Design and Application of Model Based on Bayesian Network

Ying Shi$^1$*$

$^1$Public Basic Section, Wuhan Technology and Business University, Hubei, China, 430065

*Corresponding author e-mail: shiy@wtbu.edu.cn

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 Algorithm

1. Introduction

The application of Bayesian network to solve the uncertainty problem consists of two parts: Bayesian network establishment and knowledge reasoning. The establishment of Bayesian network depends on Bayesian network learning [1-2]. Structure learning is to use sample sets, combined with prior knowledge as much as possible, to determine the appropriate Bayesian network topology; parameter learning is to determine the Bayesian network topology [3-4]. Determine the conditional probability of each node of the network corresponding to the event; the corresponding knowledge reasoning not only depends on the network structure and network parameters, and the reasoning error directly affects the correction of the network structure and network parameters [5-6].

At present, the research on Bayesian network theory and algorithm is mainly aimed at a certain separation problem between the establishment and application of Bayesian network. The effectiveness of research on local problems is often difficult to generalize to the argument of global effectiveness, and there is a lack of support for Bayesian networks [7-8]. The consistency algorithm of establishing the network structure, parameter learning, and inference error feedback correction has largely restricted the development of Bayesian network applications [9-10]. Therefore, exploring effective and rapid methods and algorithms that can support Bayesian network adaptive knowledge establishment and knowledge reasoning has become a key issue in the field of uncertainty knowledge reasoning [11].

2. Relevant conclusions of the Bayesian network model
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. Then the joint probability distribution corresponding to the nodes of the Bayesian network can be decomposed into the product of the marginal distribution of random variables, namely.

\[
P(x_1,x_2,\cdots,x_m|B) = \prod_{i=1}^{N} P(x_i|\pi_i) = \sum_{d(x)} \sum_{x_1, x_2, \cdots, x_m} P(x_1,x_2,\cdots,x_m)
\]

(1)

Where: \(x_1,x_2,\cdots,x_m\) - attribute node, \(A=\{x_1,\cdots,x_m\}\) - attribute set, \(\pi_i:i=1,2,\cdots,m\) 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 Bayesian network to calculate the relevant results with the support of the existing parameters (probability distribution) of the Bayesian network nodes. 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 \(P(K=k|H=h)\) 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. 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:

\[
P(h|D) = \frac{P(D|h)P(h)}{P(D)}
\]

(2)

\[
h_{MAP} = \arg\max_{h\in H} P(h|D)
\]

(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. 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 $C_{ij} = C_{ji} = 1$, 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 Loop.

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 Exceed 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 $C_{ij}$ from 1 to 0 or 0 to 1; after mutation, perform symmetry and acyclicity checks. If the symmetry check fails and $C_{ij} = 1$, then Make $C_{ij} = 0$. There are three situations in which this mutation can be realized:

1) If $C_{ij} = C_{ji} = 1$, it is an edge-increasing operation;
2) If $C_{ij} = 1$, it is an edge deletion operation;
3) If $C_{ij} = 0$, $C_{ji} = 1$, it is a reverse operation;

(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 cross-over 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).

(3) Change direction (i,j)->(j,i), then the parent node set π(i) of node i has one more node j, and the parent node set π(j) If one parent node i is missing, the CPT table of node i is increased. For the CPT table of node j, delete edges.

| 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      |

(5) 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. 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 \((\mu + \lambda)\).

The specific operation of the \((\mu + \lambda)\) sorting selection method selection operator is: assuming that the parent population size is \(\mu\), the generated offspring population size is \(\lambda\) \((\lambda = \lambda 1 + \lambda 2)\), where \(\lambda 1\lambda 2\) 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 \(\mu\) 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 Congres sional 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 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 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.
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.

Acknowledgments
Grant No B2018309 Application of Bayesian network in model; Grant No A2017013 Research on Application of Bayesian network.

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.