An Improved IAMB Algorithm for Markov Blanket Discovery

Yishi Zhang[1]
[1]School of Software Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China
Email: easezh@126.com

Zigang Zhang[2], Kaijun Liu[2]
[2]School of Management, Huazhong University of Science and Technology, Wuhan, 430074, China
Email: zigang_zhang@126.com

Gangyi Qian[3]
[3]School of Public Administration, Huazhong University of Science and Technology, Wuhan, 430074, China
Email: airart@126.com

***Abstract*—Finding an efficient way to discover Markov blanket is one of the core issues in data mining. This paper first discusses the problems existed in IAMB algorithm which is a typical algorithm for discovering the Markov blanket of a target variable from the training data, and then proposes an improved algorithm λ-IAMB based on the improving approach which contains two aspects: code optimization and the improving strategy for conditional independence testing. Experimental results show that λ-IAMB algorithm performs better than IAMB by finding Markov blanket of variables in typical Bayesian network and by testing the performance of them as feature selection method on some well-known real world datasets.**

***Index Terms*—data mining, classification, feature selection, Markov blanket, IAMB algorithm**

I. Introduction

As machine learning aims to address larger and more complex tasks, it is the problem of centering on the most relevant information in a potentially overwhelming quantity of data that has become increasingly important. For instance, data mining of corporate or scientific records often involves dealing with both many features and many examples. Such data present familiar "dimensional curves" for classification of the target variable, and undermine the classification accuracy due to the noise of irrelevant variables. One solution is to pre-process the dataset by selecting a minimal subset of the available variables and feed the selected variables into a preferred classifier. It demands the work on implementing a variable selection method. A good variable selection method selects an appropriate set of variables by pruning irrelevant ones from the dataset. Much effort can be seen from a large amount of literature on various variable selection approaches. See [1-3] for more details. Most variable selection methods are heuristic in nature. Several researchers [4-6] have suggested, intuitively, the Markov blanket of the variable to be classified T is a key concept in solving the variable selection problem. Markov blanket is defined as the set of variables conditioned on which all other variables are probabilistically independent of *T*. Thus, knowing the values of the Markov blanket variables renders all other variables superfluous in classifying *T* (provided one uses a powerful enough classifier) [7].

Incremental Association Markov blanket (IAMB) [8] is a basic algorithm to discover the Markov blanket to the target variable. It accepts like input a database of training and the target variable. This algorithm is classified inside of the forward strategic algorithms, because it starts with the empty set and adds the nodes one by one. The correctness of IAMB is under the assumptions that the learning database D is an independent and identically distributed sample from a probability distribution p faithful to a DAG G and that the tests of conditional independence and the measure of conditional dependence are correct, see the next section.

In this paper we propose an improvement algorithm of IAMB: λ-IAMB. We compare the new algorithm with IAMB for inducting the *MB* (*T*). The experimental results show that the λ-IAMB algorithm performs better than IAMB by testing them on the ALARM dataset. In addition, we compare the performance of classification using λ-IAMB and IAMB for variable selection with excellent results. The experiment is under Weka environment.

II. Related work

The Markov blanket of a attribute *T*, denoted by *MB*(*T*), is a minimal set of attributes (or features, variables, we use them interchangeably) conditioned on which all other attributes are probabilistically independent of the target *T* (Definition 1), thus, knowing the values of the *MB*(*T*) is enough to determine the probability distribution of *T* and the values of all other attributes become superfluous. Obviously, we can only use attributes in the *MB*(*T*)

This work was supported by China Postdoctoral Science Foundation (No.20080430968)
Corresponding author: Gangyi Qian.

doi:10.4304/jcp.5.11.1755-1761

instead of all the attributes for optimal classification. In addition, under certain conditions (faithfulness to a Bayesian Network) the *MB*(*T*) is the subset where are the parents, the children and the parents of the children of the target *T* in the Bayesian network [8] (Fig.1).

