A Bayesian Network Based Decision Support System

Wen-Hsin Chen and Wei-Hua Andrew Wang*

Department of Industrial Engineering and Enterprise Information
Tunghai University
Taichung, 40704, TAIWAN

ABSTRACT

In this highly dynamic environment, data and information are rapidly changing in time. For a decision support system, to capture continuum data and then to extract knowledge from data is in high demand. Our research work focus on developing a systematic approach for discovering causal relations among data and dynamic updating upon to the varying goals. Bayesian network is a powerful tool for representing knowledge and tackling the above inquires. In this research, the entropy-based searching mechanism designed on Bayesian network is adopted to learn the structure of a goal-specific mission. Link strength and connection strength are the measures for selecting links to be added and subtracted. However, these two measures are the relative measures instead of absolute thresholds. Initially, we designed a link/connection strength selecting method for possible adjacency matrices in tabu searching. In the second phase, we used a homogeneity test to ensure the quality of the learned structure. The results of this study will be an important stepping stone to attain the learning requirements of designing, building, and operating effective decision support systems under dynamic environment.

1. INTRODUCTION

Decision support systems (DSS) are one of the more mature systems in knowledge management research. There is a significant availability of framework and models that can be applied to the problem solving of enterprises. Enterprises must focus on the management of knowledge and to develop the decision support system for gaining competitive advantage, especially in the changing, uncertain, and complex world. The research scope is shown in Figure 1. The inputs of the system are data and knowledge, and through a learning process the output of the system is updated knowledge, which can be used to support decision making in action. Note that memory in the process is continuum learning data collected from feedback which occurs when an environment react to an action. Therefore, the process is iterative in this system, continuum learning using tabu list to implement the memory-driven mechanism.

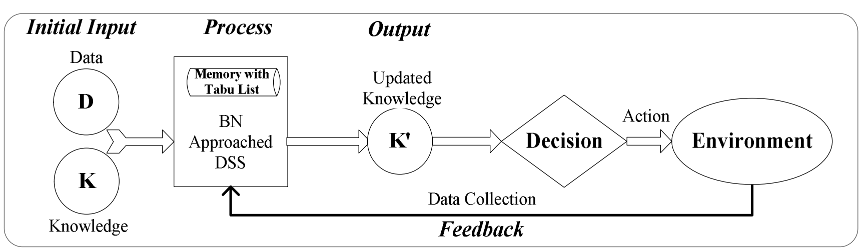

Figure 1. A Bayesian Network approached DSS.

A major challenge for research in related field is to integrate data and knowledge. This research offers a Bayesian network scheme for constructing a decision support system to integrate the data and the knowledge. A Bayesian network is a succinct and efficient tool for representing causal relations. It is a knowledge-representation tool that is capable of efficiently managing the dependence/independence of relationships among random variables. A Bayesian network provides a natural way to integrate existing knowledge with new data by learning the structure of the network.

* Corresponding author: Tel.: +886-4-23594319 Ext. 111; Fax:+886-4-23591756; E-mail: wangwh@thu.edu.tw

Focusing on link strength, we developed a new concept to perform an effective learning structure mechanism, different from traditional models where the strength of the link is equal to either zero or one, the strength of the link in this research is between zero and one and is adjustable by continuum data. This is a new concept to delineate the stochastic dependence relationships among random variables. The proposed method is applied to benchmark examples, i.e., the benchmark dataset: ASIA model [1], in learning the structure of a Bayesian network. As a result, the ASIA model tended to elucidate how the mechanism works.

Three objectives are associated with this problem, i.e., 1) to construct a stochastic model that represents data and knowledge by a Bayesian network, 2) to update the structure by evaluating the strength of the links in the Bayesian network, and 3) imply in tabu search to perform an effective Bayesian network structure learning mechanism.

