Additive Noise Model Structure Learning Based on Spatial Coordinates

Jing Yang 1,2, Youjie Zhu 1,2,*, Aiguo Wang 3

1 Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education
2 School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, Anhui, P. R. China
3 School of Electronic and Information Engineering, Foshan University, Guangdong, China
* Email: 2000800036@hfut.edu.cn

Abstract. A new algorithm named SCB (Spatial Coordinates Based) algorithm is presented for structure learning of additive noise model, which can effectively deal with nonlinear arbitrarily distributed data. This paper makes three specific contributions. Firstly, SC (Spatial Coordinates) coefficient is proposed to use as a standard of independence test and CSC (Conditional Spatial Coordinates) coefficient as a standard of conditional independence test. Secondly, it is proved that the CSC coefficient conforms to the standard normal distribution and the HSIC independence test can be regarded as a special case of the SC coefficient. Finally, based on the SC coefficient, the SCB algorithm is proposed, and the experimental comparison with some existing algorithms on seven classical networks shows that the SCB algorithm has better performance. In particular, SCB algorithm can deal with large sample, high dimensional nonlinear data, and maintain good accuracy and time performance.

1. Introduction
Causal discovery has received considerable attention in statistics and machine learning during the last decades, which has been successfully applied to many domains, including medical diagnosis, gene algorithms have been developed and proven to be effective, such as MMHC algorithm[1].

However, these linear situations often be violated in practical applications. For example, the financial market is very complex, and there is a non-linear causal dependence between crude oil prices and China’s stock market[2]; The internal relationships of complex and dynamic weather, ecology, and economy systems also show non-linearity[3]. Therefore, the nonlinear characteristics of real data bring great challenges to causal learning model.

In order to solve nonlinear problem, such as the KPC algorithm[4] and CD-NOD algorithm[5]. The above independence test methods and structure learning algorithms have relative high computational complexity and have certain limitations when used for high-dimensional and large sample data for the additive noise model., data analysis, and fault diagnosis[6][7]. For discrete data. A variety of causal structure learning.

In order to solve the causal discovery problem of high-dimensional large sample data, this paper first proposed the independence test method named SC (Spatial Coordinates) coefficient, which can
effectively judge the correlation between arbitrarily distributed data, and then proposed the SCB algorithm based on the SC coefficient.

This paper’s contributions are mainly reflected in the following three aspects:

1. We proposed the SC coefficient, then proved that it can be used as the standard of independence test. Next, we deduced the CSC coefficient based on SC coefficient, which can be used as the standard of conditional independence test.

2. We proved that the SC coefficient conforms to the standard normal distribution. Next, we further analyzed the relationship between the SC coefficient and the HSIC independence test, which concluded that the HSIC independence test is a special case of the SC coefficient.

3. We proposed the SCB algorithm based on SC coefficient and CSC coefficient, and compared it with some existing algorithms on the classical network. Through experimental inquiry, SCB algorithm has advantages in accuracy and time performance. In particular, SCB algorithm can deal with large sample and high dimensional nonlinear data, and time performance is also good while maintaining good accuracy. Finally, the validity of the algorithm is verified by real power plant data.

2. Related Work

In recent years, scholars have proposed some independence test methods and causal structure learning algorithms in the research field of processing nonlinear arbitrarily distributed data. Fukumizu et al. proposed the HSIC (Hilbert-Schmidt Independence Criterion) independence test method. However, as the total number of samples increases, the complexity of the algorithm increases rapidly[8]; Zhang et al. made a series of improvements on Tillman et al., and proposed the CD-NOD algorithm. However, the complexity does not reduce[9].

In addition to the research in the field of independence test methods, there are some other studies dealing with nonlinear methods, such as Wu P et al. established an integrated framework, namely causal Mosaic, based on nonlinear independent component analysis (ICA), which modeled causal pairs through the mixing of nonlinear models[10].

The computational complexity of most of the above algorithms is relatively large, and the time-consuming cost is relatively high. In order to reduce the complexity of the algorithm and effectively process nonlinear data, we propose SC coefficient, CC coefficient and SCB algorithm.

