

Towards Fast and Efficient Algorithm for Learning Markov Blankets

Yanying Li, Youlong Yang, Xiaofeng Zhu and Xiaoxia Dou

Abstract—Finding the Markov blanket (MB) of a target variable is an important step in many applications including Bayesian network structure learning, optimal feature subset selection and dimensionality reduction. Inspired by the algorithm IPC-MB, we present a fast and efficient algorithm FEIPC-MB to learn the MB. Unlike the previous algorithms, the condition independence(CI) tests of FEIPC-MB is implemented according to the strength of the correlation of variables, which reduces the high order CI tests significantly. Simulations illustrated that the proposed algorithm outperforms the competitors with respect to the Precision, Recall, global Euclidean Distance, run time and the number of CI test, respectively. The highest reduction percentage of CI test can reach to 100 % by FEIPC-MB compared to IPC-MB algorithm.

Index Terms—Bayesian network, Markov blanket, Learning structure

I. INTRODUCTION

Due to the wide application of the classification, optimal feature subset selection and the problem of explosion of high dimensionality data sets, local learning of Markov blanket (MB) [1] regained researchers attention after of the work of KS [2]. MB plays an important role in many domains such as feature selection for classification, casual discovery, and Bayesian network (BN) leaning. Furthermore, large data sets are becoming the norm and traditional methods designed for data sets with a modest number of variables, which will raise new challenges in practical environment. It is well known that predictive models benefit from a compact, non-redundant subset of features that improves the interpretability and generalization. A principled solution to this problem is to determine the MB of the target variable. In addition, a majority of constraint-based methods for learning BN structures fail to scale to more than a few hundred variables. A reasonable compromise to learning the full BN is to discover the local structure around a target variable or a set of interest targets. Therefore, MB is undoubtedly the best choice because the discovery of MB can be employed to induce the full BN structure.

Over the last several decades, significant progress has been made in the domain of discovering MB. The KS algorithm

is the earliest algorithm to learn MB and it is heuristic and gives no theoretical guarantee of correctness. Grow-shrink (GS) algorithm [3] is the first sound algorithm for leaning MB. As indicated by its name, it consists of the growing and the shrinking two sequential stages. Since then, several variants of GS, such as IAMB, interIAMB [4] and Fast-IAMB [5] are proposed successively to improve the speed and reliability. Given some broad assumptions and conditions, these algorithms are perfect for theory, and indeed save much time for learning MB, but they need a large number of samples to ensure the correct and reliable results in this process. In light of this case, a great quantity of efforts are made to overcome above mentioned weaknesses, including MMPC/MB [6], HITON-PC/MB [7], PCMB [8] and IPC-MB [9]. More important, we must note that MMPC/MB and HITON-PC/MB are not always correct and that both of PCMB and IPC-MB are proved to be correct, scalable and data-efficient. Besides, IPC-MB is stated with best trade-off among all other algorithms for learning MB.

However, all the algorithms mentioned above have to perform a large number of CI tests. In the worst case, some of the algorithms require the number of CI tests that increases exponentially with the number of variables. It is also worthwhile noting that large conditioning sets usually lead to errors of CI tests and result in a poorer estimation of dependent relationships for a small sample size. Thus, we tend to use CI tests of lower orders. Only in this way can we obtain more reliable results of tests. Therefore, it is vital to explore some measures to avoid or alleviate the potential problem of combinatorial explosion for CI test.

In view of above discussions, we propose a novel MB local learning algorithm, called FEIPC-MB which is based on iterative Parent-Child based search of MB (IPC-MB). Unlike the previous algorithms, the CI test of FEIPC-MB is based on the strength of the correlation of variables. For a target variable T , we first test variables with weak correlation with T and take those variables which are strong correlation with T as the conditioning sets. We can use partial correlation analysis, covariance analysis and mutual information and so on to evaluate the strength of correlation between variables. In this paper we use mutual information to sort the variables belonging to candidate $PC(T)$ of the target variable of interest T , and then, based on the sort, use CI tests to remove the (conditional) independent node X with order. Here we also alter the pattern of selecting conditioning sets for each CI test according to the sort. The ordered selection of X and the ordered selection of the conditioning sets can remove some false positive nodes as early as possible with few number of CI tests. More important, we can reduce the number of high order CI tests and increase the power of tests. Thus we efficiently alleviate the problems discussed above. Simulations illustrated that the proposed algorithms

Manuscript received October 31, 2014; revised February 14, 2015. This work was supported by the National Nature Science Foundation of China (No. 61403290, 11301408, 11401454), the Foundation for Youths of Shaanxi Province (No.2014JQ1020), the Foundation of Baoji City (No.2013R7-3) and the Foundation of Baoji University of Arts and Sciences(No.ZK15081).