This article is organized into six sections. In the second, we give a brief literature review covering the research area. The third section introduces notations and assumptions of model. In the fourth section, we illustrate the mechanism of using link strength and connection strength in learning the structure of a Bayesian network based on tabu search. The fifth section we present the experimental results of the benchmark dataset. Finally, a conclusion and perspectives will be proposed.

2. LITERATURE REVIEW

Bayesian networks have an important role in the design and analysis of machine-learning algorithms, serving as a promising way to approach present and future problems related to artificial intelligence and data mining [2]. Typically, a Bayesian network $B=(G,P)$ has two key components: a graph structure (G) and joint probability distribution(P) for a set of variables. The graphical structure illustrates the conditional independencies among the variables in a given problem. A Bayesian network is a graphical model that encodes the joint probability distribution for a set of variables, i.e., $X=\{X_1, X_2, \ldots, X_n\}, n \geq 2$. There is a finite set of vertices or nodes (V), and there is a finite set of directed edges or links (E) between the vertices in the structure of the graph, i.e., $G=(V,E)$. Bayesian networks are a specific type of graphical model with directed edges and no cycles. The directed acyclic graph (DAG) defines the structure of the Bayesian network.

Dealing with learning in a Bayesian network from empirical data, the data can be either complete or incomplete, and the structure of the network can be either known or unknown. To deal with complete data, the problem of structure learning is classified into two approaches: the constraint-satisfaction method and the scored-based method. Although the constraint-satisfaction method is efficient, it is sensitive to failures in the independence tests. Therefore, the scored-based method is a better tool for learning structure from data.

There are two procedures in the scored-based method by which algorithms learn Bayesian networks from data which are, the scoring-criterion procedure and the search procedure. The scoring-criterion procedure displays how well a DAG fits the data, and the search procedure determines a better DAG for complete data. The scoring-criterion procedure computes a score that reflects the goodness-of-fit of the structure to the data. The most common scoring functions are the Bayesian score and the likelihood score. The Bayesian score computes the posterior probability distribution $P(B \mid D)$. The best structure is the one that maximizes the posterior probability. An example of a Bayesian score is the Bayesian Dirichlet (BD) score [3]. For the likelihood score, the score of a Bayesian network B is related to the compression that can be achieved over the data D with an optimal code induced by B . Examples of such scores are the minimum description length (MDL) score [4], the Akaike information criterion (AIC) score, and the Bayesian information criterion (BIC) score. The MDL scoring function from coding theory prefers a simple structure that fits the data well. Both BIC and AIC calculate the score based on likelihood of the data and the penalty of the dimension, but the BIC score penalizes to a greater extent than the AIC score. This means that the BIC score leads to the determination of a simpler structure.

The structure can be measured by the scoring criteria introduced above. As for the search procedure, the goal is to identify the network structure with high scores. Buntine [5] proposed a Bayesian approach for updating both the parameters and the structure. He constructed an initial structure from prior probabilities associated with each possible arc in the domain. Then, he used the posterior probabilities of these arcs for updating. Lam and Bacchus [4] applied the minimal description length principle to account for the structure of the existing network. They focused on sequential updating of the parameter assuming a fixed structure. Friedman and Goldszmidt [6] modified the MDL score in conjunction with the incremental learning procedure and extended these methods to deal with incomplete data in sequential updates. The K2 algorithm maximizes the probability of an optimal graph topology, given a dataset, by using a Bayesian score to rank different graphs [3]. However, this algorithm is restricted by the order of the variables. The hill-climbing (HC) algorithms start a search from an empty, full, or random graph. The HC algorithm searches for every possible addition, removal, or reversal of links to be the candidate graph until the score of graph no longer increases [7]. Other heuristic searching methods, such as: genetic algorithms [8], simulated annealing [9], and tabu search [10] also are used for searching structure. By the tabu search, the ability to use history in creating such evaluations then becomes important for devising effective methods. This is true as well in the use of candidate list strategies. Several key forms of such strategies are highlighted in [11, 12].

