5. Hryshchenko S. (2014). Model of usage of geoinformation technologies during formation of environmental competence of future mining engineers. *Metallurgical and Mining Industry,* No 4. p.p. 8 – 9.
6. Morkun V., Tron V. (2014). Ore preparation multi-criteria energy-efficient automated control with considering the ecological and economic factors, *Metallurgical and Mining Industry,* No5, p.p. 4-7.
7. Morkun V., Morkun N., Tron V. (2015). Identification of control systems for ore-processing industry aggregates based on nonparametric kernel estimators, *Metallurgical and Mining Industry,* No1, p.p. 14-17.
8. Morkun V., Tron V. (2014). Ore preparation energy-efficient automated control multi-criteria formation with considering of ecological and economic factors, *Metallurgical and Mining Industry*, No5, p.p. 8-11.

Learning Tree Network Based on Mutual Information

Yinghan Hong[1, 2]

[1]*Physics and Electronic Engineering Department, Hanshan Normal University, Chaozhou 521041, Guangdong, China*
[2]*School of computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, Guangdong, China*

Guizhen Mai

School of Automation, Guangdong University of Technology, Guangzhou 510006, Guangdong, China

Zhusong Liu

School of computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, Guangdong, China

Abstract
Causal discovery is a fundamental problem in scientific research. Although many researchers are committed to finding causal relationships from observational data, large-scale causal discovery remains a tremendous challenge. In this work, an approach based on mutual information for tree network structure learning is proposed. In the first phase of our proposal, we present a skeleton learning method based on mutual information to discover the networks skeleton. In the second phase, an entropy-based direction learning approach is applied to distinguish the directions between nodes. It constructs the whole network of the high dimensional data sets at the end of the last iteration. Experimental results show that the algorithm outperforms the state-of-art information-geometry method especially when the dimension increases.
Key words: MUTUAL INFORMATION, STRUCTURE LEARNING, CAUSAL DISCOVERY

1. Introduction

Causal discovery is a fundamental problem in the field artificial intelligence (AI) field. In recent years, many researchers have been committed to finding causal relationships between variables based on observed data. However, the explosion of large-scale datasets in biomedicine and other fields has created a serious challenge for causal discovery algorithms, and causal discovery problems in high-dimensional data further pose a huge challenge. In this work, «scale» or «dimensionality» refers to the number of variables (or nodes) contained in a dataset (or network).

In statistics and machine learning communities, causal networks are commonly used to analyze the causal relationships among observed variables. A probabilistic graphical model can be constructed to simulate and evaluate the impact of a certain variable on other variables by running statistical significance tests on variable combinations. In general, the existing causal discovery methods can be classified into two main categories: structure learning methods and direction learning methods. Structure learning methods are based on the discovery of Markov equivalence classes [1] and include constraint-based methods and search-and-score methods.

Constraint-based methods, such as the PC algorithm [2] and the FCI algorithm [2], assume that the distribution of a given dataset is Markov and faithful with respect to the true graph G [1], meaning that all conditional independence relations in the joint distribution are entailed by the Markov condition. Constraint-based methods estimate from the data whether certain conditional independence relations hold between variables applying the d-Separation criterion [1]. For example, in the first step of the PC algorithm, the variables are assumed to be adjacent. In the next step, the dependence of two variables is tested given any other subset of other variables by conducting a conditional independence (CI) test. Using the d-Separation criterion, PC algorithm can discover the relationships in a given dataset.

Search-and-score methods [3, 4, 5], which also fall under the structure learning approach, date back even further, mainly to the works [6, 7, 8] and the references cited therein. These methods attempt to find a graph that maximizes a selected score function, such as the Bayesian Information Criterion (BIC) [9]. However, the search spaces of all directed acyclic graphs (DAGS), each of which has a certain number of variables grow super-exponentially with the number of variables [4]. Some methods use greedy search algorithms to solve Max-Score problem by obtaining the optimal value. For instance, the greedy equivalence search (GES) [4, 5] starts with an empty graph and undergoes two phases. In the first phase, edges are added until a local maximum of the score function is reached. In the second phase, the edges are removed until a local maximum is reached, and this local maximum is the output of the algorithm.

Although many encouraging results have been reported on the use of constraint-based methods and search-and-score methods, these types of methods suffer from two major drawbacks in solving large-scale causal discovery problems. In the large-scale case (i.e., when a high dimensional network has many nodes), the accuracy and reliability of a CI test with a high-order set are difficult to maintain, especially when the number of samples is insufficient. Furthermore, because of the super exponential growth of the search space, the result of the selected score function for searching a high-dimensional graph may also be unreliable when the dataset is limited. Therefore, an increase in the data dimension introduces challenges for structure learning approaches and significantly increases computational time.