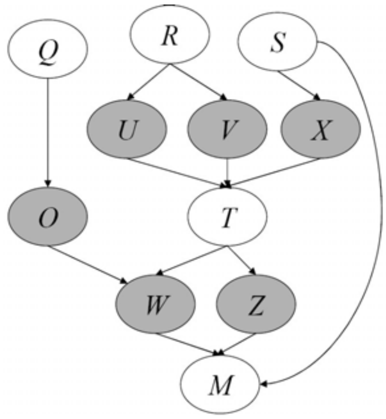

Figure 1.Markov blanket of *T*: a Bayesian network case.

Definition 1 (Markov blanket). *The Markov blanket of a target attribute T∈V, denoted as MB(T), is a minimal subset of attributes for which*

$$(T \perp V - MB(T) - T \mid MB(T)) \text{ and } T \notin MB(T),$$

where V is the set of all attributes in the domain. Symbol '⊥' denotes independency.

Pearl's book *Probabilistic Reasoning in Intelligent Systems* first mentioned the concept of Markov blanket. Koller and Sahami [6] proposed KS algorithm for attribute selection, it is the first algorithm that uses Markov blanket concept to select attributes in the domain. KS accepts two parameters *v* and *k*; *v* indicates the number of attributes that will be included in the Markov blanket subset, *k* is the number of variables which will conditioner to *v*. KS algorithm uses a backward heuristic removing strategy which is claimed unsound and the output of which is an approximate Markov blanket of the target attribute. Margaritis and Thrun [9] proposed GS algorithm and use it to build Bayesian network models. GS is claimed to be the first sound algorithm for Markov blanket discovery and will output the correct Markov blanket of target attribute under certain assumptions. GS contains two phases: growing phase and shrinking phase. By using a static forward heuristic search strategy in growing phase and false positive judgment strategy in shrinking phase it get the unique Markov blanket of target attribute. However, it is indicated that in many cases especially under the small size dataset conditions GS algorithm is not faithful and it couldn't discover the correct Markov blanket subset [8]. Tsamardinos and Aliferis [8] proposed a dynamic forward heuristic search strategic algorithm named IAMB algorithm which is similar to GS algorithm and partially solve the problems existed in GS. Like GS algorithm, IAMB algorithm also uses a two-phase approach for discovering Markov blankets. The only difference between them is that IAMB reorder the set of attributes each iteration a new attribute enters the blanket in the growing phase, which may get more accurate blankets under some conditions.

In addition, Tsamardinos and Aliferis [8] introduced some variants of IAMB such as Inter-IAMB, InterIAMBnPC, etc. Other state-of-the-art algorithms such as IAMBnPC [10], HITON [7] and Fast-IAMB [11] are also effective algorithm for searching the Markov blanket of the target in the domain under certain conditions.

However, those state-of-the-art algorithms including IAMB algorithm and its variants applied to obtain the Markov blanket of target attributes may not get the unique blanket unless some preconditions are given. Most of those algorithms are claimed that they are sound under some assumptions, listed below [8, 11]:

a) the data are generated by processes that can be faithfully represented by BNs;

b) there exist reliable statistical tests of conditional independence and measures of associations for the given variable distribution, sample size, and sampling of the data.

Note that assumption a) implies there exists a faithful Bayesian network structure that can accurately reflect and express the relations among the attributes from given dataset. It also means that for the given dataset, the value distributions of all attributes can exactly reflect a single faithful directed graphical model, which real-world datasets may not well satisfied. Next we will introduce IAMB algorithm by interpreting the process of the algorithm and analyzing the existed potential problems. After that we will introduce a hybrid two-aspect strategy which can partially solve the problems and promote performance for the algorithm.

III. IAMB ALGORITHM

The IAMB algorithm (Fig.2) discovers a unique Markov blanket of the target variable *T* once given the data instances. It takes an incremental strategy by starting an empty set and then gradually adding the Markov blanket elements. It consists of two phases: the growing phase and the shrinking phase. The growing phase looks for the nodes which have more dependence with the target variable *T* known the *MB*(*T*) as many as possible. The heuristic method used in IAMB to identify potential Markov blanket members in the growing phase is related to a heuristic function $f(X; T|MB(T))$, which should return a non-zero value for every variable that is a member of the Markov blanket for the algorithm to be sound, and typically it is a measure of association between *X* and *T* given *MB*(*T*). Generally, this operation is done using an information-theoretic heuristic function CMI (conditional mutual information) [5, 9]. To decide which variable shall be included into the Markov blanket, we use a threshold to consider if the one that owns the maximum conditional mutual information obtained means dependency or independency. The growing step terminates when no more new variables are added into the Markov blanket. It implies that the set is complete and no more variable could contribute the knowledge to the target variable T given the current blanket.

