

Root Node Vaccines for Bayesian Network Structure Learning based on Immune Algorithm

Zhiqiang Cai^{1,a}, Shudong Sun¹, Shubin Si¹ and Ning Wang¹

¹Ministry of Education Key Laboratory of Contemporary Design and Integrated Manufacturing Technology, School of Mechatronics, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China

^acaizhiqiang@nwpu.edu.cn

Keywords: Bayesian network, modeling, root node, vaccine, immune algorithm

Abstract. To facilitate the application of Bayesian network in engineering fields, learning proper structure from dataset is one of the most efficiency Bayesian network modeling technique. In this paper, the description and characteristics of Bayesian networks and immune algorithms are discussed at first. Then, the extraction method of root node vaccines is proposed to accelerate the model structure learning process. Thirdly, the immune algorithm based method is also applied to search the best Bayesian network structure. Finally, the simulation studies based on a car start BN model are carried out and the results verify that the proposed Bayesian network structure learning method can build the objective structure from dataset more effectively and more efficiently with the root node vaccines.

Introduction

Bayesian network (BN) is a directed acyclic graph (DAG) used to represent uncertain knowledge in artificial intelligence [1]. With the advantages of probability descriptions and conditional independences, BN has provided a comprehensive method for representing variable state distribution and variable relationships. So it has been widely applied in various fields to solve practical problems. Because building the objective BN model with expert experience is not easy, learning the practical model from dataset has attracted considerable attention recently [2]. The BN modeling process usually consists of two parts: learning the BN structure which represents the conditional independent relationships, and learning the BN parameters which specify the conditional probability distributions of BN.

The key problem of learning objective BN structure from dataset is to find the most proper network structure which could represent the potential relationships in dataset accurately. Since learning the BN structure from dataset is a NP-hard problem for large networks, the conditional independence tests based algorithms and the score and search based algorithms have been proposed separately to settle this challenge [3]. The first method can discover the potential conditional independence relationships of nodes from dataset with conditional independence test equation, and then the whole Bayesian network is built on the basis of such relationships [4]. In the score & search based methods, a score function is introduced as the criterion to represent how the candidate network structure fits the dataset while a searching algorithm is applied to find the best structure with the highest score in all candidate network structures. The common score functions are Copper-Herskovits function [5], Bayesian Information Criterion (BIC) function [6] and Minimum Description Length (MDL) [7]. They are proved to be effective in BN structure learning in many works. Using the searching algorithms, such as genetic algorithm (GA) [8], evolutionary programming [9], ant colony optimization [10], the candidate BN structure is usually encoded as an ordered string or a connection matrix while different operators have been designed and employed to find the one with the highest scores.

Such algorithms mentioned above have shown appropriate performance in simulations and practical applications. But with the random heuristic searching process, the best structure doesn't fit the dataset well sometimes. In the paper, we will introduce the root node vaccines into the BN structure learning process to search the optimal solution.

Bayesian Networks and Root Node Vaccines

Bayesian Networks. The BN is denoted as $\langle \mathbf{X}, \mathbf{A}, \mathbf{P} \rangle$. The $\mathbf{X} = \{X_1, X_2, X_3, \dots, X_n\}$ is the set of nodes which reflect stochastic variables in practice, where the states of nodes correspond with the states of variables. The $\mathbf{A} = \{[a_{ij}], 1 \leq i \leq n, 1 \leq j \leq n\}$ represents the set of edges which connect all the nodes in BN, where a_{ij} indicates an edge from X_i to X_j . The $\mathbf{P} = \{P(X_i | \pi(X_i)), X_i \in \mathbf{X}\}$ describes the conditional probability distributions (CPT) of the nodes in BN, where $\pi(X_i) \subseteq \{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n\}$ is a set of all the father nodes of X_i . For the node without father node, its CPT is the prior probability distribution.

As the qualitative information of associate relationship is represented by the topology of network structure of BN, the quantitative information of the dependencies is described with the conditional probability distributions.

