A NOVEL BAYESIAN NETWORK APPROACH FOR ENGINEERING RELIABILITY ASSESSMENT

Sou-Sen Leu* and Quang-Nha Bui

Department of Construction Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan
** Corresponding author (leuss@mail.ntust.edu.tw)*

ABSTRACT: Bayesian network (BN) learning is difficult task due to the limitation of available data in most engineering fields. On the other hand, BN which constructed bases on experts' judgments is hard to evaluate and often inaccurate. In this work we introduce a novel method called expert-based structural EM (Ex-SEM) for Bayesian network construction which combines limited objective data and experts' knowledge into the BN learning process. Our experimental results show that incorporating prior knowledge from experts into Bayesian learning process could help effectively discover causal relationships in the network as well as improve accuracy of learned model. Finally, Ex-SEM was adopted to assess the reliability of a Hydraulic power plant (HPP) in Taiwan. The maintenance data were collected for more than one year; nevertheless, some data were missing. To incorporate the experts' knowledge from the HPP maintenance sections, four experts were interviewed. The performance of Ex-SEM was compared with original SEM and K2, one popular BN learning algorithm. It is found that Ex-SEM outperformed SEM and K2 in terms of accuracy and efficiency.

Keywords*: Bayesian Network, Structural Learning, Expert Knowledge, Reliability Assessment*

1. INTRODUCTION

Bayesian network is established approach for handling uncertainties and have been applied in many domains, including computer vision, decision support system, construction engineering, system reliability, etc. Application of BN framework involves the determination of its structure and parameters. Once the network has completed, probability of any nodes in the established BN could be inferred for reliability assessment purpose.

One approach to construct BN is base on expert knowledge. In the studies of [1-2], experts were invited to construct BNs based on their belief about physical structure of the systems. However, constructing such BNs for application totally by domain-experts is a time-consuming task and experts normally face difficulties when working with large and complex systems.

Another approach is learning BN from observation data. Many researchers have focus on this direction [3-5]. When data are incomplete, SEM algorithm is often employed [5]. It's assumed that data are missing at random (MAR), which means missing values could be estimated by the observed ones using statistical approach. However, in the case when data are both incomplete and sparse, which frequently occur in many engineering fields, SEM is often unable to recover the "true" model and consequently bring poor prediction results.

To address the aforementioned problem, this paper proposes a novel method called Expert-based Structural EM (Ex-SEM), which combines domain knowledge in the form of prior structures with limited incomplete data to help the regular BN model learning. The major advantages of our method could be stated as follows. Firstly, the model can utilize the best information from expert knowledge and data to construct the BN efficiently. Secondly, the proposed model will generate a better and more accurate BN. At last, the ability of the proposed model can deal with missing data effectively.

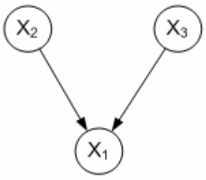
2. BACKGROUND

2.1 Bayesian network

Bayesian Network is widely used technique for representing uncertainty knowledge in Artificial Intelligence [6]. A Bayesian Network $B = < M, \Theta >$ can be represented as a directed acyclic graph (DAG) $M=<N,Z>$ where nodes N represents system components and edges Z represents casual relationships between nodes. In addition to the graph, BN has a set of conditional probability tables (CPTs) $\Theta=\{\theta_{ijk}\}$ specifying the probability of each possible state of the node Xi given each combination θ_{ijk} of states of its parents Pa_i (see Fig. 1). Given the parents Pa_i of one node Xi, this node is conditionally independent of its non-descendant in BN. Then if M and Θ are already known, a BN defines uniquely a joined probability distribution over all nodes in the graph which can be written as:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_i) \qquad \text{Eq. 1}$$

Once the joined probability distribution in Eq.1 is known, many inference algorithms could be used for reasoning efficiently probabilities of nodes that have not been observed conditionally to the values of observed nodes. A review of these inference algorithms can be found in [7]. BN therefore provides an adequate procedure for reasoning under uncertainty with incomplete data, which is often the case in the real-world application.



