International Journal for Applied Information Management
Vol. 3, No. 3, September 2023, pp. 101-108

ISSN 2776-8007

Formulation and Implementation of a Bayesian Network-Based Model

Ying Shi [1,*]

[1]Pubic Basic Section, Wuhan Technology and Business University, Hubei, China,
[1]shiying@wtbu.edu.cn*
*corresponding author

(Received: June 18, 2023; Revised: July 12, 2023; Accepted: August 31, 2023; Available online: September 10, 2023)

Abstract

At present, Bayesian networks lack consistent algorithms that support structure establishment, parameter learning, and knowledge reasoning, making it impossible to connect knowledge establishment and application processes. In view of this situation, by designing a genetic algorithm coding method suitable for Bayesian network learning, crossover and mutation operators with adjustment strategies, the fitness function for reasoning error feedback can be carried out. Experimental results show that the new algorithm not only simultaneously optimizes the network structure and parameters, but also can adaptively learn and correct inference errors, and has a more satisfactory knowledge inference accuracy rate.

Keywords: Bayesian Network, Structure Learning, Parameter Learning, Knowledge Reasoning, Genetic

1. Introduction

In the face of intricate uncertainty challenges, the application of Bayesian networks constitutes a two-fold approach: the establishment of Bayesian networks and the subsequent knowledge reasoning process. The establishment phase of a Bayesian network relies on the principles of Bayesian network learning [1-2]. This learning process can be delineated into two integral components: structure learning and parameter learning. Structure learning aims to identify the appropriate Bayesian network topology through the amalgamation of sample sets, juxtaposed with pre-existing foundational knowledge. On the other hand, parameter learning undertakes the task of associating conditional probabilities with each node within the network, thereby shaping its distinctive topology [3-4]. It is crucial to acknowledge that the accuracy of knowledge reasoning is not solely contingent upon the network's structural layout and parameter configuration, but also depends on the precision of the reasoning process itself. Any inaccuracies within this reasoning process have a direct impact on the refinement of the network's architecture and parameterization [5-6].

Currently, the realm of Bayesian network theory and algorithmic exploration is more focused on the pronounced dichotomy between the establishment and application of Bayesian networks. Research efforts aimed at addressing localized issues often encounter constraints when extrapolating their effectiveness to a broader vista of global utility. Consequently, there exists a tangible research gap in substantiating the practicality of Bayesian networks [7-8].

In this context, the state of the art in Bayesian network research plays a pivotal role. Presently, research has made impressive strides in tackling the challenges of Bayesian network establishment and application, especially in domains such as data science and artificial intelligence. Novel techniques, like the use of generative models within network structures and more efficient learning approaches, have opened new avenues to enhance the accuracy and effectiveness of Bayesian networks [9-10].

However, despite significant advancements, there are still research gaps that need to be addressed. One evident research gap is the difficulty in handling more complex uncertainty issues, including managing uncertainty associated with incomplete or uncertain data. Moreover, there is a need to develop more adaptive and responsive approaches that can accommodate changes in the environment and available resources. Additionally, when discussing the application of Bayesian networks in social, economic, and cultural contexts, emerging research gaps involve a closer integration between technical and non-technical aspects, as well as the assessment of the impact of Bayesian networks in addressing real-world societal challenges.

In line with a commitment to such explorations, it is imperative to steer efforts towards developing more holistic and inclusive models that not only consider the technical aspects of network establishment and reasoning but also delve into the broader contextual roles. For instance, approaches accommodating cultural, social, and economic factors can provide a deeper understanding of how Bayesian networks can respond to intricate dynamics within real-world contexts. In this regard, the engagement of broader domains such as social sciences and humanities can provide a more comprehensive insight into the role and impact of Bayesian networks in addressing uncertainty across various fields. Thus, sustained conceptual and applied explorations will lay the foundation for ongoing progress in the utilization of Bayesian networks as increasingly crucial tools in tackling the complex challenges we encounter.

2. Relevant conclusions of the Bayesian network

2.1. The definition of Bayesian network and its variable distribution

Bayesian network is a kind of directed acyclic graph that represents the probability dependence between variables. The Bayesian network composed of sample learning expresses a set of conditional independence assumptions: any node is independent of its non-offspring nodes given the state of its parent node [12,13]. Then the joint probability distribution corresponding to the random vector composed of the nodes of the Bayesian network can be decomposed into the product of the marginal distribution of random variables, namely.

$$P(B) = \prod_{i=1}^{N} P\left(\pi_j\right) = \frac{\sum_{A-\left(x_j\right)\cup x_j} P\left(x_1, x_2, \ldots, x_m\right)}{\sum_{A\subset x_i} P\left(x_1, x_2, \ldots, x_m\right)} \tag{1}$$

Where: x1,x2,...,xm - attribute node, A={x1,...,xm} - attribute set, πi:i=1,2,...,m represents the parent node set of node xi.

