

A Matrix-Based Genetic Algorithm for Structure Learning of Bayesian Networks

Song Ko¹, Daewon Kim¹ and Bo-Yeong Kang²

¹School of Computer Science and Engineering Chung-Ang University, 221 Heukseok-dong, Dongjak-gu, Seoul 156-756, Korea

²School of Mechanical Engineering Kyungpook National University, 1370 Sankyuk-dong, Buk-gu, Daegu 702-701, Korea

Abstract

Unlike using the sequence-based representation for a chromosome in previous genetic algorithms for Bayesian structure learning, we proposed a matrix representation-based genetic algorithm. Since a good chromosome representation helps us to develop efficient genetic operators that maintain a functional link between parents and their offspring, we represent a chromosome as a matrix that is a general and intuitive data structure for a directed acyclic graph(DAG), Bayesian network structure. This matrix-based genetic algorithm enables us to develop genetic operators more efficient for structuring Bayesian network: a probability matrix and a transpose-based mutation operator to inherit a structure with the correct edge direction and enhance the diversity of the offspring. To show the outstanding performance of the proposed method, we analyzed the performance between two well-known genetic algorithms and the proposed method using two Bayesian network scoring measures.

Keywords: Bayesian Network, Genetic Algorithm, Structure Learning, Genetic Operators.

1. Introduction

Bayesian network (BN) [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] is a directed acyclic graph (DAG) that represents a set of variables of interest based on their conditional dependencies. One of the popular methods for Bayesian structure learning is the K2 algorithm [2], but the K2 algorithm assumes that variable ordering is given a priori. In most of the real-world problems, it is usual that we do not have the prior knowledge about variable ordering, and furthermore this approach explores limited search space due to the simple greedy strategy. In this case, a Genetic Algorithm (GA) can present better structure through a more extensive searching space since a GA has few assumptions about searching strategy except natural selection [11].

The first GA approach for the structure learning is introduced by Larranaga et al. at 1996 [12]. They proposed

a sequence-based notation to represent the BN structure in GA formulation. It constructs the $n \times n$ connectivity matrix (C) where n is the number of variables. Each element (c_{ij}) is defined as follows:

$$c_{ij} = \begin{cases} 1 & \text{if a variable } i \text{ is a parent of a variable } j, \\ 0 & \text{otherwise.} \end{cases}$$

For example, the $n \times n$ connectivity matrix C can be represented as:

$$C = \begin{pmatrix} 0 & c_{12} & c_{13} \\ c_{21} & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

The C means a structure where a variable X_3 has X_1 and X_2 as parents. Thus, the chromosome for the structure is represented as the sequence like as followed:

$$C_{seq} = c_{11}c_{21} \cdots c_{n1}c_{12} \cdots c_{nn} = 001001000$$

Based on the chromosome, they employed a typical GA to infer the structure. Additionally they extended a GA method to search for better variable ordering [13]. Ross et al. presented a dynamic BN structure learning method to optimize the multiple objectives through a GA [15].

Manuscript received Aug. 3, 2011; revised Sep. 7, 2011; accepted Sep. 8, 2011.

This research was supported by Kyungpook National University Research Fund, 2009.

² Corresponding Author

The population is evolved with an assumption that every variable has M parents at most. Kabli et al. proposed a chain-model GA to search for better variable ordering [18]. The discrimination with the work of Ross et al. [13] is the evaluation method with a specific assumption that each variable has only a single parent which is just one precedent variable. Recently, Lee et al. proposed a new sequence-based genetic operator [19] where the chromosome is represented as a matrix and the matrix is divided into upper and lower triangular matrices; the two separated matrices are represented as sequences. Based on these sequences, a sequence-based crossover method is applied in two parts; crossover between the two separated sequences and crossover inside each separated sequence.