The two methods also cannot distinguish between

two (or more) Markov equivalence classes that entail exactly the same set of conditional independence relations. For instance, $X \rightarrow Z \rightarrow Y$ and $X \leftarrow Z \leftarrow Y$. Therefore, this method may fail to accurately infer the direction. Thus, this method cannot accurately complete direction inference.

From our perspective, the structure learning approaches face the following challenges: (1) the correct DAG can only be identified using the discovered Markov equivalence classes (2) the accuracy of a high-dimensional CI test, which is generally accompanied by a large conditioning set, is difficult to maintain; and (3) searching for a graph similar to the actual graph from thousands of variables is practically intractable. Thus, the distance between the inferred graph and the actual graph cannot be ascertained.

To address the difficulties in causal direction learning with existence of Markov equivalence classes, researchers have recently considered asymmetrical relationships between the cause and effect variables under various assumptions. One approach is Additive Noise Models (ANMs), which were first proposed by [10], and are proved to be effective under the assumption that the data-generating process is linear and the noise distribution is non-Gaussian. This approach was later extended to a non-linear ANM, which was applied to continuous data [11, 12] and discrete data [13, 14]. Concretely, the ANM method suggests that, for two variables, x and y, such that $y = f(x) + \varepsilon$, where f is a nonlinear function and ε is a noise term statistically independent of x. When the distribution $P(x, y)$ allows for an ANM in one direction but not in the other, i.e., x cannot be obtained as a function of y plus independent noise, the former direction is inferred to be the causal direction (i.e., $x \rightarrow y$). The Post-Non-Linear (PNL) model [15] extends the identifiability results above to allow for an additional bijective transformation on the data with the use of a functional model of the form $y = g(f(x) + \varepsilon)$, where g: R $\rightarrow$ R is bijective. However, ANM and PNL are obviously feasible only for very small networks (approximately n < 7) [11]. When the sample size of the dataset is inconsiderably large or the signal-to-noise ratio is low, these methods also tend to be unreliable. Conventional approaches to causal inference are primarily based on Markov equivalence classes and ANMs.

More recently, some direction learning methods also take into account the symmetry between cause and effect from information geometric perspective [16, 17]. These methods assume that a random cause variable is independently generated through the invertible and deterministic mapping of its effect. Thus, dependencies between the density of the former and the slope of the latter are unlikely under the correct causal direction. An example of these direction learning methods is information-geometric causal inference (IGCI) [16]. An entropy-based version of IGCI can infer the direction of two variables, x and y, by measuring the complexity loss from density $p(x)$ to density $p(y)$. The estimation of complexity loss is given by $C_{x\rightarrow y} := H(y) - H(x) = - C_{y\rightarrow x}$, where $H(x)$ and $H(y)$ denote the entropies of x and y, respectively. Then, through IGCI, the following inferences may be made: x causes y if $C_{x\rightarrow y} < 0$ or y causes x if $C_{x\rightarrow y} > 0$ (as shown in Figure 1). Unlike ANM-based methods, information geometric methods can perform well when noise is low [16]. Furthermore, information geometric methods have lower time complexities than the other existing methods. However, these methods are suitable for ultra-low-dimensional data. Thus, none of these algorithms are capable of performing causality analysis effectively on high-dimensional data.

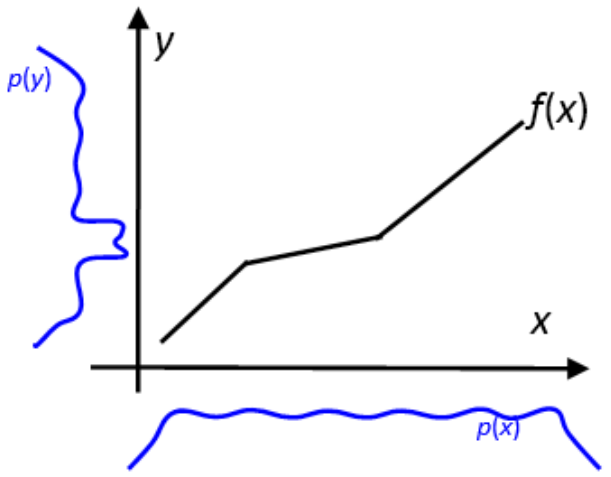


Figure 1. If the structure of the density of $p(x)$ is not correlated with the slope of f, then flat regions of f induce peaks of $p(y)$ (Taken from [16]).

