

A New Structure Learning of Bayesian Networks and Its Application to Context Aware Computing in Home Network System: Genetic Approach

JAEHUN LEE, WOYONG CHUNG, SUKHYUN YUN, SOOHAN KIM* AND EUNTAI KIM

Samsung Electronics, Network Sys. Div. Internet Infra Team*

School of Electrical and Electronics Engineering

Yonsei University

134 Shinchon-dong, Seodaemun-gu, Seoul

KOREA

etkim@yonsei.ac.kr <http://cilab.yonsei.ac.kr>

Abstract: In this paper, we suggest the model for the context aware computing of the home network systems. Our model uses the Bayesian networks (BNs) and a new approach to structure learning of BNs based on genetic algorithm is proposed. The proposed method explores the wider solution space than the previous method which also uses the genetic algorithm for structure learning of BNs. In the previous method, while the ordering among the nodes of the BNs was fixed their conditional dependencies represented by the connectivity matrix was learned, whereas, in the proposed method, the ordering as well as the conditional dependency among the BN nodes is learned. To implement this method using the genetic algorithm, we represent an individual of the population as a pair of chromosomes: The first one represents the ordering among the BN nodes and the second one represents their conditional dependencies based on the ordering of the first chromosome. To implement proposed method new crossover and mutation operations which are closed in the set of the admissible individuals are introduced. Finally, the proposed method is applied to the given context aware problem of home network system and its effectiveness is demonstrated.

Key-Words: Bayesian network, Genetic algorithms, Context awareness, Structure learning, A pair of chromosomes

1 Introduction

Recently, a home network system has received much attention from engineers as a futuristic field and has become the typical test bed of the ubiquitous computing. The prospect of home network system is bright and it can be applied to a diverse field of study. From the theoretic point of view, the context aware computing is the key issue in the home network system. It is aimed at recognizing the intention of the users from several sensors and providing the human friendly service to them, thereby upgrading the level of living [1]. In context aware computing, the uncertainty model which accommodates the probabilistic actions of the users is required. In this paper, we employ the Bayesian Networks (BNs) to model the probabilistic actions of the users.

Bayesian networks (BNs) are one of the best known formalisms to reason under uncertainty in Artificial Intelligence (AI). BN is a graphical model that denotes joint probabilistic distribution among variables of interest based on their probabilistic

relationships. The structure of BN is a directed acyclic graph, or abbreviated as DAG. In recent years, with the development of large-scale database system, BNs have become a popular knowledge representational scheme for probabilistic knowledge in data mining and knowledge discovery [2].

The problem of searching the structure of BN that best reflects the conditional dependencies in a database of cases is a difficult one because even a small number of nodes to connect lead to a large number of possible directed acyclic graph structures. To solve this problem, several methods have been reported [3], [4], [5] and the genetic algorithms (GAs) are certainly among them [6]. GAs has been successfully applied to various optimization problems in real world and it is well suited to explore complex search spaces. But the GA is not ready made optimization tool and the plain GA cannot be applied to the structure learning of BN because the plain crossover and mutation operations are not closed (not allowed) in the set of the BN structure individuals.

In this paper, we suggest a new genetic approach for structure learning of BNs from a database of cases. The proposed method searches the wider solution space than the previous method which also uses the genetic algorithm for structure learning of BNs.

This paper is organized as follows. In Section 2, a brief introduction to BNs, GAs and home network system is given. In Section 3, a new algorithm for structure learning of BNs based on genetic algorithm is presented and new genetic operations are introduced. In Section 4, the proposed method is applied to a set of human behavior data in home network system and its validity is verified through computer simulation. Finally, some conclusions are drawn in Section 5.

2 Preliminaries: Bayesian Networks, Genetic Algorithms and Home Network System

2.1 Home network system

Home network system is a typical test bed of the ubiquitous computing and sensor network and is expected to upgrade the quality of living over the coming ten years. From the implementation point of view of the home network, the context aware computing is the key issue and has received attention from engineers and scientists.

The context aware computing is a jargon coined by the computer scientists. It is aimed at 1) *gathering the information about the users from the various sensors distributed in the smart home*, 2) *recognizing the intention of the users* and 3) *providing them with human friendly services*.

It is assumed that the home appliances and the devices have information about the circumstances under which they operate and can react according to the intention of the users.

