An Efficient Kurtosis-based Causal Discovery Method for Linear Non-Gaussian Acyclic Data

Ruichu Cai *, Feng Xie*, Wei Chen*, Zhifeng Hao*†
*School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China
†School of Mathematics and Big Date, Foshan University, Foshan, China

***Abstract*—Understanding the causality behind the observational data is of great importance to a lot of real world applications, e.g., the improvement of Quality of Service. Non-Gaussianity has been exploited in numerous causal discovery methods for observational linear acyclic data. Transforming non-Gaussianity into indirect metrics is a conventional solution employed by existing methods, although this usually results in unreliable estimations or locally optimal solutions. In this work, we employs the excess kurtosis, a direct measure of non-Gaussianity, to establish a causal discovery method for linear non-Gaussian acyclic data. Firstly, we theoretically prove that an exogenous variable has the largest excess kurtosis when disturbance variables follow independent and identically distributions. Secondly, based on this property of exogenous variables, we propose an efficient exogenous variable identification algorithm, and develop a causal discovery method. Extensive experiment results verify the effectiveness and efficiency of the proposed approach.**

I. Introduction

Causality has shown its effectiveness in the design of intervention strategies, e.g., how to improve the quality of service [1], [2], [3]. Compared with the association based approaches, the causality based approach is able to locate the generic root cause among a lot of corrected variables. Take the Fig. 1 for example, all the variable is highly associated with x_4, while only x_1 is the cause of x_4, and only the intervention on x_1 can change the state of x_4.

However, in the absence of prior knowledge, causal discovery on the observational data is still a challenge problem. For example, the structural learning cannot distinguish Markov equivalence classes [4], [5]. Thus, Shimizu et al. developed Linear Non-Gaussian Acyclic Model (LiNGAM) by exploring the linear non-Gaussian acyclic data generation process [6], [7]. At present, two main approaches based on LiNGAM are available for discovering the causal structure of linear non-Gaussian acyclic data. The first approach converts the discovery process into a function optimization problem, such as is conducted in the ICA-LiNGAM [6] and Bayes-LiNGAM algorithms [8] . This approach is sensitive to initial values, and usually leads to local optima. The second approach utilizes the independence of residuals and independent variables to select exogenous variables in iterations and obtain a causal order, such as is conducted in the Direct-LiNGAM [9] and likelihood ratio (LR) algorithms [10] . This approach relies on the accuracy of the independence test method, and, as illustrated in Fig. 1, the sum of the independence indicators is an unreliable criterion for selecting exogenous variables.

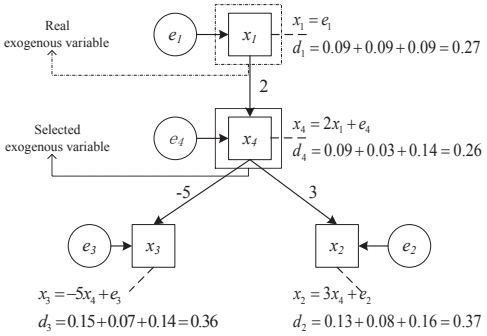

Fig. 1. An example of the erroneous selection of exogenous variable. Here, d_i is used to denote the degree of independence of the current variable x_i and the residuals obtained by the regression of variables x_i and all other variables. Then, according to the principle of selecting the minimum value, we should choose d_4, which indicates that the variable x_4 corresponding to d_4 is an exogenous variable. In fact, however, the real exogenous variable is x_1. The main reason for this false identification is that, although the value of 0.14 obtained in the d_4 summation is large (meaning that the residuals r_3^4 and variable x_3 are not independent), the value 0.3 is found to be much smaller than the other values for the average, which leads to the erroneous selection of the variable x_4.

We find that, when using the non-Gaussian characteristics of a disturbance variable, the above two approaches represent either an approximate transformation method or an indirect estimation method. However, the indirect quantization standard has the problem of estimation unreliability, particularly near the critical value (i.e., close to the Gaussian distribution). As discussed by Spirtes and Zhang [11], causal directions become increasingly difficult to distinguish as the disturbance variables increasingly approach a Gaussian distribution.

