信息与控制　　2021 年　第 50 卷　第 1 期：356 ~ 365

DOI：10.13976/j.cnki.xk.2021.0321　　文章编号：1002 - 0411(2021) - 01 - 0356 - 10

基于 HSIC-GL 的多元时间序列非线性 Granger 因果关系分析

李柏松[1]，任伟杰[1]，韩　敏[2]

1. 大连理工大学电子信息与电气工程学部，辽宁 大连　116024；
2. 大连理工大学工业装备智能控制与优化教育部重点实验室，辽宁 大连　116024

基金项目：国家自然科学基金资助项目(61773087)；中央高校基本科研业务费专项资金资助项目(DUT20LAB114，DUT2018TB06)

通信作者：韩敏，minhan@dlut.edu.cn

收稿/录用/修回：2020 - 07 - 13/2020 - 10 - 09/2020 - 12 - 25

摘要

因果分析是数据挖掘领域重要的研究课题之一．由于传统的 Granger 因果模型难以准确识别多变量系统的非线性因果关系，本文提出一种基于 Hilbert-Schmidt 独立性准则(Hilbert-Schmidt independence criterion，HSIC)的组 Lasso 模型的 Granger 因果分析方法．首先，利用 HSIC 将输入样本和输出样本映射到再生核 Hilbert 空间，克服了传统的 Granger 因果模型不能应用于非线性系统的缺陷．然后，建立具有组 Lasso 约束的回归模型，对多变量及其组派生变量进行因果分析，并采用贝叶斯信息准则进行模型选择，避免了人为设置滞后阶数和正则化参数．最后，根据 HSIC-GL 模型的回归系数和显著性检验结果，实现了多变量时间序列之间的非线性因果分析．通过对非线性和混沌系统的仿真实验，验证了该方法的有效性．最后将其应用于沈阳空气质量指数(AQI)和气象时间序列的因果关系分析．

关键词

多元时间序列

Granger 因果

Hilbert-Schmidt 独立性准则

组 Lasso

中图法分类号： TP18

文献标识码： A

Nonlinear Granger Causality Analysis for Multivariate Time Series Using HSIC-GL Model

LI Baisong[1], REN Weijie[1], HAN Min[2]

1. *Faculty of Electronic Information and Electrical Engineering*, *DaLian University of Technology*, *DaLian* 116024, *China*;
2. *Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education*, *Dalian University of Technology*, *Dalian* 116024, *China*

Abstract

Causality analysis is an important research topics in the field of data mining, but traditional Granger causality models have difficulty accurately identifying the nonlinear causality of multivariable systems. We propose a novel Granger causality analysis method based on the HSIC and group Lasso (HSIC-GL) model. Firstly, we use the Hilbert-Schmidt independence criterion (HSIC) to map the input and output samples into the Hilbert space of the reproducing kernel, which overcomes the inability to apply the traditional Granger causality model to nonlinear systems. Then, we establish a regression model with group Lasso constraints, which implements a causality analysis between multivariate and group-derived variables. The Bayesian information criterion is used for model selection, which prevents the artificial setting of the lag order and regularization parameters. Lastly, based on the regression coefficients and the results of significance tests of the HSIC-GL model, a nonlinear causality analysis is performed on the multivariable time series. The effectiveness of the proposed method is verified by the results of simulations of nonlinear and chaotic systems. We successfully applied this method to the air quality index and meteorological time series in Shenyang, China.

Keywords

multivariate time series;

Granger causality;

Hilbert-Schmidt independence criterion (HSIC);

group Lasso (GL)

0 引言

多元时间序列广泛存在于工业、气象、医学[1-3]等多个领域. 时间序列分析利用时间序列数据、应用数理统计等方法来挖掘系统潜在信息，揭示系统未来发展规律，是研究系统演化机理、建立系统模型的重要手段. 相较于单变量时间序列，多元时间序列往往包含更加丰富的系统信息，能够更加准确地揭示系统演化规律. 但是，多变量系统变量间的影响关系更加复杂，数据中难免会存在着与预测对象不相关的无关变量以及对预测对象作用相同或相似的冗余变量. 无关变量和冗余变量不仅会增加预测模型建立的难度，还会延长训练时间，对模型的预测效果产生负面的影响[4]. 因此，研究时间序列变量间的影响关系，为模型选择合适的输入变量具有重要的研究意义.

因果关系分析已经广泛应用于揭示多变量系统间的相互影响关系，可以有效识别复杂系统中无关变量和冗余变量. 在多变量时间序列的建模预测中，因果关系解释了因变量对于目标变量的影响关系. 通过因果关系分析可以有效地剔除无关和冗余变量，为模型选择合适的输入，达到建立准确的预测模型、提高预测精度的效果.

Granger 因果[5]是一种常用的因果关系分析方法. 它通过建立线性自回归(vector autoregressive，VAR)模型来揭示变量间的相互影响关系. Grange 因果基于可预测思想：对于两个时间序列 X 和 Y，如果 X 历史信息的加入有助于减少 Y 的预测误差，则可以说存在 $X \rightarrow Y$ 的 Granger 因果关系. Granger 因果一经提出就受到了学者们的广泛关注[6]. 但由于其只适用于二变量线性因果关系分析，具有很大局限性，因此国内外学者提出了大量改进模型[7]. 针对多元时间序列因果关系分析，Geweke 等向 VAR 模型中加入条件变量，提出了条件 Granger 因果(conditional Granger causality，CGC)模型[8]，可以有效区分直接和间接因果关系. Siggiridou 等在条件 Granger 因果模型的基础上，采用 Backward-in-time 方法对每个变量的滞后阶数进行限制选择，提出了 mBTS-CGCI(modify backward-in-time selection CGCI，mBTS-CGCI)方法[9]. 有效降低了 VAR 的模型阶数，可以实现高维时间序列的因果关系分析. 然而，由于条件 Granger 因果模型涉及大量参数估计，面对大规模数据集时存在较大困难. 另外，在模型求解时利用最小二乘方法求解得到的解并不稀疏. 而 Lasso 回归通过添加 l_1 范数惩罚项，收缩回归系数，可以产生稀疏的变量选择结果. 因此 Arnold 等[10]应用 Lasso 回归结合 Granger 因果模型，提出了 Lasso-Granger 因果模型(Lasso-GC)，该方法通过建立一个回归模型就可以实现系统所有变量的因果关系分析，极大地降低了计算复杂度. 此后，Lasso-GC 及其改进模型被广泛应用于因果关系分析[11].

随着研究的深入，学者们发现非线性系统广泛应用于各个领域. 而基于 VAR 模型的线性 Granger 因果模型可能无法准确获取非线性系统变量间的内在影响关系. 因此，研究学者也提出了大量非线性 Granger 因果改进模型. Ancona 等基于 Granger 因果关系的预测思想，提出基于径向基函数的非线性 Granger 因果模型[12]，实现了二变量的非线性因果关系分析. Marinazz 等[13]提出基于核方法的 Granger 因果(kernel Granger causality，KGC)模型. 该模型应用核函数将原始数据进行非线性映射，在再生核 Hilbert 空间(reproducing kernel Hilbert space，RKHS)中进行 Granger 因果关系分析. 除了基于核方法的非线性 Granger 因果模型，还有基于 Copula 函数的 Granger 因果模型[14]、基于神经网络的 Granger 因果模型[15]等. 此外，有学者基于信息理论，提出了条件熵[16]、基于混合嵌入的偏互信息(partial mutual information from mixed embedding，PMIME)[17]等因果关系分析方法. 它们都能有效地分析非线性系统中变量间的相互作用关系. 但这些基于信息理论的因果分析方法涉及概率密度函数的计算，当数据维度增加或样本量较大时，大规模数据的概率密度函数计算将变得异常困难，并且其计算量会成倍增加. 因此，随着数据维度及规模的增加，如何高效识别复杂多变量系统中的非线性因果关系，具有极大的挑战和研究价值.