In recent years, some works have been reported on extracting a variety of contexts such as user activities and the state of the environment [1]. However, the context-awareness is not fully understood and the tools and techniques for developing the context-aware applications are still in their infancy. BNs are one of the possible tools for the context aware computing.

2.2 Bayesian Networks

Bayesian networks and associated schemes constitute a probabilistic framework for reasoning under uncertainty and recently have gained popularity in the community of AI. From the structure point of view, BNs are directed acyclic graphs (DAGs), where the nodes are random variables, and the arcs specify the dependency relationships between the random variables [7].

A Bayesian network is composed of a network structure and a set of parameters associated with the structure. In general, the structure consists of nodes which are connected by directed arcs and form a directed acyclic graph. Each node represents a domain variable that can take on a finite set of values. Each arc represents a dependency between two nodes.

To specify the probability distribution of a BN, one must give prior probabilities for all root nodes (nodes with no predecessors) and conditional probabilities for all other nodes, given all possible combinations of their direct predecessors. These numbers in conjunction with the DAG specify the BN completely. The joint probability of any particular instantiation of all variables in a BN can be calculated as follows:

$$P(U) = \prod_i P(A_i | pa(A_i)) \quad (1)$$

where $U = \{A_1, A_2, \dots, A_n\}$ represents n nodes of a BN and $pa(A_i)$ is the parent set of the variable A_i .

For example, the probability of the whole network shown in Fig. 1 is given by following equation.

$$P(A_1, A_2, \dots, A_5) = P(A_1)P(A_2)P(A_3)P(A_4 | A_1, A_2)P(A_5 | A_2, A_3) \quad (2)$$

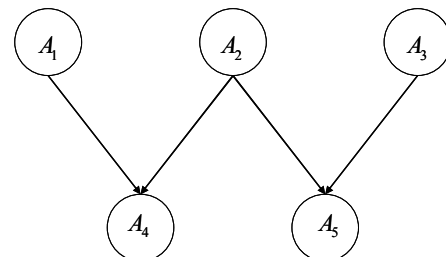


Fig. 1. An example of BN

The process of constructing BNs is called Bayesian networks learning. Learning Bayesian networks can be separated into two tasks, structure learning and parameter learning. Structure learning is to create an appropriate structure for BNs which accommodates prior knowledge from samples of data. Parameter learning is to calculate the conditional probabilistic

distribution for the given BN structure and is done by samples of data. The most popular parameter learning method is EM algorithm [8]. In this paper we focus upon structure learning of the BNs and create an appropriate structure for the BNs which best reflects the samples of given data.

2.3 Genetic Algorithms

Genetic algorithms (GAs) are based on a biological metaphor: They view learning as a competition among a population of evolving solution candidates. A “fitness” function evaluates each solution to decide which one will contribute to the next generation of solutions. Then, through operations analogous to gene transfer in sexual reproduction, a new population of candidate solutions is created [9].

The general procedure for GA is summarized as follows: An initial population of individuals (or chromosomes) is created by random selection. Next, a fitness value is assigned to each chromosome, depending on how close it actually is to solve the problem, thus arriving at the answer of the given problem. Each chromosome is a kind of solution of the problem. These “solutions” are not to be confused with “answers” of the problem, think of them as possible candidates that the system would employ in order to reach the answer. Those chromosomes with high fitness values are more likely to reproduce offspring which can mutate after reproduction. The offspring is the mixture of the father and mother and consists of a combination of their genes. This process is known as crossover. If the new generation contains a solution that produces a satisfactory performance or fitness, then the algorithm stops and the problem is said to be solved. If this is not the case, then the algorithm will be repeated until the condition is satisfied.

3 Structure Learning of Bayesian Network using Genetic Algorithms

In this section, a new approach for structure learning of BNs based on genetic algorithm is proposed. The proposed method searches the wider solution space than the previous method which also uses the genetic algorithm for structure learning of BNs.

