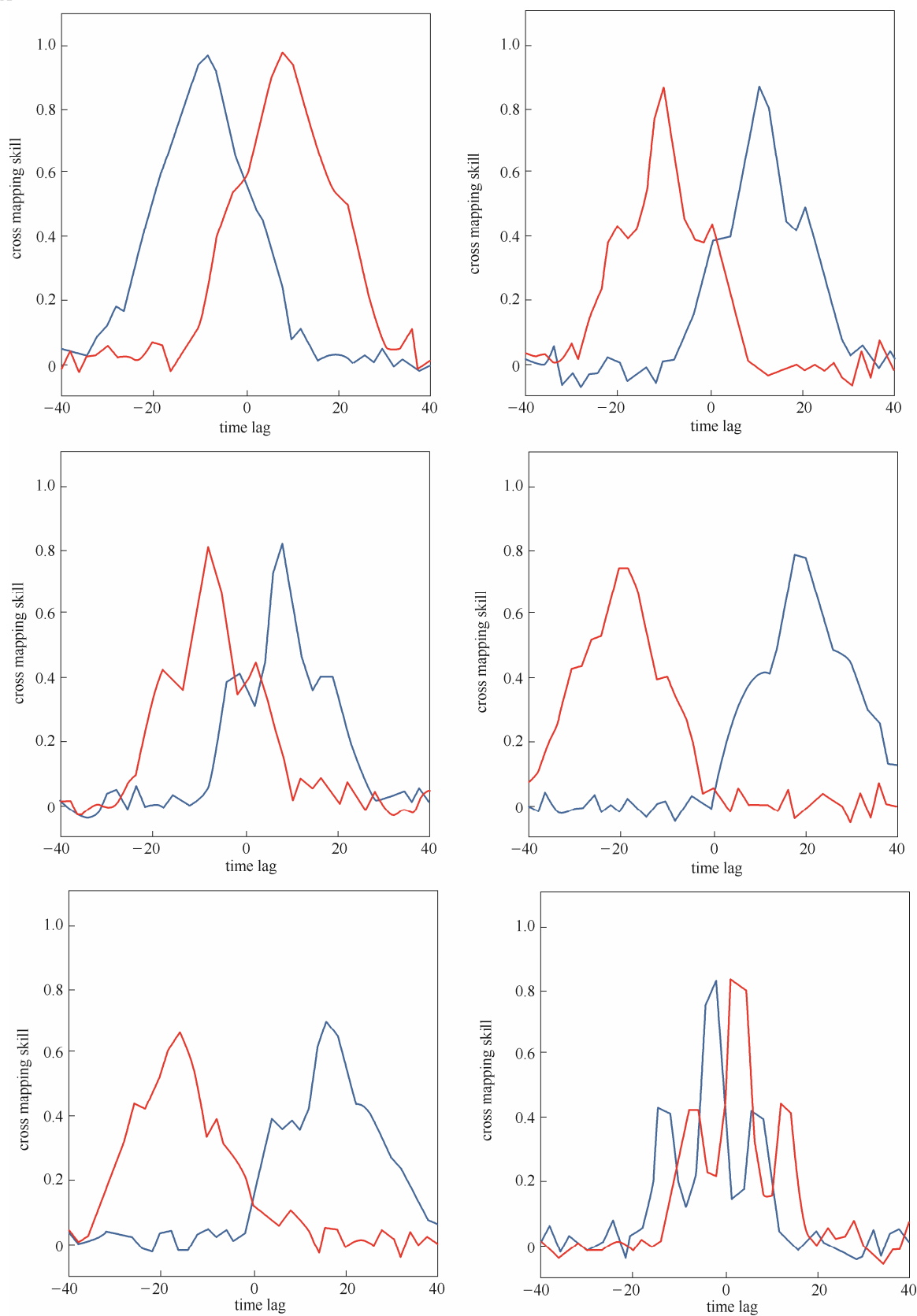


图 11 四罐串联系统时滞检测结果

Fig.11 Time-delay results of quadruple-tank model

blue line— x maps y ; red line— y maps x

变量的最强作用出现的位置是在 $\Delta t=0$ 的位置, 这与实际的输出相符合 (图中两变量呈反相关)。

3.2 DFCM 对四罐耦合系统的计算

图 9 是一个四罐耦合模型^[20], 模型输入为泵的输入电压 v_i , 其对应的流量为 $k_i v_i$, γ_i 表示流量分流比。为了维持整个系统的稳定, 分别在罐 1 和罐 3 加入了液位控制器。