For high-dimensional causal discovery problems, we realize that (1) the structure learning approach can discover the neighbors of a target variable contained from a relatively low dimensional network but cannot infer the exact direction of causalities and (2) ANM-based methods are useful to discover the causality between several variables but cannot handle high-dimensional data. Therefore, neither structure learning methods nor ANM-based methods can infer the causality in large-scale case.

Generally speaking, different causal networks have different relationships between their nodes, such as, scale-free network, sparse network, tree network and others have their own remarkable characteristics in term of the relationships between their nodes. The-

refore, it is usually difficult or expensive to find a generally method to reconstruct the corresponding structure from an arbitrary network. In other words, some remarkable characteristics of a corresponding network should be taken account for. Tree is a widely useful network structure, in which exists at most one path between any three nodes. We can see that if there is a path connecting three nodes *X*-*Y*-*Z* (there is no V-structure existing in tree structure), two disconnected nodes *X* and *Z* are doom to be separated by the middle node *Y* according to *d*-separation criteria. Motivated by this point, we present a mutual information-based structure learning method for tree network (MICN). In the first phase of our proposal, we pre- sent a skeleton learning method based on mutual information to discover the networks skeleton. We first construct the initial network with respect to the given data, by adding edges (undirected) between every two nodes (variables) to make sure each of pair nodes are directly connected. Note that, there is no triangle in tree structure. We find out all triangles and for every tree node in which, i.e., *X*, *Y* and *Z* we estimate the mutual information *I*(*X*, *Y*), *I*(*X*, *Z*) and *I*(*Y*, *Z*). We present a theoretical result in section 2 that if sets *X* and *Y* are *d*-separated by set *Z*, then $I(X; Y) \leq \min(I(X;Z), I(Y;Z))$. Because no V-structure can be contained in tree networks, every tree connecting nodes in the corresponding actual tree networks is satisfying: both sides of the nodes (*X* and *Y*) are *d*-separated by the middle one (*Z*). So one edge can be remove by mutual information test. After all triangles has been tested, finally we can obtain the tree skeleton (undirected graph) with respect to the corresponding given data. In the second phase, an entropy-based direction learning approach is applied to distinguish the directions between nodes. It constructs the whole network of the high-dimensional data sets at the end of the last iteration. Experimental results show that the algorithm outperforms the state-of-art information-geometry method especially when the dimension increases. Our idea can summarize the full map of our algorithm in Figure 2.

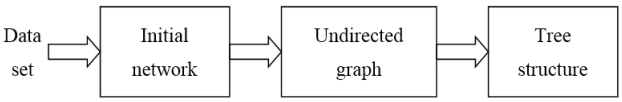


Figure 2. Sketch map of the full solution of structure learning

2. Preliminaries

2.1. *d*-Separation

Let *P* be a trail from variable x_1 to *y*. Then *P* is said to be *d*-Separated by set *Z* iff one of the following holds: (1) *P* contains a chain, $x_1 \leftarrow x_2 \leftarrow y$, where the middle node x_2 is in *Z*; (2) *P* contains a fork, $x_1 \leftarrow x_2 \rightarrow y$, where the middle node x_2 is in *Z*, or (3) *P* contains a collider x_2 that $x_1 \rightarrow x_2 \leftarrow y$, where the middle node x_2 is not in *Z* and no descendant of x_2 is in *Z*.

2.2. Conditional independence test

Conditional independence (CI) is an important concept in statistics. Let *X*, *Y* and *Z* denote sets of random variables. The CI between *X* and *Y* given *Z*, denoted by $X \perp Y|Z$, reflects the fact that given the values of Z, further knowing the values of *X* (or *Y*) does not provide any additional information about *Y* (or *X*). CI test play a central role in causal discovery [1]. One of the main reasons is that *d*-Separation criterion is always employed by CI test.

2.3. Structure learning based on mutual information

2.3.1. Mutual information and *d*-separation

Information theory [18] provides an inference method of dependency relationship between variables. Among them, mutual information is the statistical description of the correlation between two random variables. Supposing *I*(*x*; *y*) is the mutual information between two discrete random variables of *X* and *Y*, we have:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p_1(x)p_2(y)}\right)$$

Where $p_1(x)$ and $p_2(y)$ are the independent distribution probability of random variables, *x* and *y*, respectively. *p*(*x*, *y*) is the joint probability of *x* and *y*. First, we show that there are some relations between mutual information and *d*-separated as follows:

Theorem 1 In a DAG, sets *X* and *Y* are *d*-separated by set *Z* and iff *I*(*X*; *Y*|*Z*) = 0.