Therefore, the development of a direct estimation approach is necessary. The accurate measurement of the degree of non-Gaussianity is the key component of the LiNGAM, and the excess kurtosis is a significant indicator of non-Gaussianity. We find that, for any variable x, $x = \sum_{i=1}^{n} c_i e_i, c_i \neq 0$, where e_i is an independent and identically distributed disturbance variable, the excess kurtosis of x is less than that of any e_i. Based on this understanding, we propose a simple, direct criterion for identifying exogenous variables and for constructing a complete causal network quickly under a DirectLiNGAM framework.

The remainder of this paper is organized as follows. In Section II, we summarize the basic features of LiNGAM and its variants. In Section III, we present a non-Gaussianity measure based on excess kurtosis, and introduce an algorithm

978-1-5386-2704-4/17/$31.00

for identifying exogenous variables. In Section IV, we present the detailed design of our causal structure discovery method. In Section V, we present preliminary experiment results using artificial and real data. Finally, we present conclusions and a discussion of future work in Section VI.

II. Related Work

Structural equation modeling (SEM) is an important component of the theoretical background of the present work. SEM can represent the process of data generation from a mathematical model, and analyze the causal mechanism from a functional point of view [4], [12]. When the process of data generation is linear with no hidden confounders, and the disturbance variables follow a non-Gaussian distribution, the process becomes applicable to LiNGAM [6], [7].

Fundamental work has been conducted to estimate the basic LiNGAM, e.g., the independent component analysis (ICA) LiNGAM (ICA-LiNGAM) and Bayes-LiNGAM algorithms [6], [8]. Shimizu et al. cleverly transformed the problem into one of blind source separation, and obtained a solution using the ICA algorithm. However, the ICA algorithm itself has problems associated with the setting of initial values and establishing the convergence criterion, and, if these are unreasonable, the method easily falls into local optima. Similarly, the Bayes-LiNGAM algorithm is constructed as a scoring function in terms of likelihood estimates, and tends to fall into local optima when the parameters are estimated incorrectly.

To avoid falling into local optima, Shimizu presented a direct estimation algorithm, denoted as the Direct-LiNGAM algorithm [9], that eliminated parameters and provided for direct causal order estimation in a finite number of steps. The algorithm first selected a variable based on its independence from the residuals of many pairwise regressions, and removed the effect of this exogenous variable from the other variables using least squares regression. Subsequently, the algorithm iteratively selected the exogenous variables and removed their effects on the other variables. Finally, a complete causal order was obtained. In addition, the LR algorithm was proposed in the DirectLiNGAM framework [10]. The algorithm needed only to calculate the one-dimensional differential entropies of variables and residuals, and thereby avoided calculating their pairwise independence to identify exogenous variables. The complexity of the LR algorithm is reduced compared with that of the Direct-LiNGAM algorithm.

In addition to estimating the basic LiNGAM, researchers have made some improvements to the model for different scenarios. Such as, the number of observed variables is greater than the number of samples. Here, Shimizu et al. employed ridge regression rather than least squares regression, and used nonlinear correlation to measure independence for estimating the LiNGAM [13]. When the non-Gaussianity of the disturbance variables is less than that of the exogenous variables, Sogawa et al. employed negative entropy to select exogenous variables, and applied it to gene expression data [14]. Cai employed the idea of divide-and-conquer to improve the accuracy of the LiNGAM for estimating sparse networks [15]. In order to solve the time series datasets, Hyvarinen proposed the structural vector autoregressive (SVAR) model, which was solved using maximum likelihood estimation [16].

To the best of our knowledge, there is no work related to the use of a quantitative standard to directly estimate the LiNGAM.

III. Kurtosis-based exogenous variable identification

A. Preliminaries

Assume that the observational data generation process satisfies a linear SEM process, such that, without loss of generality, each observational variable x_i is assumed to have a zero mean. Then, the linear acyclic model can be expressed as

$$x_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i \tag{1}$$

where b_{ij} is the connection strength from x_j to x_i, and $k(i)$ is the causal ordering of x_i. All disturbance variables e_i are independent of each other, with zero means and non-zero variances. Note that, if the members of e_i follow a non-Gaussian distribution, then the model is a LiNGAM.