$$\frac{dh_1}{dt} = -\frac{a_1}{A_1} \sqrt{2gh_1} + \frac{a_4}{A_1} \sqrt{2gh_4} + \frac{a_3}{2A_1} \sqrt{2gh_3} + \frac{\gamma_1 k_1}{A_1} v_{10} [1 - K_C (h_1 - h_{s1})] \quad (10)$$

$$\frac{dh_2}{dt} = -\frac{a_2}{A_2} \sqrt{2gh_2} + \frac{(1-\gamma_1)k_1}{A_2} v_{10} [1 - K_C (h_1 - h_{s1})] \quad (11)$$

$$\frac{dh_3}{dt} = -\frac{a_3}{A_3} \sqrt{2gh_3} + \frac{a_2}{A_3} \sqrt{2gh_2} + \frac{\gamma_2 k_2}{A_3} v_{20} [1 - K_C (h_3 - h_{s3})] \quad (12)$$

$$\frac{dh_4}{dt} = -\frac{a_4}{A_4} \sqrt{2gh_4} + \frac{(1-\gamma_2)k_2}{A_4} v_{20} [1 - K_C (h_3 - h_{s3})] \quad (13)$$

在该模型中假设液体从上游装置流往下游需要经过一段不可忽略的时间, 用蓝色数字 τ 表示。从图 9 中很容易看出罐 3 和罐 4 应该比其他两个罐更容易受到 v_2 的扰动的影响, 它们应该处于扰动传播路径的上游位置。模型的输出由图 10 所示。

从图 11 的计算结果可以看出, h_3 始终领先于其他变量, 说明它最接近扰动的根源。之后便是 h_4 、 h_1 和 h_2 。这一计算结果与实际输出和从流程结构出发的分析相符。

4 结 论

CCM 算法是一种新提出的可适用于线性和非线性系统因果分析和时滞检测算法, 但本文发现其对某些受到外部扰动的系统并不能取得好的计算效果, 原因是在外部扰动存在的情况下无法构造原方法中的稳定的流形。针对这一问题本文对原方法进行改进, 提出了扰动过滤交叉映射 (DFCM) 方法。算例证明, 该方法适用于处于外部扰动下的化工过程中耦合变量之间时滞关系的检测。未来将会进一步完善相关工作, 期待这一方法能在化工和其他领域的因果分析和时滞检测中有进一步的应用。

References

[1] BAUER M, THORNHILL N F. A practical method for identifying the propagation path of plant-wide disturbances [J]. J. Process

Control, 2008, 18: 707-719.

