Research Article
Bayesian Network Structure Learning and Application

Liwei Zhang,1,2,3 Dayou Liu,1,2 Zheng Wang,1,2,4 and Xinlu Wang1

1College of Computer Science and Technology, Jilin University, Jilin 130012, China
2Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Jilin 130012, China
3Department of Computer, Aviation University of Air Force, Jilin 130022, China
4Engineering Training Center, Changchun Institute of Engineering, Jilin 130012, China

Correspondence should be addressed to Liwei Zhang; 2016124153@jou.edu.cn

Received 4 May 2022; Revised 25 May 2022; Accepted 31 May 2022; Published 28 June 2022

Copyright © 2022 Liwei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous development of artificial intelligence technology and information technology, a large number of background data are constantly generated. How to obtain effective and useful data in a large and complex data group becomes important and meaningful. The traditional Bayesian network can represent the probability distribution of data variables from a large number of data based on graphical models. It has relatively clear and reliable reasoning ability and decision-making mechanism. However, the traditional Bayesian network structure has serious shortcomings in the recognition accuracy of corresponding key data, so the efficiency of the corresponding algorithm is seriously low. Based on this, this study adds an adaptive genetic algorithm with causality to the original Bayesian structure, so as to optimize the strategy of its structure operation, quantitatively describe the order of the corresponding data nodes, creatively arrange the corresponding data nodes in order by using the node priority, and initialize the initial architecture of Bayesian network based on this. Finally, the network is initialized through information exchange and data score correction, so as to get the final learning results. In this study, the convolution neural network algorithm in a database is verified in the experiment. The experimental results show that the accuracy of the experimental results given by the Bayesian network structure proposed in this study is about 10% higher than the traditional accuracy, and its corresponding learning results basically cover the important algorithms, hypotheses, and verification of convolution neural network, from this level; the algorithm proposed in this study has obvious advantages in bibliometrics.

1. Introduction

With the continuous development of information technology and artificial intelligence technology, all kinds of corresponding data show geometric explosive growth. How to classify, process, find, and identify useful data in the database becomes very important and meaningful. Conventional data mining or data recognition technologies include various classical data mining and classification algorithms such as probability and statistics methods and fuzzy logic methods [1–3]. As a probability network that can graphically represent the relationship between random variables, Bayesian network data structure is essentially a causal learning network. It is essentially a directed acyclic graph. The corresponding data nodes represent the corresponding random variables, and the directed edges between the corresponding nodes represent the causal relationship between data; at the same time, the conditional probability between the corresponding data represents the relationship strength between the corresponding data nodes. The Bayesian structure combines the information of some corresponding nodes with the corresponding probability reasoning to realize the probability information of other data nodes. The recognition model with Bayesian network characteristics derived from this Bayesian model is as follows: Kalman filter, dynamic Bayesian model, and dynamic Bayesian network [4, 5].

The conventional Bayesian model has obvious advantages in data processing and identification, its corresponding Bayesian network has the characteristics of intuitive and easy to understand, its corresponding data network has strong comprehensibility and interpretability, and the dependency
between corresponding data is also very clear; conventional Bayesian network has good preciseness, its corresponding network model is based on a mathematical model, and its corresponding network reasoning process is very rigorous [6, 7]; the corresponding Bayesian network has better flexibility in processing and identifying the corresponding data information. Its corresponding Bayesian network can combine the corresponding prior knowledge and can also carry out data learning and analysis from a large number of databases. The corresponding theoretical basis is the continuous reasoning of probability theory knowledge [8]; the conventional Bayesian network has strong interpretation accuracy, which can accurately and quantitatively describe the dependencies corresponding to the corresponding data variables [9, 10]. When the conventional Bayesian network data structure expresses the uncertainty problem, it can analyze the corresponding data problem from both qualitative and quantitative levels; the conventional Bayesian network structure is essentially an important data mining tool. It can classify, cluster, and predict a large number of data. Its corresponding expressiveness, comprehensibility, interpretability, and strict logic make the Bayesian structure closer to the corresponding practical problems [11]. However, the traditional Bayesian network structure has serious shortcomings in the recognition accuracy of the corresponding key data, so the efficiency of the corresponding algorithm is seriously low.