We rewrite Eq. 1 in matrix form, as follows:

$$\mathbf{X} = \mathbf{B}\mathbf{X} + \mathbf{E} \tag{2}$$

where the vector $\mathbf{X}$ contains all x_i, the connection strength matrix $\mathbf{B}$ contains all b_{ij}, and the vector $\mathbf{E}$ contains all e_i .

Next, we introduce Lemma 1 to illustrate our starting point.

Lemma 1. *Assume that $\mathbf{X} = \{x_1, ..., x_n\}$ strictly follows the LiNGAM. Then, any x_i can be regarded as a linear sum of a set of disturbance variables.*

Proof: Assume that the causal order of $\mathbf{X}$ is exactly the order of the subscripts. Without loss of generality, the causal network is a full network (Among all the variables, the ones in front are always the fathers of the latters).

Next, we use mathematical induction to prove it.

When $n = 1$, x_1=e_1 is clearly established.

Suppose $n = k$ holds,

$$x_k = \sum_{j=1}^{k-1} b_{kj} x_j + e_k = \sum_{j=1}^{k} c_{kj} e_j \tag{3}$$

Where c_{kj} denotes the coefficient value of the e_j corresponding to the k-th x_k.

Then

$$x_{k+1} = \sum_{j=1}^{k} b_{k+1j} x_j + e_{k+1} \tag{4}$$

When $j \leq k$, due to Eq. 3, we obtain

$$x_j = \sum_{p=1}^{j-1} c_{jp} e_p \tag{5}$$

By combining Eq. 4 with Eq. 5, we obtain

$$x_{k+1} = \sum_{j=1}^{k} b_{k+1j} \sum_{p=1}^{j-1} c_{jp} e_p + e_{k+1} \tag{6}$$

In Eq. 6, for any e_i there is no two cross-multiplication terms, so x_{k+1} can be expressed as:

$$x_{k+1} = \sum_{j=1}^{k+1} c_{k+1j} e_j \tag{7}$$

Then $n = k + 1$ holds. To sum up, for any variable can be written as a set of linear sum of disturbance variables, and Lemma 1 is proven. ■

For the causal structure illustrated in Fig. 1, variables can be expressed as $x_1 = e_1$; $x_4 = 2x_1 + e_4 = 2e_1 + e_4$; $x_3 = -5x_4 + e_3 = -10e_1 - 5e_4 + e_3$; $x_2 = 3x_4 + e_2 = 6e_1 + 3e_4 + e_2$.

B. Non-Gaussianity measure of exogenous variables

Here, we investigate the non-Gaussian characteristics of disturbance variables. Non-Gaussian and Gaussian are relative concepts, where non-Gaussian variables are simply those variables not satisfying a Gaussian probability distribution function. An important feature of a non-Gaussian distribution is the size of the kurtosis in the probability distribution curve near the mean, which is defined as

$$kurt(e) = \frac{E[(e-u)^4]}{\left(E[(e-u)^2]\right)^2}$$

where u is the mean of the e.

For any Gaussian distribution, the kurtosis value is 3. It is common practice to use the excess kurtosis, which is computed as

$$\kappa(e) = kurt(e) - 3$$

for comparing distributions with the Gaussian distribution. In this paper, we employ κ to measure the degree of Gaussianity.

Next, we present two lemmas and a corollary that ensure the validity of our algorithm proposed in the following subsection C.

Lemma 2. *Assume that* $\mathbf{E} = \{e_1, ..., e_n\}$ *are independent random variables. Then for any variable* x*,* x *is generated by the sum of variable* $e_i (i = 1, ..., n)$*, and its excess kurtosis value is*

$$\kappa(x) = \frac{\sum_{i=1}^{n} \sigma_i^4 \kappa(e_i)}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^2}$$

where σ_i *is the standard deviation of* e_i*.*

Proof: Let u_i is the mean of e_i and u_x is the mean of x, then

$$\kappa(x) = \frac{E[(x-u_x)^4]}{\left(E[(x-u_x)^2]\right)^2} \tag{8}$$

$$x = \sum_{i=1}^{n} e_i \tag{9}$$

By combining Eq. 8 with Eq. 9, we obtain

$$\kappa(x) = \frac{E[(\sum_{i=1}^{n} (e_i - u_i))^4]}{\left(E[(\sum_{i=1}^{n} (e_i - u_i))^2]\right)^2} \tag{10}$$