- [2] BAUER M, COX J W, CAVENESS M H, *et al.* Finding the direction of disturbance propagation in a chemical process using transfer entropy [J]. IEEE Trans. Control Syst. Technol., 2007, 15: 12-21.
- [3] VICENTE R, WIBRAL M, LINDNER M, *et al.* Transfer entropy — a model-free measure of effective connectivity for the neurosciences [J]. J. Comput. Neurosci., 2011, 30: 45-67.
- [4] 谢中凯, 刘国华, 吴志根. 基于传递熵的梁结构损伤动力识别[J]. 浙江大学学报(工学版), 2012, 46(10): 1880-1886.
- XIE Z K, LIU G H, WU Z G. Dynamic damage identification for beam structures based on transfer entropy [J]. J. Zhejiang Univ. (Engineering Sci.), 2012, 46(10): 1880-1886.
- [5] KUEHNERT C, BEYERER J. Data-driven methods for the detection of causal structures in process technology [J]. Machines, 2014, 2: 255-274.
- [6] SHU Y, ZHAO J. Data-driven causal inference based on a modified transfer entropy [J]. Comput. Chem. Eng., 2013, 57: 173-180.
- [7] 张志森, 龚志强, 支蓉. 利用传递熵对 Lorenz 系统和 Walker 环流信息传输方向的分析[J]. 物理学报, 2013, 62(12): 548-557.
- ZHANG Z S, GONG Z Q, ZHI R. Analysis of the direction of information transfer of Lorenz system and Walker circulation with transfer entropy [J]. Acta Physica Sinica, 2013, 62(12): 548-557.
- [8] 贺丁, 赵劲松. 基于 Hopfield 网络的时滞分析故障诊断策略[J]. 化工学报, 2013, 64(2): 633-640.
- HE D, ZHAO J S. Fault strategy of time delay analysis based on Hopfield network [J]. CIESC Journal, 2013, 64(2): 633-640.
- [9] 贺丁, 舒逸聃, 赵劲松. 基于改进的时滞分析算法的化工过程故障定位[J]. 化工学报, 2012, 63(10): 3165-3172.
- HE D, SHU Y D, ZHAO J S. An improved time-delay algorithm for chemical processes fault identification [J]. CIESC Journal, 2012, 63(10): 3165-3172.
- [10] DUAN P, YANG F, SHAH L S, *et al.* Transfer zero-entropy and its application for capturing cause and effect relationship between variables [J]. IEEE Transactions on Control Systems Technology, 2015, 23(3): 855-867.
- [11] YANG F, XIAO D. Signed directed graph modeling of industrial processes and their validation by data-based methods[C]//2010 Conference on Control and Fault Tolerant Systems. Nice, France: IEEE, 2010: 387-392.
- [12] 王杭州, 陈丙珍, 何小荣, 等. 基于开源组件的 SDG 推理平台[J]. 化工学报, 2010, 61(7): 1829-1836.
- WANG H Z, CHEN B Z, HE X R, *et al.* Open source signed digraph inference framework [J]. CIESC Journal, 2010, 61(7): 1829-1836.
- [13] 卢秉南, 张贝克, 马昕, 等. 基于 SDG 模型的控制系統故障诊断方法[J]. 化工学报, 2009, 60(9): 2243-2251.
- LU B N, ZHANG B K, MA X, *et al.* Fault diagnosis of control system based on signed directed graph model [J]. CIESC Journal, 2009, 60(9): 2243-2251.
- [14] YANG F, XIAO D Y. Progress in root cause and fault propagation analysis and large-scale industrial processes [J]. Journal of Control Science and Engineering, 2012: 478373.

- [15] DUAN P, CHEN T W, SHAH L S, *et al.* Methods for root cause diagnosis of plant-wide oscillations [J]. *AIChE Journal*, 2014, **60**(6): 2019-2034.
- [16] HE B, CHEN T, YANG X H. Root cause analysis in multivariate statistical process monitoring: integrating reconstruction-based multivariate contribution analysis with fuzzy-signed directed graphs [J]. *Computers & Chemical Engineering*, 2014, **64**(7): 167-177.
- [17] WAN Y M, YANG F, LÜ N, *et al.* Statistical root cause analysis of novel faults based on digraph models [J]. *Chemical Engineering Research and Design*, 2013, **91**: 87-99.
- [18] SUGIHARA G, MAY R, YE H, *et al.* Detecting causality in complex ecosystems [J]. *Science*, 2012, **338**: 496-500.
- [19] YE H, DEYLE E R, GILATTANZ L J, *et al.* Distinguishing time-delayed causal interactions using convergent cross mapping [J]. *Sci. Rep.*, 2015, **5**: 14750.
- [20] SADATI N, RAHMANI M, SAIF M, *et al.* Two-level robust optimal control of large scale nonlinear systems [J]. *IEEE Systems Journal*, 2015, **9**: 242-251.