传统的 Granger 因果及其扩展模型只利用变量的衍生规律来解释因果关系，往往忽视了冗余信息. 另外，传统的非线性因果模型，如核 Granger 因果模型，仅仅将输入变量进行非线性映射，没有对目标变量进行非线性映射，从而可能导致缺乏同等变换的物理意义解释. 因此，为了解决上述问题，本文提出了一种新的基于 Hilbert-Schmidt 独立性准则和群组 Lasso(HSIC-GL)的 Granger 因果模型，可以有效地识别多变量系统的非线性因果关系. 该模型利用 HSIC 将输入变量和目标变量同时进行非线性映射来实现 Granger 因果模型在非线性系统中的应用，借助 HSIC 实现类似于最大相关最小冗余的效果[18]，可以有效识别无关和冗余变量. 另外，当输出中涉及大量相似特征时，基于 HSIC 关联测度的特征选择模型可能会保留冗余信息[19]. 因此，建立 HSIC-GL 模型解决上述问题. 该模型可以通过建立一个回归方程实现多变量因果分析. 此外，本文使用贝叶斯信息准则(Bayesian information criterion，BIC)来确定模型阶数和群组 Lasso 的正则化参数，排除了人为因素的干扰. 最后，根据因果邻接矩阵并运用显著性检验框架实现因果关系分析. 实验结果也验证了本文方法的有效性和优越性.

1 算法介绍

针对样本集的统计独立性评价问题，本节将介绍 Hilbert-Schmidt 独立性准则. 然后，在此基础上，介绍 HSIC-GL 模型.

1.1 Hilbert-Schmidt 独立性准则

HSIC[20]是一种评价两个样本集统计独立性的核独立性准则. Gretton 等[20]在 RKHS 中引入互协方差算子，并将 HSIC 定义为互协方差算子的 Hilbert-Schmidt 范数. 设 X 和 Y 为两个随机变量，从 X 和 Y 的概率密度函数中提取样本 (X, Y). 定义两个非线性映射 $\varphi: X \rightarrow F$, $\psi: Y \rightarrow G$, F 和 G 分别表示 X 和 Y 的 RKHS，X 和 Y 对应的核函数分别为

$$k(\boldsymbol{x}, \boldsymbol{x}') = [\varphi(\boldsymbol{x}), \varphi(\boldsymbol{x}')], \ \boldsymbol{x}, \boldsymbol{x}' \in X \tag{1}$$

$$l(\boldsymbol{y}, \boldsymbol{y}') = [\psi(\boldsymbol{y}), \psi(\boldsymbol{y}')], \ \boldsymbol{y}, \boldsymbol{y}' \in Y \tag{2}$$

$\varphi: X \to F$ 和 $\psi: Y \to G$ 的互协方差算子 $C_{xy}: G \to F$ 定义：

$$C_{xy} = E_{xy}((\varphi(\boldsymbol{x}) - \mu_x) \otimes (\psi(\boldsymbol{y}) - \mu_y)) \tag{3}$$

式中，$\otimes$表示张量积，$\mu_x = E_x\varphi(\boldsymbol{x})$，$\mu_y = E_y\psi(\boldsymbol{y})$，$E_x$，$E_y$，$E_{xy}$ 表示数学期望. HSIC[20]定义为 Hilbert-Schmidt 范数的平方：

$$\begin{aligned}\mathrm{HSIC}(P_{xy}, F, G) &= \|C_{xy}\|_{\mathrm{HS}}^2 \\ &= E_{xx'yy'}(k(\boldsymbol{x}, \boldsymbol{x}')l(\boldsymbol{y}, \boldsymbol{y}')) + \\ &\quad E_{xx'}(k(\boldsymbol{x}, \boldsymbol{x}'))E_{yy'}(l(\boldsymbol{y}, \boldsymbol{y}')) - \\ &\quad 2E_{xy}(E_{x'}(k(\boldsymbol{x}, \boldsymbol{x}'))E_{y'}(l(\boldsymbol{y}, \boldsymbol{y}')))\end{aligned} \tag{4}$$

式中，$E_{xx'yy'}$表示$(\boldsymbol{x}, \boldsymbol{y})$和$(\boldsymbol{x}', \boldsymbol{y}')$的联合数学期望. 对于一个数据对 $Z = \{(x_i, y_i) \mid i = 1, \cdots, n\}$，HSIC 的经验估计[20]为

$$\mathrm{HSIC}(Z, F, G) = \frac{1}{n^2}\mathrm{tr}(\boldsymbol{KHLH}) \triangleq \mathrm{HSIC}(\boldsymbol{K}, \boldsymbol{L}) \tag{5}$$

式中，$\mathrm{tr}(\cdot)$表示矩阵的迹运算，$\boldsymbol{K}, \boldsymbol{L} \in \mathrm{R}^{n\times n}$分别为 $K_{ij} = k(x_i, x_j)$和 $L_{ij} = l(y_i, y_j)$的核矩阵. $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}} \in \mathrm{R}^{n\times n}$ 是中心化矩阵，$\boldsymbol{1} \in \mathrm{R}^n$ 为全 1 向量.

1.2 HSIC-GL 模型

Yamada 等[18]在 Lasso 回归模型的基础上，提出了一种基于 HSIC 的非线性特征选择模型. 然而，在对回归模型的研究中，Yuan 等[21]指出 Lasso 模型虽然在计算上是高效的，但其针对选择单个输入变量，而无法选择出具有解释意义的组派生变量. 并且当直接应用于回归模型时，Lasso 倾向于根据单个输入变量的强度，而不是根据一组输入变量的强度进行选择，往往会导致选择出多余的变量. 群组 Lasso 在经验风险最小化的同时，对组内和组间采用不同的惩罚约束来进行变量选择，可以有效解决上述问题.

分析一个包含 d 维特征、样本数为 n 的矩阵 $\boldsymbol{X} \in \mathrm{R}^{d\times n}$ 与目标变量为 $\boldsymbol{Y} \in \mathrm{R}^n$ 的相关性. 首先将 d 维特征分为 G 个组，d_g 表示第 g 组的特征维数. 然后，将原始输入和输出样本映射到 RKHS，并得到 Gram 矩阵 $\boldsymbol{K} \in \mathrm{R}^{n\times n}$ 和 $\boldsymbol{L} \in \mathrm{R}^{n\times n}$，最后建立 HSIC-GL 的目标函数，如式(6)所示：

$$\begin{aligned}&\min_{\boldsymbol{\alpha}\in \mathrm{R}^d} \frac{1}{2}\Big\|\bar{\boldsymbol{L}} - \sum_{k=1}^{d}\alpha_k\bar{\boldsymbol{K}}^{(k)}\Big\|_{\mathrm{Frob}}^2 + \lambda\sum_{g=1}^{G}\sqrt{d_g}\|\alpha_g\|_2 \\ &\text{s.t. } \alpha_1, \cdots, \alpha_d \geqslant 0\end{aligned} \tag{6}$$

式中，$\|\cdot\|_{\mathrm{Frob}}$表示矩阵的 Frobenius 范数，为矩阵所有元素平方和再开方. $\sqrt{d_g}$用于区分不同组的大小. $\|\cdot\|_2$ 表示采用群组 Lasso 对组内进行 l_2 惩罚约束. $\bar{\boldsymbol{K}}^{(k)} = \boldsymbol{HK}^{(k)}\boldsymbol{H}$ 和 $\bar{\boldsymbol{L}} = \boldsymbol{HLH}$ 分别是 $\boldsymbol{X}_k$ 和 $\boldsymbol{Y}$ 经过中心化的 Gram 矩阵，$\boldsymbol{H} \in \mathrm{R}^{n\times n}$是中心化矩阵. $K_{i,j}^{(k)} = K(x_{k,i}, x_{k,j})$，$L_{i,j} = L(y_i, y_j)$，$K(\boldsymbol{x}, \boldsymbol{x}')$和 $L(\boldsymbol{y}, \boldsymbol{y}')$表示核函数.