2.2. Relevant conclusions of Bayesian network inference

The key to Bayesian network inference is to infer the possibility of results based on the events that have occurred, which is manifested in the application of the statistical knowledge of the Yeesian network to calculate the relevant results with the support of the existing parameters (probability distribution) of the Bayesian network nodes [14-16]. The point corresponds to the probability of the occurrence of the event, that is, the problem boils down to: Knowing the value h of a part of the attribute set H in the attribute set A, find the conditional probability P(K=k) that another part of the attribute set K in the attribute set A is the specified value k |H=h). By comparing the size of P(K=k|H=h) of related different nodes in the Bayesian network, the most likely result or reason can be inferred.

2.3. Minimum description length

The Brute-Force MAP algorithm is mainly based on the maximum posterior probability to obtain the event hypothesis under known conditions [17-19]. Suppose H is a finite hypothesis space defined on the instance space D. For each hypothesis h, that is, each Bayesian network, the posterior probability of the network under D can be expressed as:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \tag{2}$$

$$h_{MAP} = \arg arg\, P(h|D) \tag{3}$$

The minimum description length L(g, D) is a metric that describes the storage structure of the Bayesian network. Let G denote all possible structure spaces, then a minimum description length L(g, D) of the network structure g.

3. The Genetic Algorithm Design of Bayesian Network Knowledge Establishment and Reasoning

Select attributes from the sample set to establish the Bayesian network structure, and calculate to form learning parameters. The network correction is achieved by checking the sample feedback inference error.

3.1. Genetic Individual Design and Network Acyclicity Test

Since the Bayesian network is an acyclic directed network, every time a new individual is generated, the new individual must be checked for acyclicity [20]. In order to ensure that the algorithm is always optimized in the feasible region.

The acyclic check algorithm based on the connection matrix is mainly divided into two steps: the first step is to check the symmetry of the matrix; the second step is to check the acyclicity.

The symmetry check of the matrix, that is, through Cij=Cji=1, it is determined that the directed graph corresponding to the matrix has a loop from node i, otherwise there is no loop, which can eliminate the direct relationship between two points.

The acyclicity check algorithm is to check whether the loop appears according to the information caused by the position of the attribute node on the loop of the directed graph. If there is a loop in the network, the access sequence generated from any node will be infinite Long, because the Bayesian network is a directed acyclic network, it is easy to know that the maximum length of the access sequence cannot exceed |E|2, so it only needs to verify: whether there is a node in the network, and the access sequence starting from it exceeds the maximum length.

3.2. Genetic Population Design

Use the minimum weight span tree method to generate an initial Bayesian network connection matrix, and then randomly change each element in the connection matrix from 1 to 0 and 0 to 1 (equivalent to randomly adding edges to the directed graph And delete edges) to generate a series of different directed graphs, use the acyclicity check algorithm based on the connection matrix to select reasonable directed graphs to form the population, and then calculate each of them according to the existing connection matrix and training data The equivalent CPT table of the node. These Bayesian networks constitute the initial population of heredity.

3.3. Genetic operator design

1) Crossover operator design of connection matrix.

   The matrix column crossover operator is used, that is, an integer i ($0 \leq i \leq m$) is randomly selected to determine the crossed column, and then all the columns to the right of the selected column are exchanged, and the new connection matrix generated after the exchange needs to be performed Acyclic inspection, adding the consensus understanding to the new population.

2) Design of mutation operator of connection matrix.

   Randomly select an element in the connection matrix to mutate according to a certain probability, so that 1 becomes 0 and 0 becomes 1; in fact, the addition, deletion, and reverse of the Bayesian network can be expressed as such 1 change 0, 0 change 1 Mutation operation, such as randomly selecting a position (i, j) to change Cij from 1 to 0 or 0 to 1; after mutation, perform symmetry and acyclicity checks. If the symmetry check fails and Cij=1, then Make Cij=0.

3) The adjustment strategy of the cross-mutation operator of the connection matrix.