Proof If sets *X* and *Y* are *d*-separated by set *Z* and according to probability density definition of *d*-separation criteria, sets *X* and *Y* are independently distributed given set *Z*. On the contrary, if *X* and *Y* cannot be *d*-separated given *Z*, then we have *I*(*X*; *Y*|*Z*) ≠ 0.

It can be known from Theorem 1 that conditional mutual information can measure conditional independence between variables.

Theorem 2 If sets *X* and *Y* are *d*-separated by set *Z*, then $I(X; Y) \leq \min(I(X; Z), I(Y; Z))$.

Proof First *I*(*X*;*Y*|*Z*) = *I*(*X*;*Y*,*Z*) - *I*(*X*;*Z*), if *X* and *Y* are *d*-separated by set *Z*, according to Theorem 1, we have *I*(*X*;*Y*|*Z*) = 0, and then we obtain *I*(*X*;*Y*,*Z*) = *I*(*X*;*Z*). Moreover, *I*(*X*;*Z*|*Y*) = *I*(*X*;*Z*,*Y*) - -*I*(*X*;*Y*) and the basic inequality of conditional mutual information $I(X;Y|Z) \geq 0$, so conditional mutual information $I(X;Y,Z) \geq I(X;Y)$, so we can obtain $I(X;Y) \leq I(X;Z)$. In other words, conditional mutual informa-

tion is

$I(X;Y \mid Z) = I(Y;X \mid Z) = I(Y;X,Z) - I(Y;Z)$

Similarly, we can obtain mutual information $I(X;Y,Z) = I(X;Z)$ and conditional mutual information $I(X;Z|Y) = I(X;Z,Y) - I(X;Y)$ can be obtained. Similarly, mutual information $I(X;Y) \le I(X;Z)$ can be obtained. To sum up, we can obtain $I(X;Y) \le \min(I(X;Z), I(Y;Z))$.

According to Theorem 2, we estimate the mutual information between every two nodes. If there are two nodes X and Y are d-separated given Z, as shown in Figure 3, then we have $I(X;Y) \le \min(I(X;Z), I(Y;Z))$. X and Y are statistical independent given Z, so the edge directly connected between X and Y can be removed.

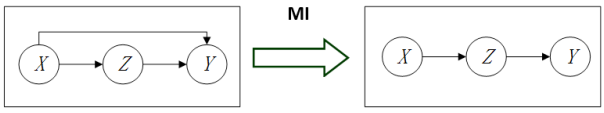


Figure 3. Mutual information test

2.3.2. Mutual information estimation based on continuous data

The traditional method to estimate mutual information from continuous data is usually based on uniform discretization and are calculated using Formula (3). However, using this method It is difficult to avoid losing some information about the continuous data. In view of this, Kraskov et al., proposed k-nearest neighbor-based algorithm to estimate mutual information from two continuous variables X and Y [19]. $I(X;Y) = \Psi(k) + \Psi(N) - <\Psi(n_x + 1) + \Psi(n_y + 1)>$. Where Ψ is bigamma function, k is the number of nearest-neighbors, N is the sample size of variables X and Y, n_x and n_y are the sums of k nearest-neighbors of variables X and Y respectively. This algorithm does not require many parameters and it also can obtain relatively reliable mutual information estimation value with a small sample size. In our work, we use this me algorithm to estimate mutual information from continuous variables.

3. The proposed method

3.1. The framework of MICN

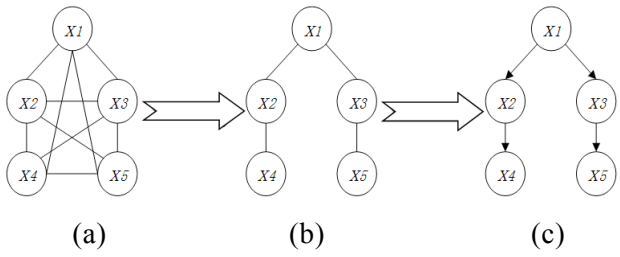


Figure 4. Processing of MICH

MICN includes two stages, shown as Figure 4.

In the first stage of the algorithm, the high-dimensional dataset is used to construct the initial network, as shown in Figure 4(a), there are edges between every two nodes. Then, mutual information test method based on d-separation criteria is present to remove redundant edges according to Theorem 2, and we obtain the skeleton of the network as shown in Figure 4(b). In the second stage of the algorithm, information-geometric method IGCI is used to orient the edges from the skeleton. Finally, the corresponding network structure of high dimensional datasets can be obtained, as shown in Figure 4(c).

