A Novel Method for Reconstructing Dynamic Bayesian Network

Ying Yu, Hao Huang, Jianye Lu, ZhenYang and Wenwen Yang

No.333 Nanchen Road, Baoshan District, Shanghai, China
Department of Mechanical and Electrical Engineering and Automation, Shanghai University
Email: yangwenwen222@163.com

Abstract. Upon data renewal, a Bayesian network might need reconstruction. In this paper, a novel method, which is based on a proposition, is proposed for such reconstruct work of a kind of symptom--disease Bayesian network scored by a K2 scoring function. The proof of the proposition for the method is given firstly, and later the performance of the method is illustrated with numerical examples.

1. Introduction

Upon data renewal, the reconstruction of a Bayesian network might be necessary, so research on dynamic Bayesian Network has attained much attention recently. Dean et al. extended the hidden Markov model and formally proposed the concept of dynamic Bayesian network in 1989[1]. Since then, dynamic Bayesian networks have developed rapidly and been applied in various fields. Friedman[2] et al. proposed a dynamic Bayesian network structure learning algorithm DPN-SEM with incomplete data in 1998. Garg[3] et al. proposed a dynamic Bayesian network learning algorithm based on AdaBoost framework and the algorithm is applied into a human-computer interaction interface. Wang[4] proposed a dynamic Bayesian network for dam monitoring analysis. Jiang[5] et al. made improvements on the dynamic Bayesian network. Su[6] et al. applied the dynamic Bayesian network model into the establishment of power distribution system and its stability analysis. Gao[7] et al. proposed a fast inference algorithm for discrete dynamic Bayesian networks with variable structures.

Dynamic Bayesian networks are also applied in medical field, Carbery[8] et al. proposed a deep dynamic Bayesian network model for medical image analysis. Jeong-Yon Shim[9] proposed a three-dimensional knowledge learning memory with dynamic Bayesian correlation matrix for medical diagnosis. Orphanou[10] et al. proposed a dynamic Bayesian network model for predicting the risk of coronary heart disease. Kourou[11] et al. proposed a method for predicting recurrence of oral cancer using a dynamic Bayesian network.

In order to adapt the built Bayesian network to data changes or renewal, this paper proposes a novel method for reconstructing a kind of symptom-disease Bayesian network. Then taking the "colitis" disease as an example, reconstruction of a symptom--"colitis" Bayesian network is made. The numerical example illustrates the validity of the proposed method.

2. Reconstruction of Dynamic Bayesian Networks

2.1. K2 Scoring Function

Scoring function is used to evaluate the fitting degree between network structure and data set, which is
an important index to judge the structure of Bayesian network[12]. K2 scoring function is one of the most popular scoring functions.

The K2 score of the structure \( p_{K2}(D, B^1_S) \) is shown in (1):

\[
p_{K2}(D, B^1_S) = p(B^1_S) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \sum_{k=1}^{q_i} N_{ijk}!
\]

(1)

Where, \( n \) is the number of variates, \( q_i \) is the number of configurations of the \( i \)-th variate \( X_i \) in the set of parent nodes \( P_a(X_i) \). \( r_i \) is the number of states taken by the discrete state variate \( X_i \). \( N_{ijk} \) is the number of samples in the sample data \( D \) when its parent node set \( P_a(X_i) \) takes the \( i \)-th configuration and the variable state is \( k \). \( N_{ij} \) is the sum of all the values of the state \( N_{ijk} \), i.e.,

\[
N_{ij} = \sum_{k=1}^{r_i} N_{ijk}.
\]

2.2. Proposition for Reconstruction

Proposition 1: In the reconstruction of a symptom-disease Bayesian network with two layers for disease \( Y \)(the first layer is made of symptom nodes, and the second layer is made of one node, namely disease \( Y \)), assuming the positive structural symptom node with the lowest posterior probability in the original structure is \( \tau_u \) (positive indicates posterior probability \( P(\tau_u | Y) > 0.5 \)). As the size of samples increases, if a non-structural symptoms \( \tau_t \) satisfies:

\[
P(\tau_t | Y) > P(\tau_u | Y) > 0.5 \text{ and } P(\tau_t | \overline{Y}) < P(\tau_u | \overline{Y}) < 0.5,
\]

\( \tau_t \) could be replaced with \( \tau_u \) to construct a new Bayesian network, and the K2 structure score of the new network can be improved.