$Pa_1=\{X_2,X_3\}$	θ_1	1	2
$X_2,X_3=1,1$	θ_{11}	θ_{111}	θ_{112}
$X_2,X_3=1,2$	θ_{12}	θ_{121}	θ_{122}
$X_2,X_3=2,1$	θ_{13}	θ_{131}	θ_{132}
$X_2,X_3=2,2$	θ_{14}	θ_{141}	θ_{142}

Fig.1. Example of Bayesian Network and CPT for X_1

2.2 SEM algorithm

Structural EM [5] is one of the most popular BN learning algorithms with the presence of missing values, which is extended from EM [8] for BN parameter learning. Different from EM, SEM is employed to find both BN structure M and parameters $E[LL(\Theta^t \mid \Theta^{t-1})]$. The parameter maximization $\{\Theta\}$ in EM is replaced by maximization in the joint space $\{M, \Theta\}$ which alternatively searches for the best structure and the best parameters corresponding to this structure. For handling the missing value, the exact score of BN is approximated by the expected value of statistics. The general framework of SEM algorithm can be described as follow:

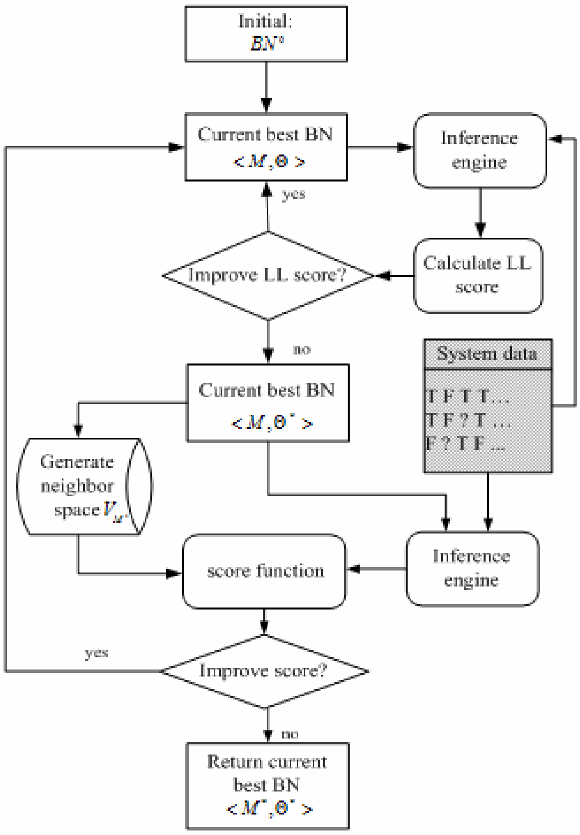


Fig.2. Generic SEM learning flowchart

As a local search method, there is also no guarantee that SEM will convert to the global optimal, but a local optimal [5]. The selection of the initial point is very important with the intent of moving the search quickly to the basin of attraction to find the true BN.

3. EX-SEM ALGORITHM

3.1 Ex-BIC score

The BIC score is often employed with the SEM algorithm for solving incomplete data cases:

$$BIC(B \mid D) = \sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} N^*_{ijk}\log(\theta_{ijk}) - \frac{1}{2}\sum_{i=1}^{n}(r_i-1)q_i\log(N) \qquad \text{Eq. 2}$$

The value of N^*_{ijk} is obtained by inference in the best current network $B=<M^*,\Theta^*>$ if $\{X_i, Pa_i\}$ are not completely measured, or else only by mere counting.

In BIC score formulation, the prior probability for each candidate network is chosen to be uninformative by assigning a uniform prior, where every candidate BN is equally likely. However, when the number of data is insufficient, there is possibility that many DAGs may fit the data well and the BIC score cannot be used to distinguish between them, thus SEM is likely to get stuck in a local optimal.

To overcome this situation, we introduce a new score Ex-BIC for BN learning. Our new score is motivated by the belief that when the data is limited and hard to fulfill, we can integrate available information from domain experts into SEM learning process to avoid local maxima. We try to maximize the probability of candidate network given data and expert BN structure:

$$\arg\max_{B}[\log(P(D \mid B)) + \log(P(B \mid BN^{Ex}))] \qquad \text{Eq. 3}$$

In Eq.3, the log-likelihood $\log(P(D \mid B))$ of data D given candidate BN B is calculated as in BIC score. Then the prior probability of a candidate BN given an expert BN $\log(P(B \mid BN^{Ex}))$ is computed by estimating how similar its structure is compares to BN^{Ex}, which depends on percent of structural difference in DAG:

$$P(B \mid BN^{Ex_l}) = \frac{1}{[n(n-1)]^{diff(BN^{Ex_l})}} \qquad \text{Eq. 4}$$

Where $diff(BN^{Ex_l}) = a_l + d_l + r_l$ is the number of different edges (add, delete, or reverse edges) between candidate BN and expert BN l and $n(n-1)$ is possible directed edge in an n nodes BN.