Firstly, we analysis the denominator in Eq. 10. $e_1 ... e_n$ are independent of each other, then $e_1 - u_1, ..., e_n - u_n$ are also independent of each other. For the expansion of any two $e_i - u_i$ and $e_j - u_j$ multiplication is 0, that is

$$E[(e_i - u_i)(e_j - u_j)] = E[(e_i - u_i)]E[(e_j - u_j)] = 0$$

So, we obtain

$$E[(\sum_{i=1}^{n} (e_i - u_i))^2] = \sum_{i=1}^{n} E[(e_i - u_i)^2] = \sum_{i=1}^{n} \sigma_i^2 \tag{11}$$

Similarly, for the numerator in Eq. 10

$$\begin{aligned} E[(\sum_{i=1}^{n} (e_i - u_i))^4] &= \sum_{i=1}^{n} E[(e_i - u_i)^4] \\ &= (E[(\sum_{i=1}^{n} (e_i - u_i))^2])^2 \frac{E[(\sum_{i=1}^{n} (e_i - u_i))^4]}{\left(E[(\sum_{i=1}^{n} (e_i - u_i))^2]\right)^2} \\ &= \sum_{i=1}^{n} \sigma_i^4 \kappa(e_i) \end{aligned} \tag{12}$$

By combining Eq. 11 with Eq. 12, we obtain

$$\kappa(x) = \frac{\sum_{i=1}^{n} \sigma_i^4 \kappa(e_i)}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^2}$$

Lemma 2 is proven. ■

Lemma 3. *For any variable* e*,* $\kappa(ae)$*, for any real number* $a \neq 0$*, is equivalent to* $\kappa(e)$*, that is,* κ *has linear scale-invariant properties.*

Proof: $\kappa(ae)$ is calculated as

$$\begin{aligned} \kappa(ae) &= \frac{E[(a(e-u))^4]}{(E[(a(e-u))^2])^2} - 3 = \frac{a^4 E[((e-u))^4]}{a^4 (E[((e-u))^2])^2} - 3 \\ &= \frac{E[((e-u))^4]}{(E[((e-u))^2])^2} - 3 = \kappa(e) \end{aligned}$$

Where u is the mean of e. Therefore, $\kappa(ae)$ is equivalent to $\kappa(e)$, and Lemma 3 is proven. ■

Corollary 1. *Assume that* $\mathbf{E} = \{e_1, ..., e_n\}$ *are independent and identically distributed random variables, and any variable* $x = \sum_{i=1}^{n} a_i e_i, (a_i \neq 0)$*. Then,* $\kappa(x)$ *must be less than or equal to any* $\kappa(e_i)$*.*

Proof: $\kappa(x)$ is calculated as

$$\kappa(x) = \kappa(\sum_{i=1}^{n} a_i e_i) \tag{13}$$

From Lemma 2, $\kappa(x)$ can be expressed as

$$\kappa(x) = \frac{\sum_{i=1}^{n} a_i^4 \sigma_i^4 \kappa(a_i e_i)}{(\sum_{i=1}^{n} a_i^2 \sigma_i^2)^2} \tag{14}$$

where σ_i is the standard deviation of x_i.

κ has linear scale-invariant properties in Lemma 3, we can express $\kappa(x)$ as

$$\kappa(x) = \frac{\sum_{i=1}^{n} a_i^4 \sigma_i^4 \kappa(e_i)}{(\sum_{i=1}^{n} a_i^2 \sigma_i^2)^2} \tag{15}$$

Since all $e_i (i = 1, ..., n)$ follow the same distribution, we can be obtained that they have the same excess kurtosis value, which is denoted $\kappa(e)$. Then, $\kappa(x)$ can be expressed as

$$\kappa(x) = \frac{\sum_{i=1}^{n} a_i^4 \sigma_i^4}{(\sum_{i=1}^{n} a_i^2 \sigma_i^2)^2} \kappa(e) \tag{16}$$

We find $\frac{\sum_{i=1}^{n} a_i^4 \sigma_i^4}{(\sum_{i=1}^{n} a_i^2 \sigma_i^2)^2}$ constant is less than 1, except when $n = 1$.so when $n \geq 2$,

$$\kappa(x) < \kappa(e)$$

This shows that $\kappa(x)$ must be less than any $\kappa(e_i)$. When $n = 1$, obviously $\kappa(x) = \kappa(e)$. Corollary 1 is proven. ■

C. Exogenous variable identification

Before stating the algorithm, we first state the theorem upon which our algorithm is based.