式中的第一项可以进行展开求解，如式(7)所示：

$$\begin{aligned}&\frac{1}{2}\Big\|\bar{\boldsymbol{L}} - \sum_{k=1}^{d}\alpha_k\bar{\boldsymbol{K}}^{(k)}\Big\|_{\mathrm{Frob}}^2 \\ &= \frac{1}{2}\|\bar{\boldsymbol{L}}\bar{\boldsymbol{L}}\|_{\mathrm{Frob}} - \sum_{k=1}^{d}\alpha_k\|\bar{\boldsymbol{K}}^{(k)}\bar{\boldsymbol{L}}\|_{\mathrm{Frob}} + \frac{1}{2}\sum_{k,l=1}^{d}\alpha_k\alpha_l\|\bar{\boldsymbol{K}}^{(k)}\bar{\boldsymbol{K}}^{(l)}\|_{\mathrm{Frob}} \\ &= \frac{1}{2}\mathrm{tr}(\bar{\boldsymbol{L}}\bar{\boldsymbol{L}}) - \sum_{k=1}^{d}\alpha_k\mathrm{tr}(\bar{\boldsymbol{K}}^{(k)}\bar{\boldsymbol{L}}) + \frac{1}{2}\sum_{k,l=1}^{d}\alpha_k\alpha_l\mathrm{tr}(\bar{\boldsymbol{K}}^{(k)}\bar{\boldsymbol{K}}^{(l)}) \\ &= \frac{n^2}{2}\mathrm{HSIC}(\bar{\boldsymbol{L}}, \boldsymbol{L}) - n^2\sum_{k=1}^{d}\alpha_k\mathrm{HSIC}(\bar{\boldsymbol{K}}^{(k)}, \boldsymbol{L}) + \\ &\quad \frac{n^2}{2}\sum_{k,l=1}^{d}\alpha_k\alpha_l\mathrm{HSIC}(\bar{\boldsymbol{K}}^{(k)}, \boldsymbol{K}^{(l)})\end{aligned} \tag{7}$$

式中，$\mathrm{HSIC}(\bar{\boldsymbol{L}}, \boldsymbol{L})$是一个常数，与分析结果无关. 如果第 k 个特征 $\boldsymbol{X}_k$ 与 $\boldsymbol{Y}$ 具有很强的相关性，则 $\mathrm{HSIC}(\bar{\boldsymbol{K}}^{(k)}, \boldsymbol{L})$和 α_k 都取得较大值，而式(7)取得最小值. 反之如果 $\boldsymbol{X}_k$ 与 $\boldsymbol{Y}$ 相互独立，则 HSIC 的值接近于 0，对应的系数 α_k 也接近于零. $\sum_{k,l=1}^{d}\alpha_k\alpha_l\mathrm{HSIC}(\bar{\boldsymbol{K}}^{(k)}, \boldsymbol{K}^{(l)})$可以衡量 $\boldsymbol{X}_k$ 与 $\boldsymbol{X}_l$ 之间的冗余信息，冗余特征得到较大的 $\sum_{k,l=1}^{d}\alpha_k\alpha_l\mathrm{HSIC}(\bar{\boldsymbol{K}}^{(k)}, \boldsymbol{K}^{(l)})$值，从而获得较小的 HSIC 值. 因此，可以看到该方法可以最大化相关特征，最小化冗余信息，达到去除无关和冗余信息，选择相关变量的目的.

2 基于 HSIC-GL 的 Granger 因果关系分析

本节将详细描述基于 HSIC-GL 的 Granger 因果分析模型(HSIC-GL-GC)，并推导出群组 Lasso 的求解算法. 结尾还将给出 HSIC-GL-GC 的伪代码和复杂度分析.

2.1 平稳性分析

Granger 因果模型是针对平稳时间序列的因果关系分析方法. 因此，在进行建模之前，需要对数据进行平稳性分析. 平稳性分析最常用的方法为单位根检验，如果时间序列具有单位根，则表明该时间序列是非平稳的，反之则为平稳时间序列. 本文采用 ADF (augmented Dickey-Fuller)[22]检验方法来检验时间序列平稳性，ADF 如式(8)所示：

$$\Delta X_t = \alpha + \beta t + \delta X_{t-1} + \sum_{i=1}^{m}\beta_i\Delta X_{t-i} + \varepsilon_t \tag{8}$$

式中，α 是一个常数，β 是趋势项系数，m 是自回归模型的阶数，ε_t 是误差项. 在零假设下进行单位根检验，如果接受零假设 $H_0: \delta = 0$，则表明时间序列是非平稳的. 如果时间序列是非平稳的，则需要采用差分方法实现平稳化. 反之如果拒绝原假设，而接受备择假设 $H_1: \delta < 0$，则表明时间序列是平稳的，可以直接进行 Granger 因果关系分析.

2.2 模型选择

在进行模型拟合时，模型阶数对拟合精度影响较大. 如果 VAR 模型中解释变量的最大滞后阶数 p 太小，则可能存在残差自相关，导致参数估计不一致. 适当增加 p 可以解决上述问题. 但是，如果 p 过大，则会导致估计参数过多，自由度严重降低，直接影响模型参数估计的有效性[23]. 另外，在群组 Lasso 模型中，正则化参数 λ 对变量的选择结果影响很大，如果正则化参数值过大，可能会错误剔除相关特征. 反之，如果正则化参数值过小，则可能保留不相关或冗余的特征. 因此，本文采用 BIC[23]来自动选择合适的模型阶数和正则化参数 λ，以消除人为选择的不确定性干扰. BIC 如式(9)所示：

$$\mathrm{BIC} = k\ln n - 2\ln L \tag{9}$$

式中，L 表示似然函数，k 是模型有效参数的个数，n 是样本个数. 其中，对于线性回归模型 $y = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X} + \varepsilon$，即 $P(y \mid \boldsymbol{X}, \theta) \sim N(y \mid \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}, \sigma^2)$，其似然估计函数为

$$L(\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}_i)^2 \tag{10}$$

式中，$\sum_{i=1}^{n}(y_i-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}_i)^2$ 为模型残差平方和 RSS($\boldsymbol{W}$). 设模型残差为 RSS，去掉式(10)常数项 $-\dfrac{n}{2}\ln(2\pi\sigma^2)$ 后进行简化. BIC 可以近似改写为

$$\mathrm{BIC}=k\ln n+n\ln \mathrm{RSS} \tag{11}$$

式中，第一项表示模型拟合的优良性，第二项表示对模型复杂度的惩罚力度，可以有效刻画构建模型的准确度.

2.3　HSIC-GL-GC 求解

考虑一组长度为 n 的矩阵 $\boldsymbol{X}\in\mathrm{R}^{n\times d}$，$\boldsymbol{Y}\in\mathrm{R}^{n\times 1}$. 本文通过建立如式(6)所示的 HSIC-GL 模型，然后求解群组 Lasso 获得稀疏的变量选择结果.

本文采用高斯核函数对 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 进行非线性映射，如下所示：

$$K(\boldsymbol{x},\boldsymbol{x}')=\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{x}')^2}{2\sigma_x^2}\right) \tag{12}$$

$$L(\boldsymbol{y},\boldsymbol{y}')=\exp\left(-\frac{(\boldsymbol{y}-\boldsymbol{y}')^2}{2\sigma_y^2}\right) \tag{13}$$

式中，σ_x，σ_y 为核宽，$\boldsymbol{x}'$，$\boldsymbol{y}'$为核中心. 通过式(12)、式(13)将其进行非线性映射，然后应用 HSIC 得到 Gram 矩阵，并利用 Shooting 算法[21]求解如式(6)所示的群组 Lasso 模型. 文[21]指出式(6)的第一项可以利用矢量化算子改写为如式(14)：

$$\frac{1}{2}\left\|\mathrm{vec}(\bar{\boldsymbol{L}})-[\mathrm{vec}(\bar{\boldsymbol{K}}^{(1)}),\cdots,\mathrm{vec}(\bar{\boldsymbol{K}}^{(d)})]\boldsymbol{\alpha}\right\|_2^2 \tag{14}$$

式中，vec(·)表示矢量化算子. 然后，根据 Shooting 算法[21]，式(6)解的充分必要条件如式(15)和式(16)所示.