3.1 The previous method

In this subsection, the previous genetic method for structure learning of BNs is explained briefly [6]. The BN structure with n variables is represented by an $n \times n$ connectivity matrix C of which element c_{ij} is represented as

$$c_{ij} = \begin{cases} 1 & \text{if } i \text{ is a parent of } j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

and each individual of the population is encoded as a chromosome:

$$c_{11}c_{12} \cdots c_{1n}c_{21}c_{22} \cdots c_{2n} \cdots c_{n1}c_{n2} \cdots c_{nn} \quad (4)$$

With this representation, the general crossover and mutation operators would produce the illegal BN structures [6]. In the previous method, to overcome this problem, the connectivity matrix was confined to being triangular and the chromosome was encoded as

$$x = c_{12}c_{13} \cdots c_{1n}c_{23}c_{24} \cdots c_{2n} \cdots c_{n-2n-1}c_{n-2n}c_{n-1n} \quad (5)$$

for the connectivity matrix

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix} \quad (6)$$

3.1 The proposed method

In the previous method, the ordering between the nodes was predetermined and their connections were confined to being triangular. So, the search space was not fully explored to find the fittest structure of BNs. If we do not have the prior knowledge about the ordering between variables (it is usual in most of the actual problems), the resulting structure will be not satisfactory. In this paper, we propose a new genetic method which can be applied to the structure learning of the BNs. In the proposed method, the connectivity matrix is assumed to be upper triangular, but it is not exactly the same as that of the previous method. The ordering of the nodes is not fixed and it is also learned by the GA. Thus, the entire solution space is searched for the fittest structure of the BN.

To implement this strategy, a BN structure is represented as a pair of chromosomes (ordering chromosome and the connection chromosome): The first one denotes the ordering between the BN nodes and the second one is the connectivity matrix which does not need to be triangular. The ordering

chromosome consists of indices of variables. If there are m root nodes (nodes with no parents), the first m genes of the chromosome are the indices of the root nodes. The next genes are the indices of the children nodes of the root nodes without overlaps. This structure is repeated until the whole variables show up. The connection chromosome denotes the connectivity matrices which defines the dependency relation between the variables. The ordering chromosome X_o and the connectivity chromosome X_c are encoded as follows:

$$X_o = x_1 x_2 \dots x_n \quad (7)$$

$$X_c = c_{12} c_{13} \dots c_{1n} c_{23} c_{24} \dots c_{2n} \dots c_{n-2n-1} c_{n-2n} c_{n-1n}$$

where

x_i is an integer and $x_i \in [1, n]$, $x_i \neq x_j$ iff $i \neq j$

$$C = \begin{pmatrix} 0 & c_{12} & c_{13} & \dots & c_{1n-1} & c_{1n} \\ 0 & 0 & c_{23} & \dots & c_{2n-1} & c_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & c_{n-1n} \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix} \quad (8)$$

$$c_{ij} = \begin{cases} 1 & \text{if } x_i \text{ is a parent of } x_j \\ 0 & \text{otherwise} \end{cases}$$

For the evolution of the connectivity chromosome, the plain GA operators are applied. For the evolution of the ordering chromosome, however, the plain GA operators do not work and some genetic operators used in traveling salesman problem (TSP) are applied to find the best ordering of the variables [9]. In the proposed method, the GA operators similar to those of [10], [11] are used. The following Example 1 shows a simple example of the suggested crossover operation.

Example 1

Consider the two BNs in Fig. 2. Assume that the crossover operation is applied to the ordering chromosome at the point between the second and the fifth gene. The parent ordering chromosomes are

$$x_1 = 126354 \quad (9)$$

$$x_2 = 543126$$

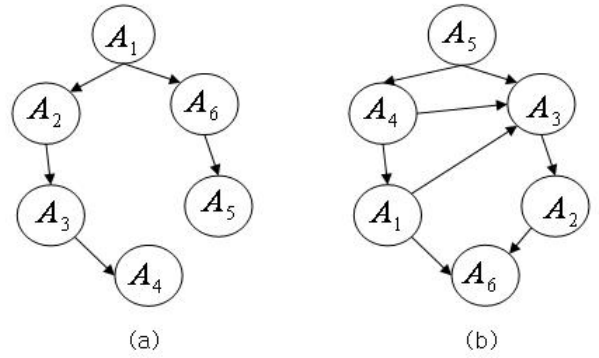


Fig. 2. Two parent BNs

The offspring chromosomes are

$$x'_1 = xx63xx \quad (10)$$

$$x'_2 = xx31xx$$

Since x'_1 contains “6” and “3,” we delete the two digits from x'_1 and the remaining sub-chromosome is “54” and “12.” The concatenation of two parts yields an offspring ordering chromosome $x'_1 = 546312$. In the same way, another offspring ordering chromosome becomes $x'_2 = 263154$.

$$x'_1 = 546312 \quad (11)$$

$$x'_2 = 263154$$

The offspring BNs are depicted in Fig. 3.