However, the existing GA methods for BN structure learning have common limitations. First, although a good chromosome representation can be a help to develop efficient genetic operators that maintain a functional link between parents and their offspring, and it is highly recommended to select an intuitive structure derived from the problem as a good representation [14], the sequence-based representation used in the conventional methods is not intuitive form to represent the BN, the DAG structure. Second, it is difficult to enhance the diversity of the offspring through previous reproduction methods that employ cut-point-based crossovers between chromosomes to create offspring. In case of two cut-point crossover, the possible cases of the offspring are 2^2 if the cut-point area is given. Considering the cut-point area is determined randomly, the possible cases become $n^2 2^2$ where the n is the number of variables. The problem is that the possible cases of the offspring are not large enough compared to the entire searching space [12].

Therefore, in this paper, we propose a matrix-based GA algorithm, which represents a chromosome as a matrix and thus, enables us to develop more efficient genetic operators to inherit a structure with the correct edge direction and enhance the diversity of the offspring.

2. The Proposed Method

In the proposed method, we denoted the chromosomes for the BN structure as the same matrix with a connectivity matrix as follows: we constructs the $n \times n$ matrix (C^k) for a k -th chromosome C^k , where n is the number of variables. Each element c_{ij} is defined as 1 if a variable i is a parent of a variable j , 0 otherwise. Unlike the previous works[12, 13, 15, 18, 19], our matrix does not go through the process of the sequence-based manipulation for a chromosome representation.

$$C^k = \begin{pmatrix} 0 & c_{12} & c_{13} \\ c_{21} & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{pmatrix}$$

Using the matrix-based representation, first, we designed a probability matrix (P) that indicates the probability distribution of each edge to link variables among several chromosomes, and represents how frequently each edge appears in the selected chromosomes.

Definition 2.1 (Probability Matrix). Let's suppose the following notations. λ represents the population size; $C = \{C^1, C^2, \dots, C^\lambda\}$ is the population; $\{T^1, T^2, \dots, T^t\}$ is a set of randomly selected chromosomes from C ($2 \leq t \leq \lambda$); n is the number of the variables. Based on these notations, we defined the P as:

$$P = \begin{pmatrix} 0 & P_{12} & \dots & P_{1n} \\ P_{21} & 0 & \dots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \dots & 0 \end{pmatrix},$$

where

$$P_{ij} = \frac{1}{t}(T_{ij}^1 + T_{ij}^2 + \dots + T_{ij}^t) \quad (1)$$

The P is constructed via the following two steps. First, it randomly selects t chromosomes among the population based on their probabilities which are calculated according to their ranking [12]. Second, it calculates the averages per element using the selected t chromosomes. Thus each P_{ij} ($0 \leq P_{ij} \leq 1$) represents the average strength of edge between variables i and j , which is observed from the selected t chromosomes.

The diversity of the offspring through the P is enhanced for the following three reasons. One is that it is possible to create a new offspring using more than two chromosomes. The number of chromosomes t used to construct the P is determined dynamically. Second is that the offspring is evolved with all elements of the matrix simultaneously, which is same as the evolution of all variables n . As a result, the possible cases of the offspring through the P become t^n . Third is that it creates offspring by probability based on the probability distribution. Enhancing searching space is a meaningful factor to infer the appropriate structure considering the limited search space, which depends on the initial population.

New genetic operators are designed to create offspring from P : a reproduction method substituting for the typical crossover, and transpose-based mutation. The reproduction method creates offspring based on the P . To create an element (O_{ij}) of an offspring O , it compares two probability values, i.e. P_{ij} and $rand[0, 1]$, as follows:

$$O_{ij} = \begin{cases} 1 & \text{if } P_{ij} \geq rand[0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

One offspring(O) is created after comparing all elements. We designed the proposed method to create one

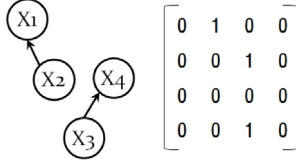


Figure 1: The optimal structure and its matrix representation.