3.2. Skeleton learning

The first stage of MICN is to learning the skeleton from given data, the method we presented is mainly based on Theorem 2, and is as follows:

Step 1 Construct the initial network X: we add edges between every two nodes to make sure each of pair nodes are directly connected.

Step 2 For network X, and its node set $X = \{x_1, x_2, \ldots, x_3\}$, select arbitrary three connected nodes x_1, x_2 and x_3 to test where the relation between them satisfies Theorem 2, $I(x_1;x_3) \le \min\{I(x_1;x_2), I(x_3;x_2)\}$ If so, nodes x_1 and x_3 is d-Separated by node x_2, so there is no causality between x_1 and x_3, then we can remove the edge between them. Similarly, we calculate $I(x_1;x_2) \le \min\{I(x_1;x_3), I(x_2;x_3)\}$ and $I(x_2;x_3) \le \min\{I(x_2;x_1), I(x_3;x_1)\}$ to decide whether the edge x_1-x_2 *or* x_2 x_3 can be removed.

Step 3 Repeat Step 2 until every three nodes in network X are tested.

Through the above three steps, we can remove all the redundant edges and finally obtain the corresponding skeleton. This method estimate mutual information between only two nodes, thus is very fast and efficient. The detailed of the method is shown as Algorithm 1:

Algorithm 1 Skeleton learning

Input: Datasets $X = \{x_1, x_2, \ldots, x_n\}$.
Output: The corresponding skeleton of X.
1: Construct the initial network X: add edges between every two nodes.
2: For $i = 1$ to n
3: For $j = 1$ to n
4: For $k = 1$ to n
5: If $i \ne j$, $i \ne k$, $k \ne j$ and x_i, x_j, x_k form a triangle
6: find the two nodes of x_i, x_j, x_k meet the condition: $Min\{I(x_i;x_j), I(x_i;x_k), I(x_j;x_k)\}$, e.g., i and j.
7: remove the edge between x_i and x_j
8: End if
9: End for
10: End for
11: End for

3.3. Direction learning

The second stage of MICN is to infer direction using an information geometric IGCI. IGCI assume that a random cause variable is independently generated through the invertible and deterministic mapping of its effect. Thus, dependencies between the density of the former and the slope of the latter are unlikely under the correct causal direction. IGCI can infer the direction of two variables, x and y, by measuring the complexity loss from density $p(x)$ to density $p(y)$. The estimation of complexity loss is given by $C_{x\to y} := H(y) - H(x) = -C_{y\to x}$, where $H(x)$ and $H(y)$ denote the entropies of x and y respectively. Then, through IGCI, the following inferences may be made: x causes y if $C_{x\to y} < 0$ or y causes x if $C_{x\to y} > 0$. Unlike ANM-based methods, information geometric methods can perform well when noise is low. Furthermore, information geometric methods have lower time complexities than the other existing direction learning methods. Here IGCI is used to infer the directions of the skeleton. The detailed of this stage is shown as algorithm 2.

Algorithm 2 Direction learning

Input: variables x and y.
Output: the causal direction between x and y

1: Computing the entropy $H(x)$ and $H(y)$ of x and y, respectively.
2: If $H(x) > H(y)$
3: return $x \to y$;
4: End if
5: If $H(x) < H(y)$
6: return $y \to x$;
7: End if

4 Experiments

In this section, we describe some experiments to show that our approach can discover causality on various dimensions in both simulated and real-world networks. In which, we also provide comparative results of MICN with other causal inference method IGCI. The efficient and accurate estimation of mutual information plays a significant role in the performance of MICN . All the data in simulated experiments are continuous and the mutual information estimator is employed.

4.1. Experimental setup

The precise form of the data generating process consists of two phases: 1) structure generating phase and 2) data generating phase. In the structure generating phase, we specifically generate the trees with {100; 300; 500; 700; 900} nodes to evaluate the effect of our algorithm in different dimensions, and the default sample size is 500.

In the data generating phase, for the topological order of the structure generated above can be easily obtained, then the rest of variables are generated with the function: $y = f(x) + \varepsilon$ according to the topological order of a certain skeleton. In these experiments, the default function is $y = w_i * \sin(x_i) + w_j * \sin(x_j) + \varepsilon_i$, in which w is a weight of the corresponding function and randomized at $[-0.3,-0.7] \cup [0.3,0.7]$, the default additive noise ε_i will follow uniform distribution and the weight is fixed at 5%.