Here we make the assumption that each expert is expected to provide his BN independently, and does not depend on the opinions of others. Then the probability of BN given experts BNs is provided as follows:

$$P(B \mid BN^{Ex}) = \prod_{l=1}^{L} P(B \mid BN^{Ex_l}) \qquad \text{Eq. 5}$$

Each expert l will be assigned with an important weight w_l, which implies the degree of expert confidence. The influence of experts' BNs to learning process can be adjusted by including a weight α .Finally we end up with Ex-BIC score:

$$ExBIC(B \mid D, BN^{Ex}) = BIC - \alpha\log[n(n-1)]\sum_{l=1}^{L}(w_l diff(BN^{Ex_l})) \qquad \text{Eq. 6}$$

$\alpha \in [0-1]$ is referred as the prior weight of expert information. $\alpha = 0$ corresponds to uniform prior while $\alpha = 1$ corresponds to full log prior probability given by experts.

3.2 Ex-SEM algorithm

The Ex-SEM algorithm is the modification of original SEM by adapting expert knowledge into learning process. This model of Ex-SEM integrates experts' BN structures into learning process. Similar to the original SEM, the proposed Ex-SEM algorithm also employs a hill-climbing search procedure in which each iteration only searches in the neighbor space. In essence, Ex-SEM algorithm includes

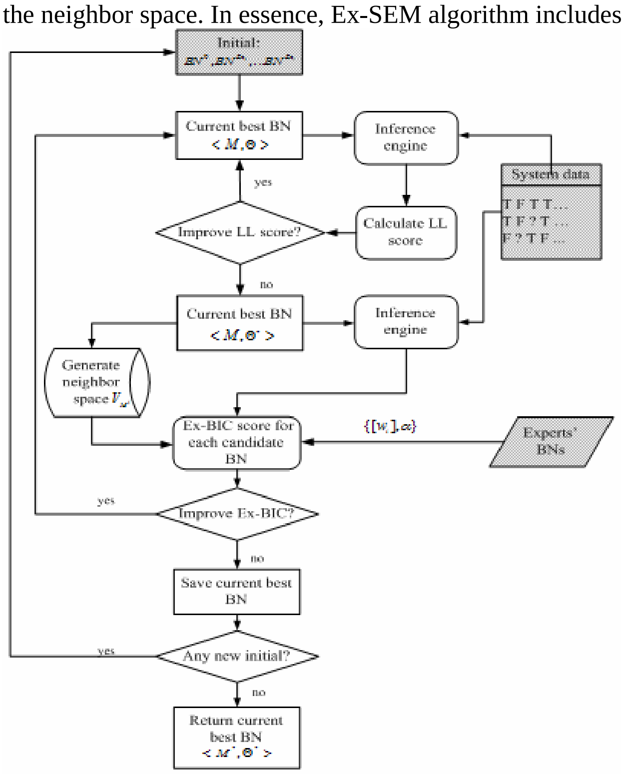


Fig.3. Ex-SEM algorithm for BN learning

two loops: inner EM loop for optimizing BN parameters and outer EM loop for optimizing BN structure. The procedure of Ex-SEM algorithm is provided in Fig. 3.

One disadvantage of hill-climbing search is that it easily gets stuck into local optimal. One way to avoid trapping into local optimal is by providing SEM with a good initial. The initial BN^0 is suggested to start from existent BN model in literature or conversed from other equivalent model. Here we provide a more convenient procedure to avoid local maxima by restarting initial BN to other experts' BNs as start points whenever highest score reached. By taking advantage of experts' BNs, we assume that each expert would provide a "good" enough BN that we take as initial start points. This does not only improve the search result, but help reducing the required time for search process.

4. EXPERIMENTS

4.1 Experimental setup

The experiments were implemented using the Bayesian network toolbox (BNT) which developed by Murphy [7]. To simulate the missing data, we sampled from a ground truth BN using logic sampling method [9].Then we randomly delete a certain percent of values in the data set, which corresponds to missing part. For the simplicity, we only deal with binary state variables, that is, $X_i = \{0,1\}$ for every $i = 1:n$. To simulate BN structure provided by experts, we follow the method in [10] to construct the current BN as well as the experts' BNs by adding noise to the ground truth BN. By this, we can control the amount of noises in those simulated BNs with parameters matrix η, in which η_0 is noise of current BN and η_i ($i = 1:L$) is added noise of expert BN i.

To compare the performance of Ex-SEM with SEM and K2, two famous performance criteria were used: The first criterion is structural difference, which reflects degree to which the learned structure has captured causal relationships. The second criterion is KL divergence which reflects how close a model's JPD is to the ground truth BN's JPD. Further reading about these two criteria is referred to [10].