The crossovers and forks in the connection of the attribute nodes in the network-the CPT table. Among these effects, some of them can ensure that they will not destructively destroy their own parameter learning results, and take good care of them. The properties of the attribute nodes are retained in a new generation of individuals. Some of them will only destroy the effective parameter information in the attribute nodes, making it important to relearn the parameters according to the training set after each cross-mutation of the connection matrix. Therefore, it is necessary to analyze the destructiveness and destruction degree of parameter learning by the crossover and mutation operators of the connection matrix.

Therefore, the cross-connection operation of the connection matrix will have the following effects on parameter learning: row crossover will change the parent node of most attribute nodes; and the CPT table of all attribute nodes in column cross-over does not need to be updated, only exchange The CPT table of the corresponding attribute. Therefore, the direct use of column crossover in this paper can completely ensure that the results of parameter learning are not lost.

The mutation operator adjustment strategy can be achieved through the following strategies:

1) Delete the edge (i, j), the parent node set π(i) of node i remains unchanged, and the parent node set π(j) of node j is missing one parent node i, at this time Since one parent node is missing, the corresponding CPT table can be directly reduced, that is, the column where i is first deleted, and then the same rows are merged together. It is like the change from Table 1 before deleting edge (2,0) to Table 2.
2) Add an edge (i, j), then the parent node set π(i) of node i has no change, and the parent node set π(j) of the node has one more parent node at this time. Insert the parent node i, then add rows to the CPT table according to the rules, and at the same time, split the previous row. The split method is temporarily equal. The average score is regarded as the prior value, combined with the frequency of the data, and finally the Bayes posterior probability. For example, the changes from Table 2 before the edge increase (2,0) to Table 1 after the edge increase (2,0).

Table 1. Node 0's parent node set is P(0,{1,2})

Probability	Node 0	Node 1	Node 2
0.2	a	a	a
0.05	b	a	a
0.15	a	b	a
0.1	b	b	a
0.2	a	a	b
0.05	b	a	b
0.15	a	b	b
0.1	b	b	b

Table 2. P(0,{1}) whose parent node set of node 0 is {1}

Probability	Node 0	Node 1
0.4	a	a
0.1	b	a
0.3	a	b
0.2	b	b

3) Select operator design.

In this paper, each individual of each generation has a huge amount of calculation and it is not easy to satisfy the condition of acyclic directed network [21]. Therefore, the individual generated by cross mutation should not directly replace the parent individual, but should remain with the parent Carrying out competition, on the one hand, better individuals can be found to avoid the phenomenon of regression of the fitness of the population as a whole, and on the other hand, the use of retaining intermediate results can save the amount of calculation. According to the above analysis, this paper adopts the sort selection operator based on (μ+ λ).

The specific operation of the (μ+ λ) sorting selection method selection operator is: assuming that the parent population size is μ, the generated offspring population size is λ (λ=λ1+λ2), where λ1λ2 is generated by the crossover operator and the mutation operator respectively Then let all parent individuals and all offspring individuals sort from good to bad according to the priority order of the target, and take the best first μ after sorting as the next-generation population.

4. Bayesian network adaptive learning algorithm analysis

4.1. Experimental design and results

This article uses the 1984 United States Congressional Voting Records Database instance data in the UCI database. There are 17 attributes in this instance data set. The attribute with serial number 0 is a classification attribute, and the values of all attributes in this database are binary. The experiment uses 162 complete records in this database for learning, and uses 1/10 of the samples as inference test samples to compare the adaptive learning algorithm with the maximum likelihood algorithm and the minimum description length algorithm in the establishment of the Bayesian network structure. Process-related indicators, the test results in Table 3 are obtained.

Table 3. Algorithm results

Algorithm	Adaptive Algorithm	Maximum Likelihood Method	MDL method
MAP value	1.60×10140	2.84×10143	1.30×10122
MDL value	412.05	465.83	392.01
Correct rate	97.84%	86.63%	96.12%
Reduce node	{2,10,11,15}	{2,9}	{2,10,11}
Number of connected edges	15	16	17
Number of points with connected edges greater than 3	4	6	6