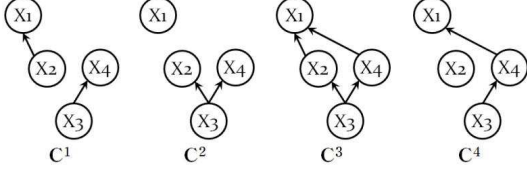


Figure 2: An initial population composed of four chromosomes (C^1, C^2, C^3, C^4)

offspring per one P to induct more enhanced diversity of the offspring. And it repeats until the number of reproduced offspring becomes λ . After applying mutation, we randomly eliminated edges, that is a popular repair operator in Bayesian network, that invalidated DAG if it does not met DAG condition[12].

Next, we introduce a new transpose-based mutation. Basically, we employed the previous mutation method which was changing the element of the matrix from 1 to 0 or vice versa. Additionally, we adjusted the transpose-based mutation on the reproduced offspring to enhance the possibility to infer the structure with the correct edge.

Definition 2.2 (Edge Type). The ‘correct edge’ means the edge of the same direction with the edge in the answer BN. The ‘reverse edge’ means that the edge of an inverse direction, and the ‘wrong edge’ means the edge that should be in the BN, but is deleted, or vice versa.

Several edges may be inferred as reverse edges since the score of the reverse edge is similar with the score of the correct edge due to the likelihood equivalence assumption[10]. To prevent the inferred structure having reverse edges, we proposed a transposed-based mutation to keep both of the edge directions within the population. Based on the enhanced diversity of the offspring, it can produce various combinations of the causalities and infer the optimal combinations of the causalities. It can also infer the optimal structure.

A main scheme of the proposed method is shown in Algorithm 1. And to illustrates the working process of the proposed method, we show how the proposed method can infer the optimal structure in Figure 1 between four variables, starting from the given population composed of four chromosomes in Figure 2. Let us denote the k -th parents and k -th offsprings by C^k and O^k respectively. The required parameters are set to be $n = 4$, $\lambda = 4$, and $t = 2$. Let

Algorithm 1 Main scheme of proposed method.

$D_{l=0} \leftarrow$ Generate λ individuals randomly for the first generation

repeat

$D_{l-1} \leftarrow$ Select $T \leq$ individuals from D_{l-1}

$p_l(x) = p(x|D_{l-1}) \leftarrow$ Estimate a joint probability distribution of selected T individuals

$D_l \leftarrow$ Sample λ individuals from $p_l(x)$

until for $l = 2 \rightarrow \dots$

us assume that the t chromosomes are selected as $T^1 = C^1$ and $T^2 = C^3$.

Step 1: The initial population is represented as follows:

$$C^1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad C^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$C^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad C^4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Step 2: Constructs P by the selected $T^1 = C^1$ and $T^2 = C^3$.

$$P = \frac{1}{2}(T^1 + T^2) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Step 3: Creates the k -th new offspring O^k (initially, $k = 1$).

$$O_{ij}^k = \begin{cases} 1 & \text{if } P_{ij} \geq \text{rand}[0, 1] \\ 0 & \text{otherwise.} \end{cases}$$

Step 4: Alter O^k by the typical mutation (flipping $0 \rightarrow 1$ or $1 \rightarrow 0$)

$$O_{ij}^k = \begin{cases} (O_{ij}^k + 1) \% 2 & \text{if mutation rate} \geq \text{rand}[0, 1] \\ O_{ij}^k & \text{otherwise.} \end{cases}$$

Step 5: Alter O^k by the proposed transpose-based mutation

$$O^k = \begin{cases} [O^k]^T & \text{if mutation rate} \geq \text{rand}[0, 1] \\ O^k & \text{otherwise.} \end{cases}$$

Step 6: Iterates Step.2-5 until $O = \{O^1, \dots, O^\lambda\}$ is generated.

Here, when we assume that the all $\text{rand}[0, 1]$ are 0.5, then we can get the first ($k=1$) new offspring:

$$O^1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (2)$$

Moreover, if the O^1 is transposed in the Step 5, then the O^1 becomes to be the optimal structure in Fig. 1:

$$O^1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

From this simple example, we see that the proposed method can easily infer the optimal structure.