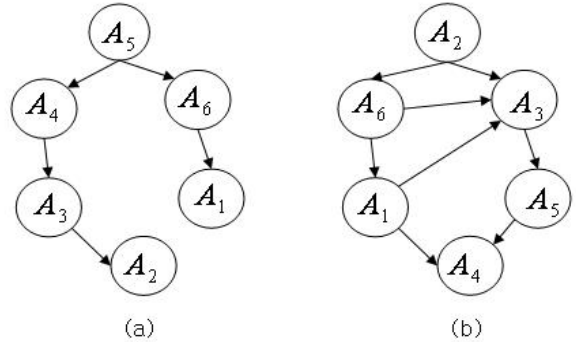


Fig. 3. Offspring BNs after crossover

Next, let us consider the mutation. In the mutation, randomly chosen two genes of the ordering chromosome are exchanged.

Example 2

Consider a BN structure of Fig. 4(a) which consists of three variables. Using the proposed representation, the structure is represented by

$$X_o = 123 \quad (12)$$

$$X_c = 110$$

Suppose that the first and the second genes of the ordering chromosome are exchanged by mutation and the resulting one becomes

$$\begin{aligned} X_o &= 213 \\ X_c &= 110 \end{aligned} \tag{13}$$

Fig 4(b) depicts the resulting BN. Like the crossover, the mutation is also closed in the set of BNs.

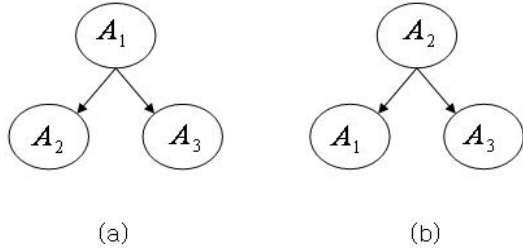


Fig. 4. BNs before and after mutation

4 Simulation Results

In this section, the proposed method is applied to context aware computing of the home network systems. We developed a BN to estimate the context of a single user who lives in the home network system. Forty two random variables are used to implement the context awareness BN. Twenty one variables indicate the state of the sensors embedded in home appliances such as TV, light of living room or refrigerator, etc. The twenty one variables also include the useful information such as the time and the location of the user. The other twenty one variables represent the activities of a user, for example, he (or she) is studying, he (or she) is watching a TV, or he (or she) is sleeping.

To collect the data, we developed a web based virtual action simulator. The context aware simulator was developed in FLASH and C and the behaviors of a single user over six days were collected. The proposed method is applied to learn the structure of the BN in such a way that the BN reflects the probabilistic relationships between the variables as well as possible. Shown in Fig. 5 is an example of BN structure for the data.

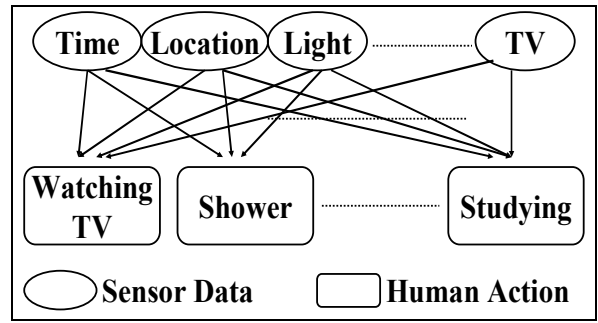


Fig. 5. An example of BN structure for a Home network system

To apply the new genetic encoding and operations and evaluate the effectiveness of the structure of the given BN, we need theorem in [12]. This theorem means that given a database D the probability of the certain BN structure B_s with the database D is represented by the following equation (14).

$$P(B_s, D) = P(B_s) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \tag{14}$$

Therefore, in this paper, we use

$$P(D | B_s) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \tag{15}$$

as a fitness function to evaluate the structure of the given BN.