Yanying Li is with the School of Mathematics and Statistics, Xidian University, Xi'an, Shannxi, 710126 P.R. China and the College of Mathematics and Information Science, Baoji University of Arts and Sciences, Baoji, Shannxi, 721013, P.R.China e-mail: liyanying81@163.com.

Youlong Yang and Xiaofeng Zhu are with the School of Mathematics and Statistics, Xidian University, Xi'an, Shannxi, 710126, P.R. China e-mail:ylyang@mail.xidian.edu.cn, zxfq115@163.com

Xiaoxia Dou is with the College of Mathematics and Information Science, Baoji University of Arts and Sciences, Baoji, Shannxi, 721013, P.R.China e-mail:34398882@qq.com

outperform their competitors with respect to computational complexity, run-time and accuracy, respectively. The highest reduction percentage of CI test is 100 % by FEIPC-MB compared to IPC-MB algorithm.

The structure of this paper is as follows. In Section 2, some important notations and definitions are given. Section 3 describes the details of our improvement on the IPC-MB algorithm for learning MB of the target variable. Section 4 presents some simulation results and compares the improved method with its closest competitor in details. Final section provides conclusion.

II. DEFINITIONS AND PRELIMINARIES

In this section, we introduce some concepts and theorems closely related to this paper. Some well-known concepts can refer to books on Bayesian network and articles, e.g. [1], [5], [10], [11], [16]. Let capital letters denote discrete random variables and lower case denote states of variables. We denote a set of variables by upper-case bold-face \mathbf{Z} , and we apply the corresponding lower-case bold-face \mathbf{z} to denote configurations of state to each variable in the set. In specific distribution P , we denote the conditional independence of the variables X and Y given \mathbf{Z} with $Ind_p(X; Y|\mathbf{Z})$.

A graph $G = (\mathbf{U}, \mathbf{E})$ consists of a set of nodes $\mathbf{U} = \{X_1, X_2, \dots, X_n\}$ and a set of directed edges $\mathbf{E} \subseteq \mathbf{U} \times \mathbf{U}$. We denote the assertion that X is d-separated from Y given \mathbf{Z} with $D_{sep}(X; Y|\mathbf{Z})$ in a given G . If $(X_i, X_j) \in \mathbf{E}$ and $(X_j, X_i) \notin \mathbf{E}$, then the edge $(X_i, X_j) \in \mathbf{E}$ is called directed, which can be denoted by $X_i \rightarrow X_j$. If both $(X_i, X_j) \in \mathbf{E}$ and $(X_j, X_i) \in \mathbf{E}$, the edge is called undirected, we indicate it with $X_i - X_j$. Markov condition: a node X is conditionally independent of its non-descendants given its $Pa(X)$. Bayesian networks (G, P) , also known as probabilistic belief networks or casual networks, are graphical models that cover two components that respectively codify qualitative and quantitative knowledge: (i) graphical structure, more precisely, a directed acyclic graph (DAG). As above mentioned, the nodes correspond to random variables from problem studied by us and the edges represent dependence relationship among the variables, and (ii) A set of numerical parameters Θ consists a series of conditional probability distributions $P(X_i|Pa(X_i))$ which can be drawn from the structure of graph, where $Pa(X_i)$ denotes any combination of the values of the parents of variable X_i . Combine these conditional distribution and Markov condition, we can recover the joint probability distribution over \mathbf{U} :

$$P(X_1, X_2, \dots, X_n) = \prod P(X_i|Pa(X_i))$$

This property gives rise to important savings in storage requirements and also facilitates performance of probability inference. A node or variable X is called a collider in a path π if X having two incoming edge, denoted as $Y \rightarrow X \leftarrow W$. A collider node is also known as a head to head vertex. A v-structure in a DAG is a triple of nodes X, Y, Z such that there exist directed edges $X \rightarrow Z, Y \rightarrow Z$, and X and Y are not connected. The skeleton of a DAG is the undirected graph which can be obtained from G by dropping the arrows of directed edges.

It is well known that for the same distribution P generated by a DAG G , there is a whole equivalence class of DAGs

corresponding to P . In some extent, even there exist infinitely many observations, we cannot distinguish among the difference of DAGs of an equivalence class. We can characterize equivalent classes: Two DAGs are equivalent if and only if they are equipped with the same skeleton and the same v-structures. Therefore, we can use a partially directed acyclic graph (PDAG) to represent the underlying DAG.