3. Experiments

To show the effectiveness of the proposed method, we compared the performance of the proposed method with the method of Larranaga et al. [12] and the method of Ross et al. [15] based on 10 repeated experiments. Three well-known data sets are employed. The Asia data set, which structure is depicted in Fig. 5(a), is a small piece of fictitious qualitative medical knowledge related to the shortness of breath (dyspnoea). The ALARM(A Logical Alarm Reduction Mechanism) data set is constructed to provide an alarm message of a state of a patient using 37 variables. The Car-start data set is designed to represent the car start problem with 18 variables[23]. We generated 10,000 patterns, respectively, based on the optimal structure with conditional probabilities. The parameters for GA were set to be 50 chromosomes, 1,000 generations, and 0.2 mutation rate; since the performances on mutation rate(= 0.2) are generally outperformed among several mutation rates.

3.1 Performance comparison by scoring measure

We compared the performance of each method using well-known two scoring measures: the Bayesian score and Bayesian Information Criterion (BIC). Table 1 shows the Bayesian scores for the inferred structures through the methods on the three data sets. A lower value of the Bayesian score represents better BN structure. We can see that the proposed method outperforms the previous methods entirely. Especially, the scores of the Asia data and Car-start data are nearly approximated to the optimal scores; a difference with one score and a difference with nine scores, respectively. However, we can see that the previous methods inferred structures which had lower scores. Among the results of the previous methods, the method of the Ross et al. inferred a structure which score was similar with the optimal score at the Asia data.

Table 2 shows the BIC scores for the inferred structures through the methods on the three data sets. We can see that the previous methods inferred structures which had lower scores. About 20 - 30 scores are different with the optimal scores of the Asia data and Car-start data. However, the proposed method improved the performance at the three data sets. Especially, the score of the proposed method is

Table 1: Performance comparison of the Bayesian score by each method for three data sets. The asterisk(*) represents the score of the optimal structure for each data set.

Method	Asia (-22,405*)	ALARM (-95,188*)	Car-start (-23,149*)
Larranaga	-22,428	-98,028	-23,181
Ross	-22,412	-98,463	-23,176
Proposed	-22,406	-96,574	-23,158

Table 2: Performance comparison of the BIC score by each method for three data sets. The asterisk(*) represents the score of the optimal structure for each data set.

Method	Asia (* -22,421)	ALARM (* -95,965)	Car-start (* -23,319)
Larranaga	-22,447	-101,778	-23,410
Ross	-22,454	-99,882	-23,389
Proposed	-22,424	-98,418	-23,316

approximated to the optimal score at the Asia data. In here, additional noticeable point is that the score of the proposed method overtook the optimal score at the Car-start data. This result comes from the likelihood equivalence of the scoring measures.

3.2 Performance comparison by edge counting

We analyzed the experimental results through edge counting-based evaluation, which was used in the experiments of Campos and Blanco et al. for the BN structure learning [10, 21]. In general, it is difficult to figure out how perfectly the proposed method inferred the structure through the conventional score-based evaluations since it is not easy to interpret the meaning of their scores. Therefore, we additionally evaluated the inferred structures by an intuitive method how many edges are correctly and wrongly inferred. This proposed simple and intuitive evaluation is performed by counting the number of inferred edges according to the three edge types; correct edges, reverse edges and wrong edges.

Figure 3 shows the comparison results by edge counting among the three methods for the Asia data. Above all, we can see that the proposed method outperformed the previous methods at all edge types. The previous two methods commonly inferred below half of the correct edges and about half of reverse edges. Furthermore, the previous methods inferred several wrong edges. On the other hand, the proposed method inferred a structure with almost correct edges; there is a little wrong and reverse edges. Obviously, we can see that inferring a structure with correct edges reduces the number of reversed edges and wrong edges.

