the UNIVERSITY of GREENWICH

Proceedings

11th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD2010)

9-11 June 2010

The University of Greenwich, London, United Kingdom

Editors:

Dr. Jixin Ma, The University of Greenwich, United Kingdom

Dr. Liz Bacon, The University of Greenwich, United Kingdom

Dr. Wencai DU, Hainan University, China

Dr. Miltos Petridis, University of Greenwich, United Kingdom

Sponsored by the International Association for Computer & Information Science (ACIS)

2010 11th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing

SNPD 2010

Table of Contents

Bayesian Network Inference with Qualitative Expert Knowledge for Decision Support Systems

Nipat Jongsawat
Graduate School of IT in Business
Siam University
Bangkok, Thailand
nipatj@yahoo.com

Wichian Premchaiswadi
Graduate School of IT in Business
Siam University
Bangkok, Thailand
wichian@siam.edu

***Abstract*— In this paper, we consider a methodology that utilizes qualitative expert knowledge for inference in a Bayesian network. The decision-making assumptions and the mathematical equation for Bayesian inference are derived based on data and knowledge obtained from experts. A detailed method to transform knowledge into a set of qualitative statements and an "a priori" distribution for Bayesian probabilistic models are proposed. We also propose a simplified method for constructing the "a prior" model distribution. Each statement obtained from the experts is used to constrain the model space to the subspace which is consistent with the statement provided. Finally, we present qualitative knowledge models and then show a full formalism of how to translate a set of qualitative statements into probability inequality constraints.**

Keywords-Bayesian network, Bayesian network inference, decision-support systems, qualitative expert knowledge, probability inequality constraints Bayesian network, Bayesian network inference, decision-support systems, qualitative expert knowledge, probability inequality constraints

I. Introduction

One of the most difficult obstacles in the practical application of probabilistic methods is the effort that is required for model building and, in particular, for indentifying the causal relationships among variables in the model and quantifying graphical models with numerical probabilities. In general, the causal structure and the numerical parameters of a Bayesian Network (BN) can be obtained from an expert or can be learned from a data set. This paper only focuses on acquiring the causal structure and the numerical parameters of a Bayesian Network from an expert. In BN analysis, users can perform Bayesian inference in the model and they can also compute the impact by observing values of a subset of the model variables on the probability distribution over the remaining variables. A set of qualitative statements and numerical parameters obtained from an expert are very important for inference in a Bayesian network. Thus, the method to transform expert knowledge, represented by a set of qualitative statements, into an a priori distribution for Bayesian probabilistic models is an important consideration. The set of available information should be classified into an available set of data and a body of nonnumeric expert knowledge should be included for performing Bayesian inference in a model. Several notations have been adopted for a Bayesian model class. A Bayesian model is determined by a graph structure and by the parameter vector needed to specify the conditional probability distributions given that structure. The BN model distribution is represented by a mathematical formula. Therefore, this paper provides a full formalism of how to translate a set of qualitative statements into probability inequality constraints, which is described in section 4. Several cases are provided of how Bayesian influence is classified and the probability inequality constraints are shown for each case.

This paper is organized as follows: Section 2 presents more detail about the background of Bayesian networks and some perspectives of qualitative causal relationships in the Bayesian approach. Section 3 addresses the methods to transform expert knowledge into an a priori distribution for Bayesian probabilistic models in more detail. Section 4 describes the probabilistic representation of a qualitative expert knowledge model and section 5 presents a conclusion and discusses some perspectives and ideas for future work.

II. Background

This section is intended to describe the background of Bayesian networks and some perspectives of qualitative causal relationships in the Bayesian approach. Bayesian networks (also called belief networks, Bayesian belief networks, causal probabilistic networks, or causal networks) are acyclic directed graphs in which nodes represent random variables and arcs represent direct probabilistic dependencies among the nodes [1]. Bayesian networks are a popular class of graphical probabilistic models for research and application in the field of artificial intelligence. They are motivated by Bayes' theorem [2] and are used to represent a joint probability distribution over a set of variables. This joint probability distribution can be used to calculate the probabilities for any configuration of the variables. In Bayesian inference, the conditional probabilities for the values of a set of unconstrained variables are calculated given fixed values of another set of variables, which are called observations or evidence.