Immune Algorithm (IA). Because the crossover and mutation operators of GA can only change individuals randomly and indirectly during the evolution process, it is deficient to simulate the ability of human beings to deal with practice problems. So the IA introduces the theory of immunity of biotic science into the traditional GA to guide the evolution [11]. The key advantage of the IA is to raise fitness by adding the vaccination and verification operators. Because the biology immune system has the characteristics of vaccine diversity and artificial vaccination, the IA not only inherits the advantage of GA but also overcomes the weakness of prematurity and degeneracy.

Root Node Vaccines. In the IA, the vaccine is the key to improve convergence speed as it contains the useful information to guide the evolution direction. A proper vaccine could lead to a better result and improve the convergence speed in a great deal. The vaccines are usually extracted by expert knowledge from characteristics of the pending problem. For the BN structure learning problem, the vaccine is a kind of coded father node set of a certain node in BN which is used to replace the father node set of corresponding node in individuals.

According to the definition of BN, there must be at least one root node in the network structure. Because the objective of BN structure learning is to search the best structure that reflects the dataset accurately, the root node vaccine set $\{0, 0, \dots, 0; 0, 0, \dots, 0; \dots; 0, 0, \dots, 0\}$ is put forward to approximate a few parts of the best structure. After the vaccination of this type of vaccine, if the vaccinated node is the root node in the best structure, the fitness value of new structure will rise. Otherwise, the fitness value will decrease and the verification process will replace this new structure with the best structure of last population.

Bayesian Network Structure Learning

For the dataset with variables $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$, the detailed algorithm of BN structure learning based on root node vaccines (BNRN) is shown in Fig. 1.

Step 1. Coding. In the BNRN, the network structure is coded with the fix length matrix $\mathbf{T} = (t_{ij})_{n \times (n-1)}$, $t_{ij} \in \{0, 1, 2, \dots, n\}$, where each row in \mathbf{T} describes the father node set of a node, t_{ij} represents the j th father node ID in the father node set of node i . If $t_{ij} = 0$, it means that there is no father node for the j th father node.

Step 2. Population initialization. The number of initial individuals in the population is I_m and each individual represents a candidate BN structure. These individuals are usually generated at random.

Step 3. Fitness calculation. The fitness value represents the quality of each individual in the population. In the BNRN, the BIC score is chosen to calculate the quality of each candidate structures, as shown in Eq. 1.

$$V_{\mathbf{X}}^{BIC} = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \log \frac{m_{ijk}}{m_{ij}} - \sum_{i=1}^n q_i \frac{(r_i - 1)}{2} \log m \quad (1)$$

In equation (1), n represents the number of nodes in network structure; q_i represents the number of candidate state space of the father nodes of the i th node; r_i represents the number of candidate states of the i th node; m_{ijk} represents the number of records which satisfy the request that the i th node is in the k th state and its father node set is in the j th state; m_{ij} represents the number of record which matches that the father node set of the i th node is in the j th state; m represents the whole number of all the dataset records.

Because the fitness value must be positive in BNRN, the applied fitness function is shown as Eq. 2 based on the BIC score, where K represents a constant.

$$V_x^{LA} = -\frac{K}{V_x^{BIC}} = -\frac{K}{\sum_{i=1}^n V_{x_i}^{BIC}} \quad (2)$$

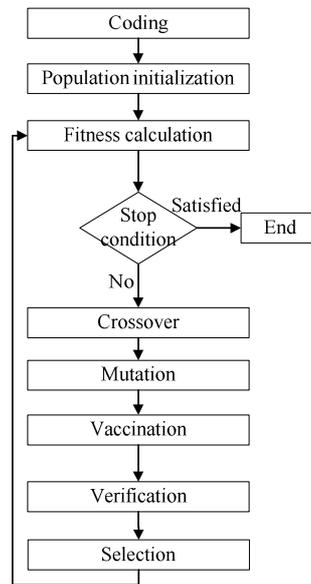


Fig. 1 Process of BNRN to learn BN structure

Step 4. Stop condition. The BNRN will stop when it has run for the scheduled number of iterations I_t . If the number of current generation meets I_t , output the best structure; if not, turn to Step 5 for the population crossover.