3.3 Analysis according to the mutation types

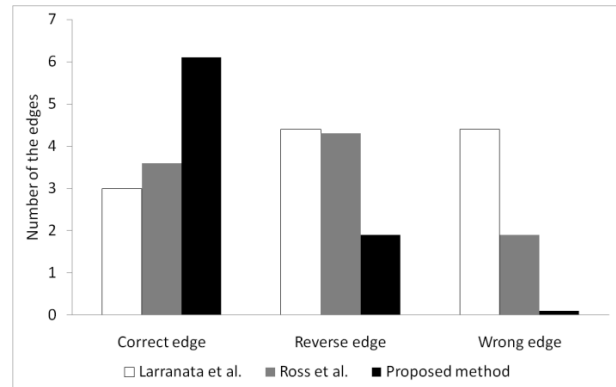
The proposed method contains two types of mutations: the typical flipping mutation and the transpose-based mutation. In this section, we present that the correct edge direction can be found by transpose-based mutation through an additional experiment with three cases: (1) only typical mutation used, (2) only transpose-based mutation used, and (3) both mutations used. The structure learning progressed with the Bayesian scoring metric.

Figure 4 shows the number of inferred edges by the proposed method using three-cases of mutations for the Asia data. Case 1 showed that the typical flipping mutation has a limitation to improve the number of the correct edges and to reduce the number of the reverse edges and wrong edges. However, we can see that the proposed method with transpose-based mutation only (case 2) reduced the number of the reverse edges and it means that the number of the correct edges was increased. Finally, the proposed method with both mutations (case 3) shows the best evaluation result with the most correct edges and least reverse edges. From the results, it is observed that the inferred structure via the proposed method can represent the better relationship between variables with a few wrong relationship.

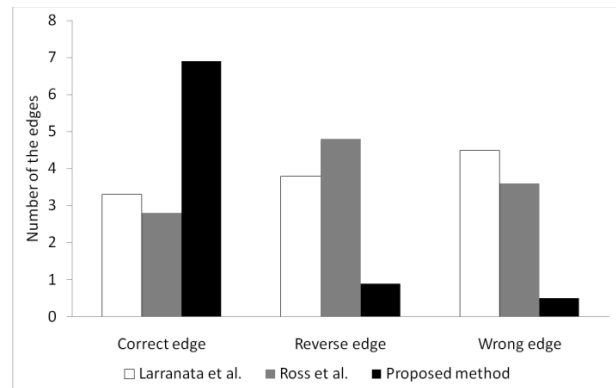
3.4 Evolution progresses of the three methods

To visualize the evolution progress of the compared methods, we depicted the intermediate inferred structures obtained at four generation point from 1,000 generations for the Asia data; the evolution progress of the Larranaga et al. at 10th, 20th, 30th, 40th generations are shown, and the progress of the Ross et al. and the proposed method are shown at 25th, 50th, 75th, 100th generations. Figure 5 shows the optimal structure for the Asia and the elitist chromosome of the 1st population. Figures 6-8 show the evolution progresses of the three methods respectively. In each figure, a solid line represents a correct edge (C), a dotted line represents a reversed edge (R), and the broken-dotted line represents a wrong edge (W). The initial populations of all methods are identical. Among them, the elitist chromosome is depicted in Figure 5.(b) which contains two correct edges, two reversed edges and 11 wrong edges which are summed of four non-inferred correct edges and seven faulty-inferred edges. And the score of the elitist chromosome of the first population is $-23,462$.

Starting from this common first generation, the progress of all the methods are depicted in Figure 6.(a)~(c). We can see that both of the previous methods have a common limitation to increase the number of the correct edges and to reduce reverse edges. Although both of the previous methods increased scores during the evolutions, the reason is from the deletion of the incorrectly inferred edges. Moreover, the two previous methods can not find the correct edges if



(a) The result after inferring through the Bayesian score.



(b) The result after inferring through the BIC score.

Figure 3: Performance comparison by edge counting for the Asia data.

