The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Bayesian Network Structure Learning: The Two-Step Clustering-Based Algorithm

Yikun Zhang
School of Mathematics
Sun Yat-sen University
Guangzhou, China 510275
email: yikunzhang@foxmail.com

Jiming Liu,**Yang Liu**
Department of Computer Science
Hong Kong Baptist University
Kowloon Tong, Hong Kong
email: jiming@comp.hkbu.edu.hk, csygliu@comp.hkbu.edu.hk

Abstract

In this paper we introduce a two-step clustering-based strategy, which can automatically generate prior information from data in order to further improve the accuracy and time efficiency of state-of-the-art algorithms for Bayesian network structure learning. Our clustering-based strategy is composed of two steps. In the first step, we divide the potential nodes into several groups via clustering analysis and apply Bayesian network structure learning to obtain some pre-existing arcs within each cluster. In the second step, with all the within-cluster arcs being well preserved, we learn the between-cluster structure of the given network. Experimental results on benchmark datasets show that a wide range of structure learning algorithms benefit from the proposed clustering-based strategy in terms of both accuracy and efficiency.

Introduction

Bayesian network models, first introduced into artificial intelligence by (Pearl 1982), have been applied to miscellaneous fields of science. To tackle the Bayesian network structure learning problem, some constraint-based and score-based algorithms have been proposed in the last two decades ((Margaritis 2003)). (Chickering 1996) proved that this problem is NP-hard and might produce some undesired network structures that can hardly describe the original datasets. Some prior information helps to ameliorate computational costs and accuracy of existing algorithms. (Friedman, Nachman, and Peér 1999) applied clustering to figure out the candidate parents of a variable in a Bayesian network. This idea inspires us to resort to clustering analysis in order to obtain some prior information about the existence of arcs based on correlations between variables.

Method

Our *two-step clustering-based* (TSCB) strategy, which generates some pre-existing arcs in the first step, can be applied to any structure learning algorithms. We first apply the clustering analysis to group strongly dependent variables and learn the arcs among them, which work as the prior information for the second step of structure learning. To learn the arcs in each cluster and combine clusters, we apply the same structure learning algorithm. See Algorithm 1 for details.

Algorithm 1 *Two-step Clustering-based* Bayesian Network Structure Learning Algorithm

Input:

- Data set: $\mathcal{D} = \{X_1, X_2, ..., X_N\}$ with N variables
- The number of clusters: K

Step 1:

1: Compute the dissimilarity matrix.
2: Carry out clustering analysis via *average linkage agglomerative clustering method* with the pre-assigned method and cut the dendrogram into K groups (clusters).
3: Learn Bayesian network structures within each cluster using Algorithm A.[1]

Step 2:

1: Apply Algorithm A again on all variables with the retained arcs to combine clusters.

Output: Bayesian network structure learned from the dataset $\mathcal{D}$.

We utilize the correlations between variables as the distance metric for clustering analysis, as they can well-represent dependencies among variables. Here the well-known Pearson correlation is used. To compute the correlations between discrete variables in real-world data, we introduce a technique to transform them.

1. **Converting**: Label attributes of discrete (or categorical) variables by nonnegative integers
2. **Centralization**: Shift the variables such that their attributes are central at 0

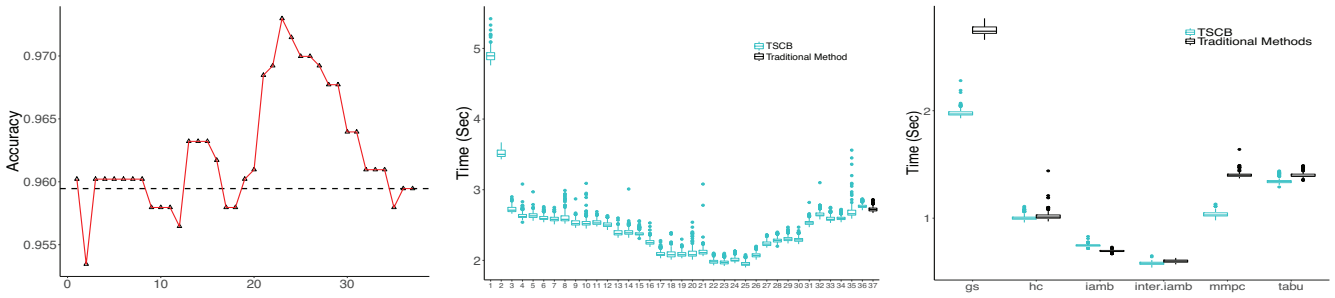
To determine the accuracy of a learned network structure, we use the following accuracy metric.

$$Accuracy = \frac{\sum True\,positive + \sum True\,negative}{\sum Total\,population} \quad (1)$$

Results

To justify the effectiveness of our method, experiments on accuracy and time efficiency have been conducted. First,

[1]This could be any traditional structure learning algorithm, like the grow-shrink algorithm. See (Margaritis 2003) for details.