Definition 1(blocked path) A path between node X and Y is blocked by a set of vertices \mathbf{Z} , if there exists a node W on the path for which one of the following conditions holds: (i) W is not a collider and $W \in \mathbf{Z}$, or (ii) W is a collider and neither W nor its descendants are in \mathbf{Z} ;

Of course, a path that is not blocked is active or open. In DAG, two nodes X and Y are d-separated by a set \mathbf{Z} if and only if every path from X to Y is blocked by \mathbf{Z} , which is denoted as $D_{sep_G}(X; Y|\mathbf{Z})$.

Definition 2(Conditional independence) Two variables X and Y are conditionally independent given \mathbf{Z} , denoted as $Ind_P(X; Y|\mathbf{Z})$, if and only if $\forall x, y, z$ and $P(\mathbf{Z} = \mathbf{z}) > 0$, the following formula holds,

$$\begin{aligned} P(X = x, Y = y|\mathbf{Z} = \mathbf{z}) \\ = P(X = x|\mathbf{Z} = \mathbf{z})P(Y = y, \mathbf{Z} = \mathbf{z}) \end{aligned}$$

Definition 3(Faithfulness) A directed acyclic graph G is faithful to a joint probability distribution P over variable set \mathbf{U} if and only if every independence present in P is entailed by G and the Markov condition. A distribution P is faithful if and only if there exists a directed acyclic graph G such that G is faithful to P . [9]

Definition 4(Markov blanket) Given the faithfulness assumption, the Markov blanket of a target T , denoted as $MB(T)$, is a minimal set conditioned on which all other variables are independent of T .

This concept implies that $MB(T)$ can shield all effect of other nodes on T . Therefore the information of the $MB(T)$ is enough to determine the probability distribution of T and the knowledge of all rest of nodes become superfluous, then the variables in the $MB(T)$ are adequate for global optimal feature combination for classification. From graphical viewpoint, $MB(T)$ consists of the parents, children and spouses of T .

Theorem 1 A Bayesian network(BN) satisfies the faithfulness condition, then

$$D_{sep_G}(X; Y|\mathbf{Z}) \Leftrightarrow Ind_P(X; Y|\mathbf{Z})$$

In the paper, according to Theorem 1, the terms conditional independence and d-separation are used interchangeably.

III. FAST AND EFFECTIVE ALGORITHM: FEIPC-MB

This section lists the Pseudo code of FEIPC-MB and describes the major steps in details.

It is worthwhile to noting that the iterative Parent-Child based on search of MB (IPC-MB) [9] algorithm for learning Markov blanket of variable T is stated with the best trade-off among all published works of this family in terms of soundness, time and sample efficiency. In our proposed algorithm of fast and efficient IPC-MB (FEIPC-MB), we alleviate the complexity of IPC-MB and improve its accuracy. What is important is that we try to avoid CI tests

Algorithm RecognizePC

Input: target variable T adjacency set ADJ_T
 dataset D threshold ε ;
Output: $PC(T)$.

1. $NonPC(T) = \emptyset$.
2. $Cand_PC = ADJ_T \setminus \{T\}$
3. Rank the variables $X \in Cand_PC(T)$ in ascending order according to $I(X, T)$ and remove X from $Cand_PC(T)$ if $I(X, T) \leq \varepsilon$
4. $Test_conditioning_set$ consists of variables $X \in Cand_PC(T)$ in reverse order based on step 3
5. $Sepsetsize = 0$
6. while $|Cand_PC| > Sepsetsize$
7. for every $X \in Cand_PC(T)$
8. for each $S \subseteq Test_conditioning_set \setminus \{X\}$ with $|S| = Sepsetsize$
9. if $Ind_p(X \perp T | S)$, then
10. $NonPC = NonPC \cup \{X\}$
11. $Sepset_{T, X} = S$
12. break
13. end if
14. end for
15. end for
16. if $|NonPC| > 0$, then
17. $Cand_PC(T) = Cand_PC(T) \setminus NonPC$ and $Test_conditioning_set = Test_conditioning_set \setminus NonPC$
18. $Sepsetsize = Sepsetsize + 1$
19. $NonPC = \emptyset$
20. else
21. break
22. end if
23. end while
24. return $Cand_PC(T)$

with large conditioning sets and reduce the number of CI tests by using a greedy-search method. According to the strength of the correlation with a target variable T , we sort the variables in candidate $PC(T)$ and then reduce the conditional independent variables from current $PC(T)$ with smaller and compact conditioning sets. We use the mutual information to evaluate the correlation of variables. The following procedure describes the details.