$$(\mathrm{vec}(\bar{\boldsymbol{K}}^{(g)}))^{\mathrm{T}}(\mathrm{vec}(\bar{\boldsymbol{L}})-\mathrm{vec}(\bar{\boldsymbol{K}})\boldsymbol{\alpha}_{-g}-\mathrm{vec}(\bar{\boldsymbol{K}}^{(g)})\boldsymbol{\alpha}_g)+\frac{\lambda\sqrt{d_g}\boldsymbol{\alpha}_g}{\|\boldsymbol{\alpha}_g\|}=\boldsymbol{0},\ \forall\boldsymbol{\alpha}_g\neq\boldsymbol{0} \tag{15}$$

$$\left\|-(\mathrm{vec}(\bar{\boldsymbol{K}}^{(g)}))^{\mathrm{T}}(\mathrm{vec}(\bar{\boldsymbol{L}})-\sum_{k\neq g}(\mathrm{vec}(\bar{\boldsymbol{K}}^{(k)}))\boldsymbol{\alpha}_g)\right\|\leqslant\lambda\sqrt{d_g},\ \forall\boldsymbol{\alpha}_g=0 \tag{16}$$

其中，$(\mathrm{vec}(\bar{\boldsymbol{K}}^{(g)}))^{\mathrm{T}}\mathrm{vec}(\bar{\boldsymbol{K}}^{(g)})=\boldsymbol{I}$. 由式(15)可以求得

$$\boldsymbol{\alpha}_g=\left(1-\frac{\lambda\sqrt{d_g}}{\|\boldsymbol{S}_g\|}\right)+\boldsymbol{S}_g,\ g\in\{1,\cdots,G\} \tag{17}$$

式中，$\boldsymbol{S}_g=(\mathrm{vec}(\bar{\boldsymbol{K}}^{(g)}))^{\mathrm{T}}(\mathrm{vec}(\bar{\boldsymbol{L}})-\mathrm{vec}(\bar{\boldsymbol{K}})\boldsymbol{\alpha}_{-g})$，$\boldsymbol{\alpha}_{-g}=[\boldsymbol{\alpha}_1^{\mathrm{T}},\cdots,\boldsymbol{\alpha}_{g-1}^{\mathrm{T}},\boldsymbol{0},\boldsymbol{\alpha}_{g+1}^{\mathrm{T}},\cdots,\boldsymbol{\alpha}_G^{\mathrm{T}}]$. 然后，通过迭代即可求得全部解，并根据该解得到因果邻接矩阵.

2.4　显著性检验

在因果关系分析中进行显著性检验能够有效消除偶然因素导致的虚假因果，使结果更具可信度. 进行显著性检验时，首先对原始时间序列 $X_k(k=1,2,\cdots,d)$进行随机移位，得到置换时间序列，然后进行置换检验，分析原始序列的因果关系与置换序列的因果关系的检验统计性. 最后，根据显著检验结果推断原始序列因果关系的有效性. 置换时间序列按式(18)构造.

$$\tilde{X}_k=\{x_{s+1},\cdots,x_n,x_1,\cdots,x_s\} \tag{18}$$

式中，s 表示移位因子. 显著性检验的完整性步骤如下：

步骤 1　计算原始时间序列的因果关系，得到因果邻接矩阵.

步骤 2　根据式(18)对原始时间序列移位，获得 100 组不同置换时间序列.

步骤 3　分别对这 100 组置换时间序列进行因果关系分析，获得每一组置换时间序列的因果邻接矩阵.

步骤 4　提出 H_0 假设：置换时间序列的因果关系邻接矩阵与原时间序列因果关系邻接矩阵相同. H_1 假设：置换时间序列的因果关系邻接矩阵与原时间序列因果关系邻接矩阵不同.

步骤 5　计算 p_{value}. 如果 $p_{\mathrm{value}}>0.95$，即原始时间序列的因果系数大于 95% 置换时间序列因果系数，则拒绝显著性水平为 $\alpha=0.05$ 的原假设，证明了原始时间序列因果关系的有效性. 反之接受原假设，则表明得到的原始因果关系为虚假因果关系.

2.5　总结与讨论

首先，总结 HSIC-GL-GC 的伪代码，然后对其计算复杂度进行了分析. 基于 HSIC-GL 的 Granger 因果关系分析伪代码如算法 1 所示.

算法 1　HSIC-GL-GC

输入：时间序列{$\boldsymbol{X}$, $\boldsymbol{Y}$}，其中 $\boldsymbol{X}=[X_1,X_2,\cdots,X_d]^{\mathrm{T}}\in\mathrm{R}^{d\times n}$，$\boldsymbol{Y}=[y_1,y_2,\cdots,y_n]\in\mathrm{R}^{1\times n}$；
输出：因果关系邻接矩阵.

```
平稳性检验：ΔX_t = α + βt + δX_{t-1} + Σ_{i=1}^{m} β_i ΔX_{t-i} + ε_t；
for p = 1 : p_max (p_max是设定的最大滞后阶数)
  利用 HSIC 将 X 和 Y 映射到 RKHS，分别得到对
  应的 Gram 矩阵 K̄ 和 L̄；
  for λ = λ_min : λ_max
    根据式(6)构建回归模型；
    根据计算 BIC；
  end for
  根据计算得到的最小 BIC 值确定模型阶数和惩
  罚系数，建立模型并计算 α；
  end for
for k = 1 : d
    if p_value(k) > 0.95，即通过显著性检验
      存在 X_k→Y 的因果关系；
    else if
      不存在 X_k→Y 的因果关系；
    end if
end for
输出：因果关系分析结果.
```

在实际应用中，算法的计算复杂度是一个重要的评价标准，它决定了算法在实际中的应用普及程度. 本节分析了 HSIC-GL-GC 的时间复杂度，并与 mBTS-CGCI 和 Lasso-GC 方法进行对比. 对于 n 个样本的时间序列数据{$\boldsymbol{X}$, $\boldsymbol{Y}$}，首先对 d 维输入变量进行相空间重构，得到 b 维输入特征. 然后建立一个包含 b 维特征的 VAR 模型，其计算复杂度介于 $O(bn)$到 $O(b^2n^2)$之间[10]. 设 P 为设定的模型阶数的选择范围，则确定模型阶数，并创建 VAR 模型的计算复杂度介于 $O(Pb^2n^2)$到 $O(Pbn)$之间.

对于 mBTS-CGCI，因为需要对 d 维输入变量进行两两计算，所以其时间复杂度介于 $O(Pdb^2n^2)$ 到 $O(Pdbn)$ 之间.

Lasso-GC 模型常用最小角回归算法求解，通过建立一个回归模型就可以实现所有输入与输出的特征分析. 因此 Lasso-GC 的时间复杂度介于 $O(Pb^2n^2)$ 到 $O(Pbn)$ 之间[10].

对于 HSIC-GL-GC，计算 Gram 矩阵的时间复杂度为 $O(bn^2)$，而求解 HSIC-Lasso 模型的时间复杂度为 $O(n^3)$[24]. 假设将原始特征分为 G 组，则求解 HSIC-GL 的时间复杂度为 $O(Gn^3)$，因此 HSIC-GL-GC 的计算复杂度为 $O(bn^2+PGn^3)$.

从上述分析可以看出，本文方法的时间复杂度低于 mBTS-CGCI 方法，高于 Lasso-GC 模型. 虽然时间复杂度高于 Lasso-GC 模型，但仍处于合理范围之内. 相比于 Lasso-GC，本文方法增加了核映射和群组 Lasso 约束，扩展了模型的应用范围，并且提高了算法的分析精度，可以广泛应用于多变量系统的非线性因果关系分析.

3　仿真实验与分析

为了验证本文方法的有效性，选取两个多变量非线性系统进行仿真. 然后利用 HSIC-GL-GC 分析了中国沈阳的空气质量指数和气象时间序列. 将仿真结果与 mBTS-CGCI[9]、Lasso-GC[10]、KGC[13] 和 PMIME[17] 进行了对比.

3.1　多变量非线性系统

第一组标杆数据是一个包含 4 维变量的 5 阶非线性 VAR 系统 $\mathrm{VAR}_4(5)$[25]. 由式(19)产生：

$$\begin{aligned}
x_{1,t}&=0.8x_{1,t-1}+0.65x_{2,t-4}+\varepsilon_{1,t}\\
x_{2,t}&=0.6x_{2,t-1}+0.6x_{4,n-5}^2+\varepsilon_{2,t}\\
x_{3,t}&=0.5x_{3,t-3}-0.6x_{1,t-1}^2+0.4x_{2,t-4}+\varepsilon_{3,t}\\
x_{4,t}&=1.2x_{4,t-1}-0.7x_{4,t-2}+\varepsilon_{4,t}
\end{aligned}\tag{19}$$

式中，$\varepsilon_i(i=1,2,3,4)$ 表示高斯白噪声序列. 可以看到，系统中真实存在的因果关系为 $X_1\to X_3$，$X_2\to X_1$，$X_2\to X_3$ 和 $X_4\to X_2$，其因果影响关系如图 1 所示.