In the experiments of real-world, we use Cancer network, Asia network and Gene Link network to evaluate the effect of our algorithm in different dimensions. And node that all the three networks are not tree. We evaluate the performance of our approach by three criteria, Precision, Recall and F1, which are defined below:

$$\text{Precision} = \frac{\{\text{discovered causality}\} \cap \{\text{actual causality}\}}{\{\text{discovered causality}\}}$$

$$\text{Recall} = \frac{\{\text{discovered causality}\} \cap \{\text{actual causality}\}}{\{\text{actual causality}\}}$$

$$\text{F1} = \frac{2 * \{\text{discovered causality}\} * \{\text{actual causality}\}}{\{\text{discovered causality}\} + \{\text{actual causality}\}}$$

One can see that Precision is the actual fraction of inferred causality with respect to a true graph. Similarly, Recall is the fraction of actual causality found by the algorithm. F1 is the organic combination of Precision and Recall which can be a criterion of accuracy to our algorithm.

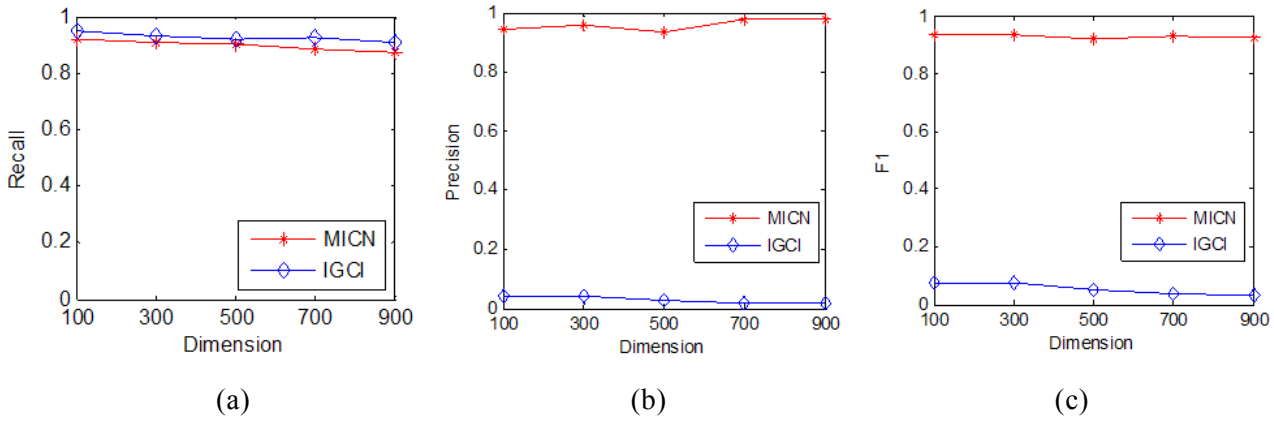
4.2. Results on simulated data

In this section, we compare our method against IGCI, the mainstream causal inference method, under {100; 300; 500; 700; 900} dimensional networks, and the default sample size is 500.

Figure 5 (a) shows that the Recall of IGCI is slightly higher than that of MICN. Both the two method can discover most of the actual causality.

Figure 5 (b) shows that the Precision of MICN is significantly better than IGCI. Most of the edges discovered by MICN are accurate. While IGCI cannot do that. The main reason of that is IGCI cannot remove the redundant edges during direction learning without a corresponding skeleton.

Figure 5 (c) shows that the F1 of MICN is significantly better than IGCI. It can be seen from this group of experiments, for the high-dimensional tree network, MICN owns excellent structure learning capacity and the performance of which is obviously better than that of IGCI.



(a) (b) (c)

Figure 5. Performance of MICN on different dimension

4.3. Demonstration on real data

4.3.1. Sample data set

In this subsection, we use tree real-word networks to evaluate our proposed method MICN, and the three real-word networks are as follows.

Cancer network: a very small causal network for a fictitious medical example about whether a patient has metastatic brain cancer based on headaches, coma and serum calcium. It has 5 nodes, Metastasis Cancer, Brain Tumor, serum calcium, coma and headaches.

Asia network: a small causal network for a fictitious medical example about whether a patient has tuberculosis, lung cancer or bronchitis, related to their *X*-ray, dyspepsia, visit-to-Asia and smoking status. Also called «Chest Clinic». It has 8 nodes, respectively VisitAsia, Tuberculosis, Smoking, Cancer, TborCa, Xray, Bronchitics and Dyspnea.