Theorem 1. *Assume that the observational variable* $\mathbf{X} = \{x_1, ..., x_n\}$ *strictly follows the LiNGAM, and* $\mathbf{E} = \{e_1, ..., e_n\}$ *is independent and identically distributed random variables. Then, the* x_i *with the largest* κ *value is the exogenous variable.*

Proof: From Lemma 1, we know that x_i is a linear combination of a set of independent and identically distributed random variables e. As presented in Corollary 1, $\kappa(x_i)$ must be less than any $\kappa(e_i)$ unless $x_m = e_m$ for some m. Obviously, this indicates that x_m is an exogenous variable, which necessitates that the variable with the largest κ value must be the exogenous variable, and Theorem 1 is proven. ■

Based on Theorem 1, we propose a new algorithm, denoted as the kurtosis-based identification of exogenous variables (KIEV) algorithm, that employs the excess kurtosis to identify exogenous variables.

Algorithm 1 KIEV

1: Input: variable set $\mathbf{X} = \{x_1, ..., x_n\}$
2: Output: exogenous variable x_{exo}
3: **for** each variable $x_i \in \mathbf{X}$ **do**
4: Calculate $\kappa(x_i)$ and add to the k set;
5: **end for**
6: $j \leftarrow$ Find the subscript with the maximum value from the set k;
7: $x_{exo} \leftarrow x_j$;
8: Return exogenous variable x_{exo};

IV. Kurtosis-based causal network discovery

Once exogenous variables have been identified, we can then construct the causal order of the entire causal network. According to Shimizu et al., once the effect of the exogenous variable has been removed from the other variables using least squares regression, the remaining residual data sets must still satisfy the basic LiNGAM [9]. Therefore, exogenous variables need only be identified iteratively, and the cause and effect order is the actual causal order (a detailed proof can be found in Lemmas 2 and 3 in Shimizu et al. [9]). Once the causal order of variables is identified, the values of b_{ij} are easily estimated using some edge pruning methods (we selected the Prune-Edges [1]). A complete constructive causal network is obtained using Algorithm 2.

Algorithm 2 KFCI

1: Input: variable set $\mathbf{X} = \{x_1, ..., x_n\}$
2: Output: a complete causal network G
3: Initialize subscript set of variables $U = \{1, 2, ..., n\}$ and causal order $K = \emptyset$;
4: **while** $n - 1$ **do**
5: $x_m \leftarrow$ KIEV(X);
6: Perform least squares regressions of x_i on x_m for all $i \in U \setminus K (i \neq m)$, obtain the residual r_i^m and residual dataset $\mathbf{R^m}$;
7: $\mathbf{X} \leftarrow \mathbf{R^m}$;
8: Add m into K;
9: **end while**
10: Add the last variable into K ;
11: The connection Matrix $\mathbf{B}$=Prune-Edges($\mathbf{X}, K$);
12: Return G with the connection Matrix $\mathbf{B}$;

Next, we analyzed the computational complexity of the KFCI algorithm, and compared that with the computational complexities of three established algorithms: the ICA-LiNGAM algorithm, Direct-LiNGAM algorithm, and LR algorithm. As shown in Table I, the proposed algorithm demonstrated the lowest computational cost, which is only quadratic with respect to the variable size n. However, we note that the ICA-LiNGAM and Direct-LiNGAM algorithms have complexities reaching n^4. We note a computational complexity of M^3, although M is much smaller than m, and the complexity of the Direct-LiNGAM algorithm is greatly increased after successive multiplications.