(a) Accuracy Variations (b) Time Distributions (c) Time Comparisons

Figure 1: Experimental results on the dataset "alarm" ((Beinlich et al. 1989)). Fig 1a represents the variation of accuracies with regard to the number of clusters when we embed the Margaritis's grow-shrink (constraint-based) algorithm. The horizontal dash line indicates the accuracy of the grow-shrink algorithm. Fig 1b displays time distributions of 200 repeated experiments with regard to the number of clusters when we embed the grow-shrink algorithm. The rightmost boxplot represents the performance of the traditional algorithm. Fig 1c presents the time comparison between the TSCB algorithm and six traditional algorithms. For each pair of boxplots, the left is for our TSCB method while the right is for the embedded traditional algorithm.

Methods	**Classical**	**TSCB (Mean)**	**TSCB (Optimal)**
GS	0.9594595	0.9646772	**0.9729730**
IAMB	0.9677177	0.9733483	**0.9767267**
Inter-IAMB	0.9804805	0.9791667	**0.9812312**
MMPC	**0.9624625**	0.9611111	**0.9624625**
HC	0.9587087	0.9603228	**0.9699700**
TABU	0.9662162	0.9632883	**0.9729730**

Table 1: Accuracies of some classical methods with and without the TSCB strategy on the dataset "alarm" ((Beinlich et al. 1989)). Constraint-based: GS, IAMB, Inter-IAMB, MMPC; Score-based: HC, TABU.

we inspect the variation of accuracies on synthetic datasets with regard to K. Moreover, we utilize six classical structure learning algorithms as the baseline to evaluate the adaptability of our method. We also analyze the variation of total running times of our algorithm with regard to the parameter. In addition, total elapsed times of our algorithm with the choices of the parameter corresponding to the best accuracies on synthetic datasets are also tested.

The experiments based on the *average linkage agglomerative clustering* on synthetic datasets show that our TSCB strategy can improve the performance of the embedded structure learning algorithms in terms of the accuracy and time efficiency with a wide range of K, which demonstrates the robustness of the proposed method. Moreover, the improvement of constraint-based algorithms is more significant. The reason is that we utilize the correlations between variables as the distance metric to conduct clustering analysis, which coincides with the principle of the conditional independence test used in constraint-based algorithms. Interestingly, we observe that nearly all the clusters contain no more than three variables, which implies that the group of two or three nodes might be the primitive unit of the network structure. The phenomenon is consistent with the concept of "network motifs", proposed by (Milo et al. 2002).

Conclusion

In this paper we have proposed a *two-step clustering-based* strategy for Bayesian network structure learning. By dividing the original set into clusters and learning the network structure within and between clusters, the performance of a wide range of Bayesian network structure learning methods have been further improved. In our future work, we are particularly interested in investigating the physical meaning of each detected cluster, which will give us more insight into the performance improvement.

References

Beinlich, I.; Suermondt, H.; Chavez, R.; and Cooper, G. 1989. The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks. In *Proceedings of the 2nd European Conference on AI in Medicine*, 247–256. Springer-Verlag.

Chickering, D. M. 1996. *Learning Bayesian Networks is NP-Complete*. New York: Springer New York. 121–130.

Friedman, N.; Nachman, I.; and Peér, D. 1999. Learning Bayesian Network Structure from Massive Datasets: The "Sparse Candidate" Algorithm. In *Proceedings of the 15th Conference on UAI*, 206–215. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Margaritis, D. 2003. *Learning Bayesian Network Model Structure from Data*. Ph.D. Dissertation, Pittsburgh, USA.

Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; and Alon, U. 2002. Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298(5594):824–827.

Pearl, J. 1982. Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach. In *Proceedings of the 2nd AAAI Conference on Artificial Intelligence*, 133–136. AAAI Press.