3. REPRESENTATION AND ASSUMPTIONS

We use a binary relation R between the finite sets X and Y to represent knowledge. The knowledge of the model can be represented by an adjacency matrix A_{ij} with R, the row and column indices of which are given by X and Y.

$$A_{ij} = \begin{cases} 1 & (i,j) \in R \\ 0 & (i,j) \notin R \end{cases} \tag{1}$$

A_{ij} is also the directed acyclic graph of Bayesian network B, and the assumptions made in this research are: 1) The data are complete and sufficient. 2) All nodes, i.e., $V = \{V_1,\ldots,V_k\}$, in the Bayesian network are fixed. This means we are not to add or subtract nodes; rather, we only add or subtract links while updating. 3) The behavior of the structure is consistent and the proportion of observations is the same for each population. 4) The variables and the initial adjacency matrix are defined by domain expert. However, the completeness of the adjacency matrix is not a major issue since the matrix will be improved along the learning process.

4. BAYESIAN NETWORK BASED DECISION SUPPORT SYSTEM

Two different approaches have been used to construct Bayesian network based decision support system. The first approach is to implement the Bayesian network structure learning with tabu search. The second approach is to test the homogeneity of the real data and the data generated from the learned Bayesian network.

4.1. THE PROPOSED STRUCTURE LEARNING MECHANISM

To implement the process of the integration of data and knowledge, we provided a systematic approach for learning the structure of the Bayesian network. Figure 2 shows the tabu search procedure of the Bayesian network structure learning mechanism. The tabu search is introduced to go beyond the locally optimal termination point. Learning is starting from defining the initial structure of the Bayesian network, and then to do the parameter estimation. Before making the next move, we evaluate the Bayesian network by scoring method. If the score is not meeting the stopping criteria, we make a move or choose from the candidate list. The move can be either to add a link or subtract a link depending on which move get higher score. For choosing the move for next iteration, we use two measures: link strength and connection strength. Note that, tabu lists contain moves, which have been previously made but are prohibited for a certain number of iterations. The searching process is iterative and to ensure convergence, we keep the best score structure.

The variables and the initial adjacency matrix A_{ij} are defined by domain expert. To construct a Bayesian network, we must estimate the parameters based on the existing structure and collect complete data. This is done by using the maximum likelihood parameter estimation in the equations below to estimate θ. Parameter estimation is to find θ^*_{ML} that maximizes the likelihood given the data D.

$$\begin{aligned} \theta^*_{ML} &= \arg\max\left[P(D|\theta)\right] \\ P(D|\theta) &= \prod_{k=1}^{n} P(x^k|\theta) \end{aligned} \tag{2}$$