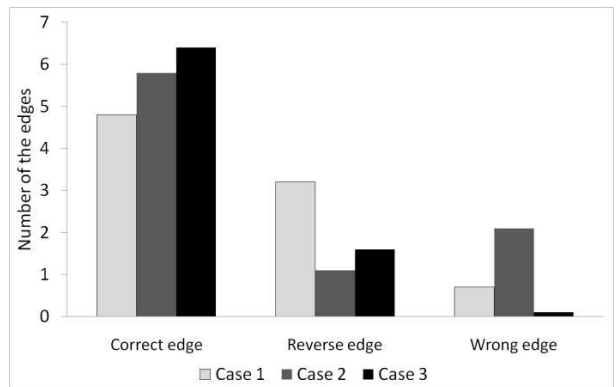
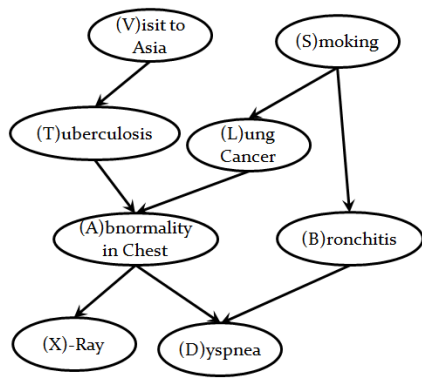
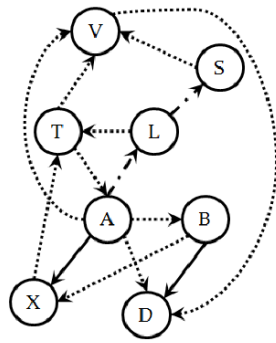


Figure 4: Number of inferred edges by the proposed method using three-cases of mutations for the Asia data. (1) Case 1: Only typical mutation used, (2) Case 2: Only transpose based mutation used, (3) Case 3: Both mutations used.

there were inferred with the reverse edges. On the other hand, we can see that the proposed method inferred the structure with the most correct edges. The noticeable point is that the inferred structure via the proposed method has



(a) The structure for the Asia data



The first generation
Bayesian score = -23,462
C = 2, R = 2, W = 11

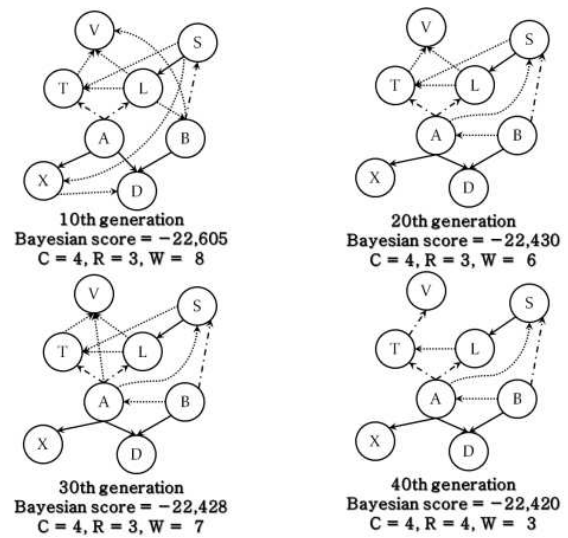
(b) The structure of the elitist chromosome

Figure 5: The optimal structure for the Asia and the elitist chromosome of the first population.