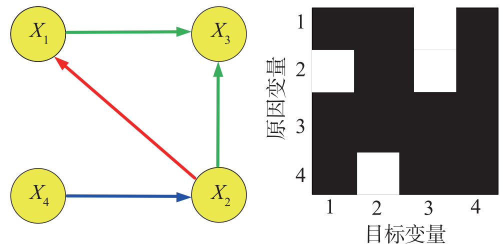
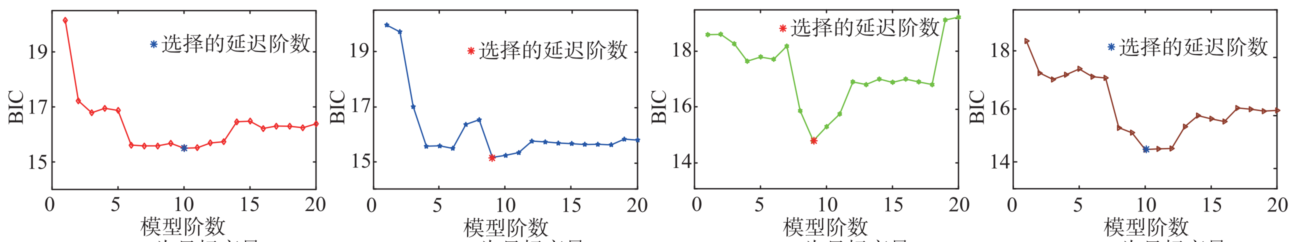


图 1　$\mathrm{VAR}_4(5)$ 真实因果关系（黑色表示存在因果关系）

Fig.1　The real causality diagram of $\mathrm{VAR}_4(5)$ (Black indicates that there is causality)

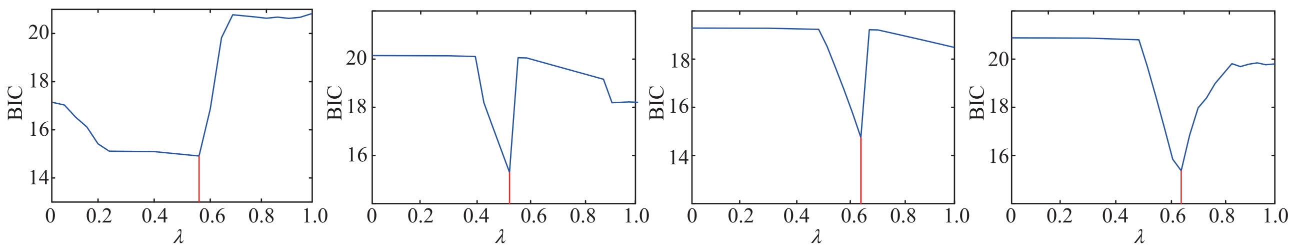
图 2 和图 3 是利用 BIC 算法选择模型阶数和惩罚系数的三维图像的二维平面展示. 子图(a)、(b)、(c)、(d)分别表示目标变量为 X_1、X_2、X_3、X_4 的实验结果. 最小的 BIC 值对应着最优参数，即实验中选择的模型阶数和正则化参数. 图 4 是不同方法得到的因果邻接矩阵. 从图中可以看到，Lasso-GC 和 mBTS-CGCI 得到了虚假因果 $X_3\to X_1$ 的非零因果系数. 此外，mBTS-CGCI 没有识别出真正的非线性因果关系 $X_1\to X_3$ 和 $X_4\to X_2$. 类似地，Lasso-GC 没有识别出的因果关系有 $X_2\to X_3$ 和 $X_4\to X_2$. 这是因为这两种方法只适用于线性系统的因果关系分析，对于非线性系统可能产生错误因果识别. 同样的虚假因果非零系数也发生在 PMIME 和 HSIC-GL-GC 上. 但是，不同的是这两种方法的虚假因果关系邻接系数远远小于真实因果关系的邻接系数，因此可以进行有效区分.



(a) X_1为目标变量　(b) X_2为目标变量　(c) X_3为目标变量　(d) X_4为目标变量

图 2　$\mathrm{VAR}_4(5)$ 模型阶数选择

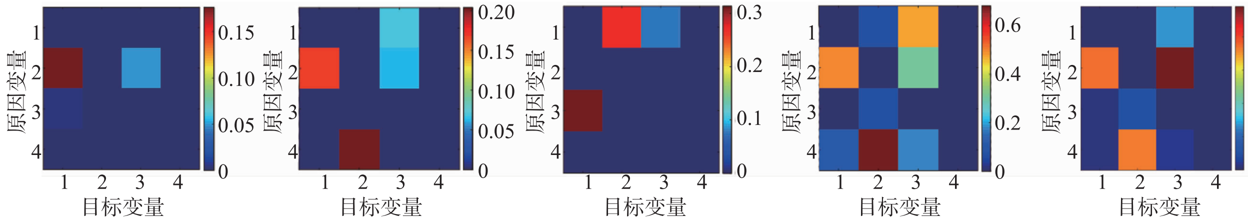
Fig.2　The model order selection of $\mathrm{VAR}_4(5)$



(a) X_1为目标变量　(b) X_2为目标变量　(c) X_3为目标变量　(d) X_4为目标变量

图 3　$\mathrm{VAR}_4(5)$ 惩罚系数选择

Fig.3　The penalty coefficients selection of $\mathrm{VAR}_4(5)$



(a) mBTS-CGCI　(b) Lasso-GC　(c) KGC　(d) PMIME　(e) HSIC-GL-GC

图 4　$VAR_4(5)$ 的因果邻接矩阵

Fig.4　The causal adjacent matrix of $VAR_4(5)$

表 1 为 $VAR_4(5)$ 的显著性检验结果．从表 1 可以看到，KGC、PMIME 和 HSIC-GL-GC 的真实因果均通过了显著性水平为 0.05 的检验，而 mBTS-CGCI 在 $X_1 \to X_3$ 和 $X_4 \to X_2$，以及 Lasso-GC 在 $X_2 \to X_1$、$X_2 \to X_3$ 和 $X_4 \to X_2$ 的检验结果均未通过显著性水平为 0.5 的显著性检验，而虚假因果关系也存在类似的结果．

表 1　$VAR_4(5)$ 显著性检验结果

Tab.1　The significance test results of $VAR_4(5)$

方法	mBTS-CGCI	Lasso-GC	KGC	PMIME	HSIC-GL-GC
$X_1 \to X_2$	2	95	0	37	0
$X_1 \to X_3$	15	96	98	98	100
$X_1 \to X_4$	2	4	5	1	3
$X_2 \to X_1$	98	2	100	100	99
$X_2 \to X_3$	95	4	96	99	96
$X_2 \to X_4$	0	0	1	0	0
$X_3 \to X_1$	2	16	3	0	1
$X_3 \to X_2$	1	2	0	62	3
$X_3 \to X_4$	0	1	2	3	0
$X_4 \to X_1$	0	2	0	4	0
$X_4 \to X_2$	45	68	97	98	96
$X_4 \to X_3$	2	0	0	25	3

3.2　Henon 非线性耦合系统

第二组仿真数据来自于 Henon 系统．Henon 耦合系统是一个具有复杂关联的多变量混沌系统．三变量非线性 Henon 耦合系统[26]的方程如下：

$$\begin{aligned} x_{1,t} &= 1.4 - x_{1,t-1}^2 + 0.3x_{1,t-2} \\ x_{2,t} &= 1.4 - cx_{1,t-1}x_{2,t-1} - (1-c)x_{2,t-1}^2 + 0.3x_{2,t-2} \\ x_{3,t} &= 1.4 - cx_{2,t-1}x_{3,t-1} - (1-c)x_{3,t-1}^2 + 0.3x_{3,t-2} \end{aligned} \tag{20}$$

式中，c 表示耦合强度．在实验中，分别设置 $c=0$ 和 $c=0.5$进行仿真实验．当 $c \neq 0$ 时，存在的因果关系为 $X_1 \to X_2$、$X_2 \to X_3$，其真实因果关系如图 5 所示．