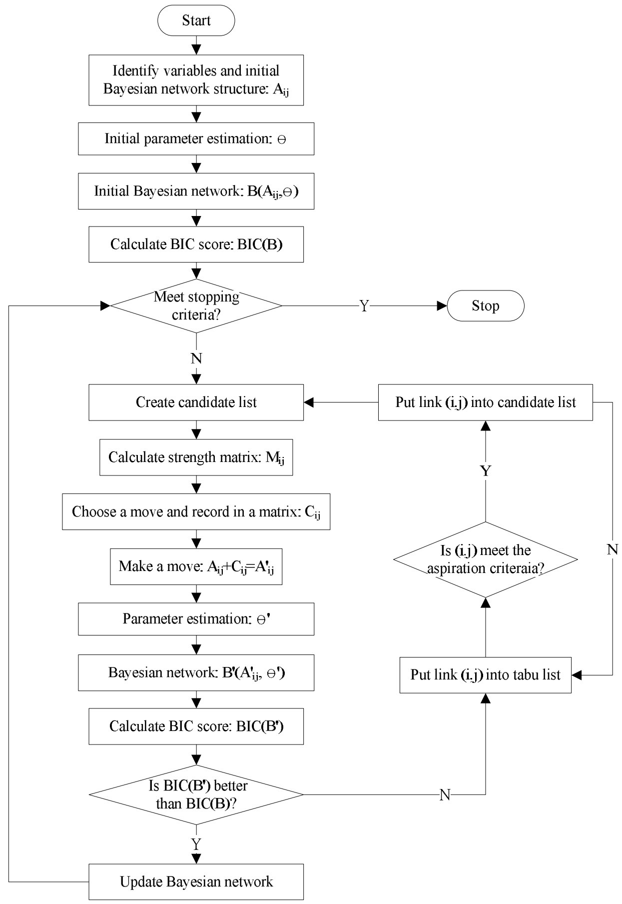

Figure 2. Procedure of proposed structure learning mechanism.

The estimation of the maximum likelihood seeks the solution that can best explain dataset D. Here we get the initial Bayesian network, $B_{i,j} = B(A_{ij},\theta)$. Here we use the Bayesian information criterion in the equation below to evaluate networks:

$$BIC(B,D) = \log P\left(D \middle| B,\theta^*_{ML}\right) - Penalty = \log P\left(D \middle| B,\theta^*_{ML}\right) - \frac{1}{2} Dim(B) \log V \tag{3}$$

In the above formula, $P\left(D \middle| B,\theta^*_{ML}\right)$ is the likelihood of the data D according to estimated parameter θ_{ML} and structure B. V is the sample size of the dataset, and $Dim(B)$ is the dimension of the data. The BIC score is the sum of the log likelihood term and the penalty term. The penalty term penalizes structures that have many links. The BIC score leads to the determination of a simpler structure. Based on the concept of Occam's razor, the simplest model is preferable when all the models have the same score. Therefore, the model with higher dimensions is penalized, as follows:

$$\begin{aligned} &Dim(B) = \sum_{k=1}^{n} Dim(V_k, B) \\ &Dim(V_k, B) = (r_i - 1)q_i \text{, } where\ q_i = \prod_{N_k \in Pa_k(V_k)} r_j \end{aligned} \tag{4}$$

If the score does not meeting the stopping criteria, we do the structure learning. In this step, we choose a move from candidate list. Candidate lists generate new solutions from the current solution. It narrows the examination of elements. To choose a move, we use two measures for calculating strength of links: 1) connection strength and 2) link strength. Connection strength is the measure for selecting one link to be added from all possible links. Connection strength of the link $X \rightarrow Y$ denotes the reduction of uncertainty in Y by having the information of state X. The formula of mutual information is used to determine the relevance of the nodes to others. Link strength is the measure for selecting one link to be subtracted from all possible links; it is calculated by conditional mutual information. Focusing on the link $X \to Y$, the formula of link strength adjusts the mutual information by conditioning on the set $\mathbf{Z}$ of all other parents of Y. The definition of connection strength and link strength are [13]:

$$\begin{aligned} CS_{XY} &= MI(X;Y) = H(X) + H(Y) - H(X,Y) = \sum_{x,y} P(x,y)\log_2\left(\frac{P(x,y)}{P(x)P(y)}\right) \\ LS_{X\to Y} &= MI(X,Y|\mathbf{Z}) = \sum_{x,\mathbf{z}} P(x,\mathbf{z}) \sum_{y} P(y|x,\mathbf{z}) \log_2 \frac{P(y|x,\mathbf{z})}{P(y|\mathbf{z})} \end{aligned} \tag{5}$$