4.2 Experimental result

We compared our learning algorithm results with two other learning methods: original SEM and K2 algorithm. To evaluate the sensitivity of each learning algorithm given different parameters setup, we range the number of data from 200 to 1000 and BN's nodes from 9 to 15. Optimal parameters about number of experts and prior weight for Ex-SEM are adapted from prior tests $L = 4$, $\alpha = 1$. We repeat each case 5 times and calculate the mean and standard deviation of the results.

Fig.9 shows results comparison. For structural difference, one can observe from charts (a), (b), and (c) that Ex-SEM totally dominate SEM and K2 results, in both terms of mean and standard deviation. It shows stable results given different number of nodes and available data. For KL divergence criterion as we can see in chart (d), (e), and (f), Ex-SEM once again performs better than the other two rivals K2 and SEM. For examples in chart (d), when number of data increases from 200 to 1000, KL mean value for Ex-SEM decreases from 0.125 to 0.014 whereas for K2 and SEM, KL mean value decreases from 0.21 to

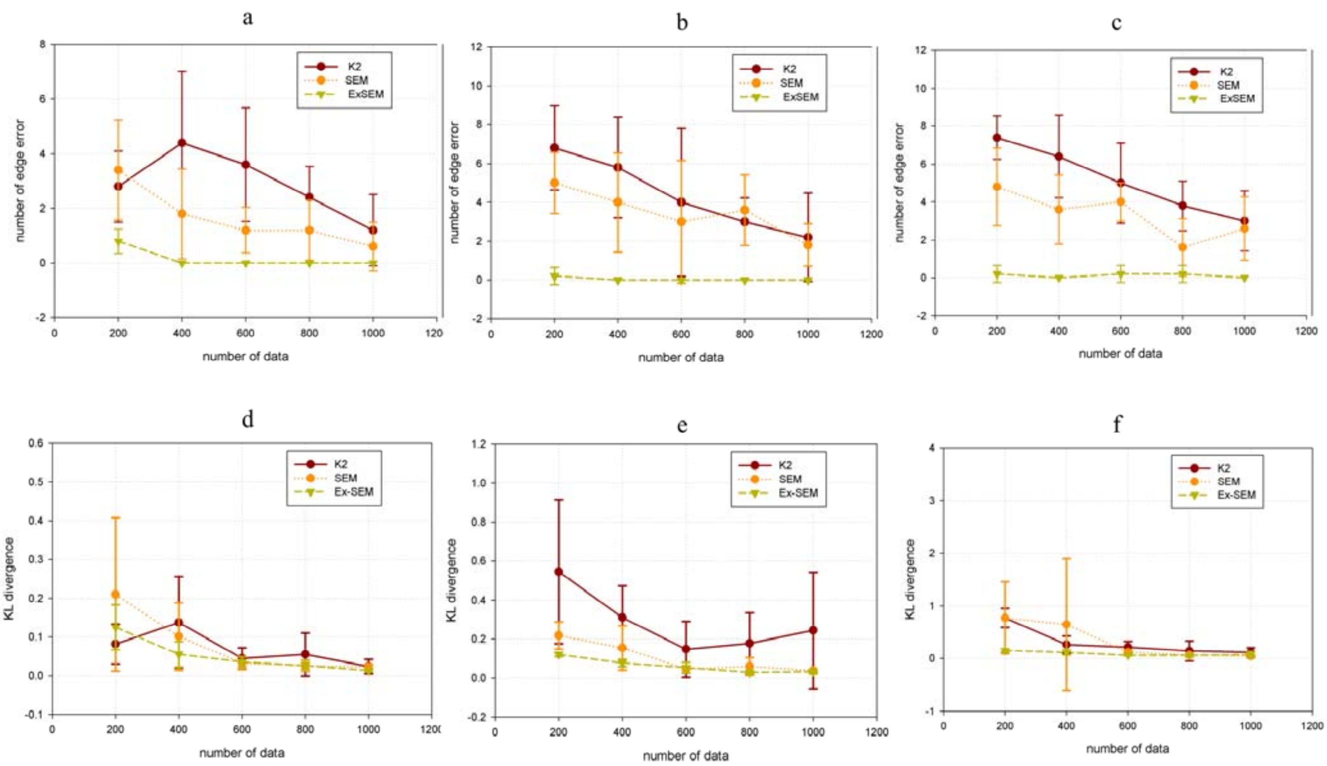


Fig. 4. Learning results vs. number of training samples: (a), (b), (c) structural difference for 9, 12, 15 nodes BN, respectively. (d), (e), (f) KL divergence for 9, 12, 15 nodes BN, respectively.