Fig. 6 compares the fitness of the previous and the proposed methods. The vertical axis denotes the minus logarithm of (15) and the horizontal axis denotes the number of generations. So as the BN is close to the given database, the probability (15) goes to 1 and the minus logarithm of (15) goes to zero. It can be seen that the proposed method starts to outperform the previous method after only two generations and keeps much outperforming for the next almost one hundred generations.

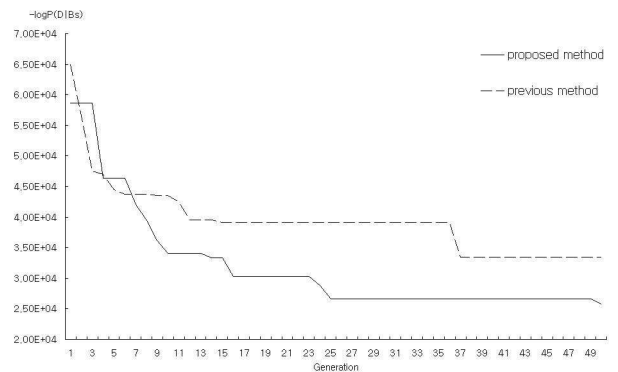


Fig. 6. Comparison of the performances of the previous and proposed methods

The reason for the excellence of the proposed method might be that wider search space is explored compared with the previous method.

5 Conclusion

In this paper, a new approach for the structure learning of Bayesian Networks from databases of cases has been suggested. In the proposed method, the connectivity matrix of the BN is not confined to being triangular and wider solution space is explored by encoding the BN as a pair of ordering chromosome and connection chromosome. The proposed method was applied to the context aware computing of the home network system and the effectiveness of the proposed method is verified through the simulation.

Acknowledgments

This work was supported by the Ministry of Commerce, industry and Energy of Korea (HISP).

References:

- [1] P. Korpipää, M. Koskinen, J. Peltola, S. Mäkelä, T. Seppänen, "Bayesian Approach to Sensor-Based Context Awareness," *Personal and Ubiquitous Computing J.*, vol. 7, no. 4, 2003
- [2] F.V. Jensen, "Introduction to Bayesian Networks," Technical Report IR 93-2003, Dept. of Mathematics and Computer Science, Univ. of Aalborg, Denmark, 1993.
- [3] S. Acid, L.M. De Campos, A. Gonzalez, R. Molina, and N. Perez de la Blanca, "Learning with CASTLE," Symbolic and Quantitative Approaches to Uncertainty, R. Kruse and P. Siegel, eds., *Lecture Notes in Computer Science* 548. Springer-Verlag, 1991.
- [4] D.M. Chickering, D. Geiger, and D. Heckerman, "Learning Bayesian Networks: Search Methods and Experimental Results," *Fifth Int'l Workshop Artificial Intelligence and Statistics*, pp. 112-128, 1995.
- [5] G.M. Provan, "Model Selection for Diagnosis and Treatment Using Temporal Influence Diagrams," *Proc. Fourth Int'l Workshop Artificial Intelligence and Statistics*, pp. 469-480, 1995.
- [6] P. Larrañaga, M. Poza, Y. Yurramendi, R. Murga, and C. Kuijpers, "Structure Learning of Bayesian Network by Genetic Algorithms: A Performance Analysis of Control Parameters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 912-926, Sept. 1996.
- [7] F.V. Jensen, *Bayesian Networks and Decision Graphs*, Springer, 2001.
- [8] S.Z. Zhang, Z.N. Zhang, N.H. Yang, J.Y. Zhang, and X.K. Wang, "An improved EM algorithm for Bayesian networks parameter learning," *Machine Learning and Cybernetics*, vol. 3, pp. 1503 - 1508, Aug. 2004
- [9] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, 1999.
- [10] P. Larrañaga, C. Kuijpers, and R. Murga, "Learning Bayesian network structures by searching for the best ordering with genetic algorithms," *IEEE Trans. Systems, Man, Cybernetics, Part A*, vol. 26, pp. 487-493, July 1996
- [11] J. Lee, W. Chung and E. Kim, "Structure Learning of Bayesian Networks Using Dual Genetic Algorithm," *IEICE Trans. Inf. & Syst.*, 2007 (accepted to publish)
- [12] G.F. Cooper and E.A. Herskovits, "A Bayesian Method for the Induction of Probabilistic Networks from Data," *Machine Learning*, vol. 9, no. 4, pp. 309-347, 1992