The Algorithm **RecognizePC**, learning the parents and children set of variables, is central to the whole FEIPC-MB algorithm, because the identification of spouses of target T is based on its parents and children. So we try our best to take some measures to determine the correct $PC(T)$ as far as possible. The algorithm RecognizePC consists of two shrinking phases to remove the non-PC nodes from the candidate PC set of T . The first phase computes the mutual information values between T and the nodes in its adjacency set ADJ_T (initial state is all nodes except T), and then rank these variables in ascending order by the mutual information values and remove X from the current candidate PC set ($Cand_PC(T)$), if $I(X, T) \leq \varepsilon$. The sort is vital important for the next shrinking phase. It is well known that larger value of mutual information between X and T implies that X is more likely to be directed neighborhood of T for graphical models. In contrast, smaller mutual information between X and T means that T is weakly

Algorithm FEIPC-MB

Input: dataset D threshold ε
Output: $MB(T)$.

1. $CanADJ_T = U \setminus T$
2. Call $RecognizePC(T, ADJ_T, D, \alpha)$ to get $PC(T)$
3. $MB = PC(T)$
4. for every $X \in PC(T)$ do
5. $CanADJ_X = U \setminus X$
6. $Cand_SP = RecognizePC(X, CanADJ_X, D, \varepsilon)$
7. if $T \notin Cand_SP$, then
8. $MB = MB \setminus \{X\}$
9. continue
10. end if
11. for each $Y \in Cand_SP$ and $X \notin MB$ do
12. if $\neg Ind_p(T, Y | Sepset_{Y, T} \cup X)$ then
13. $MB = MB \cup \{Y\}$
14. end if
15. end for
16. end for
17. return MB

related to X or X is marginal or conditional independent of T . Thus, in the second shrinking phase, we first check the independence relationship between target node and its adjacent variables with smaller mutual information values in $Cand_PC(T)$. In this process, we use variable(s) with larger mutual information value(s) with target T from ADJ_T as the conditioning sets for CI tests. In other words, we first select these nodes sorted at the end of the sequence as the conditioning sets, which is different from the general methods that randomly select node or sets from the candidate PC set. The ordered selection of X and the ordered selection of the conditioning sets can remove some false positive nodes as early as possible with few number of CI tests. More important, we can reduce the number of high order CI tests and increase the power of tests.

Next, the second stage, see the Algorithm **FEIPC-MB**, which determines the spouses of target variable is identical to the original phase of the IPC-MB. Note that we don't change the criterion for determining the condition independence relation of two variables in the proposed algorithm, so the soundness of IPC-MB still stands in FEIPC-MB.

IV. EXPERIMENTS AND ANALYSIS

In this section, we compare the FEIPC-MB method with other state-of-the-art and prototypical algorithms which deal with the same problems, respectively. The experiments are run on a Pentium 3.19GHz with 1.96GB RAM using Windows XP system and Matlab in version R2009a.

Because IPC-MB has been proved to be the best trade-off among the family of methods to learn Markov blanket of target node, we only compared FEIPC-MB with IPC-MB. We test both of them on three well-known networks. The first is Alarm network [12] which is a widely accepted benchmark for evaluating the performance of many algorithms. The second is Insurance network [13] which is for estimating the expected claim costs for a car insurance policyholder, consists of 27 vertices. And the last is Hailfinder network [14], a BN with 56 variables used for modeling and predicting the weather. To guarantee the reliability of the experimental

results and fair for comparison, we use different size of samples which are randomly generated by true networks, respectively. Meanwhile, the number of data sets with the same size and the same network are ten. We introduce the common and classical criterions such as Precision, Recall and Distance to evaluate the performance of IPC-MB and FEIPC-MB. The Precision is the ratio of the number of true positive in the MB(T) over the size of MB(T) in the output; Recall is the number of the true positives in the MB(T) in the output divided by the size of true positives in the true BN; and the global Euclidean distance metric assessing the quality of the learned Markov blanket, was expressed as

$$\sqrt{(1 - \text{precision})^2 + (1 - \text{recall})^2} \quad (1)$$

In addition, the complexity of the involved algorithms is measured by the number of CI test and run time.

Notice that, the significance level α for the conditional independence test is set to 0.05 and threshold $\varepsilon = 0.001$ for all algorithms used in this paper. Besides, our implementation is based on the Bayesian network toolbox written by Murphy [15].

Foremost, the experiments are carried out on the database of the Alarm network. The average accuracy of different aspects employed by each algorithm is reported in Fig. 1. Of course, the value of every point is the average over 10 databases with the same size. The size of the data sets is varied from 1000 to 15000. $k = 1000$ is the basic of the horizontal axis in each subgraphs in Fig.1. Fig. 1.(a) and Fig. 1.(b) show the values of precision and recall returned by FEIP-MB compared to that of the IPC-MB. Obviously, FEIPC-MB achieves a satisfactory performance in above metrics, individually.