Step 5. Crossover. Match each two individuals in the population randomly and exchange the corresponding father node set of each pair of individuals with the probability p_c .

Step 6. Mutation. Change the father node ID in the father node set of a random node in the individual with a probability p_m .

Step 7. Vaccination. Replace a father node set randomly with the root node vaccine in the individual with a probability p_t .

Step 8. Verification. For the individual that has fitness decrease, replace the individual with the best individual in last generation.

Step 9. Selection. Supposing that the fitness of the i th individual is $V_x^{LA}(i)$, then it will be chosen into the next generation with the probability of $P(i) = e^{V_x^{LA}(i)/T_k} / \sum_{j=1}^{I_m} e^{V_x^{LA}(j)/T_k}$, where $T_k = \ln(\frac{100}{k} + 1)$ and k represents the sequence number of current generation. After the new population is generated, turn to Step 3 for the next generation of evolution.

Simulation study

Simulation dataset. The car start BN model [12] is applied as the original model which is used to generate the simulation dataset. The network structure of the car start BN is shown in Fig. 2 and the nodes information is shown in Table 1.

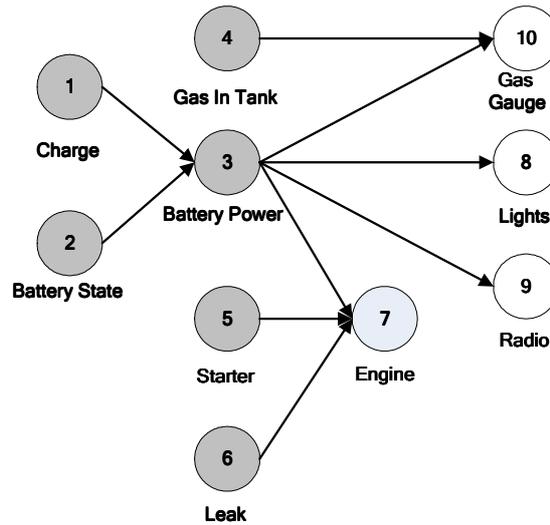


Fig. 2 Structure of a car start BN model

Table 1 Nodes of the car start BN model

NO	Variable	States	States description
1	Charge	0,1	Good, Low
2	Battery State	0,1	OK, Weak
3	Battery Power	0,1	Good, Poor
4	Gas In Tank	0,1	Gas, No Gas
5	Starter	0,1	OK, Faulted
6	Leak	0,1	True, False
7	Engine	0,1	Cranks, No Crank
8	Lights	0,1	Work, No Light
9	Radio	0,1	Works, Dead
10	Gas Gauge	0,1	Gas, No Gas

In this model, nodes “Charge” and “Battery State” affect node “Battery Power” together. The failure of node “Battery Power” will lead to “Engine” failure with the impact of “Starter” and “Leak”. The “Battery Power” also causes the states of “Radio” and “Lights” separately while it also affects “Gas Gauge” based on the state of “Gas In Tank”. Then, 4000 failure records are generated from the BN with a random sampling algorithm and the failure record dataset of car start is prepared.

Simulation Results. Based on the failure record dataset, the BNRN algorithm is applied to learn the network structure. The coding parameters used in BNRN are shown in Table 2. According to the root node vaccine extraction method described above, the practical root node vaccines of dataset are also prepared.