As the computation of conditional information relies on the set of Markov blanket that is formed as far (line 3), the false positives occurs in the growing phase. A variable that owns the maximum conditional mutual information given the known Markov blanket is included in the current step; however, it would not be the one when the Markov blanket evolves over time. Thus, it is necessary to have the shrinking step to remove false positives from the current blanket.

The shrinking step tests the conditional independence. We remove one-by-one the features that do not belong to the MB(T) by testing whether a variable X from MB(T) is independent of T given the remaining MB(T). As mentioned above, we consider $X \perp T \mid \mathrm{MB}(T)$ *iff* CMI(X ; $T \mid \mathrm{MB}(T)$) < threshold. Then the algorithm returns MB(T) as the Markov blanket of the target variable T.

IAMB(dataset D; target T)
1: $MB(T) = \varnothing$
2: V = Set of features in D
▷ **Growing Phase: Add true positives to MB(*T*)**
3: Repeat Until $MB(T)$ does not change
4: Find X_{max} in $V\text{-}MB(T)\text{-}\{T\}$ that maximizes CMI(X_{max} ; $T \mid MB(T)$)
5: If $X_{max} \not\perp T \mid MB(T)$ Then
6: $MB(T) = MB(T) \cup \{ X_{max} \}$
7: End If
▷ **Shrinking Phase: Remove false positives from MB(*T*)**
8: For each $X \in MB(T)$ do
9: If $X \perp T \mid MB(T) - \{X\}$ Then
10: $MB(T) = MB(T) - \{X\}$
11: End if
12: End For
13: Return $MB(T)$

Figure 2. IAMB algorithm, where we consider $X \perp T \mid \mathrm{MB}(T)$ *iff* CMI(X_{max} ; $T \mid \mathrm{MB}(T)$) > threshold.

We know that the correctness of the output of IAMB is based on the assumption that the training sample is enough for performing reliably the statistical tests required by the algorithm. But the ‘enough’ counts for the training samples are not known figures for us, and if the training samples are not enough for the reliability, some of the samples may obstruct the way that algorithm used to find a correct Markov blanket of the target T, that is, once IAMB puts the variable X_{max} that maximizes CMI(X_{max} ; $T \mid MB(T)$) but actually is a false positive in MB(T) in the current iteration, it may change the order of the values of the conditional mutual information in later iterations and enhance the opportunity that other false positives be added in MB(T) by the algorithm. In addition, similar problem also exists in the shrinking phase, so that the false positives may not finally be removed from MB(T). In that case the IAMB will be of low inefficiency and will not output the correct Markov blanket.

IV. Two-aspect improving strategies for IAMB

In this section we propose a two-aspect improving strategy consisting of two improving approaches in order to partially resolve the problem existed in IAMB. The strategy contain two aspects which are: a) using entropy instead of conditional mutual information to measure the conditional independence between two variables to reduce the actual computational complexity. b) Refining the dynamic forward heuristic searching strategy to partially avoid the phenomenon that many false positives are added into the blanket occurs at some time.