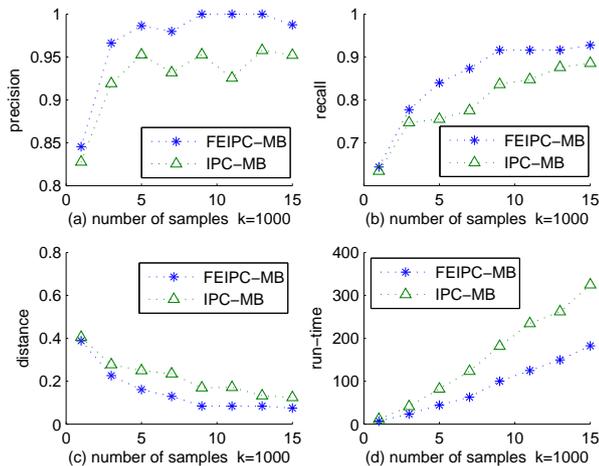


Fig. 1. Results relative to the average precisions, recalls, distances and time consumption returned by FEIPC-MB and IPC-MB for learning Markov blankets of Alarm network. Higher values imply a better performance for both sub-figures (a) and (b), on the contrary, lower values indicate outstanding performance for (c) and (d).

Fig. 1.(c) depicts the distances between true BN and output obtained by IPC-MB and FEIPC-MB. Its worth noting that the smaller value of distance, the closer the result of an algorithm is to the true local neighborhood. In addition, these measured metrics were calculated for each variable of the Alarm network independently and then averaged. As

shown Fig. 1.(d), FEIPC-MB saves much time than IPC-MB. Overall, we can draw conclusions that FEIPC-MB is data efficient and FEIPC-MB is super to IPC-MB in all aspects in our experiments.

It is common to consider the number of CI test and the highest orders of tests performed by an algorithm as the main criterion of computational complexity and a major contributors to the algorithm run-time. This is because that run-time relies on the orders of CI tests and the number of states of each nodes included in the conditioning set. Fig. 2 shows the average number of CI tests on eight sample sizes returned by FEIPC-MB and IPC-MB, respectively. The horizontal axis is the the order of CI test in each subgraph. The percentages of saved CI tests by FEIPC-MB compared to IPC-MB with increasing orders are depicted in Fig. 3 on different sizes of samples. The horizontal axis is the the order of CI test in each subgraph in Fig. 3. And the sample sizes of the eight subgraphs are ranged from 1000 to 15000 in Fig. 3.

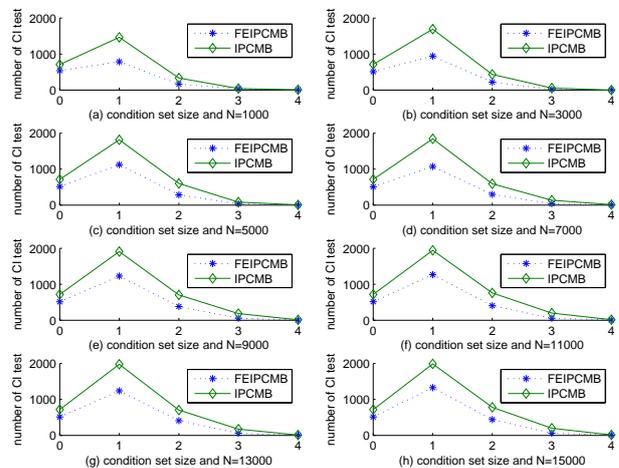


Fig. 2. Results relative to the average number of CI tests returned by FEIPC-MB and IPC-MB on Alarm network, respectively. The horizontal axis is the the order of CI test in each subgraph, and the sample sizes of the eight subgraphs are ranged from 1000 to 15000

Furthermore, since many algorithms just perform well for Alarm, it is necessary to examine the performance on other networks. Hence, we conduct similar experiments on Insurance network. As shown below, we show similar results of IPC-MB and FEIPC-MB.

Fig. 4 (a), Fig. 4 (b) and Fig. 4 (c) show the values of precision, recall and the distances between true BN and output obtained by IPC-MB and FEIPC-MB, respectively. Results of precision and recall returned by FEIPC-MB are higher than that of IPC-MB and distance is lower again. As indicated in Fig. 4 (d), it can be seen that the time complexity of the FEIPC-MB algorithm is lower than that of IPC-MB. Obviously, FEIPC-MB achieves a satisfactory performance in above metrics individually. Besides, they show that the run-time consumed by the FEIPC-MB algorithm increase more slowly with the size of samples compared to that of the IPC-MB algorithm and that this advantage is much more obvious for large database.