3. SC and CSC Coefficients

3.1. SC Coefficient

Hypothetical characteristics $x_i = a_i \phi(t) + b_i$, $t$ is the assumed independent variable, $\phi$ is some kind of mapping. If data sets $X$ and $Y$ are related, it can be concluded that $y_i = a_i \phi(t) + b_i$, that is, the mapping function should be consistent. Then take $x_i$, can get the equation $x_i - x_i = a_i (\phi(t) - \phi(t)) + b_i$, take $y_i$, can get the equation $y_i - y_i = a_i (\phi(t) - \phi(t)) + b_i$, observe $x_i - x_i, ..., x_m - x_i$ and $y_i - y_i, ..., y_m - y_i$, corresponding position $x_i - x_i = a_i (\phi(t) - \phi(t)) + b_i$ and $y_i - y_i, ..., y_m - y_i$ should be both positive number or both negative number. It is concluded that $(x_i - x_i) * (y_i - y_i) \neq 0$, then $X_i * Y_i \neq 0$. This is because they are basically the sum of values greater than 0. Obviously, if the dataset and are not related, that is, the mapping function is inconsistent, corresponding position $x_i - x_i = a_i (\phi(t) - \phi(t)) + b_i$ and $y_i - y_i, ..., y_m - y_i$ should be irregular. Both positive number and negative number values should exist, then $X_i * Y_i$ should tend to 0. To exclude contingency, We can choose any $x_i$ as the subtraction to get the following definition.

Definition 3.1 [26] "analogous" discretization: The total number of samples in data sets $X$ and $Y$ is $m$, and the total number of features of samples is $n$, the process of choosing any $x_i$ as the subtraction is called class discretization.
Definition 3.1 [26] SC coefficient: The result of multiplying the corresponding rows of two discrete variables $X$ and $Y$ and then normalizing them is called spatial coordinates SC (spatial coordinates) coefficient, which is expressed by $r_{sc}$.

$$r_{sc} = \frac{X_1Y_1^T + X_2Y_2^T}{\sqrt{(X_1X_1^T + X_2X_2^T)(Y_1Y_1^T + Y_2Y_2^T)}}$$

Above $X_i = (x_i - x_1, ..., x_m - x_1)$, $Y_i = (y_i - y_1, ..., y_m - y_1)$.

Theorem 3.1: When $r_{sc}$ tends to 0, there is no functional relationship between variables $X$ and $Y$, that is independence. On the contrary, when $r_{sc}$ is greater than 0, there is a functional relationship, between variables $X$ and $Y$, that is correlation.

Therefore, $r_{sc}$ can also be regarded as the cosine value of variables $X$ and $Y$. When the vectors represented by $X$ and $Y$ are vertical, they tend to 0, on the contrary, when they are greater than 0.

3.2. CSC Coefficient

Theorem 3.2: "analogous" discretization variable: $X_1$, $X_2$ and $X_3$, let component $X_1 - X_3$ represent the component $X_1$ posed perpendicular to $X_3$, and component $X_2 - X_3$ represent the component $X_2$ posed perpendicular to $X_3$, then the SC coefficient between component $X_1 - X_3$ and component $X_2 - X_3$ represents the independence of $X_1$ and $X_2$ after removing the influence of, that is, conditional independence.

From theorem 3.2, we can deduce the definition of CSC coefficient.

Definition 3.3: "analogous" discretization variable: $X_1$ and $X_2$, then measure the SC coefficient between them after removing other conditions, it is called CSC (Conditional Spatial Coordinates) coefficient and is represented by $r_{csc}$.

$$\cos \theta_{23|4...n} = \frac{\cos \theta_{23|4...n-1} - \cos \theta_{34...n-1} \cos \theta_{23|4...n-1}}{\sqrt{1 - \cos \theta_{34...n-1}^2} \sqrt{1 - \cos \theta_{23|4...n-1}^2}}$$

In addition, we also prove that HSIC is a special case of SC coefficient.