4.2. Result analysis

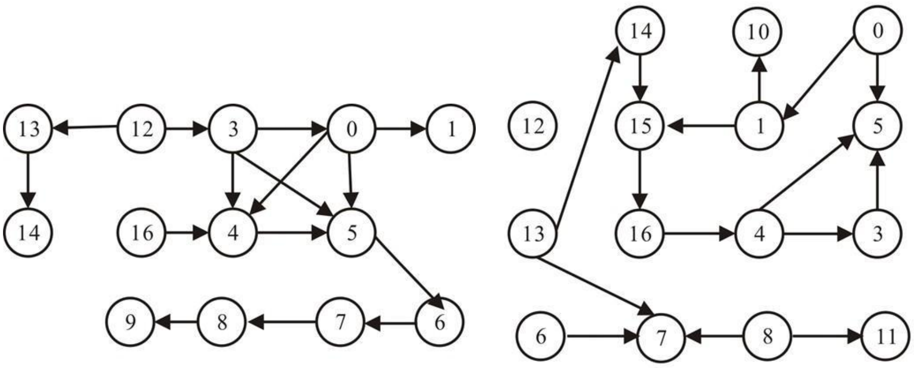
It can be seen from Figure 3 that the Bayesian network obtained by the adaptive learning algorithm combined with the adjustment strategy reduces the 4 nodes 2, 10, 11, and 15, and the network reduction obtained by the maximum likelihood method learning After the two nodes 2, 9 are used, the 3 nodes 2, 10, and 11 are finally reduced by the network learning with the minimum description length. From the number of nodes reduced in Figure 1 and Table 3, the number of connected edges, and the number of points with connected edges greater than 3, it can be seen that the

ISSN 2776-8007

Bayesian network obtained is better than the network obtained by the maximum likelihood method and the minimum description length learning Be concise.

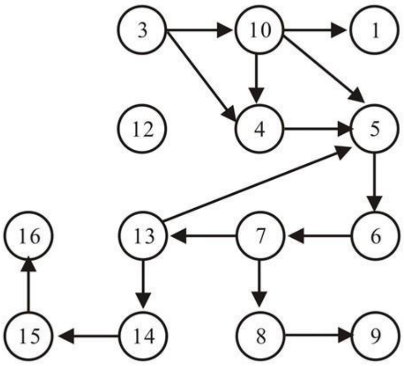
It can be seen from Table 3 that the correct classification of the Bayesian network is obtained, and its network maximum likelihood value is second only to the maximum likelihood of the network obtained by the maximum likelihood method. Likelihood value, the corresponding minimum description length value is second only to the network generated by the minimum description length algorithm.

As mentioned above, the Bayesian network obtained by the adaptive learning algorithm has a higher inference accuracy rate than the Bayesian network generated by the maximum likelihood method and the minimum description length method, and the maximum likelihood value of the network And the minimum description length also have satisfactory values. In addition, the adaptive learning algorithm makes up for the shortcomings of the maximum likelihood and minimum description length algorithms that can only apply structure learning, parameter learning and reasoning error feedback, and truly realize Bayesian.



(a) Adaptive learning algorithm

(b) Maximum likelihood method



(c) Minimum description length method

Figure 1. Bayesian network connections produced by three methods

5. Conclusion

The new Bayesian network adaptive knowledge establishment and reasoning algorithm based on genetic algorithm proposed in this paper is applied to the experiment of establishing the Bayesian network structure by using the maximum likelihood algorithm and the minimum description length algorithm respectively to compare the corresponding structure Indexes such as maximum likelihood value, minimum description length, classification accuracy rate, reduced nodes, number of connected edges, etc., indicate that the new adaptive learning algorithm not only has a simpler connection structure and higher knowledge inference accuracy, but also can be inferred by reasoning. The error synchronization optimizes the network structure and learning parameters, breaking through the problem that the Bayesian network can only be studied locally but cannot be checked and corrected by feedback as a whole. It provides a theoretical framework for the application of Bayesian networks in the field of knowledge reasoning.

References

[1] Marin, J. M, Pillai, N. S, Robert, C. P, & Rousseau, J. (2015). Relevant statistics for bayesian model choice. Journal of the Royal Statistical Society, 76(5), 833-859.

[2] Rachel, C, Lawson, A. B, Christel, F, Kirby, R. S, Mehreteab, A, & Kevin, W. (2016). Spatio - temporal bayesian model selection for disease mapping. Environmetrics, 27(8), 531-55.

[3] Yunqing, Xia, Erik, Cambria, Amir, & Hussain, et al. (2015). Word polarity disambiguation using bayesian model and opinion-level features. Cognitive Computation, (34), 1104-1109.

[4] Holmes, C. C, & Walker, S. G. (2017). Assigning a value to a power likelihood in a general bayesian model. Biometrika, 66(2), 419-424.

[5] Lam, H. F, Yang, J, & Au, S. K. (2015). Bayesian model updating of a coupled-slab system using field test data utilizing an enhanced markov chain monte carlo simulation algorithm. Engineering Structures, 102(1), 144-155.