Table 2 Coding parameters of dataset

Node number	10
The largest number of father node	9
Individual length	90
Available value for a bit in individual	0-10

With the parameters of $I_m = 20, I_l = 400, K = 20000, p_c = 0.5, p_m = 0.01, p_i = 0.3$, the BNRN algorithm learns the same dataset 10 times separately. The highest fitness value and corresponding convergence iteration of each run are listed in Table 3. Another algorithm is also introduced by deleting the vaccination process, which is named as BNGA, to verify the efficiency and effectiveness of the BNRN. According to the comparison results, the BNRN algorithm could find the highest fitness value with the lowest number of iterations.

By comparing the root node vaccines with the original car start BN model in Fig. 2, there are five root node vaccines (node 1, 2, 4, 5, 6) which have the same father nodes. The similarity between the vaccines and the original structures may explain why this type of vaccine can improve the performance of BNRN.

Table 3 Learning result of each run

Runs	BNRN		BNGA	
	Iteration	Fitness	Iteration	Fitness
1	99	1.793	93	1.801
2	87	1.789	100	1.79
3	87	1.792	86	1.793
4	87	1.785	98	1.789
5	96	1.8	91	1.787
6	96	1.785	87	1.79
7	88	1.792	100	1.8
8	74	1.796	81	1.79
9	95	1.804	96	1.784
10	98	1.778	94	1.795

Conclusion

The paper has presented a root node vaccine based method (BNRN) to learn the Bayesian network structure from dataset. A car start BN model is applied to generate the simulation dataset. And the BN structure learning process is implemented. Experimental results of the simulation study prove that these proposed vaccines greatly enhance the convergence speed by comparing with the algorithms without vaccination. For the future research, more vaccines about BN structure leaning will be put forward and their effect will be studied.

Acknowledgment

The authors gratefully acknowledge the financial support for this research from the China Civil Aircraft Advanced Research Project, the China Aeronautical Science Foundations (Grant No. 2009ZE53052), the Science and Technology Project of Shaanxi Province (Grant No. 2010K8-11).

References

- [1] F. Jensen, An Introduction to Bayesian Networks, UCL Press, London, 1996.
- [2] T. Silander, T. Roos, P. Myllymaki, Learning locally minimax optimal Bayesian networks, International Journal of Approximate Reasoning (2010) doi:10.1016/j.ijar.2010.01.012.
- [3] D. Chickering, C. Meek, Large-sample Learning of Bayesian Networks Is NP-hard, Journal of Machine Learning Research 5 (2004) 1287-1330.
- [4] J. Cheng, R. Greiner, J. Kelly, et al, Learning Belief Networks from Data: An Information-theory Approach, The Artificial Intelligence Journal 137 (2002) 43-90.
- [5] G. Cooper, E. Herskovits, A Bayesian Method for the Induction of Probabilistic Networks from Data, Machine Learning 9 (1992) 309-347.
- [6] G. Schwarz, Estimating the Dimension of a Model, The Annals of Statistics 6 (1978) 461-464.
- [7] W. Lam, F. Bacchus, Learning Bayesian Belief Networks An approach based on the MDL Principle, Computational Intelligence 10 (1994) 269-293.
- [8] P. Larranaga, M. Poza, Y. Yurramendi, et al, Structure learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters, IEEE Transactions Pattern Analysis and Machine Intelligence 18 (1996) 912-925.
- [9] M. L. Wong, Y. Y. Guo, Learning Bayesian networks from incomplete databases using a novel evolutionary algorithm, Decision Support Systems 45 (2008) 368-383.
- [10] J. Z. Ji, H. X. Zhang, R. B. Hu, C. N. Liu, A Bayesian Network Learning Algorithm Based on Independence Test and Ant Colony Optimization, ACTA Auto. Sinica 35 (2009) 281-288.
- [11] L. C. Jiao, L. Wang, A novel genetic algorithm based on immune, IEEE Trans. on System, Man, and Cybernetics-Part A 30 (2000) 1-10.
- [12] D. Heckerman, J. Breese, K. Rommelse, Troubleshooting Under Uncertainty, Microsoft Research Technical Report MSR-TR-94-07, Microsoft Corporation, 1994.