4. SCB Algorithm

SCB algorithm uses SC coefficient and CSC coefficient for structural learning constraints, and greedy learning for structural learning search. The algorithm framework is as follows.
Algorithm 1 Restriction Stage of the SCB Algorithm

/*Restriction Stage: Get the constrained Bayesian adjacency matrix.*/

/* Input: */

dataset: $X = \{X_1, ..., X_n\}$

Threshold: $K$

/* Output: */

A potential neighbors matrix:

1: matrix: $SCM(i, j) = 0$ (i=1 to n, j=1 to n)

2: for i=1 to n
   for j=1 to n
      Calculate the SC coefficient of dataset $\{X_i, X_j\}$ and get the $r_{SC}$.
      $SCM(i, j) = r_{SC}$
   end
end

3: $C = SCM^{-1}$

4: matrix: $R(i, j) = 0$ (i=1 to n, j=1 to n)

5: for i=1 to n
   for j=1 to n
      Calculate $p$ value
      if $p$ value $< K$
         $CC(i, j) = 1$
      else
         $CC(i, j) = 1$
      end
   end
end

6: matrix $CC_{con}(i=1$ to n, $j=1$ to n)

7: for i=1 to n
   for j=1 to n
      Calculate $p$ value
      if $p$ value $< K$
         $CC(i, j) = 1$
      else
         $CC(i, j) = 1$
      end
   end
end

return $CC_{con}$

5. Experimental Analysis

We have carried out experiments: Compare SCB algorithm with CD-NOD and KPC algorithm in the same experimental environment. Due to limited space, some results are shown below:

![Fig. 1 Part of experimental results of KPC algorithm, CD-NOD algorithm and SCB algorithm](image)

We can see from Figure 1: the accuracy and time efficiency of SCB algorithm are better than CD-NOD and KPC algorithm. In terms of time: HSIC independence test tool is adopted for CD-NOD and KPC algorithms. Its operation is matrix multiplication. With the increase of samples, the time increases exponentially, while SCB algorithm is linear operation and time is also linear superposition; In terms of the number of structural errors: CD-NOD and KPC algorithms are single square forms, without the difference between positive and negative, which is equivalent to amplifying the error,
while SCB algorithm retains the difference between positive and negative, which is more accurate.

6. Conclusions and prospects
The algorithm has room for improvement in accuracy and the same time performance, and more efforts are needed to perfect our work. Therefore, we plan to continue the theoretical research and strive to further improve the accuracy, time performance and application scope of the algorithm, so that our algorithm can be accurately and effectively applied in real life.

References
[1] Schmidt M, Niculescu-Mizil A, Murphy K 2007 Learning graphical model structure using L1-regularization paths. 22nd national conference on Artificial intelligence on Proceedings pp 1278–1283
[2] Wen F, Xiao J, Xia X, Chen B, Xiao Z and Li J 2018 Oil Prices and Chinese Stock Market: Nonlinear Causality and Volatility Persistence Emerging Markets Finance and Trade pp 1-17
[3] Wang W, Lai C and Grebogi C 2016 Data based identification and prediction of nonlinear and complex dynamical systems Physics Reports Finance and Trade 644 pp 1–76
[4] Tillman R E, Gretton A, Spirtes P 2009 Nonlinear directed acyclic structure learning with weakly additive noise models 22nd International Conference on Neural Information Processing Systems on Proceedings pp 1847–1855
[5] Zhang K, Huang B, Zhang J 2017 Causal Discovery from Nonstationary/Heterogeneous Data 26th International Joint Conference on Artificial Intelligence on Proceedings pp 1347–1353
[6] Yang J, An N and Alterovitz G 2016 A Partial Correlation Statistic Structure Learning Algorithm Under near Structural Equation Models IEEE Transactions on Knowledge and Data Engineering 28(10) pp 2552–2565
[7] Wang Y, Yang H, Yuan X and Cao Y 2018 The max-min hill-climbing Bayesian network structure learning algorithm Machine Learning 51(21) pp 341–346
[8] Fukumizu K, Gretton A, Sun X, Schlkopf, B 2008 Kernel Measures of Conditional Dependence. 20th International Conference on Neural Information Processing Systems on Proceedings pp 489–496
[9] Zhang H, Zhou S, Zhang K, Guan J 2017 Causal Discovery Using RegressionBased Conditional Independence Tests 31st national conference on Artificial intelligence on Proceedings pp 1250–1256
[10] Wu P and Fukumizu K 2020 Causal Mosaic: Cause-Effect Inference via Nonlinear ICA and Ensemble Method 23rd International Conference on Artificial Intelligence and Statistics on Proceedings pp 1157–1167