[6] Huang, G, Lee, D, & Scott, M. (2015). An integrated bayesian model for estimating the long- term health effects of air pollution by fusing modelled and measured pollution data: a case study of nitrogen dioxide concentrations in scotland. Spatial and Spatio-temporal Epidemiology, 14-15, 63-74.

[7] Kim, B. C. (2015). Integrating risk assessment and actual performance for probabilistic project cost forecasting: a second moment bayesian model. IEEE Transactions on Engineering Management, 62(2), 158-170.

[8] Riad, Stella, Bergstroem, & Johannes. (2015). Bayesian model comparison of higgs couplings. Physical Review D Particles Fields Gravitation & Cosmology, 39(6), 1006-1011.

[9] Liu, C, & Sharma, A. (2018). Using the multivariate spatio-temporal bayesian model to analyze traffic crashes by severity. Analytic Methods in Accident Research, 17, 14-31.

[10] Daniel, T. (2015). Comparison of the frequentist mata confidence interval with bayesian model- averaged confidence intervals. Journal of Probability and Statistics,2015,(2015-10-8), 2015, 1-9.

[11] Fer, I, Kelly, R, Moorcroft, P. R, Richardson, A. D, & Dietze, M. C. (2018). Linking big models to big data: efficient ecosystem model calibration through bayesian model emulation. Biogeosciences Discussions, 1-30.

[12] I. Ntzoufras, V. Palaskas and S. Drikos, "Bayesian models for prediction of the set-difference in volleyball," in IMA Journal of Management Mathematics, vol. 32, no. 4, pp. 491-518, Aug. 2021, doi: 10.1093/imaman/dpab007.

[13] S. Yıldırım, M. B. Kurutmaz, M. Barsbey, U. Şimşekli and A. T. Cemgil, "Bayesian Allocation Model: Marginal Likelihood-Based Model Selection for Count Tensors," in IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 3, pp. 560-573, April 2021, doi: 10.1109/JSTSP.2020.3045297.

[14] Tian, Huaming, Dian-Qing Li, Zi-Jun Cao, and Wenqi Du. "Auxiliary Bayesian updating of embankment settlement based on finite element model and response surface method." Engineering Geology 323 (2023): 107244

[15] J. Cui, F. Feng, W. Na and Q. -J. Zhang, "Bayesian-Based Automated Model Generation Method for Neural Network Modeling of Microwave Components," in IEEE Microwave and Wireless Components Letters, vol. 31, no. 11, pp. 1179-1182, Nov. 2021, doi: 10.1109/LMWC.2021.3087163.

[16] N. Calik, M. A. Belen, P. Mahouti and S. Koziel, "Accurate Modeling of Frequency Selective Surfaces Using Fully-Connected Regression Model With Automated Architecture Determination and Parameter Selection Based on Bayesian Optimization," in IEEE Access, vol. 9, pp. 38396-38410, 2021, doi: 10.1109/ACCESS.2021.3063523.

[17] J. Cai et al., "Bayesian Inference-Based Behavioral Modeling Technique for GaN HEMTs," in IEEE Transactions on Microwave Theory and Techniques, vol. 67, no. 6, pp. 2291-2301, June 2019, doi: 10.1109/TMTT.2019.2906304.

[18] Perepolkin, Dmytro, Benjamin Goodrich, and Ullrika Sahlin. "The tenets of quantile-based inference in Bayesian models." Computational Statistics & Data Analysis (2023): 107795.

[19] J. Dong, W. Lyu, D. Zhou and W. Xu, "Variational Bayesian and Generalized Approximate Message Passing-Based Sparse Bayesian Learning Model for Image Reconstruction," in IEEE Signal Processing Letters, vol. 29, pp. 2328-2332, 2022, doi: 10.1109/LSP.2022.3221344.

[20] Xiao, Fei, Xiao-kang Wang, Wen-hui Hou, Xue-yang Zhang, and Jian-qiang Wang. "An attention-based Bayesian sequence to sequence model for short-term solar power generation prediction within decomposition-ensemble strategy." Journal of Cleaner Production (2023): 137827.

[21] S. Bourouis, R. Alroobaea, S. Rubaiee, M. Andejany, F. M. Almansour and N. Bouguila, "Markov Chain Monte Carlo-Based Bayesian Inference for Learning Finite and Infinite Inverted Beta-Liouville Mixture Models," in IEEE Access, vol. 9, pp. 71170-71183, 2021, doi: 10.1109/ACCESS.2021.3078670.