A. Using entropy instead of CMI

IAMB algorithm uses mutual information and conditional mutual information to determine whether a variable is independent of T. The mutual information of T and X is defined as:

$$MI\left(T;X\right)=\sum_{x\in X,\,t\in T} p\left(t,x\right)\log\frac{p\left(t,x\right)}{p\left(t\right)p\left(x\right)}\ , \quad (1)$$

It can be shown with the entropies:

$$MI\left(T;X\right)=H\left(T\right)-H\left(T\mid X\right). \quad (2)$$

And the conditional mutual information of T and X given Z is defined as:

$$CMI\left(T;X\mid Z\right)=$$
$$\sum_{z\in Z} p\left(z\right)\left[\sum_{x\in X,\,t\in T} p\left(t,x\mid z\right)\log\frac{p\left(t,x\mid z\right)}{p\left(t\mid z\right)p\left(x\mid z\right)}\right], \quad (3)$$

It can be shown with the entropies:

$$CMI\left(T;X\mid Z\right)=H\left(T\mid Z\right)-H\left(T\mid X,Z\right). \quad (4)$$

Under the entropy-form (shown in equation (4)) of the conditional mutual information, it’s not difficult to find that the attribute X which owns the maximum conditional mutual information value with target T also owns the minimum entropy value with that. Note that in each iteration H ($T \mid \mathrm{MB}(T)$) is calculated repeatedly, so we could just calculate H ($T \mid X$, MB(T)) in each iteration and only calculate H ($T \mid \mathrm{MB}(T)$) once at the beginning of all the iterations for current attribute and then we could find out the attribute $X_{\min}$ which is the same with $X_{\max}$ that IAMB owns. Here we have a theorem to describe this phenomenon.

Theorem 1 1) X maximizes MI (T ; X) if and only if X minimizes H ($T \mid X$).
2) X maximizes CMI (T ; $X \mid Z$) if and only if X minimizes $H(T \mid X , Z)$.

Proof. If X maximizes $MI(T; X)$, then for any Y in the domain V, we have

$$MI(T;X) \geq MI(T;Y)$$

that is

$$H(T) - H(T \mid X) \geq H(T) - H(T \mid Y)$$

then we have

$$H(T \mid Y) \geq H(T \mid X). \quad (5)$$

Thus, X minimizes $H(T \mid X)$, which proves 1).

The proof of 2) is similar to that of 1). We do not list the proof here because of space limitation. We will use these results to reduce the real computational complex in practice in our algorithm.

B. *Refining the dynamic forward heuristic searching strategy*

IAMB algorithm considers whether X and T are independent by calculating the conditional mutual information between them. An existing problem is that variable that maximizes the conditional mutual information may not be the correct one which should be a member of the Markov blanket of target T. It means that the variable is a false positive. Even more unfortunately, once a false positive X_{false} enters the $MB(T)$, it may obstruct the next comparison by increasing the conditional mutual information of target T and some variables which have strong association with X_{false}. Thus the true positive may not be the one that maximizes the conditional mutual information and other false positive may enter the $MB(T)$. This effect will continue or even be magnified in latter iterations. What we can do in the heuristic is to impair the influence brought about by the false positive, especially by X_{false}. We note that, in general, the more true positives exist in the $MB(T)$, the smaller the conditional mutual information between target T and other outside-$MB(T)$ variables are. In particular, if all the true positives are in $MB(T)$, then given $MB(T)$, target T and other outside-$MB(T)$ variables are conditional independent, and the conditional mutual information between them should be zero in theory. Thus we consider that "group-enter" strategy may be an appropriate way to impair the negative impact brought about by those false positives, which means that algorithm allows one or more variables (more true positives, hopefully) to be added in the blanket in each iteration. If we see the false positives as noise of the dataset, then although the false positive may own the maximum CMI value with target variable, it may be more possible (hopefully) that the variable owns the second maximum CMI is a true positive. Based on our experience and the operational convenience, for each iteration we give the first two variables which own the maximum CMI with T the opportunity to enter the $MB(T)$.

V. Λ-IAMB ALGORITHM

In this section we present a new algorithm, called λ-IAMB algorithm, which can partially solve the problem existed in IAMB using the two improving approaches. The pseudo-code of the λ-IAMB algorithm is shown in Fig.3.