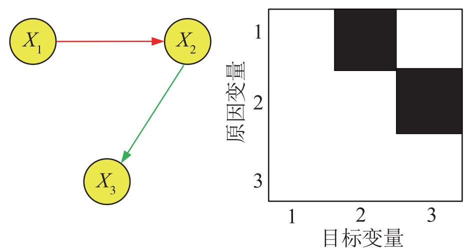
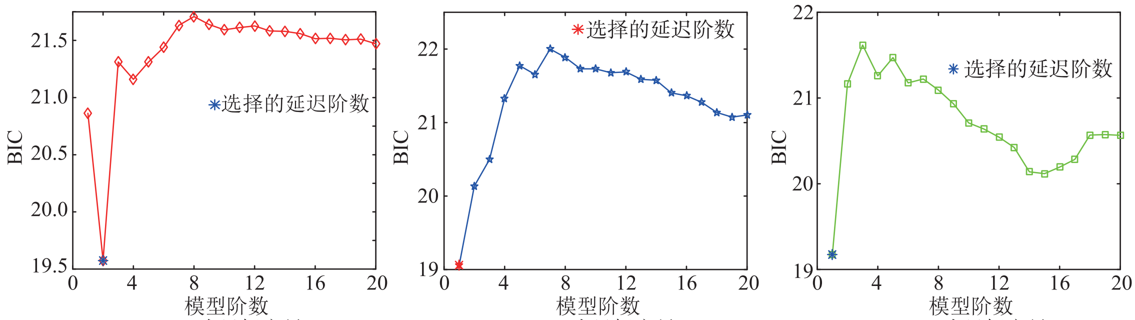


图 5　Henon 系统真实因果关系图（黑色表示存在因果关系）

Fig.5　The real causality diagram of Henon system (Black indicates that there is causality)

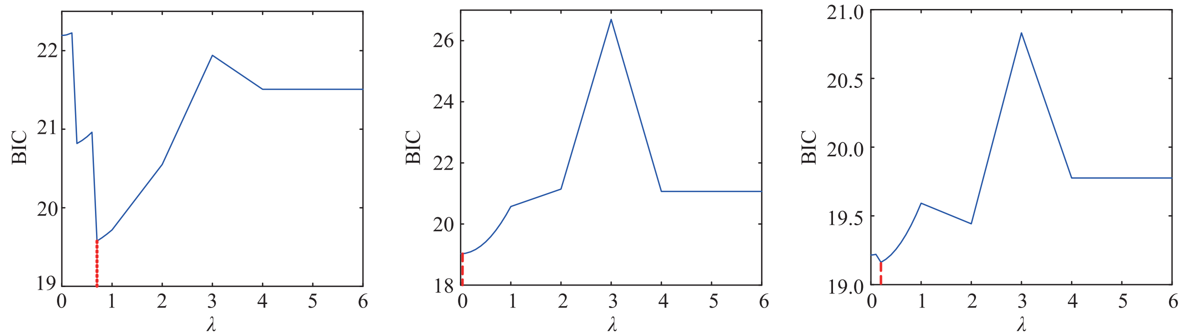
图 6 和图 7 分别是当耦合强度 $c=0.5$ 时，本文方法采用 BIC 选取模型阶数和惩罚系数的三维图像的二维平面展示．选取 BIC 最小值对应的 p 和 λ 作为模型的阶数和群组 Lasso 的惩罚系数．图 8 是对耦合强度 $c=0.5$ 的 Henon 系统因果关系分析的邻接矩阵图．从图中可以看出，Lasso-GC 对虚假因果 $X_2 \to X_1$ 和 $X_3 \to X_1$ 产生了非零因果系数，mBTS-CGCI 则存在着 $X_2 \to X_1$ 的非零系数．这是因为这两种方法适用于多变量的线性因果关系分析，无法准确识别非线性系统的因果关系，并且其对于真实的因果关系的识别效果不明显，识别能力较弱．而 KGC、PMIME 和本文方法都能够正确识别系统中存在的因果关系．相比之下，本文方法计算得到的因果关系邻接矩阵数值更大，对因果关系的识别效果更佳．虽然本文方法得到了 $X_3 \to X_2$ 的非零因果系数，但其数值远远小于真实因果 $X_1 \to X_2$ 和 $X_2 \to X_3$ 的因果关系系数，因此能够进行有效区分识别．当耦合强度 c 取其它非零值时，得到了类似的因果关系分析结果．



(a) X_1为目标变量　(b) X_2为目标变量　(c) X_3为目标变量

图 6　耦合系数为 0.5 的 Henon 系统模型阶数选择

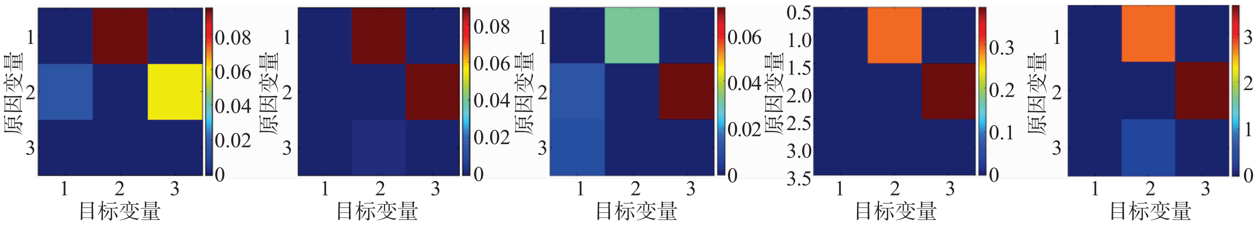
Fig.6　The model order selection of Henon system with coupled strength 0.5



(a) X_1为目标变量 (b) X_2为目标变量 (c) X_3为目标变量

图 7 耦合系数为 0.5 的 Henon 系统惩罚系数选择

Fig.7 The penalty coefficients selection of Henon system with coupled strength 0.5



(a) mBTS-CGCI (b) Lasso-GC (c) KGC (d) PMIME (e) HSIC-GLasso-GC

图 8 耦合强度为 0.5 的 Henon 系统因果邻接矩阵

Fig.8 The causal adjacent matrix of Henon system with coupled strength 0.5

表 2 是耦合强度分别为 0 和 0.5 时的显著性检验结果. 从表 2 可看出, mBTS-CGCI 对虚假因果 $X_1 \rightarrow X_2$ 和 $X_1 \rightarrow X_3$ 具有统计显著性, 其它方法的显著性检验结果都验证了当 $c=0$ 时 Henon 系统变量间的非因果性. 当耦合强度为 0.5 时, 系统存在着 $X_1 \rightarrow X_2$ 和 $X_2 \rightarrow X_3$ 的因果关系, 根据检验结果可以看出, mBTS-CGCI 和 Lasso-GC 都识别出了虚假因果关系. 相比之下, 本文方法能够正确识别变量间的因果关系, 通过了显著性检验, 证明了本方法的有效性. 同时, 相比于 PMIME 方法, 本文方法在计算难度及计算成本具有更加明显的优势, 尤其是当数据维度增加或规模扩大时, PMIME 涉及边缘概率密度的计算, 将面临着极大的困难. 相比于 KGC 方法, 本文方法避免了对于核参数和滞后阶数的人为设定, 从而减少了分析方法的人为偶然因素影响. 综上可以看出, 本文方法具有明显的优越性.