978-0-7695-4088-7/10 $26.00 © 2010 IEEE
DOI 10.1109/SNPD.2010.10

Bayesian models have been widely used for efficient probabilistic inference and reasoning [1], [3], and numerous algorithms for learning the Bayesian network structure and parameters from data have been proposed [4], [5], [6]. The causal structure and the numerical parameters of a Bayesian network can be obtained using two distinct approaches [7], [15]. First, they can be obtained from an expert. Second, they can also be learned from a dataset or data residing in a database. The structure of a Bayesian network is simply a representation of independencies in the data and the numerical values are a representation of the joint probability distributions that can be inferred from the data [8, 9]. In practice, some combination of these two approaches is typically used. For example, the causal structure of a model is acquired from an expert, while the numerical parameters of the model are learned from the data in a database.

For realistic problems, the data base is often very sparse and hardly sufficient to select one adequate model. This is considered as model uncertainty. Selecting one single model can lead to strongly biased inference results. On the other hand, in science and industry, there is an enormous amount of qualitative knowledge available. This knowledge is often represented in terms of qualitative causal relationships between two or more entities. For example, in the statement: "smoking increases the risk of lung cancer," the two entities: smoking and lung cancer are related to each other. Moreover, the smoking entity positively influences the lung cancer entity since lung cancer risk is increased in the case of smoking. It is therefore desirable to make use of this body of evidence in probability inference modeling.

III. Methods

In this section, we describe a methodology to use qualitative expert knowledge for inferencing in a Bayesian network. We proceed from the decision-making assumptions and the general equation for Bayesian inference based on data and knowledge obtained from an expert to a detailed method to transform knowledge, represented by a set of qualitative statements, into an a priori distribution for Bayesian probabilistic models.

Consider a simple case of decision making in which the body of expert knowledge ω consists of a single statement ω = "making decision A causes B". We know that there are 2 random events or variables A and B, which we assume are binary, and we need to consider the set of all possible Bayesian models on A and B. The set of possible model structures are described in the following categories: 1) S_1: A and B have no causal relationship between them, 2) S_2: A and B have some causal relationship between them but the direction of influence cannot be identified, 3) S_3: A causes B, and 4) S_4: B causes A. "decision making in an event A activates an event B" directly states a causal influence of A on B. We use the statement "A activates B" to constrain the space of structures: $P(S_3|\omega) = 1$; $P(S_n|\omega)=0$, n=1,2,4). The ω is represented as a qualitative statement described by the expert, A causes B. The graph structure (S_3) encodes the probability distribution

$$P(A,B) = P(B|A)P(A) \tag{1}$$

No further information on P(A) is available; however, P(B|A) can be further constrained. The corresponding Conditional Probability Table (CPT) is shown in Table I.

TABLE I. Conditional Probability Table

A	P(B=1|A)
0	θ_0
1	θ_1

The values of the conditional probabilities from the components of the parameter vector $\theta = (\theta_0, \theta_1)$ of the model class with structure S_3. θ_0 is the probability of B is active when A is not active. θ_1 is the probability of B is active when A is active. From the statement, we now can infer that the probability of B is active when A is active is higher than the same probability with A inactive. The P(B|A) when P(A) is available is higher than the P(B|A) when P(A) is not available. The inequality relationship is obtained as follows:

$$P(B=1|A=1) \geq P(B=1|A=0),\ \theta_1 \geq \theta_0 \tag{2}$$

Hence, the set of model parameters consistent with that statement is given by

$$\Theta_3 = \{(\theta_0, \theta_1)|\ 0 \leq \theta_0 \leq 1 \wedge \theta_0 \leq \theta_1 \leq 1\} \tag{3}$$

and the distribution of models in the structure-dependent parameter space becomes

$$P(\theta \mid S_3, \omega) = \begin{cases} 1, \theta \in \Theta_3 \\ 0, \text{else} \end{cases} \tag{4}$$

A Bayesian model m represents the joint probability distribution of a set of variables $X = X_1, X_2, X_3, \ldots, X_D$. The model is defined by a graph structure, which determines the structures of the conditional probabilities between variables, and a parameter vector θ, the components of which define the entries of the corresponding conditional probability tables (CPTs). Hence, a Bayesian network can be written as $m = \{s, \theta\}$. Given some observations or evidence E, reflected by fixed measured values of a subset of variables, the conditional probability given the evidence in light of the model is described as P(X|E, m).