λ-IAMB algorithm use a coefficient λ as a threshold which effectively reduce the possibility that the second-max-CMI variable is a false positive. We know that IAMB employs conditional mutual information to test independence, the CMI values of the first attribute and the second attribute should satisfy a certain inequality. Algorithm uses λ to describe this inequality relationship:

$$\mathrm{CMI}(X^{2}_{max}; T \mid \mathrm{MB}(T)) > \lambda * \mathrm{CMI}(X^{1}_{max}; T \mid \mathrm{MB}(T)) \quad (6)$$

Moreover, it may make the algorithm more robust when face to the negative situation made by false positives which have entered *MB(T)*, and the first-max-CMI variable which is a false positive due to the unreliable of the dataset is an example in current iteration. Notice that if line 8 is false in the pseudo-code, it is completely the same situation that compared to the iterative process of IAMB. It means that for $\lambda = 1$, the algorithm degenerates to IAMB algorithm when the second-max-CMI variable is not admitted into *MB(T)*. So we can see λ-IAMB as an extension of IAMB.

λ-IAMB (dataset D; target T)
1: $MB(T) = \varnothing$
2: V = Set of variables in D
3: Calculates $H(T)$;
▷ **Growing Phase: Add true positives to MB(T)**
4: Repeat Until $MB(T)$ does not change
5: Find $X^{1}_{\min}$ in $V\text{-}MB(T)\text{-}\{T\}$ that minimizes $H(T \mid X^{1}_{\min}, MB(T))$
6: Find $X^{2}_{\min}$ in $V\text{-}MB(T)\text{-}\{T\}\text{-}\{X^{1}_{\min}\}$ that minimizes $H(T \mid X^{2}_{\min}, MB(T))$
7: If $X^{1}_{\min} \not\perp T \mid MB(T)$ then
8: If $H(T \mid X^{2}_{\min}, MB(T)) - \lambda * H(T \mid X^{1}_{\min}, MB(T)) < (1-\lambda) H(T \mid MB(T))$ and $X^{2}_{\min} \not\perp T \mid MB(T)$ then
9: $MB(T) = MB(T) \cup \{X^{1}_{\min}\}$
10: $MB(T) = MB(T) \cup \{X^{2}_{\min}\}$
11: End If
12: Else $MB(T) = MB(T) \cup \{X^{1}_{\min}\}$
13: End If
▷ **Shrinking Phase: Remove false positives from MB(T)**
14: For each $X \in MB(T)$ do
15: If $X \perp T \mid MB(T) - \{X\}$ then
16: $MB(T) = MB(T) - \{X\}$
17: End If
18: End For
19: Return $MB(T)$

Figure 3. λ-IAMB algorithm, note that line 8 is the variant of equation (6).

From the pseudo-code we know the worst-case complexity of the algorithm is $O(|V|^2)$. But when considering the calculation of the conditional mutual information, the complexity is different from $O(|V|^2)$. In our experiment we statistic the "actual conditions" instead of exhausting every combination of variables' values and we calculate the conditional entropy instead of conditional mutual information (shown in equation (5)). For each instance (consisting of a set of particular values of each variable, including t, x and $mb(T)$)[1] we statistics distribution for the correlative values of variables and then get $H(t\ |x, mb(T))$, After traversing the dataset within $O(n^2|V|)$ times we get the value of $H(T\ |X, MB(T))$, and if we add a comparison operation at the end of each traversal, we can get the first two variables that minimizes the $H(T|X,\ MB(T))$, so the worst-case complexity in our experiment is $O(n^2|V|^2)$, where n denotes the total counts of the dataset.

VI. Experimental Results

Experiment Set 1: BNs from real diagnostic systems (Table1). We tested the algorithm on the ALARM Network [12] which represents the uncertain relationships among some relevant proposition in the intensive-care unit (ICU) having 37 variables. We randomly sampled 10,000 training instances from the joint probability that each network specifies. The task was to find the Markov blanket of certain target variables. For ALARM we report the performance of λ-IAMB and IAMB with the same threshold for each variable shown in Fig.6.

Experiment Set 2: A well-known real world dataset from Dutch data mining company Sentient Machine Research: TICDATA[2]. The TICDATA contains the complete set of possible board configurations at the end of tic-tac-toe games. It has 86 variables and 5822 instances and the target variable is CARAVAN. We use IAMB, λ-IAMB and a well known feature selection method-InfoGain to select variables for classification and comparing the performance. In addition, we show the classification accuracy when no any variable is pruned from the dataset. We use several general classifiers such as Naïve Bayes [14], Id3 [15], C4.5 [16] and K2 [17]. The results are shown in TABLE II and Fig.4.