0.02 and 0.137 to 0.023, respectively. The results from above comparison reinforce our belief that integrating experts' knowledge does increase Ex-SEM learning accuracy.

5. CASE STUDY

In this section, Ex-SEM was applied to assess reliability of a Hydraulic power plant in Taoyuan County, Taiwan ROC. To evaluate learning performance, we firstly use Ex-SEM to train the BN model, and then use this model to predict system failure probability. Percentage of successful predictions for each algorithm is measured for comparison. Data for learning is maintenance data which were collected for about one year, in which 5 percent of data are missing. 32 possible factors were taken into investigation and 4 experts were asked to provide BN structures based on their belief about system performance. We use 10-folds cross-validation in order to obtain a balance between bias and variance of the estimation. Because of the computational cost of the tests, we only compare Ex-SEM with K2 algorithm.

For BN model training, Ex-SEM algorithm recovered most of underlying causal relations which did not appear in the initial BN. It has 11 modifications compared to the initial BN, in which there are 9 added edges and 2 deleted edges. As compare to K2, it only recovered few causal relations between factors in the network and lost most of the underlying causal relations between factors which could be causes to system failure.

From Fig. 5, it concluded that Ex-SEM totally outperform K2 in MSE indicator. Since K2 can only discover few causal relationships between factors, it is hardly possible to expect a better result for K2.We also want to make note that when the estimated failure probability is small, as in the case of predicting reliability, the estimation results may be less accurate due to the constraint of limited data which hardly to represent the distribution over variables. According to the experiment, we can explain why K2 as well as most of other learning algorithms which based merely on observed data often brings poor results when data are limited.

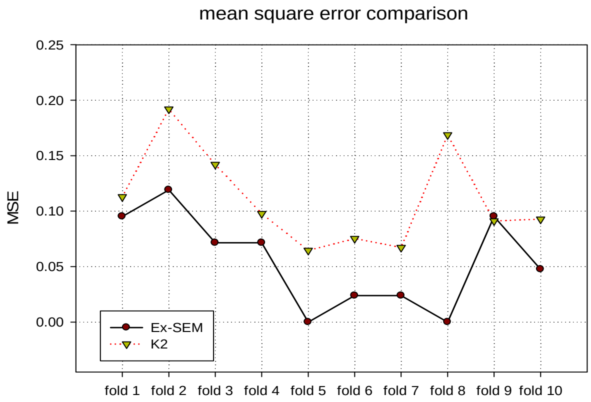


Fig. 5. Comparison of Ex-SEM and K2

6. CONCLUSION

Bayesian learning often requires a lot of information in order to achieve a reasonable BN, especially in the case when a large amount of data is missing, thus combination of several information resources is necessary. Nevertheless, there are few researches in literature focus on this direction. In this study, we have proposed a new approach for BN construction by combining domain knowledge from experts with limited incomplete data. On the aspect of theory, our algorithm is built on existing theory of learning, and we further improve such idea to devise a proper BN learning algorithm. On the aspect of practical implementation, we show rigorous results in our experiments with synthetic data as well as application in a case study. While this study is exploratory in nature, further research needs to continue in this area. This study is the first attempt to apply a new strategy for Bayesian network construction. we believe that this study would open a new direction for further research in the future.

REFERENCES

[1] Luu, V. T., Kim, S.-Y., Tuan, N. V. and Ogunlana, S. O. Quantifying schedule risk in construction projects using Bayesian belief networks. *International Journal of Project Management*, 27, 1 2009), 39-50.

[2] Hu, X.-X., Wang, H. and Wang, S. *Using expert's knowledge to build Bayesian networks*. Institute of Electrical and Electronics Engineers Inc., City, 2007.

[3] Doguc, O. and Ramirez-Marquez, J. E. A generic method for estimating system reliability using Bayesian networks. *Reliability Engineering & System Safety*, 94, 2 2009), 542-550.

[4] Schulte, O., Frigo, G., Greiner, R., Wei, L. and Khosravi, H. *A new hybrid method for Bayesian network learning With dependency constraints*. City, 2009.

[5] Friedman, N. *The Bayesian structural EM algorithm*. City, 1998.

[6] Pearl, J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., 1988.

[7] Murphy, K. The Bayes Net Toolbox for MATLAB. *Computing Science and Statistics*, 332001).

[8] Dempster, A. P., Laird, N. M. and Rubin, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1 1977), 1-38.

[9] Henrion, M. *Propagating uncertainty in bayesian networks by probabilistic logic sampling*. City, 1986.

[10] Heckerman, D., Geiger, D. and Chickering, D. M. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20, 3 1995), 197-243.