The full Bayesian network model does not attempt to approximate the true underlying distribution. Instead, all available information is used in an optimal way to perform inference, without taking one single model for granted. To formalize this statement for our purposes, let us classify the set of available information into an available set of data D and a body of nonnumeric expert knowledge ω. The probability distribution of model m is given by

$$P(m|D, \omega) = \frac{P(D|m)P(m|\omega)}{P(D, \omega)} \tag{5}$$

The first parameter value D of P(D, ω) is the likelihood of the data given the model, which is not directly affected by nonnumeric expert knowledge ω, the second parameter value ω denotes the model a prior, whose task is to reflect the background knowledge. For simplicity, the numerator P(D, ω) of P(m|D, ω) will be omitted from the equation (5). The term P(D|m) contains the constraints of the model space by the data, and the term P(m|ω) contains the constraints imposed by the expert knowledge. Hence, given some observation or evidence E, the conditional distribution of the remaining variable X is performed by integrating over the models.

$$P(X|E, D, \omega) = \int P(X|E, m)P(m|D, \omega)dm \tag{6}$$

$$= \int P(X|E, m)P(D|m)P(m|\omega)dm \tag{7}$$

In this paper, we consider the case of no available quantitative data; D is assigned a null value. The term D and P(D|m) will be omitted from equation (6) and (7). Even in this case, it is still possible to perform a proper Bayesian inference.

$$P(X|E, \omega) = \int P(X|E, m)\, P(m|\omega)dm \tag{8}$$

Now, the inference is based on the general information (contained in ω) obtained from experts, and the specific information provided by the measurement E. In order to determine P(m|ω), we need a formalism to translate the qualitative expert knowledge into an a priori distribution over Bayesian models. The following notations are adopted for a Bayesian model class. A Bayesian model is determined by a graph structure s and by the parameter vector θ needed to specify the conditional probability distributions given that structure. The parameter vector θ is referred to by one specific CPT configuration. A Bayesian model class is then given by 1) a discrete set of model structures $S = \{s_1, s_2, s_3, \ldots, s_K\}$ and for each structure s_k, a set of CPT configurations Θ_k. The set of member Bayesian models m Є M of that class is then given by $m = \{(s_k, \theta)|k \in \{1, \ldots, K\}, \theta \in \Theta_k\}$. The model distribution is shown in (9).

$$P(m|\omega) = P(s_k, \theta|\omega)$$

$$= \frac{P(\theta|s_k, \omega)P(s_k|\omega)}{\sum_{a=1}^{K} \int_{\Theta_a} P(\theta|s_a, \omega)d\theta\, P(s_a|\omega)} \tag{9}$$

In (9), the set of allowed structures is determined by means of ω, followed by the distributions of the corresponding CPT configurations. The model's a posterior probability P(m|ω) is calculated as shown in (9). Inference is carried out by integrating over the structure space and the structure-dependent parameter space.

$$P(X|E, \omega) = \sum_{k=1}^{K} \int_{\Theta_3} P(X|E, s_k, \theta)P(s_k, \theta|\omega)d\theta \tag{10}$$

It is common to express nonnumeric expert knowledge in terms of qualitative statements about a relationship between entities. The ω is represented as a list of such qualitative statements. The following information is essential to determine the model a priori (10): First, each entity which is referenced in at least one statement throughout the listed is assigned to one variable X_i. Second, each relationship between a pair of variables constrains the likelihood of an edge between these variables being presented. Last, the quality of the statement such as activates or inactivates affects the distribution over CPT entries θ given the structure. The statement can be used to shape the joint distribution over the class of all possible Bayesian models over the set of variables obtained from ω in the general case.

We propose a simplified method for constructing the a priori model distribution. Each statement is used to constrain the model space to that subspace which is consistent with that statement. In other words, if a statement describes a relationship between two variables, only structures s_k which contain the corresponding edge are assigned a nonzero probability $P(s_k|\omega)$. Likewise, only parameter values on that structure, which are consistent with the contents of that statement, are assigned a nonzero probability $P(\theta|s_k, \omega)$. If no further information is available, the distribution remains constant in the space of consistent models.