[1]Matlab codes can be downloaded at http://www.cs.helsinki.fi/group/neuroinf/lingam/lingam.tar.gz

TABLE I
COMPARISON OF FOUR ALGORITHMS' COMPLEXITY

Algorithm	Computational Complexity
ICA-LiNGAM	$O(mn^3 + n^4)$
DirectLiNGAM	$O(mn^3M^2 + n^4M^3)$
LR	$O(mn^3 + mn^4)$
KFCI	$\mathbf{O(mn^2)}$

Note: n is the number of variables, m is the sample size, and $M(\ll m)$ is the maximal rank found by the low-rank decomposition used in the kernel-based independence measure.

V. EXPERIMENTS AND DISCUSSION

In this section, our proposed algorithm is studied on articial data and real data. The experiments were compiled and run using MATLAB 2012b on a windows PC equipped with a dual-core 3.10 GHz CPU and 4 GB RAM.

A. Experiments on Simulated Data

We randomly generated 10 data sets according to the LiNGAM based on sparse networks, where the maximum indegree was 2, the number of samples = 5000, the sample dimensions varied from 20 to 200 in increments of 20, the disturbance variables followed a Laplace distribution or an approximate Gaussian distribution with a mean of 0 and variance of 1, and the non-zero elements in $\mathbf{B}$ were randomly generated according to a normal distribution. To evaluate the advantages and disadvantages of the constructed causal networks, we evaluated the performance according to the recall, precision, and F1 score . Specifically, F1 score is calculated as $\frac{2(R\times P)}{R+P}$, which R and P are recall and precision respectively. Because the LR algorithm is an improved version of the Direct-LiNGAM algorithm, we compared the performance of the proposed algorithm with the performances of the ICA-LiNGAM and LR algorithms.

1) For Laplace and approximate Gaussian distributions: The performances of the three algorithms in terms of the three indexes are presented in Fig. 2 and Fig. 3 with respect to the sample dimension for disturbance variables following a Laplace distribution and an approximate Gaussian distribution, respectively. The figures indicate that the performance of the KFCI algorithm is superior to those of the other two algorithms in terms of all three indexes with increasing sample dimension. This is particularly the case when the non-Gaussianity is very weak (i.e., for the approximate Gaussian distribution in Fig. 3). The recall value of the KFCI algorithm attains a value greater than 90%, and the F1 value is greater than 70% for sample dimensions less than 140 regardless of the degree of non-Gaussianity. In contrast, the F1 values of the ICA-LiNGAM and LR algorithms are less than 40% when the non-Gaussianity is weak (Fig. 3). The reason of the low recognition rate of ICA-LiNGAM and LR algorithms is that they are an indirect estimate of the exogenous variables and are sensitive to non-Gaussianity intensity. Thus, the two algorithms can not correctly select exogenous variables in weak non-Gaussianity cases. In contrast, our algorithm is a direct estimation method, so it can aquire a high recognition rate under weak non-Gaussianity.

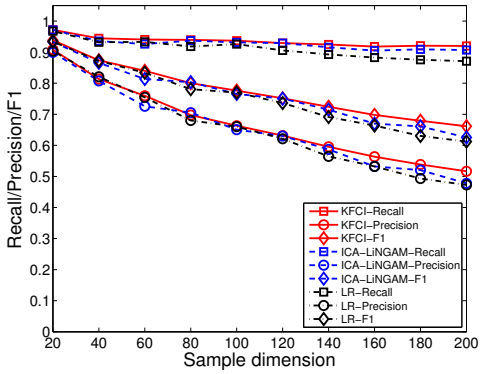

Fig. 2. The performance of the three algorithms in terms of the three indexes under different sample dimensions with a number of samples = 5000 and a maximum indegree of 2, where the disturbance variables follow a Laplace distribution.

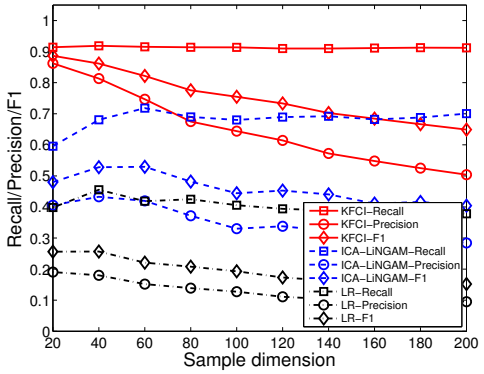

Fig. 3. The performance of the three algorithms in terms of the three indexes under different sample dimensions with a number of samples = 5000 and a maximum indegree of 2, where the disturbance variables follow an approximate Gaussian distribution.