Experiment Set 3: Real world dataset form the UCI repository [13]: ADULT dataset [3]. The data set contains demographic information about individuals gathered from the Census Bureau database. The original data set had 15 attributes in total, 9 of which were discrete and 6 continuous. We use the discretization method in Weka to discretize numeric variables. Like what we do with TICDATA, We also use λ-IAMB, IAMB and InfoGain to select variables for classification and comparing the performance with ADULT, meanwhile we show the classification accuracy when no any variable is pruned from the dataset. We use the same classifiers as what we used in experiment set 2. The results are shown in TABLE III and Fig.5.

Both TICDATA and ADULT are well-known real world datasets. Description of them are listed below. All experiment sets in our paper are under Weka 3.5.7 environment.

TABLE I. Description of the real-world datasets: TICDATA and ADULT.

Dataset	Target	Type	Missing Value?	Var. No.	Ins. No.
TICDATA	'CAVARAN'	nominal	no	86	5822
ADULT	'≤50K'	nominal numeric	yes	15	32561

TABLE II. Performance of classification on TICDATA without pruning the variables or using either the IMAB or λ-IAMB as the variable selection method (threshold = 0.018, λ = 0.5).

	All (%)	**InfoGain (%)**	**IAMB (%)**	**λ-IAMB (%)**
Naïve Bayes	81.35	88.92	92.80	93.13
Id3	87.50	86.00	86.74	85.90
C4.5	93.98	94.02	94.02	94.02
K2	90.85	93.63	92.89	93.01

TABLE III. Performance of classification on ADULT without pruning the variables or using either the IAMB or λ-IAMB as the variable selection method (threshold = 0.025, λ = 0.5).

	All (%)	**InfoGain (%)**	**IAMB (%)**	**λ-IAMB (%)**
Naïve Bayes	83.45	81.59	85.14	85.86
Id3	84.62	85.76	86.87	90.56
C4.5	87.86	85.41	85.92	86.45
K2	84.09	81.60	85.36	84.91

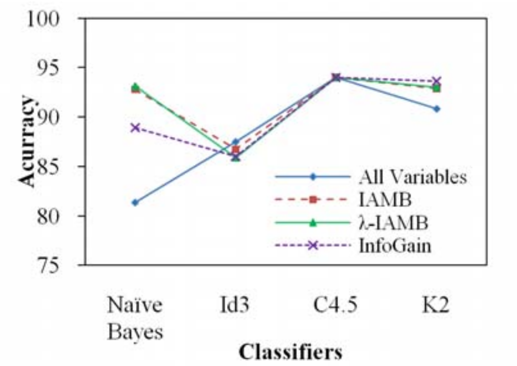

Figure 4. Performance of classification on TICDATA without pruning the variables or using either the IMAB or λ-IAMB as the variable selection method (threshold = 0.018, λ = 0.5).

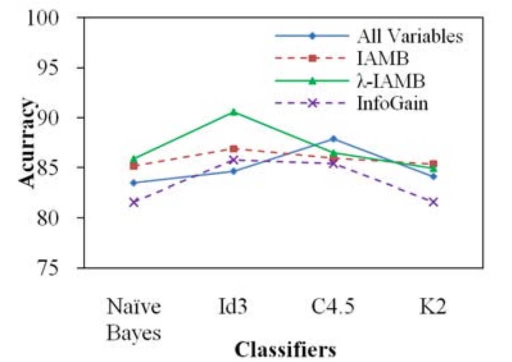

Figure 5. Performance of classification on ADULT without pruning the variables or using either the IMAB or λ-IAMB as the variable selection method (threshold = 0.025, λ = 0.5).