Based on the results of calculating the connection strength and link strength, we calculated the strength matrix by $M_{ij} = \text{strength of links} \cdot A_{ij}$. The result of connection strength matrix and link strength matrix are calculated in the following matrices:

$$\begin{aligned} M_{ij}(CS_{XY}) &= CS_{XY} \cdot A_{ij} = \begin{bmatrix} 0 & CS_{12} & \cdots & CS_{1Y} \\ CS_{21} & 0 & \cdots & CS_{2Y} \\ \vdots & \vdots & \ddots & \vdots \\ CS_{X1} & CS_{X2} & \cdots & 0 \end{bmatrix}, 0 \le CS_{XY} \le 1, \forall i,j \\ M_{ij}(LS_{X\to Y}) &= LS_{X\to Y} \cdot A_{ij} = \begin{bmatrix} 0 & LS_{1\to 2} & \cdots & LS_{1\to Y} \\ LS_{2\to 1} & 0 & \cdots & LS_{2\to Y} \\ \vdots & \vdots & \ddots & \vdots \\ LS_{X\to 1} & LS_{X\to 2} & \cdots & 0 \end{bmatrix}, 0 \le LS_{X\to Y} \le 1, \forall i,j \end{aligned} \tag{6}$$

Note that the elements of connection strength matrix and link strength matrix are between 0 and 1. The 'from-zero-to-one relationship' is transferring back to the 'either-zero-or-one relationship' in the next step. The transferring process is to define a matrix C_{ij} that records the chosen link for moving. For adding a link, we use connection strength with a roulette wheel selection mechanism. Consider a population $RW = \{CS_{12}, CS_{13}, \ldots, CS_{XY}\}$ with size n from connection strength matrix $M_{ij}(CS_{XY})$. Here $n = k(k-1)$, where k is the number of nodes. The probability of each link being selected is:

$$\mathrm{P}(CS_{XY}) = \frac{CS_{XY}}{\sum_{n=1}^{k(k-1)} CS_n}, n = 1,2,\ldots,k(k-1) \tag{7}$$

By using roulette wheel selection, we record the link to be added in the matrix C_{ij}^{+}. The matrix C_{ij}^{+} and updated adjacency matrix A_{ij}^{+} are as follows:

$$\begin{aligned} C_{ij}^{+} &= \begin{cases} 1 & \textit{when (i, j) is the link to be added} \\ 0 & \textit{otherwise} \end{cases} \\ A_{ij}^{+} &= A_{ij} + C_{ij}^{+} \end{aligned} \tag{8}$$

We must look more carefully into the link-adding process because of the DAG constraint of a Bayesian network. Therefore, we must conduct a topological check with adjacency matrix A_{ij}^{+}. The matrix must pass the checking process in order to be the suggested adjacency matrix by adding A_{ij}^{+}. If it does not, we must choose another link by using the roulette-wheel selection. The checking process creates a linear ordering of nodes. If the link (x,y) appears in the graph, then x comes before y in the ordering. The graph must be a directed acyclic graph in order to pass the checking process. But, if x comes after y in the ordering, then the graph is not a DAG.

For subtracting a link, we use link strength as the measure. The link with smallest $LS_{X\to Y}$ is chosen from connection strength matrix $M_{ij}(LS_{X\to Y})$. We recorded the link to be subtracted in the matrix C_{ij}^-. The matrix C_{ij}^- and updated adjacency matrix A_{ij}^- are as follows:

$$C_{ij}^- = \begin{cases} 1 & \textit{when (i, j) is the link to be subtracted} \\ 0 & \textit{otherwise} \end{cases}$$
$$A_{ij}^- = A_{ij} + C_{ij}^- \tag{9}$$

With two updated adjacency matrices A_{ij}^+ and A_{ij}^-, we conduct the parameter estimation and then get two candidate Bayesian networks: $B^+ = B(A_{ij}^+, \theta^+)$ and $B^- = B(A_{ij}^-, \theta^-)$. To find the best network during this iteration, we used the Bayesian information criterion to compare the two candidate Bayesian networks. Here, we obtained $BIC(B^+, D)$ and $BIC(B^-, D)$ as follows:

$$\begin{aligned} BIC(B^+, D) &= \log P(D|B^+, \theta_{ML}^*) - \frac{1}{2} Dim(B^+) \log V \\ BIC(B^-, D) &= \log P(D|B^-, \theta_{ML}^*) - \frac{1}{2} Dim(B^-) \log V \end{aligned} \tag{10}$$