Having derived the Bayesian model class (s_3, Θ_3) consistent with the statement, we can now perform inference by using an equation (10). Under the condition of A is set to active (E = {A = 1}), let us ask what is the probability of having B active. We can determine this by integrating over all models with nonzero probability and averaging their respective inferences, which can be done analytically in this simple case.

$$P(B = 1|E, \omega) = \sum_{k=1}^{K} P(s_k, \omega) \int_{\Theta_3} P(B|A, s_k, \theta)P(\theta|s_k, \omega)d\theta$$

$$= \omega \int_{\Theta_3} P(B{=}1|A{=}1, \theta)d\theta \tag{11}$$

$$= \omega \int_0^1 \int_{\theta_2}^1 \theta_1 d\theta_1\, d\theta_0 = 2/3$$

where $\omega = 2$ is the normalizing factor in the parameter space of $\theta = (\theta_0, \theta_1)$ such that

$$\omega \int_{0}^{1} \int_{\theta_0}^{1} d\theta_1 d\theta_0 = 1 \qquad (12)$$

It is worth noting that, as long as simple inequalities are considered as statements, the problem remains analytically tractable even in higher dimensions. In general, integration during Bayesian inference can become intractable using analytical methods.

IV. Probabilistic Representation of a Qualitative Expert Knowledge Model

The model from the previous section is derived to provide a full formalism of how to translate a set of qualitative statements into probability inequality constraints. Several qualitative models have been proposed in the context of qualitative probabilistic networks. Qualitative knowledge models describe the process of transforming qualitative statements into a set of probability constraints. The proposed Bayesian inference method outlined above is independent of the qualitative knowledge model. The model's a posterior probability is independent of the set of qualitative statements used, once the set of probabilistic inequality constraints which are translated from qualitative statements is determined. Three existing qualitative models are the Wellman approach [10], the Neufeld approach [11], and the orders of magnitude approach [12]. In this paper, we utilize the Wellman approach, where qualitative expert knowledge involves influential effects from parent variables to child variables which are classified according to the number of inputs from parent to child and their interaction. For reasons of simplicity, binary-valued variables are used in our examples. The values of a variable or node defined as "present" and "absent" or "active" and "inactive" are represented as logical values "1" and "0" (as synonyms A and A). For multinomial variables, similar definitions can be applied.

Qualitative influences with directions can be defined based on the number of influences imposed from parent to child. There are three cases of influences, namely, single influence, joint influence, and mixed joint influence. We will also discuss the qualitative influence derived from recurrent or conflicting statements based on three cases of influence. The definitions of influence presented in this article are refined based on the QPN in [10]. They are used to translate the qualitative expert statements into a set of constraints in the parameter space which can be used to model the parameter distribution given the structure.

A. Single Influence

In the statement, "*investing in project A increases the profit of the entire project in such good economic situations*," investing in project A is the parent node which has a single positive influence on child node the profit of the entire project. The "$\overline{\text{Invest A}}$" or "not invest in A" means that the company may invest in other projects rather than the project A.

$$P(\text{Entire Project Profit}|\text{Invest A}) \geq P(\text{Entire Project Profit}|\overline{\text{Invest A}})$$

In another statement, "*investing in project A reduces the profit of the entire project in such a severe economic crisis*," investing in project A is the parent node which imposes a single negative influence on child node the profit of the entire project.

$$P(\text{Entire Project Profit}|\text{Invest A}) \leq P(\text{Entire Project Profit}|\overline{\text{Invest A}})$$

The graphical representation of the above qualitative statements from an expert is shown in Fig.1.

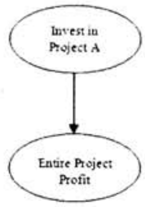

Figure 1. Example of single positive and negative influence.

B. Joint Influence

Let us consider credit worthiness of individual causes. Several risk factors have been identified for credit worthiness. According to the Thai credit bureau report, the three most prominent risk factors are reliability, future income, and age. The chance of getting credit worthiness increases as an individual gets higher future income, age, and reliability. This knowledge about credit worthiness factors can be encoded by a qualitative causality model. According to the statements, the main risk factors that influence credit worthiness by positive synergy as shown in Fig. 2.