[1] t denotes a value of target T. Similarly to $mb(T)$ and x.
[2] http://hcs.science.uva.nl/benelearn99/datadescription.html
[3] http://archive.ics.uci.edu/ml/datasets/Adult

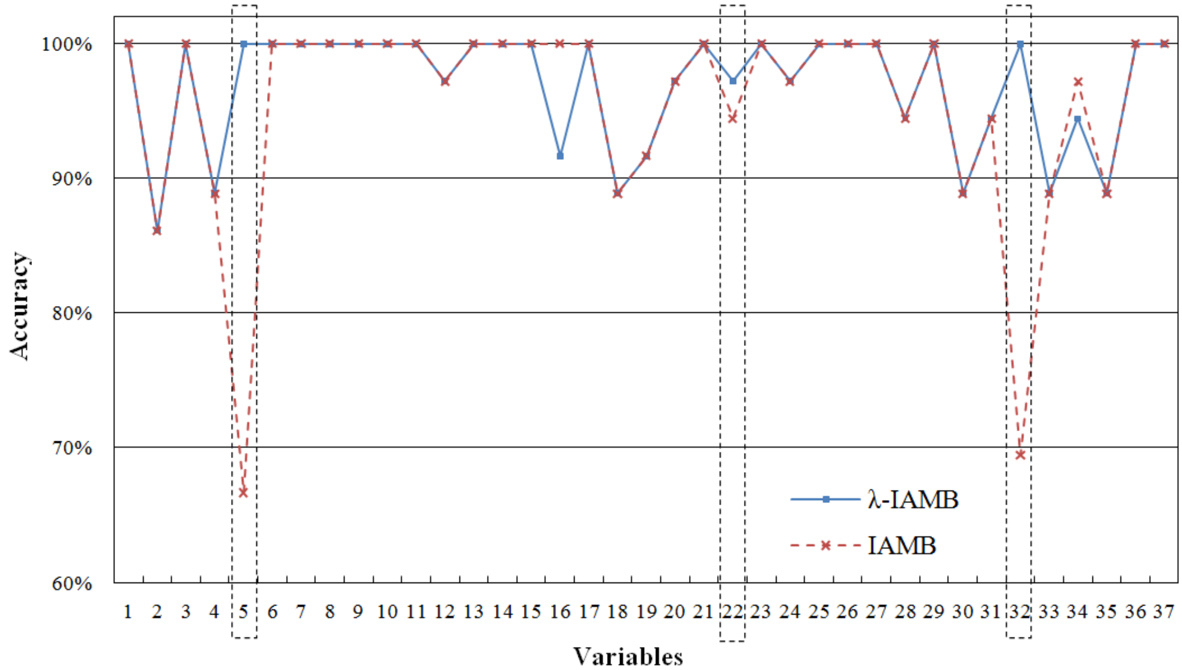

Figure 6. Performance of λ-IAMB and IAMB with the same threshold on ALARM dataset. The notations in the chart show the most significant difference between λ-IAMB and IAMB.

The result of the experiment set 1 (Fig.6) shows that although there exists some 'little-worse' variables, λ-IAMB performs obviously better than IAMB with No.5, 22 and 32 variables (as is shown by the notations in Fig.5). It illustrates that λ-IAMB owns a more efficient Markov blanket discovery strategy than IAMB.

TABLE II and Fig.4 show that λ-IAMB owns the best performance when using the Naïve Bayes as the classifier, and owns nearly the same performance and better one than InfoGain and IAMB, separately, when using the classifier K2. While the results in TABLE III and Fig.5 show that the performance of λ-IAMB is the best when using Id3 and Naïve Bayes as the classifiers.

In average, it's obvious that λ-IAMB algorithm owns the best performance among those feature selection methods including IAMB.

VII. Conclusion and Future Research

Our work, including proposing λ-IAMB algorithm based on a two-aspect strategy and the empirical study on typical Bayesian network and some well-known real-world datasets, shows that λ-IAMB algorithm is a useful and efficient algorithm for Markov blanket discovery to select most relevant features to reduce the dimension of the domain and promote the accuracy for the output of the classifiers.

Our empirical results show that IAMB algorithm may induce amount of false positives into $MB(T)$. However, λ-IAMB algorithm has obviously reduced these negative cases by not only inducing first-max-CMI variable into MB(T) but also conditionally inducing the second-max-CMI variables into $MB(T)$. Meanwhile, λ-IAMB owns the potential faster running time than IAMB for that it allows one or more variables being added in the MB(T) at each iteration.