Gene Link network: a high-dimensional (724 nodes) causal network for the linkage between two genes. One of the genes is the human LQT syndrome (a rare heart disease) and the other is a genetic marker.

4.3.2. Experimental results of real network

In this subsection, we use Cancer, Asia and Gene Link networks to evaluate our proposed method MICN, we generate 2000 samples from the corresponding CPTs (conditional probability tables). Note that the three networks are not tree structure, thus it can evaluate the performance of our proposed method in more general cases. As shown in Figure 6, the performance of MICN is significantly better than that of IGCI. In Cancer and Asia networks, the Recall of IGCI is better than MICN, which means that most of the actual edges are inferred correctly by IGCI. But the performances of IGCI on other two criteria Precision and F1 seem to deviate from MICN, because the accuracy of IGCI is extremely lower if it works on multi-dimensional data without discover the causal skeleton. Furthermore, as shown in Figure 7, the performance of MICN on the high-dimensional network Gene Link is not good enough, one of the main reason is that Gene Link in which 45.21% nodes have more than one father node, and is very different from tree structure. In addition, it also exactly shows that the MICN has good robustness on a wide range of different structures.

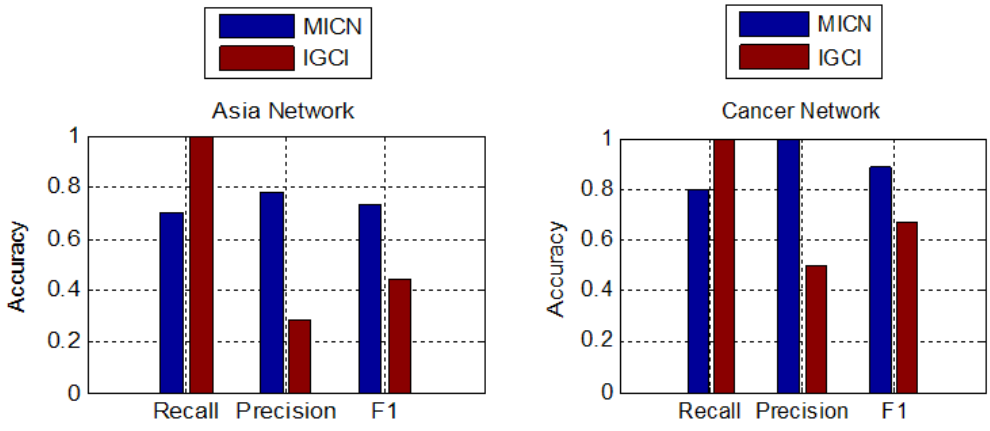


Figure 6. Comparison between MICN and IGCI assessment parameters in Cancer network (left) and Asia Network (right)

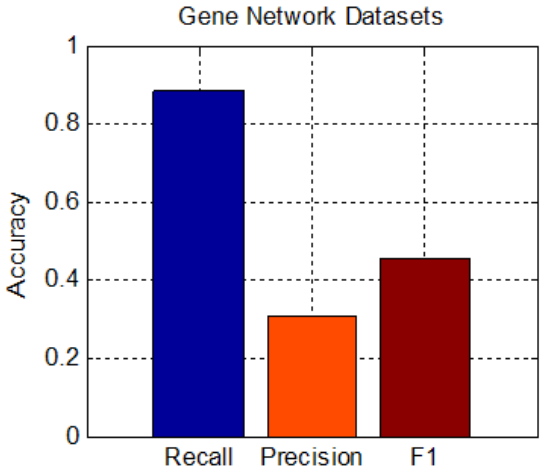


Figure 7. MICN performance assessment parameters in Gene Link network

5. Conclusions

Using traditional Bayesian network structure learning algorithms to construct network structure from the corresponding high-dimensional data is often inefficient and unreliable. In this paper, we proposed a fast and efficient algorithm for tree network structure learning, from the perspective of information. We first estimated mutual information between every two nodes to construct the skeleton with respect to the data. And then an information-geometry method IGCI was employed to infer the directions of edges. Both of the two steps are very fast and efficient, since it is very easy to estimate mutual information or entropy from two (or one) variables. Moreover, as the show of the experimental results, the proposed methods can maintain high inference accuracy with the data dimension increases, which outperformed the other learning methods.

Acknowledgements

This work was supported by Science and Technology Planning Project of Guangdong Province, China (2015A030401101), (2012B040500034) and by National Natural Science Foundation of China(61572144).