Fig. 5 shows the average number of CI tests of each order required by FEIPC-MB and IPC-MB for learning MB of all

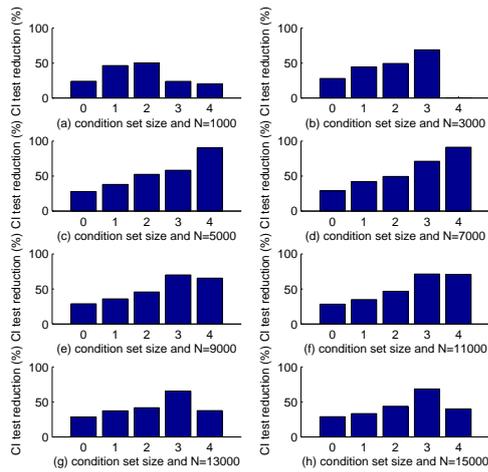


Fig. 3. Results relative to the average percentage of the number of CI tests reduced by FEIPC-MB compared to IPC-MB on Alarm network. The horizontal axis is the the order of CI test in each subgraph, and the sample sizes of the eight subgraphs are ranged from 1000 to 15000

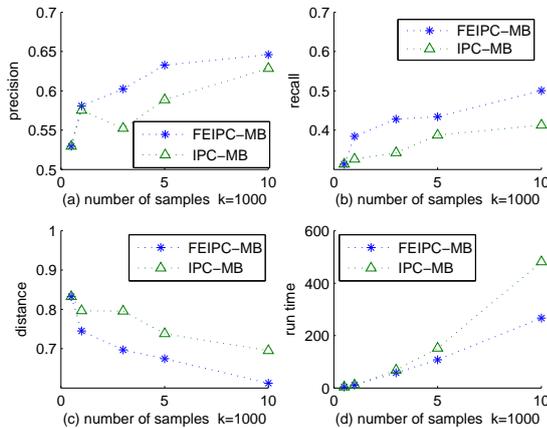


Fig. 4. Results relative to the average precisions, recalls, distances and time consumption returned by FEIPC-MB and IPC-MB for learning Markov blankets of Insurance network. Higher values imply a better performance for both sub-figures (a) and (b), on the contrary, lower values indicate outstanding performance for (c) and (d).

nodes in Insurance network with sample sizes $N = 500$, $N = 1000$, $N = 3000$, $N = 5000$ and $N = 10000$. The FEIPC-MB algorithm performs less tests than the IPC-MB algorithm. Fig. 6 demonstrates the average reduction percentage on CI tests obtained by the FEIPC-MB algorithm compared with the IPC-MB algorithm on the five different sample sizes.

In order to further compare the complexity of our proposed method with original method, we conduct simulations on Hailfinder network. In Fig. 7, the results of the four metrics precision, recall the distances and the run-time are showed. They indicate the good performance of FEIPC-MB again. As are shown in Fig. 8, we can see that FEIPC-MB did not employ the CI test of orders over 4 compared with IPC-MB in the same experimental environment. In addition, FEIPC-MB performs fewer CI tests of orders from 0 to 4 than that of IPC-MB.

All the experimental results reported above demonstrate that FEIPC-MB use relatively fewer or conquer CI tests and

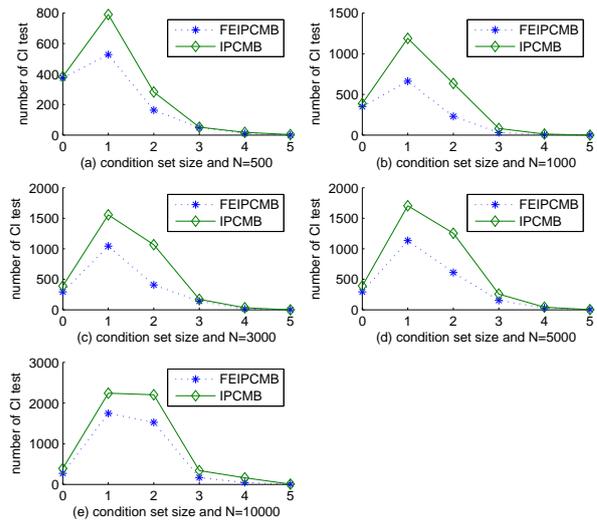


Fig. 5. Results relative to the average number of CI tests returned by FEIPC-MB and IPC-MB on Insurance network, respectively. The horizontal axis is the the order of CI test in each subgraph, and the sample sizes of the five subgraphs are ranged from 500 to 10000