We compared the two BIC scores, i.e., $BIC(B^+, D)$ and $BIC(B^-, D)$, and then the Bayesian network with the higher BIC score is selected to be the updated Bayesian network $B'_{i,j}$ as shown below:

$$B'_{i,j} = \begin{cases} B^-(A_{ij}^-, \theta^-), & \textit{when } BIC(B^-, D) > BIC(B^+, D) \\ B^+(A_{ij}^+, \theta^+), & \textit{when } BIC(B^+, D) > BIC(B^-, D) \end{cases} \tag{11}$$

Learning is conducted iteratively. We use tabu lists to record the moves. The aspiration criteria is the size of memory; follow the rule of first-in, first-out. Also, in each iteration, we use breeder selection to keep the best structure. Breeder selection is helping the searching methods converge. The stopping criteria for iteratively structure learning is to stop when the deviation is less than the threshold.

4.2. HOMOGENEITY TEST

Because the model of the original complete data is unknown, it is difficult to evaluate the validity of the structure between an unknown structure and the learned structure. In order to compare the behavior of the learned structure to the original model, we conducted the homogeneity test for I populations. The null hypothesis and the alternative hypothesis of the homogeneity test between $data'$ and $data$ are as follows:

$$\begin{aligned} &H_0 : p_{1j} = p_{2j}, j = 1, 2, \ldots, J \\ &H_1 : H_0 \textit{ is not true} \end{aligned} \tag{12}$$

The test statistical approach is the use of a chi-squared random variable, χ^2, which is defined by the following equation:

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - E_{ij})^2}{E_{ij}}, \textit{where } E_{ij} = n_i \frac{N_{\bullet j}}{n} \tag{13}$$

Here, $N_{\bullet j}$ is the number of observations among the n samples that fall into category j .The result is to reject H_0 if $\chi^2 \geq \chi^2_{\alpha,(I-1)(J-1)}$. Otherwise, if not to reject H_0, the result states that the proportion of observations in category j is the same for each population.

5. A Numerical Example

The ASIA model [1], a benchmark Bayesian network, is a lung cancer model of 8 nodes and 8 links. Although the ASIA model belongs to the medical related domain, the relations and probability property between variables in ASIA model are the same in the industry domain where we can also implement this mechanism to support decision making. The ASIA model shows a combination of diseases represented in a Bayesian network: tuberculosis (Node 3), lung cancer (Node 4), and bronchitis (Node 5), together with the presentation of dyspnea (Node 8) and X-ray (Node 7). A recent visit to Asia (Node 1) increases the risk of tuberculosis, while smoking (Node 2) is known to be a risk factor for both lung cancer and bronchitis. And Node 6 is the OR logic node of cancer or tuberculosis. In this model, each node has two states.