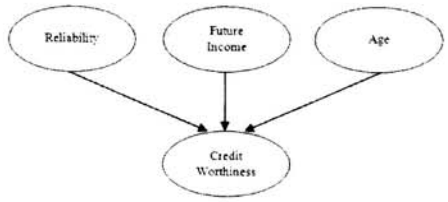

Figure 2. Example of plain synergy influence. Reliability, future income, and age synergically influence credit worthiness.

The joint influence of these three factors together is more significant than individual influences from any of these factors alone. We can represent this synergy by the inequalities

$$P(CW|R, FI, A) \geq \begin{Bmatrix} P(CW|R, \overline{FI}, \overline{A}) \\ P(CW|\overline{R}, FI, \overline{A}) \\ P(CW|\overline{R}, \overline{FI}, A) \end{Bmatrix}$$

$$P(CW|R, FI, A) \geq \begin{Bmatrix} P(CW|R, FI, \overline{A}) \\ P(CW|R, \overline{FI}, A) \\ P(CW|\overline{R}, FI, A) \end{Bmatrix}$$

and

$$P(CW|R, FI, A) \geq P(CW|\overline{R}, \overline{FI}, \overline{A})$$

If we assume these risk factors pair wise symmetric, we can further derive the following inequalities:

$$\begin{matrix} P(CW|R, FI, \overline{A}) \\ P(CW|R, \overline{FI}, A) \\ P(CW|\overline{R}, FI, A) \end{matrix} \geq \begin{Bmatrix} P(CW|R, \overline{FI}, \overline{A}) \\ P(CW|\overline{R}, FI, \overline{A}) \\ P(CW|\overline{R}, \overline{FI}, A) \end{Bmatrix}$$

where CW, R, FI, and A stands for Credit Worthiness, Reliability, Future Income, and Age. Note that often but not always, the combined influence refers to the sum of independent influences from each parent node to each child node. Assume that parent nodes R and FI impose negative individual influence on child node CW, then the knowledge model can be defined as

$$P(\overline{CW}|R, FI) \geq \begin{Bmatrix} P(\overline{CW}|R, \overline{FI}) \\ \\ P(\overline{CW}|\overline{R}, FI) \end{Bmatrix} \geq P(\overline{CW}|\overline{R}, \overline{FI})$$

C. *Mixed Joint Influence*

Generally, the extraction of a probability model is not well defined if the joint affect on a child is formed by a mixture of positive and negative individual influences from its parents. Therefore, we adopted the following scheme: If there are mixed influences from several parent nodes on a child node, and no additional information is given, then these are treated as independent and with equal influential strength. Assume that parent node A imposes a positive single influence on child node C and parent node B imposes a negative single influence on child node C, then the joint influence can be represented by

$$\begin{aligned} &P(C|A, B) \geq P(C|\overline{A}, B), \\ &P(C|A, \overline{B}) \geq P(C|\overline{A}, \overline{B}), \\ &P(C|A, \overline{B}) \geq P(C|A, B), \\ &P(C|\overline{A}, \overline{B}) \geq P(C|\overline{A}, B). \end{aligned}$$

For example, future income imposes a positive single influence on credit worthiness and debt imposes a negative single influence on credit worthiness, then the joint influence can be represented by

$$\begin{aligned} &P(CW|FI, D) \geq P(CW|\overline{FI}, D), \\ &P(CW|FI, \overline{D}) \geq P(CW|\overline{FI}, \overline{D}), \\ &P(CW|FI, \overline{D}) \geq P(CW|FI, D), \\ &P(CW|\overline{FI}, \overline{D}) \geq P(CW|\overline{FI}, D). \end{aligned}$$

A credit worthiness case study for a mixed joint influence is shown in Fig. 3.

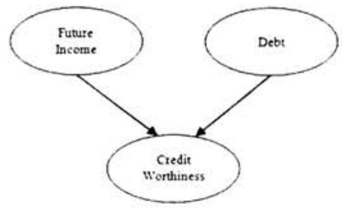

Figure 3. Example of mixed joint influence. Debt and future income influence on credit worthiness.

Once formulated, we can use a Monte Carlo sampling procedure to make sure that all inequalities are satisfied for valid models. Any additional structure can be brought into the CPT of the corresponding structure as soon as the dependencies between influences are made explicit by further qualitative statements.