Proof:

Assuming \( B^1_S \) to be the current network structure, as \( \tau_u \) is the positive symptom node in the current Bayesian network, the formula shown in Eq.(1) could be rewritten as Eq.(2):

\[
p_{K2}(D, B^1_S) = p(B^1_S) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \sum_{k=1}^{q_i} N_{ijk}!
\]

(2)

Replacing the symptom node \( \tau_u \) with \( \tau_t \), which is a positive symptom word with higher posterior probability, the K2 structural score of the Bayesian network is as shown in Eq.(3):

\[
p_{K2}(D, B^1_S) = p(B^1_S) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \sum_{k=1}^{q_i} N_{ijk}!
\]

(3)

According to the hypothesis of K2 scoring function, \( p(B^1_S) = p(B^2_S) \), then Eq. (4) could be obtained:
\[
p_{K2}(D, B_3^2) = \frac{\prod_{i=1}^{n} \left( \frac{(r_i - 1)!}{(N_i + r_i - 1)!} \right) \prod_{k=1}^{g} N_{i_k}!}{\prod_{i=1}^{n} \left( \frac{(r_i - 1)!}{(N_i + r_i - 1)!} \right) \prod_{k=1}^{g} N_{i_wk}!}
\]

(4)

Since \( N_{ij} = \sum_{k=1}^{g} N_{ijk} \), which is the number of times disease Y occurs or does not occur with all kinds of symptom configurations, so \( N_{iw} = N_{it} \) could be obtained. Let \( N_{iw} = N_{it} \), Eq.(5) could be obtained:

\[
p_{K2}(D, B_3^2) = \frac{\prod_{i=1}^{n} \left( \frac{(r_i - 1)!}{(N_i + r_i - 1)!} \right) \prod_{k=1}^{g} N_{i_k}!}{\prod_{i=1}^{n} \left( \frac{(r_i - 1)!}{(N_i + r_i - 1)!} \right) \prod_{k=1}^{g} N_{i_wk}!}
\]

\[
= \frac{\prod_{i=1}^{n} \left( \frac{(r_i - 1)!}{(N_i + r_i - 1)!} \right) \prod_{k=1}^{g} N_{i_k}!}{\prod_{i=1}^{n} \left( \frac{(r_i - 1)!}{(N_i + r_i - 1)!} \right) \prod_{k=1}^{g} N_{i_wk}!}
\]

(5)

As the numerator of the left fraction has \( N_{i_k} - N_{i_wk} + 1 \) items, and the denominator of the left item has \( N_{i_wk} - N_{i_k} + 1 \) items, and \( N_{it} = N_{i_k} + N_{i_{tk}} \), \( N_{iw} = N_{i_wk} + N_{i_{wtk}} \), \( N_{it} = N_{iw} \), we have: \( N_{i_k} - N_{i_wk} + 1 = N_{i_wk} - N_{i_{tk}} + 1 \). Accordingly, the items of the numerator and the denominator are the same.

According to the prerequisite: \( P(\tau_i | Y) > P(\tau_w | Y) > 0.5 \), we have: \( \frac{N_{i_{tk}}}{N_{it}} > \frac{N_{i_{wtk}}}{N_{iw}} > 0.5 \), and \( N_{i_{tk}} > N_{i_{wtk}} > 0.5N_1 \), \( N_{i_{tk}} < N_{i_{wtk}} < 0.5N_1 \) (\( k_1, k_2 \) are two states of Y. \( k_1 \) denotes that disease Y is diagnosed, and \( k_2 \) denotes that disease Y is not diagnosed). Besides, as \( P(\tau_i | Y) > P(\tau_w | Y) < 0.5 \), then \( N_{2tk_1} < N_{2wtk_1} < 0.5N_2 \), \( N_{2tk_2} < N_{2wtk_2} < 0.5N_2 \). Obviously, \( \frac{N_{i_{wtk}} + l}{N_{i_{tk}} + l} > 1, \frac{N_{2wtk_1} + l}{N_{2tk_1} + l} > 1, (l = 1, 2, \ldots) \) could be obtained. So we have: \( \frac{p_{K2}(D, B_3^2)}{p_{K2}(D, B_3^2)} > 1 \).

3. Dynamic Reconstruction of a Symptom-disease Bayesian Network

According to the proposition 1, if a non-structural symptom word with higher posterior probability is used to replace the positive symptom word node with the lowest posterior probability in the original structure, the K2 score of the overall Bayesian network can be improved.

So, upon the data renewal, a method is proposed for dynamically reconstructing a symptom-disease
Bayesian network: Assume $\tau_u$ is a symptom node with the lowest positive posterior probability of the current Bayesian network, and $\tau_t$ is some other symptom word that is not in the network structure, and target disease is $Y$, and the initial sample size is $M_0$. When the size of the sample increases, or $\Delta n$ new data is added, comparing the posterior probabilities of the symptom words outside the network structure with that of $\tau_u$, if for some other symptom word $\tau_t$ which is not in the network structure, $$P'_{M_0+\Delta n}(\tau_t \mid Y) > P'_{M_0+\Delta n}(\tau_u \mid Y) > 0.5 \text{ and } P(\tau_t \mid \bar{Y}) < P(\tau_u \mid \bar{Y}) < 0.5$$ is satisfied, replace the current node $\tau_u$ with the new node $\tau_t$, otherwise keep the current network unchanged.