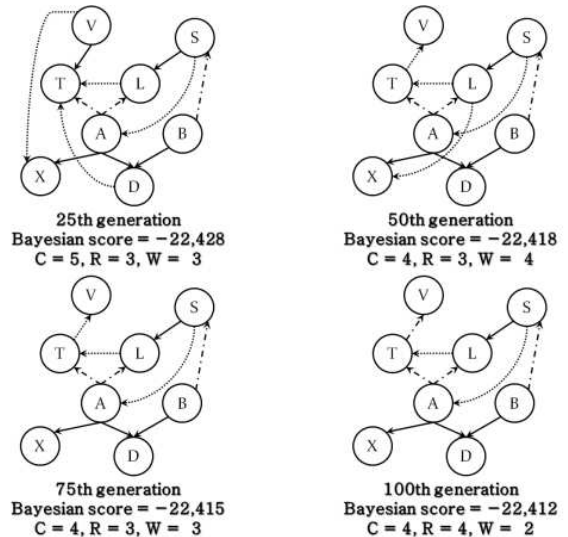
only one reverse edge, which is a relation between (S)moke and (L)ung cancer without any wrong edges. However, it is difficult to infer the reverse edge to be correct since the scores of the correct edge and the reverse edge between the two variables are the same. As a result, the score of the inferred structure is equal to the score of the optimal structure.

4. Conclusions

In this paper, we showed that the matrix based genetic operators outperforms the previous methods. We think that the transpose based mutation inducts to infer a better structure which has a higher score as well as more many number of the correct edges. Although the proposed method inferred much better structure in small size data such as Asia and Car datasets, it does not inducts significant results at the large size data sets such as the ALARM dataset. It seems that there is too extensive search space since the transpose based mutation makes too many possible cases, especially, on the large data. Therefore, our further research issue is a separating method of the large data into several smaller part.



(a) The evolution process of Larranaga et al.



(b) The evolution process of Ross et al.

References

- [1] Judea Pearl, *Probability Reasoning in intelligent systems; Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, California, 1989.
- [2] Gregory F.Cooper and Edward Herskovits, "A Bayesian Method for the Induction of Probabilistic Networks from Data," *Machine Learning*, vol.9, no.4, pp.309-347, 1992
- [3] Steffen L.Lauritzen and David J.Spiegelhalter, "Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems," *Journal of the Royal Statistical Society. Series B(Methodological)*, vol.50, no.2, pp.157-224, 1988.

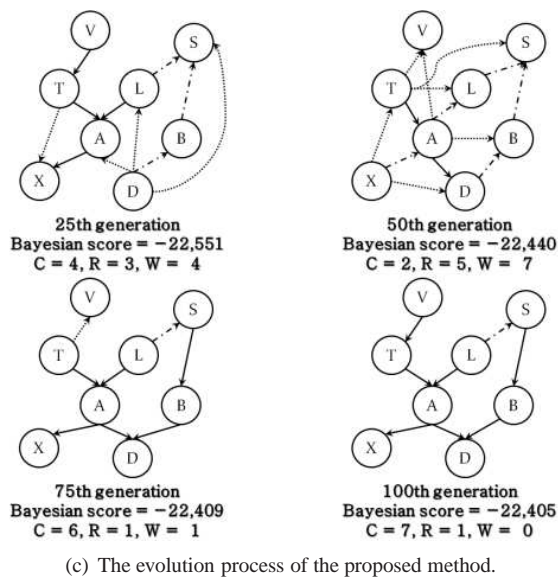


Figure 6: The evolution progress according to the methods. The solid line means correct edge, the dotted line means reverse edge and the broken-dotted line means wrong edge.

[4] Edward Herskovits and Gregory Cooper, "Kutato: An Entropy-Driven System for Construction of Probabilistic Expert Systems from Databases," *Report KSL-90-22, Knowledge Systems Laboratory, Medical Computer Science, Stanford Univ*, 1990.

[5] Remco R. Bouckaert, "Probabilistic Network Construction Using the Minimum Description Length Principle," *Lecture Notes in Computer Science*, pp.41-48, 1993.

[6] Remco R. Bouckaert, "Properties of Bayesian Belief Networks Learning Algorithms," *Tenth Conference on Uncertainty in Artificial Intelligence*, pp.102-109, 1994.

[7] David M. Chickering, Dan Geiger and David Heckerman, "Learning Bayesian Networks: Search Methods and Experimental Results," *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, pp.112-128, 1995.