D. *Recurrent Statements*

Recurrent statements such as "A activates itself" or "A activates B"; "B activates A" cannot directly be represented in Bayesian models. This problem can be solved by transforming the initial Bayesian model class into the equivalent class of a Dynamic Bayesian Network (DBN) [13], [14]. A DBN is a BN extended with a temporal dimension to enable us to model dynamic systems. A DBN unfolds the variable space in time, i.e., each variable exists in two replicas, one in a "cause" layer, the second one in the "effect" layer. A causal relationship from variable X_i to variable X_j is transformed to an edge from X_i in the cause layer to X_j in the effect layer. As a consequence, the resulting DBN automatically obtains a directed acyclic graph structure.

V. Conclusion and Future Work

In this paper, a methodology to use qualitative expert knowledge for inference in a Bayesian network is proposed. We establish the decision-making assumptions and the general equation for Bayesian inference based on data and knowledge obtained from experts. We also describe a detailed method to transform knowledge, represented by a set of qualitative statements, into an a priori distribution for Bayesian probabilistic models. The set of model parameters consistent with the statements and the distribution of models in the structure-dependent parameter space are presented. We propose a simplified method for constructing the a prior model distribution. Each statement is used to constrain the model space to a subspace which is consistent with the statements. Next, we provide a full formalism of how to translate a set of qualitative statements into probability inequality constraints. Several cases of Bayesian influence are classified and the probability inequality constraints presented in each case are described.

For future research based on this study, we intend to apply Bayesian network inference with qualitative expert knowledge for group decision making. The qualitative knowledge from different experts in group work will be considered as a case study. Group decision knowledge will further be transformed into a set of qualitative statements and then converted into probability inequality constraints. We will apply the proposed method to a specific case study using a set of group decision making statements and report the simulation results.

REFERENCES

[1] J. Pearl, J, "Probabilistic Reasoning in Intelligent Systems," Networks of Plausible Inference, San Mateo, CA, Morgan Kaufmann Publishers.

[2] T. Bayes, "An Essay Towards Solving a Problem in the Doctrine of Chances," Philosophical Trans. Royal Soc. Of London, 1763.

[3] S.L. Lauritzen and D.J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems, " J. Royal Statistical Soc., 1988.

[4] D. Heckerman, "A Tutorial on Learning with Bayesian Networks," Technical Report MSR-TR-95-06, Microsoft, http://research.microsoft.com/research/pubs/view.aspx?msr_tr_id=MSR-TR-95-06, 1996.

[5] D. Heckerman, "Learning Bayesian Networks: The Combination of Knowledge and Statistical Data, "Proc. KDD Workshop, 1994.

[6] N. Friedman and M. Goldszmidt, "Learning Bayesian Networks with Local Structure, " Learning in Graphical Models, 1999.

[7] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu, "Learning Bayesian Networks from Data: An Information-Theory Based Approach," Department of Computing Sciences, University of Alberta, Faculty of Informatics, University of Ulster, 2001.

[8] M. Sin and M. Valtorta, "Construction of Bayesian, Network Structures from Data: a Brief Survey and an Efficient Algorithm," Dept. of Computer Science, University of South Carolina, Columbia, USA, 1995.

[9] P. Spirtes and C. Meek, "Learning Bayesian Networks with Discrete Varibales from Data," Proceedings of the Conference on Knowledge Discovery & Data Mining, 1995.

[10] M. P. Wellman, "Fundamental Concepts of Qualitative Probabilistic Networks," Artificial Intelligence, 1990.

[11] E. Neufeld, "A Probabilistic Commonsense Reasoner," Int'l J. Intelligent Systems, 1990.

[12] J. Cerquides and R. Lopez de Mantaras, "Knowledgge Discovery with Qualitative Influences and Synergies, " Proc. Second European Symp. Principle of Data Mining and Knowledge Discovery (PKDD), 1998.

[13] T. Dean and K. Kanazawa, "A Model for Reasoning about Persistence and Causation," Artificial Intelligence, 1989.

[14] K. Murphy, "Dynamic Bayesian Networks: Representation, Inference, and Learning," PhD dissertation, University of California, Berkeley, 2002.

[15] J. Nipat and P. Wichian, "A SMILE Web-based Interface for Learning the Causal Structure and Performing a Diagnosis of a Bayesian Network," Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, USA , 2009