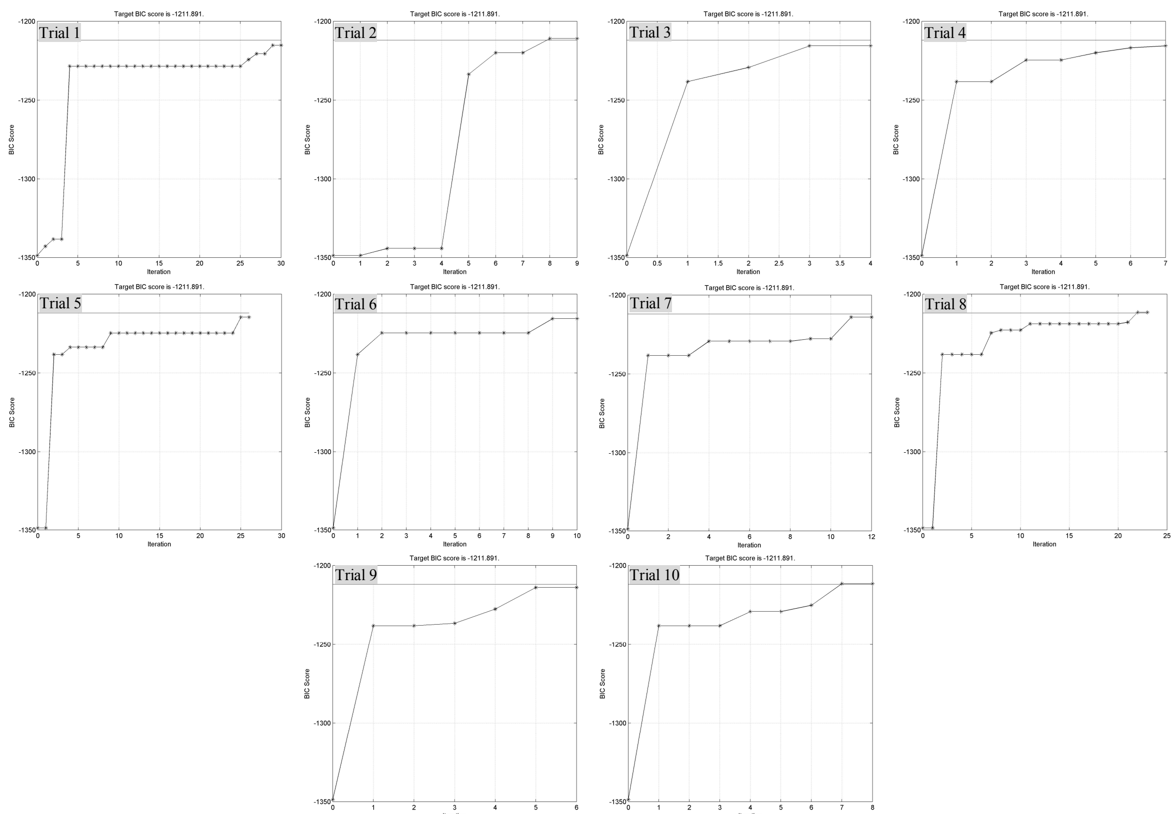

Figure 3. Plots of BIC score per iteration.

Table 1. Homogeneity test result of experiment.

Trials	BIC(B')	Iterations	Result of Homogeneity Test
1	-1215.17	30	$\chi^2 = 8.5399 < \chi^2_{0.05.14} = 23.6848$, Not reject H_0
2	-1210.91	9	$\chi^2 = 16.9303 < \chi^2_{0.05.14} = 23.6848$, Not reject H_0
3	-1215.50	4	$\chi^2 = 16.7254 < \chi^2_{0.05.13} = 22.3620$, Not reject H_0
4	-1215.50	7	$\chi^2 = 17.0935 < \chi^2_{0.05.15} = 24.9958$, Not reject H_0
5	-1214.59	26	$\chi^2 = 12.6328 < \chi^2_{0.05.14} = 23.6848$, Not reject H_0
6	-1214.59	10	$\chi^2 = 13.4950 < \chi^2_{0.05.15} = 24.9958$, Not reject H_0
7	-1213.93	12	$\chi^2 = 12.9777 < \chi^2_{0.05.14} = 23.6848$, Not reject H_0
8	-1211.52	23	$\chi^2 = 8.6177 < \chi^2_{0.05.13} = 22.3620$, Not reject H_0
9	-1213.93	6	$\chi^2 = 22.3298 < \chi^2_{0.05.16} = 26.2962$, Not reject H_0
10	-1211.52	8	$\chi^2 = 10.3579 < \chi^2_{0.05.15} = 24.9958$, Not reject H_0

The target BIC score of our experiment is $BIC(\mathrm{T}) = -1211.89$. The tabu list aspiration criterion is the size of seven [8] and the learning stopping criteria is the deviation of five. The results of learning are shown in Figure 3. The use of tabu search avoids being trapped at local optimum, i.e, the scores plot of trial 1, 5, and 8 in Figure 3. From the efficiency of the learning process, the learning procedure converged in an average of 13.5 iterations. The homogeneity test results (Table 1) show that H_0 is not rejected in any of the 10 trials. This result states that the proportions of observations in categories are the same for two structures. As a result, we may not reject that the behavior of learned structure is the same as original model.