Based on the above analyses, we find that our proposed algorithm is more stable and runs faster compared with the ICA-LiNGAM and LR algorithms.

TABLE II
THE RUNTIME OF THE THREE ALGORITHMS FOR DIFFERENT SAMPLE DIMENSIONS WITH A NUMBER OF SAMPLES = 5000 AND A MAXIMUM INDEGREE OF 2

dim	ICA-LiNGAM	LR	KFCI
20	0.14 sec.	15.99 sec.	**0.04 sec.**
40	0.29 sec.	2.25 min.	**0.24 sec.**
60	0.55 sec.	7.75 min.	**0.54 sec.**
80	1.04 sec.	18.47 min.	**1.00 sec.**
100	1.85 sec.	36.16 min.	**1.80 sec.**
120	3.20 sec.	1.04 hrs.	**2.74 sec.**
140	5.21 sec.	1.67 hrs.	**3.82 sec.**
160	8.93 sec.	2.54 hrs.	**4.96 sec.**
180	15.05 sec.	3.57 hrs.	**6.51 sec.**
200	28.06 sec.	4.91 hrs.	**8.12 sec.**

2) Runtime with different sample dimensions: From Table II, we note that the runtime of the KFCI algorithm is the smallest, followed by the ICA-LiNGAM algorithm and the LR algorithm, which runs the slowest. In addition, the runtime of the KFCI algorithm increases gradually with increasing sample

dimension, while the runtime of the LR algorithm increases obviously. Moreover, when the dimension equaled 200, the runtime of the LR algorithm reached 4.9 hours, indicating that ours was almost 2000 times faster.

B. Application on Quality of Service

We also test the proposed approach on a real world data set collected in a 3G network base station of Guangzhou city. The data set consists of 30406 samples, with each sample contains 7 different measures of the network service quality. The measures include, *TCH Available Rate*, *SDCCH Allocation Success Rate*, *TCH Allocation Success Rate*, *Random Access Success Rate*, *Interference Factor*, *Average Strength of Uplink Signal* and *Uplink Signal Coverage Rate*.

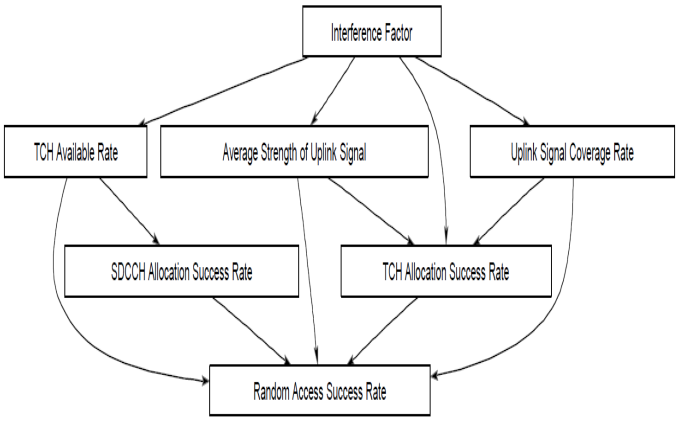

Fig. 4. Wireless network local causal diagram.

Fig. 4 shows the discovered causal structure on the 7 parameters. Among the discovered results, the following direct casual relations have been verified by industry experts, 1) *Interference Factor* is the direct cause of *TCH Available Rate*, *Average Strength of Uplink Signal* and *Uplink Signal Coverage Rate*, 2) *SDCCH Allocation Success Rate* and *TCH Allocation Success Rate* are direct causes of *Random Access Success Rate*. These results not only verify the effectiveness of our algorithm on the real world application, but also provide helpful instructions for the improvement of network service quality. For example, reducing the *Interference Factor* can increase the *TCH Available Rate* .

VI. Conclusion and future work

In this paper, we proposed the KIEV algorithm to identify exogenous variables based on the excess kurtosis, and provided a detailed theoretical proof. We also employed the KIEV algorithm to estimate a complete causal network using the DirectLiNGAM framework, which we denoted as the KFCI algorithm. The KFCI algorithm was demonstrated to be more stable for disturbance variables following distributions of varying non-Gaussianity. Through the local causal network analysis of wireless network data, it had the guidance of significance for wireless network performance optimization.