References

1. Pearl J. (2009) *Causality: models, reasoning and inference*. Cambridge: The MIT press.
2. Spirtes, Glymour C, Scheines R. (2000) *Causation, Prediction, and Search*. The MIT Press.
3. Herskovits, Edward (1991) *Computer-based probabilistic-network construction*. Diss. Stanford University.
4. Chickering, David Maxwell (2003) Optimal structure identification with greedy search. *The Journal of Machine Learning Research*, 3(3), p.p.507-554.
5. Meek, Christopher (1997) *Graphical Models: Selecting causal and statistical models*. Diss. PhD thesis, Carnegie Mellon University.
6. Geiger, Dan, and David Heckerman (1994) Learning Gaussian networks. *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc.
7. Heckerman, David, Christopher Meek, and Gregory Cooper (1999) A Bayesian approach to causal discovery. *Studies in Fuzziness and Soft Computing*, 194, p.p.1-28.
8. Chickering, David Maxwell (2003) Optimal structure identification with greedy search. *The Journal of Machine Learning Research*, 3(3), p.p.507-554.
9. Schwarz, Gideon (1978) Estimating the dimension of a model. *The annals of statistics* 6(2), p.p.461-464.
10. Shimizu, Shohei, et al. (2006) A linear non-Gaussian acyclic model for causal discovery. *The Journal of Machine Learning Research*, 7(6), p.p.2003-2030.
11. Hoyer, Patrik O., et al. (2009) Nonlinear causal discovery with additive noise models. *Advances in neural information processing systems*. p.p.689-696.
12. Peters, Jonas, et al. (2014) Causal discovery with continuous additive noise models. *The Journal of Machine Learning Research*, 15(1), p.p.2009-2053.
13. Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf (2010) Identifying cause and effect on discrete data using additive noise models. *International Conference on Artificial Intelligence and Statistics*. p.p. 597-604.
14. Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf (2011) Causal inference on discrete data using additive noise models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12), p.p.2436-2450.
15. Zhang, Kun, and Aapo Hyvärinen (2008) Distinguishing causes from effects using nonlinear acyclic causal models.» Journal of Machine Learning Research, *Workshop and Conference Proceedings*, p.p.203-230.
16. Janzing, Dominik, et al. (2012) Information-geometric approach to inferring causal directions. *Artificial Intelligence*, 182(12s), p.p. 1-31.
17. Daniusis, Povilas, et al. (2012) Inferring deterministic causal relations. *Artificial Intelligence*, 182(12s), p.p. 1-31.
18. Cover, Thomas M., and Joy A. Thomas. (2012)

Elements of information theory. John Wiley & Sons.
19. Kraskov, Alexander, Harald Stögbauer, and Peter Grassberger (2004) Estimating mu- tual information. *Physical Review E*, 69(6), p.p.066138-066154.

Discussion on the Experiment Teaching Reform for Electronic Measurement Technology

Li Shihong, Chen Bin

Information Technology Department, Wenzhou Academy of Agricultural Sciences, Wenzhou 25300, Zhejiang, China

Abstract
Targeting at the existing problems in current electronic measurement course practice teaching, a teaching method of implementing comprehensive experiment practice program is proposed to better develop students' problem analysis ability and independent innovation capability. Through progressive practice, the students are required to construct a complete digital signal generator step by step: welding and debugging single chip processor's minimum system and AD9850 module, checking its circuit functions, testing the waveforms and parameters of AD9850 module, and assembling and debugging the entire digital signal generator. This teaching method not only mobilizes students' learning enthusiasm and achieves good teaching effect, but also cultivates their practical engineering application ability and improves their measuring skills and innovation capabilities, fully satisfying the needs of the enterprises.
Key words: ELECTRONIC MEASUREMENT TECHNOLOGY, EXPERIMENT TEACHING, PRACTICE ABILITY

1. Introduction

Electronic measurement is an electronic technology-based measuring technique. The development of effectively promotes the application of electronic measurement and allows it to play a vital role in people's production and living activities. As a specialized course for students majoring in Electronic Information Engineering Technology, Electrical Engineering and Automation, Telecommunications Engineering Technology, Measuring and Controlling Technology, electronic measurement is a highly practical course that develops students' production practice ability, comprehensive application competence and adaptive capability.

The following problems can be found in the practice teaching of electronic measurement in many colleges and universities at home and abroad:

(1) The experimental teaching instruments are backward and worn-out and update slowly, failing to keep up with the industry's development.

(2) The experimental practice links mostly involve fundamental replication experiments and lack com-