[8] David Heckerman, Dan Geiger and David M. Chickering, "Learning Bayesian Networks: The Combination of Knowledge and Statistical data," *Machine Learning*, vol.50, pp.95-126, 1995.

[9] Wai Lam and Fahiem Bacchus, "Learning Bayesian Belief Networks. An Approach Based on the MDL Principle," *Computational Intelligence*, vol.10, no.3, pp.269-293, 1994.

[10] Luis M. de Campos, "A Scoring Function for Learning Bayesian Networks based on Mutual Information

and Conditional Independence Tests," *Journal of Machine Learning Research*, vol.7, pp.2149-2187, 2006.

[11] John H. Holland, *Adaptation In Natural and Artificial Systems*, University of Michigan Press, 1975.

[12] Pedro Larranaga, Mikel Poza, Yosu Yurramendi, Roberto H. Murga and Cindy M. H. Kuijpers, "Structure Learning of Bayesian Networks by Genetic Algorithms: A Performance Analysis of Control Parameters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.18, no.9, pp.912-926, 1996.

[13] Pedro Larranaga, Cindy M. H. Kuijpers, Roberto H. Murga and Yosu Yurramendi, "Learning Bayesian Network Structure by Searching for the Best Ordering with Genetic Algorithms," *IEEE Transactions on System, Man, and Cybernetics - PART A: Systems and Humans*, vol.26, no.4, pp.487-493, 1996.

[14] Zbigniew Michalewicz and David B. Fogel, *How to Solve It: Modern Heuristics*, Springer Verlag, 2010.

[15] Brian J. Ross and Eduardo Zuviria, "Evolving Dynamic Bayesian Networks with Multi-objective Genetic Algorithms," *Applied Intelligence*, vol.26, no.1, pp.13-23, 2005.

[16] Kevin Patrick Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning*, Doctor of Philosophy of the University of California, 2002.

[17] Abdullah Konak, David W. Coit, Alice E. Smith, "Multi-objective optimization using genetic algorithms: A tutorial," *Reliability Engineering and System Safety*, vol.91, pp.992-1007, 2006.

[18] Ratiba Kabli, Frank Herrmann and John McCall, "A Chain-Model Genetic Algorithm for Bayesian Network Structure Learning," *Proceedings of the 9th annual conference on Genetic and evolutionary computation (GECCO'07)*, pp.1264-1271, 2007.

[19] Jaehun Lee, Wooyong Chung, Euntai Kim and Soohan Kim, "A New Genetic Approach for Structure Learning of Bayesian Networks: Matrix Genetic Algorithm," *International Journal of Control, Automation and Systems*, vol.8, no.2, pp.398-407, 2010.

[20] Shummet Baluja, "Population-Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning," *Carnegie Mellon University, Pittsburgh, PA*, 1994.

[21] Rosa Blanco, Inaki Inza and Pedro Larranaga, "Learning Bayesian Networks in the Space of Structure by Estimation of Distribution Algorithms," *International Journal of Intelligent Systems*, vol.18, pp.205-220, 2003.

- [22] Shulin Yang and Kuo-Chu Chang, "Comparison of Score Metrics for Bayesian Network Learning," *IEEE Transactions on System, Man, and Cybernetics - PART A: Systems and Humans*, vol.32, no.3, pp.419-428, 2002.
- [23] Finn V. Jensen, *Bayesian Networks and Decision Graphs*, Springer Verlag, 2007.



Daewon Kim

Professor, School of Computer Science and Engineering at Chung-Ang University

Research Area: Artificial Intelligence, Data Mining, Biomedical Informatics, Affective Computing.

E-mail : dwkim@cau.ac.kr



Song Ko

Ph.D candidate of the Chung-Ang University

Research Area: Artificial Intelligence, Data Mining, Knowledge Representation, Bayesian Networks.

E-mail : sko22.cau@gmail.com



Bo-Yeong Kang

Professor of the Kyungpook National University

Research Area: Machine Intelligence Technology, Robotics, Biomedical System.

E-mail : kby09@knu.ac.kr