For future work, we will analyze exogenous variables from the perspective of entropy, and compare the differences and links obtained using non-Gaussianity analysis and entropy analysis. In addition, we will extend our approach to nonlinear models.

Acknowledgment

This work was supported by the NSFC-Guangdong Joint Found(U1501254), Natural Science Foundation of China (61472089, 61572143), Natural Science Foundation of Guangdong (2014A030306004, 2014A030308008), Science and Technology Planning Project of Guangdong (2013B051000076, 2015B010108006, 2015B010131015), Guangdong High-level personnel of special support program (2015TQ01X140), Pearl River S&T Nova Program of Guangzhou (201610010101), and Science and Technology Planning Project of Guangzhou (201604016075).

References

[1] H. Yan, L. Breslau, Z.H. Ge, D. Massey, D. Pei, and J. Yates, "G-rca: a generic root cause analysis platform for service quality management in large ip networks," *IEEE/ACM Transactions on Networking*, vol. 20, no. 6, pp. 1734-1747, 2012.

[2] X.H. Nie, Y.J. Zhao, K.X. Sui, D. Pei, Y. Chen, and X.P. Qu, "Mining causality graph for automatic web-based service diagnosis," *in Proceedings of In Performance Computing and Communications Conference* , Las Vegas, Nevada, USA, Dec, 2016.

[3] Z.F. Hao, W. Chen, R.C. Cai, R.H. Huang, W. Wen, and L.J. Wang, " Performance optimization of wireless network based on canonical causal inference algorithm," *Biological psychiatry*, vol. 36, no. 8, pp. 2114-2120, 2016.

[4] J. Pearl, "Causality: models, reasoning and inference," *Econometric Theory*, vol. 19, pp. 675-685, 2003.

[5] P. Spirtes, C.N. Glymour and R. Scheines, *Causation, prediction, and search*, MIT press, 2000.

[6] S. Shimizu, P.O. Hoyer, A. Hyvärinen, and A. Kerminen, "A linear non-Gaussian acyclic model for causal discovery," *Journal of Machine Learning Research*, vol. 7, no. Oct, pp. 2003-2030, 2006.

[7] S. Shimizu, "LiNGAM: non-Gaussian methods for estimating causal structures," *Behaviormetrika*, vol. 41, no. 1, pp. 65-98, 2014.

[8] P.O. Hoyer, and A. Hyttinen, "Bayesian discovery of linear acyclic causal models," *in Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, Canada, Montreal, June, 2009.

[9] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P.O. Hoyer,and K. Bollen, "DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model," *Bournal of Machine Learning Research*, vol. 12, no. Apr, pp. 1225-1248, 2011.

[10] A. Hyvärinen, and S.M. Smith, "Pairwise likelihood ratios for estimation of non-Gaussian structural equation models," *Journal of Machine Learning Research*, vol. 14, no. Jan, pp. 111-152, 2013.

[11] P. Spirtes,and K. Zhang, "Causal discovery and inference: concepts and recent methodological advances," *Applied Informatics*, vol. 3, no. 3, pp. 1-28, 2016.

[12] K.A. Bollen, *Structural equations with latent variables*, Wiley, 2009.

[13] S. Shimizu, "Learning LiNGAM based on data with more variables than observations," *arXiv preprint arXiv:1208.4183*, 2012.

[14] Y. Sogawa, S. Shimizu, T. Shimamura, A. HyväRinen, T. Washio, and S. Imoto, "Estimating exogenous variables in data with more variables than observations," *Neural Networks*, vol. 24, no. 8, pp. 875-2880, 2011.

[15] R.C. Cai, Z.J. Zhang, and Z.F. Hao, "SADA: A General Framework to Support Robust Causation Discovery," *in Proceedings of ICML(2)* , Atlanta, United States, June, 2013.

[16] A. Hyvärinen,K. Zhang, S. Shimizu, and P. Hoyer, "Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity," *Journal of Machine Learning Research*, vol. 11, pp. 1709-0731, 2010.