6. CONCLUSION

In this research, we used link strength and connection strength to describe the link of a Bayesian network. This research provided a stochastic model that can be used to integrate data and knowledge by a Bayesian network. Also, the structure and the parameters can be updated by a series iterative tabu search procedures that avoid being trapped at local optimum. The experimental results of the benchmark datasets (the ASIA model) show that the Bayesian network based decision support system demonstrated homogeneous behavior of learned structure. The experimental results also display that the proposed approach is able to search for the Bayesian network structure in an effective way.

The application of Bayesian network based decision support system in healthcare-related fields should be considered in the future. Despite the limitations and the need to improve the algorithm, the results of this research can be used to establish a decision support system for preventing disease and promoting healthier lifestyles. In the healthcare-related field, there are huge quantities of available data that are highly dimensional and uncertain. Regarding the issue of large and highly-dimensional datasets in real-world applications, future researchers should consider the use of divide-and-conquer techniques to deal with the magnitude and complexity of the situation.

REFERENCES

[1] S. L. Lauritzen and D. J. Spiegelhalter: "Local computations with probabilities on graphical structures and their application to expert systems", Journal of the Royal Statistical Society, Series B (Statistical Methodology), Vol.50, No.2, pp.157–224, 1988.

[2] P. Doshi, L. Greenwald and J. Clarke, "Towards Effective Structure Learning for Large Bayesian Networks", AAAI Workshop on Probabilistic Approaches in Search, pp.16-22, Edmonton, Canada, 2002.

[3] G. F. Cooper and E. Herskovits: "A Bayesian Method for the Induction of Probabilistic Networks from Data", Machine Learning, Vol.9, No.4, pp.309–347, 1992.

[4] W. Lam and F. Bacchus: "Learning Bayesian belief networks: An approach based on the MDL principle", Computational Intelligence, Vol.10, pp.269–293, 1994.

[5] W. Buntine, "Theory refinement on Bayesian networks", In D'Ambrosio B. D., Smets P., and Bonissone P. P. (Eds.), Proceedings of the Seventh conference on Uncertainty in Artificial Intelligence (UAI'91), pp. 52–60. San Francisco, CA, USA, Morgan Kaufmann, 1991.

[6] N. Friedman and M. Goldszmidt, "Sequential update of Bayesian network structure", In D. Geiger and P. Shanoy (Eds.), Proceedings of Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI '97). San Francisco, CA, USA, Morgan Kaufmann, 1997.

[7] D. Margaritis, "Learning Bayesian network model structure from data", PhD thesis. Pittsburgh: Carnegie-Mellon University, School of Computer Science, 2003.

[8] P. Larranaga: "Structure learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.18, pp.912–926, 1996.

[9] D. M. Chickering, Learning Bayesian networks is NP-Complete. In Fisher, D. and Lenz, H., (Eds), Learning from Data:Artificial Intelligence and Statistics V, Springer-Verlag, 1995.

[10] P. Muntenau and D. Cau, "Efficient score-based learning of equivalence classes of Bayesian networks", In Lecture Notes in Artificial Intelligence, pp. 96–105, Springer, 2000.

[11] F. Glover: "Tabu Search Part I", ORSA Journal on Computing, Vol. 1, No. 3, pp. 190–206, 1989.

[12] F. Glover: "Tabu Search Part II", ORSA Journal on Computing, Vol. 2, pp. 4–32, 1990.

[13] I. Ebert-Uphoff, "Tutorial on How to Measure Link Strengths in Discrete Bayesian Networks", Research Report, GT-ME-2009-001, 2009.