4. Numerical Examples

Using disease "colitis" as an example to construct a “symptom-disease” Bayesian network, the proposed method and the K2 scoring method are used respectively to construct a two-layer network (the first layer is made of symptom nodes, and the second layer is made of one node, which is disease $Y$), and the initial sample size $M_0 = 50$. As the size of incremental sample data $\Delta n$ increased, the score of network structure changes as shown in Table 1:

| $\Delta n$ | 0    | 60   | 80    | 100   |
|-----------|------|------|-------|-------|
| Network structure | $C_1 \rightarrow Y$ | $C_{10} \rightarrow Y$ | $C_{10} \rightarrow Y$ | $C_{10} \rightarrow Y$ |
| Structural score | -17.9967 | -49.732 | -55.3881 | -62.2265 |
| $P'_{M_0+\Delta n}(C_1 \mid Y)$ | 0.62 | 0.53 | 0.6083 | 0.6 |
| $P'_{M_0+\Delta n}(C_{10} \mid Y)$ | 0.52 | 0.54 | 0.6167 | 0.6071 |

$\Delta n$ is the size of incremental sample. It can be seen from Table 1 that as the sample size increases, $P'_{M_0+\Delta n}(C_1 \mid Y)$ and $P'_{M_0+\Delta n}(C_{10} \mid Y)$ also change, so are the network structure and the relative scores.

If no changes are made on the Bayesian network structure as the size of the sample increases, it is a static Bayesian network. The scores of the dynamic Bayesian network and the static Bayesian network are compared as shown in Table 2.

| Network sample size | Static network structure | Dynamic network structure |
|---------------------|--------------------------|--------------------------|
| Structural score    | Network structure        | Structural score          | Network structure        |
| 50                  | -17.9967                 | $C_1 \rightarrow Y$      | -17.9967                 |
| 110                 | -50.2138                 | $C_1 \rightarrow Y$      | -49.732                 |
| 130                 | -56.0112                 | $C_1 \rightarrow Y$      | -55.3881                |
| 150                 | -62.8337                 | $C_1 \rightarrow Y$      | -62.2265                |
As shown in Table 2, compared with the static Bayesian network, the dynamic Bayesian network can adjust the network with the new data, and obtain a higher score.

5. Summary
This paper proposes a novel method for dynamically reconstructing Bayesian network structure with less calculating effort comparing with the method of global reconstruction. A proposition is proposed and proved to support the method, and an example about a "symptom-colitis" dynamic Bayesian network is given to show the validity of the method.

6. References
[1] Dean T, Kanazawa K. A model for reasoning about persistence and causation[J]. Computational Intelligence, 1989, 5(2):142-150.
[2] Friedman N. Learning the Structure of Dynamic Probabilistic Networks[C]// 14th International Conference on Machine Learning (ICML), 1997. 1997.
[3] Garg A, Pavlovic V, Rehg J M. Boosted learning in dynamic Bayesian networks for multimodal speaker detection[J]. Proceedings of the IEEE, 2003, 91(9):1355-1369.
[4] Wang Shuyu. Bayesian dynamic model in dam monitoring analysis[J]. Journal of Hydraulic Engineering, 1998, 29(7): 73-77.
[5] Jiang Jinfeng, Liu Fusheng, Zhang Xiaoling. Combined forecasting of Bayesian dynamic linear model[J]. Journal of Shandong University of Science and Technology(Natural Science), 2002, 21(3): 16-17.
[6] SU Ao-xue, FAN Ming-ming, LI Zhong-lai, et al. Reliability Analysis of Distribution System Based on Dynamic Bayesian Network[J]. East China Electric Power, 2012(11):1912-1916.
[7] Gao Xiaoguang, Chen Haiyang, Shi Jianguo. Research on the Mechanism of Variable Structure Dynamic Bayesian Networks[J]. Acta Automatica Sinica, 2011, 37(12):1435-1444.
[8] Carbery C M, Marshall A H, Woods R. Proposing the Deep Dynamic Bayesian Network as a Future Computer Based Medical System[C]// IEEE International Symposium on Computer-based Medical Systems. IEEE, 2016.
[9] Shim J Y. Three dimensional knowledge learning memory with dynamic Bayesian associative matrix for the medical diagnosis[C]// International Joint Conference on Neural Networks. IEEE Xplore, 2001.
[10] Orphanou K, Stassopoulou A, Keravnou E. DBN-extended: A Dynamic Bayesian network model extended with temporal abstractions for coronary heart disease prognosis[J]. IEEE Journal of Biomedical and Health Informatics, 2015:1-1.
[11] Kourou K, Rigas G, Exarchos K P, et al. Prediction of oral cancer recurrence using dynamic Bayesian networks[C]// Engineering in Medicine & Biology Society. IEEE, 2016.