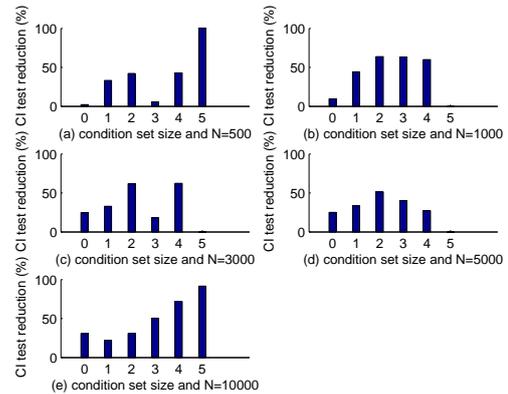


Fig. 6. Results relative to the average percentage of the number of CI tests reduced by FEIPC-MB compared to IPC-MB on Insurance network. The horizontal axis is the the order of CI test in each subgraph, and the sample sizes of the five subgraphs are ranged from 500 to 10000

lower orders to obtain better results than that of IPC-MB. In fact, it is well known that tests of high orders have significant effect on computational time and accuracy of results obtained by CI test. Therefore, our proposed method is superior to IPC-MB in all aspects such as complexity of time, reliability of tests, precision of results and so on.

V. CONCLUSIONS

In this paper, we proposed the algorithm FEIPC-MB which successfully avoid CI tests with large conditioning sets and use as few CI tests as possible. Here using mutual information to sort the nodes in the candidate PC set of T is the key steps for the algorithm. By the sort, we can not only select the possible non-PC nodes X in a specific sort, but also select the conditioning sets for X and T in a fixed sort. These non-random selection can efficiently reduce the orders CI tests and the number of high order of CI tests and remove false positive nodes as early as possible.

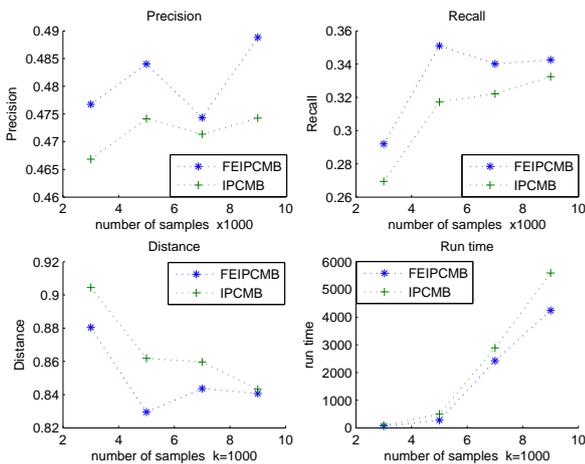


Fig. 7. Results relative to the average precisions, recalls, distances and time consumption returned by FEIPC-MB and IPC-MB for learning Markov blankets of Hailfinder network. Higher values imply a better performance for both sub-figures (a) and (b), on the contrary, lower values indicate outstanding performance for (c) and (d).

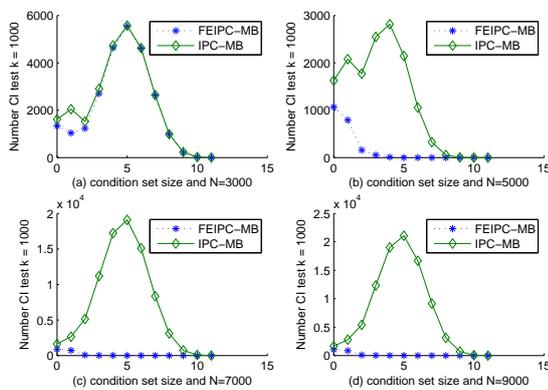


Fig. 8. The average number of CI tests for each order required by both of the algorithms with sample size 3000, 5000, 7000 and 9000 for learning the MB of all nodes of the Hailfinder Network.

In addition, we compared the proposed algorithm with IPC-MB on some standard networks. Simulations results demonstrate that the proposed method outperforms its competitive algorithm with respect to accuracy, run-time and complexity.

We plan to extend our study on several aspects. Firstly, we explore how to determine appropriate conditioning sets for CI tests. Secondly, we intend to study the ordering of nodes adjacent to target variable for every tests.

ACKNOWLEDGMENT

I would like to thank the anonymous reviewers for their help with improving this paper.

REFERENCES

[1] P. Spirtes, C. N., Glymour and R. Scheines, *Causation, prediction, and search*, MIT press, 2000.
 [2] D. Koller, M. Sahami, "Toward optimal feature selection," in *Proc. of International Conference on Machine Learning*, (ICML, 1996), pp. 284–292.
 [3] D. Margaritis and S. Thrun, "Bayesian network induction via local neighborhoods," in *Advances in Neural Information Processing Systems*, 12(1999).