In λ-IAMB algorithm how to select the coefficient λ is a worth-concerned work. We expect that the algorithm which owns the ability to select λ with self-adapting method will work better for variable selection task. Furthermore, another direction of potential future research is to make the algorithm to deal with missing and abnormal values properly.

Acknowledgment

We are grateful for the helpful comments of the reviewers and editors that led to many improvements in the paper.

References

[1] H. Liu, and L. Yu. "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, 2005, Vol. 17, pp. 491–502.

[2] M. Dash, and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, 1997, Vol. 1, pp. 131–156.

[3] P. Langley, "Selection of relevant features in machine learning," Proceedings of the AAAI Fall symposium on relevance. AAAI Press, 1994, pp. 140–144.

[4] G.F. Cooper, et al. "An evaluation of machine learning methods for predicting pneumonia mortality," *Artif Intel Med*, 1997, No.9, pp. 107-138.

[5] J. Cheng, and R. Greiner, "Comparing bayesian network classifiers," Proceedings of Uncertainty in AI, 1999.

[6] D. Koller, and M. Sahami, "Toward optimal feature selection," International Conference on Machine Learning, Bari, 1996, pp. 284-292.

[7] C. Aliferis, I. Tsamardinos and A. Statnikov. "HITON, a novel Markov blanket algorithm for optimal variable Selection," Proceedings of the 2003 American Medical

Informatics Association Annual Symposium, 2003, pp. 21-25.

[8] I. Tsamardinos, C.F Aliferis, and A. Statnikovs. "Algorithms for large scale Markov blanket discovery," Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference, 2003, pp. 376-381.

[9] D. Margaritis, and S. Thrun, "Bayesian network induction via local neighborhoods," Carnegie Mellon University. *Technical Report CMU-CS-99-134*, August 1999.

[10] C. Aliferis, I. Tsamardinos. "Algorithms for large-scale local causal discovery and feature selection in the presence of small sample or large causal neighborhoods," *Technical Report DSL 02-08*, Department of Biomedical Informatics, Vanderbilt University, Nashville, USA, 2002

[11] S. Yaramakala, and D. Margaritis, "Speculative Markov blanket discovery for optimal feature selection," Proceedings of the Fifth IEEE International Conference on Data Mining, Washington, DC, USA, 2005, pp. 809–812.

[12] I.A., Beinlich, et al. "The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks," Proceedings of the Second European Conference on Artificial Intelligence in Medicine, London, England. 1989.

[13] UCI Machine Learning Repository, [http://archive.ics.uci.edu].

[14] P. Domingos and P. Michael. "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, 1997, Vol.29, pp.103-137

[15] J. Quinlan. "Simplifying Decision Trees," *International Journal of Man-Machine Studies*, 1987, Vol.27, pp. 221-234

[16] J. Quinlan. *C4.5: Programs for machine learning*, San Francisco: Morgan Kaufmann Publishers, 1993

[17] G. Cooper and E. Herskovits. "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, 1992, Vol.9, pp. 309-347

Yishi Zhang was born in Wuhan, Hubei Province, China, on July 3, 1986. M.S. candidate. His main research interests include data mining, pattern recognition and knowledge management.

Zigang Zhang was born in Changsha, Hunai Province, China, on October 12, 1947. He received his B.S degree from TsingHua University. Now he is a professor at School of Management, Huazhong University of Science and Technology. He has visited and worked at Stanford University and Toronto University as a senior scholar. His research interests include decision support system, knowledge management and supply chain management.

Kaijun Liu is a post doctor of Huazhong University of Science and Technology. He holds a Ph.D. degree in Management Science and Engineering. His research interests include supply chain management, knowledge management and E-commerce.

Gangyi Qian was graduated from Huazhong University of Science and Technology in 1996 (Mechatronics), completed his doctorate in 2006 (Management Sience and Engineering), and became a postdoctor in School of Public Administration, Huazhong University of Science and Technology. His research interest is data mining for emergency management.