表 2 耦合强度为 0 和 0.5 的 Henon 系统因果关系显著性检验结果

Tab.2 The significance test results of Henon system with coupled strengths 0 and 0.5

方法	mBTS-CGCI		Lasso-GC		KGC		PMIME		HSIC-GL-GC	
	$c=0$	$c=0.5$	$c=0$	$c=0.5$	$c=0$	$c=0.5$	$c=0$	$c=0.5$	$c=0$	$c=0.5$
$X_1 \rightarrow X_2$	100	100	50	51	22	99	0	100	2	100
$X_1 \rightarrow X_3$	100	100	0	0	0	100	0	0	9	23
$X_2 \rightarrow X_1$	0	4	65	100	0	0	0	0	21	4
$X_2 \rightarrow X_3$	0	100	0	100	0	98	0	100	4	100
$X_3 \rightarrow X_1$	0	0	14	100	0	0	0	0	31	0
$X_3 \rightarrow X_2$	0	1	0	100	13	0	0	0	6	10

3.3 沈阳 AQI 和气象时间序列

本节将对沈阳市的 AQI(air quality index)和气象时间序列进行因果关系分析, 并建立预测模型对因果分析结果进行验证. 数据来源于 UCI 数据库, 采样时间为 2014 年 3 月 24 日至 2015 年 2 月 22 日, 采样间隔为 8 h, 共计 11 维, 每一维变量包含 886 组数据, 包括 6 维空气污染数据和 5 维气象数据, 其变量含义如表 3 所示. 利用因果分析方法找出 PM2.5 浓度变化的主要影响因素, 剔除无关和冗余变量, 保留相关变量, 并将该相关变量作为预测模型的输入进行建模预测, 可以有效降低预测模型的规模, 提高预测精度.

表 3 沈阳 AQI 和气象时间序列变量表

Tab.3 AQI and meteorological time series variables of Shenyang

编号	物理含义
1	PM2.5
2	PM10
3	O_3
4	NO_2
5	SO_2
6	CO
7	温度
8	露点
9	气压
10	湿度
11	风速

首先，采用 ADF 检验和差分平稳化方法对数据进行预处理．然后以 PM2.5 为分析目标，建立 HSIC-GL 模型，其模型阶数和惩罚系数均采用 BIC 确定，选择的结果如图 9 所示．最后，求解 HSIC-GL-GC，根据模型的回归系数和显著性检验结果来确定变量间的因果关系，并建立预测模型，选择相关变量作为预测模型的输入子集．

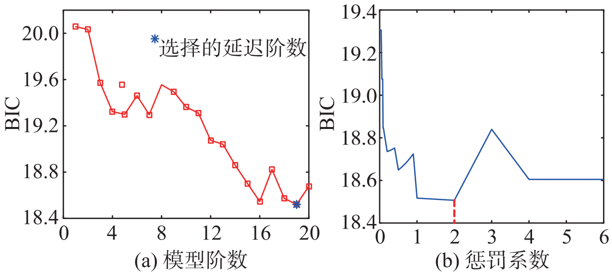


图 9　模型阶数和惩罚系数选择

Fig.9　The model order and coefficient penalty selection

几种因果分析方法的结果如表 4 所示．从表 4 可以看出本文方法识别出对 PM2.5 具有因果关系的变量分别是 PM 10、O_3、SO_2、CO 和风速．沈阳市位于辽河平原中部，是我国重要的重工业基地．同时，随着汽车的普及，当地的空气易受到工业、汽车尾气排放及燃煤等污染．工业及汽车尾气排放主要是 CO、NO_x、SO_2 和烟尘微粒等．PM2.5 是空气中直径小于 2.5 μm 的一类微粒的统称，其产生途径主要包括直接排放以和二次产生．二次产生是指大气中的气态物质，如 SO_2、NOx 和 VOCs 等经过一系列复杂的物理化学反应产生 PM2.5 的过程．在沈阳地区，由于工业发展和汽车的普及，空气质量主要受到工业废气及汽车尾气排放的影响．PM2.5 往往由空气中的其它气态排放物经过一系列物理化学变化形成，其浓度变化更易受到如 SO_2、NOx 等的影响．此外，风在污染物的扩散和累积中起着至关重要的作用．一般来说，风速越大，越有利于 PM2.5 的扩散，PM2.5 的浓度越低．反之，风速越小，PM2.5 浓度越高．所以 PM2.5 的浓度很大程度上也会受到风速的影响．综上所述，可以看到 HSI-GLasso-GC 的分析结果与沈阳当地 PM2.5 浓度变化的影响关系是一致的，进一步验证了本文方法的有效性．

在进行因果关系分析之后，选择出与目标变量具有因果关系的原因变量作为模型的输入进行建模预测，对分析结果进行进一步的验证．本文建立回声状态网络(echo state networks，ESN)预测模型[27]，来分别分析每一种方法所选择因变量的建模预测效果，并进行 30 次实验取平均值以消除偶然因素的影响．采用均方根误差(RMSE)，平均绝对百分误差(MAPE)和对称平均绝对百分比误差(SMAPE)三个指标来定量评价预测精度，定义如下：

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{21}$$

$$\mathrm{MAPE}=\sum_{i=1}^{n}\left|\frac{\hat{y}_i-y_i}{y_i}\right| \tag{22}$$

$$\mathrm{SMAPE}=\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{\frac{(|y_i|+|\hat{y}_i|)}{2}} \tag{23}$$

式中，y_i 和 $\hat{y}_i$ 分别表示真实值和预测值，n 表示样本数．上述 3 个指标越小表示预测效果越好．

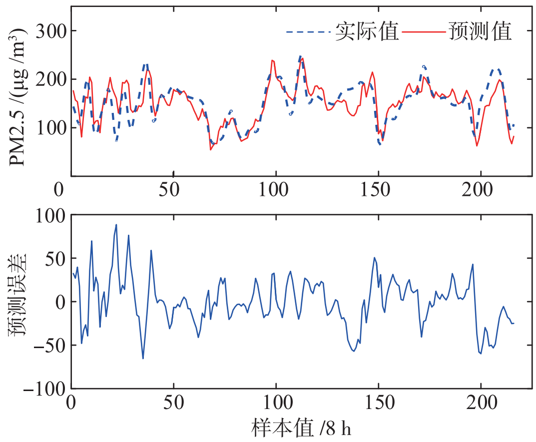


图 10　不进行因果关系分析的 PM2.5 预测图

Fig.10　Prediction graph of PM2.5 without causality analysis

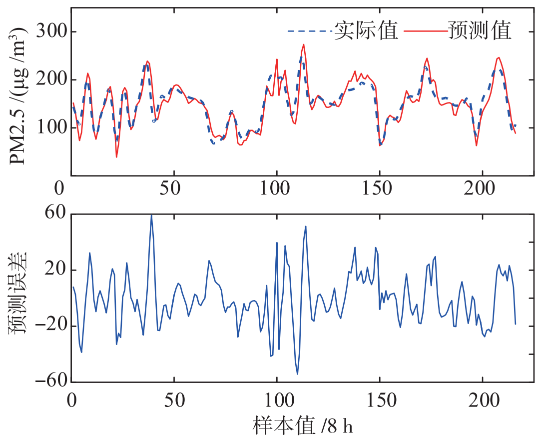


图 11　基于 HSIC-GL-GC 的预测图

Fig.11　Prediction graph based on HSIC-GL-GC

图 10 和图 11 是 PM2.5 的预测效果图．可以看到本文方法能够更加准确跟踪 PM2.5 的变化趋势，拟合效果更好．表 4 是不同方法预测 PM2.5 的精度比较．可以看到，本文方法在 3 个预测指标都取得了最小值，表明其预测精度最高，具有明显的优越性．由于实际系统各个变量之间存在着复杂的非线性关系，基于线性模型的 mBTS-CGCI 和 Lasso-GC 在面对复杂的非线性系统时，无法正确地排除所有无关及冗余变量的影响，难以提高预测精度．而 KGC、PMIME 和本文的 HSIC-GL-GC 方法能够识别复杂系统的非线性关系，但面对类似的更为复杂的多变量系统时，PMIME 方法的计算难度及计算成本往往会成倍增加，这会大大限制其应用范围，难以推广使用．相比之下，本文方法能够有效应用于多变量系统的非线性因果关系分析，并取得良好的分析结果．综上，可以看出本文所提方法能够有效地识别多变量复杂系统的非线性因果关系，为预测模型选择合适的输入变量，提高预测精度，具有良好的现实应用价值．

表4 PM2.5 预测结果

Tab.4 The prediction results of PM2.5

方法	因变量(编号)	RMSE	MAPE	SMAPE
–	全部	41.822 8	0.111 1	0.106 1
mBTS-CGCI	4、9、10、11	35.861 8	0.093 0	0.089 3
Lasso-GC	2、3、7、8、10	45.114 1	0.083 6	0.080 0
KGC	2、3、4、9、10	31.858 0	0.074 6	0.072 3
PMIME	11	34.456 5	0.109 3	0.104 8
HSIC-GL-GC	2、3、5、6、11	**26.725 3**	**0.065 3**	**0.063 8**

4 结论

针对多变量系统的非线性因果关系识别问题，本文提出了 HSIC-GL-GC 方法，可以有效识别多变量系统的非线性因果关系. 该方法具有以下特点：

1）利用 HSIC 对原始时间序列进行非线性映射，将 Granger 因果进行了非线性扩展，克服了传统 Granger 因果模型难于应用于非线性系统的不足.