[4] I. Tsamardinos and C. F. Aliferis, "Towards principled feature selection: Relevancy, filters and wrappers," *Proceedings of the ninth international workshop on Artificial Intelligence and Statistics*, Morgan Kaufmann Publishers: Key West, FL, USA, 2003.
 [5] S. Yaramakala and D. Margaritis, "Speculative Markov blanket discovery for optimal feature selection," *Data mining, fifth IEEE international conference on. IEEE, Fifth IEEE International Conference on Data Mining*, (Houston, Texas, 2005), pp. 809–812.
 [6] I. Tsamardinos, C. F. Aliferis and A. Statnikov, "Time and sample efficient discovery of Markov blankets and direct causal relations," *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2003) pp. 673–678.
 [7] C. F. Aliferis, I. Tsamardinos and A. Statnikov, "HITON: a novel Markov Blanket algorithm for optimal variable selection," *AMIA Annual Symposium Proceedings. American Medical Informatics Association*, (2003), pp. 21–25.
 [8] J. M. Peña, R. Nilsson and J. Björkegren, "Towards scalable and data efficient learning of Markov boundaries," *International Journal of Approximate Reasoning* 45(2007) 211–232.
 [9] Shunkai Fu, M.C. Desmarais, "Fast Markov blanket discovery algorithm via local learning within single pass," in *Advances in Artificial Intelligence*, Windsor, Canada, (2008), 96–107.
 [10] J. Pearl, *Causality: models, reasoning and inference*, Cambridge, MIT press, 2000.
 [11] Y. Yang and Y. Wu, "On the Properties of Concept Classes Induced by some Multiple-valued Bayesian Networks," *Information Sciences*, 184 (2012) 155–165.
 [12] I. Beinlich, G. Suermondt, R. Chavez, G. Cooper, "The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks," *Proc. Second European Conf. Artificial Intelligence in Medicine*, Springer, Berlin Heidelberg, (1989), pp. 247–256
 [13] J. Binder, D. Koller, S. Russell, and K. Kanazawa, "Adaptive probabilistic networks with hidden variables," *Machine learning*, 29(2-3) (1997) 213–244.
 [14] J. Abramson, A. Brown, Murphy, and R. L. Winkler. "A Bayesian system for forecasting severe weather," *International Journal of Forecasting*, 12 (1996) 57–71.
 [15] K. Murphy, "The bayes net toolbox for matlab," *Computing science and statistics* 33 (2010) 1024–1034.
 [16] A. Hassan, M. Kerrie, "Bayesian estimation of the time of a decrease in risk-adjusted survival time control charts," *IAENG International Journal of Applied Mathematics* 41(4) (2011) 360–366.

Yanying Li was born in Songyuan, Jilin Province, China in 1981. She received her B.S. degree, M.S. degree in the Department of Mathematics from Jilin Normal University in 2004 and 2007 respectively. And began work for a doctorate at Xidian University in 2011. She is interested in Probabilistic Graphical Models and Time Series.

Yanying is a Ph.D in Xidian University and a lecturer in Baoji University of Arts and Sciences.

Dr. Li has published 15 papers in English and in Chinese.

Youlong Yang was born in Pucheng, Shaanxi, China in 1967. He received his B.S. degree, M.S. degree in the Department of Mathematics from Shaanxi Normal University, in 1990 and 1993, respectively. And received the Ph.D. degree at Northwestern Polytechnical University in 2003. In 2006, he was out bounded post-doctoral mobile stations in Xidian University, and then was sent to the United States University of Rochester as a national public school student in 2007. His scientific interest is Probabilistic Graphical Models and Time Series.

Youlong is full professor, PhD supervisor in Xidian University and executive director of the Mathematical Association of Shaanxi Province.

Prof. Yang has published more than 40 papers. His papers were published as the first-named author in journals such as Information Sciences, International Journal of Approximate Reasoning, Acta Application Mathematic, Control Theory and Applications. He has received one reward of Shaanxi Provincial Science and Technology and two Bureau departmental level research awards and so on.

Xiaofeng Zhu was born in Xinzhou, Shanxi, China in 1988. He received his B.S. degree from Xinzhou Normal College in 2008 and begin work for a M.S. degree in the Department of Mathematics at Xidian University in 2012. His scientific interest is Probabilistic Graphical Models and Time Series.

Xiaoxia Dou was born in Fengxiang, Shaanxi, China in 1979. She received her B.S. degree, M.S. degree in the Department of Mathematics from Northwest University, in 2002 and 2005, respectively. Her scientific interest is number theory.

Xiaoxia is a lecturer in Baoji University of Arts and Sciences.