2）利用群组 Lasso 进行约束，通过建立一个模型就可以实现多变量因果关系分析. 另外，群组 Lasso 可以有效地对组派生变量进行选择，获得稀疏的变量选择结果，避免选择出无关和冗杂变量.

3）利用贝叶斯信息准则对 HSIC-GL 的模型阶数和惩罚系数进行选择，避免了人为设定的干扰，可以有效消除人为因素产生的偶然误差.

但是，该方法仍存在不能直接应用于非平稳时间序列的缺陷. 当原始序列非平稳时，平稳化的过程可能会使原始序列的实际意义发生变化，从而导致因果分析方法的解释意义发生改变. 未来，将研究因果关系分析的非平稳模型，致力于采用时变模型将 HSIC-GL-GC 进行非平稳扩展等研究.

参考文献

[1] 谢婷玉，徐德刚，阳春华，等. 基于重要点双重评价的时间序列趋势提取[J]. 信息与控制，2018，47(6)：730 –736.
Xie T Y, Xu D G, Yang C H, et al. Trend feature extraction method for time series based on double evaluation factors of important points[J]. Information and Control, 2018, 47(6): 730 –736.

[2] Han M, Ren W, Xu M, et al. Nonuniform state space reconstruction for multivariate chaotic time series[J]. IEEE Transactions on Cybernetics, 2018, 49(5): 1885 –1895.

[3] 李琨，韩莹，黄海礁. 基于 IBH-LSSVM 的混沌时间序列预测及其在抽油井动液面短期预测中的应用[J]. 信息与控制，2016，45(2)：241 –247.
Li K, Han Y, Huang H J. Chaotic time series prediction based on IBH-LSSVM and its application to short-term prediction of dynamic fluid level in oil wells[J]. Information and Control, 2016, 45(2): 241 –247.

[4] 韩敏，张瑞全，许美玲. 一种基于改进灰色关联分析的变量选择算法[J]. 控制与决策，2017，32(9)：1647 –1652.
Han M, Zhang R Q, Xu M L. A variable selection algorithm based on improved grey relational analysis[J]. Control and Decision, 2017, 32(9): 1647 –1652.

[5] Granger C W J. Investigating causal relations by econometric models and cross-spectral methods[J]. Econometrica: Journal of the Econometric Society, 1969, 37(3): 424 –438.

[6] Bressler S L, Seth A K. Wiener-Granger causality: A well established methodology[J]. Neuroimage, 2011, 58(2): 323 –329.

[7] Barnett L, Seth A K. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference[J]. Journal of Neuroscience Methods, 2014, 223: 50 –68.

[8] Geweke J. Measurement of linear dependence and feedback between multiple time series[J]. Journal of the American Statistical Association, 1982, 77(378): 304 –313.

[9] Siggiridou E, Kugiumtzis D. Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model[J]. IEEE Transactions on Signal Processing, 2015, 64(7): 1759 –1773.

[10] Arnold A, Liu Y, Abe N. Temporal causal modeling with graphical granger methods[C]//13th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, USA: ACM, 2007: 66 –75.

[11] Bolstad A, Van Veen B D, Nowak R. Causal network inference via group sparse regularization[J]. IEEE Transactions on Signal Processing, 2011, 59(6): 2628 –2641.

[12] Ancona N, Marinazzo D, Stramaglia S. Radial basis function approach to nonlinear Granger causality of time series[J]. Physical Review E, 2004. DOI: 10.1103/PhysRevE. 70.056221.

[13] Marinazz D, Pellicoro M, Stramaglia S. Kernel method for nonlinear Granger causality[J]. Physical Review Letters, 2008. DOI: 10.1103/PhysReVLett. 100.144103.

[14] Hu M, Liang H. A copula approach to assessing Granger causality[J]. NeuroImage, 2014, 100: 125 –134.

[15] Montalto A, Stramaglia S, Faes L, et al. Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality[J]. Neural Networks, 2015, 71: 159 –171.

[16] Faes L, Nollo G, Porta A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique[J]. Physical Review E, 2011. DOI: 10.1103/PhysRevE. 83.051112.

[17] Kugiumtzis D. Direct-coupling information measure from nonuniform embedding[J]. Physical Review E, 2013. DOI: 10.1103/PhysRevE. 87.062918.

[18] Yamada M, Jitkrittum W, Sigal L, et al. High-dimensional feature selection by feature-wise kernelized Lasso[J]. Neural Computation, 2014, 26(1): 185 -207.
[19] Poignard B, Yamada M. Sparse Hilbert-Schmidt independence criterion regression[C]//International Conference on Artificial Intelligence and Statistics. Piscataway, USA: IEEE, 2020: 538 -548.
[20] Gretton A, Bousquet O, Smola A, et al. Measuring statistical dependence with Hilbert-Schmidt norms[C]//International Conference on Algorithmic Learning Theory. Berlin, Germany: Springer, 2005: 63 -77.
[21] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2006, 68(1): 49 -67.
[22] Paparoditis E, Politis D N. The asymptotic size and power of the augmented Dickey-Fuller test for a unit root[J]. Econometric Reviews, 2018, 37(9): 955 -973.
[23] Barber R F, Drton M. High-dimensional Ising model selection with Bayesian information criteria[J]. Electronic Journal of Statistics, 2015, 9(1): 567 -607.
[24] He D, Rish I, Parida L. Transductive HSIC Lasso[C]//SIAM International Conference on Data Mining. Piscataway, USA: IEEE, 2014: 154 -162.
[25] Jia Z Y, Lin Y, Liu Y X, et al. Refined nonuniform embedding for coupling detection in multivariate time series[J]. Physical Review E, 2020. DOI: 10.1103/PhysRevE.92.062113.
[26] Papana A, Kyrtsou C, Kugiumtzis D, et al. Simulation study of direct causality measures in multivariate time series[J]. Entropy, 2013, 15(7): 2635 -2661.
[27] 李晓华，李军. 基于 ESN 网络的连续搅拌反应釜(CSTR)辨识[J]. 信息与控制，2014，43(2)：223 -228.
Li X H, Li J. Identification of continuous stirred tank reactor based on echo state network[J]. Information and Control, 2014, 43(2): 223 -228.

作者简介

李柏松(1995 -)，男，硕士生．研究领域为多元混沌时间序列因果关系分析.
任伟杰(1990 -)，男，博士生．研究领域为多元混沌时间序列分析与变量选择.
韩　敏(1959 -)，女，博士，教授，博士生导师．研究领域为复杂系统建模与混沌时间序列预测.

(上接第 355 页)

[25] Aroussi M E, Hassouni M E, Ghouzali S, et al. Local steerable pyramid binary pattern sequence LSPBPS for face recognition method[J]. International Journal of Signal Process, 2009, 5(4): 281 -284.
[26] Zhou L, Liu W, Lu Z M, et al. Face recognition based on curvelets and local binary pattern features via using local property preservation[J]. Journal of Systems & Software, 2014, 95(9): 209 -216.
[27] Muqeet M A, Holambe R S. Local binary patterns based on directional wavelet transform for expression and pose-invariant face recognition[J]. Applied Computing and Informatics, 2019, 15(2): 163 -171.
[28] Cao Y, Zhang Y, Wen T, et al. Research on dynamic nonlinear input prediction of fault diagnosis based on fractional differential operator equation in high-speed train control system[J]. Applied Computing and Informatics, 2019, 15(2): 163 -171.

作者简介

牛冰川(1994 -)，男，硕士生．研究领域为模式识别，计算机视觉.
文成林(1963 -)，男，博士，教授，博士生导师．研究领域为传感器网络信息融合理论，模式识别，多目标跟踪，复杂系统和设备故障诊断，可靠性评估和健康控制，高超声速飞行器识别和跟踪.