Based on the analysis of the current situation of Bayesian network structure, this study will add an adaptive genetic algorithm with causality to the original Bayesian structure, so as to optimize its structure operation strategy, quantitatively describe the order of corresponding data nodes, innovatively arrange the corresponding data nodes in order using node priority, initialize the initial architecture of Bayesian network based on this, and finally modify the initialization network through information reciprocity and data scoring, so as to obtain the final learning results. In this study, the convolution neural network algorithm in a database is verified in the experiment. The experimental results show that the accuracy of the experimental results given by the Bayesian network structure proposed in this study is about 10% higher than the traditional accuracy, and its corresponding learning results basically cover the important algorithms, hypotheses, and verification of convolution neural network, from this level; the algorithm proposed in this study has obvious advantages in bibliometrics.

The structure of this study is arranged as follows: the second section of this study will analyze and study the current research status of Bayesian network structure; the third section analyzes the principle of causal adaptive genetic algorithm in the improved Bayesian network structure; the fourth section of this study will experiment with the algorithm proposed in this study and analyze the experimental results; finally, this study will be summarized.

2. Research Status of Sports Video Athlete Detection Technology

Based on the application of Bayesian network in the uncertainty of network data, a large number of scholars and research institutions have carried out research and analysis on it. The corresponding research results mainly focus on Bayesian network theory learning, the combination of expert domain learning and Bayesian network theory learning, and the research on the practical application of Bayesian network. The relevant scientists in the United States have carried out systematic research and analysis on the theory of Bayesian network, and their corresponding segmentation principles play an important role in the follow-up classical Bayesian learning [12, 13]; at the level of Bayesian network structure model research, the main theories focus on the corresponding modeling processing based on domain expert knowledge and determine the probability learning distribution of each variable according to parameter learning method. However, this construction method greatly shows the disadvantages of the Bayes network framework [14]; the relevant European research institutions and corresponding scientists have proposed the Bayesian construction algorithm based on dependency analysis. It mainly detects variables based on the interrelationship between variables, thus determining the corresponding direction and other parameters among different variables. Finally, based on this, the corresponding Bayesian network architecture diagram is constructed. The corresponding algorithms integrate statistical theory, information theory, and other relevant algorithms, but the Bayesian network construction efficiency of the algorithm is low [15]; based on the above network, the corresponding researchers propose a conditional independence detection algorithm to improve the construction strategy of Bayesian network, but the construction method requires a high sample capacity of training [16, 17]; the corresponding researchers proposed a hybrid construction method based on the disadvantages of Bayesian network constructed on independent conditions. It mainly used an independence test to get a fuzzy Bayesian network and then the heuristic search algorithm based on the scoring function algorithm, which solved the drawbacks existing in the previous two algorithms when they were running separately [18–20]. Based on the above analysis and research, the Bayesian network structure has obvious advantages in the information uncertainty expression and probability distribution, but it still has more or less recognition accuracy and multidate and complex node identification problem.

3. Bayesian Network Analysis Based on Improved Genetic Algorithm

The main idea of this study is to add an adaptive genetic algorithm with a causal relationship into the original Bayesian structure, so as to optimize the strategy of its structure operation, quantitatively describe the corresponding data nodes in order, arrange the corresponding data nodes in order by node priority, and initialize the Bayesian network initial architecture based on this. Finally, the network is initialized through information exchange and data score correction, so as to get the final learning results. The corresponding Bayesian network construction framework is shown in Figure 1. From the figure, we can see the corresponding core modules in the algorithm proposed in this study.
In view of the increase of causality factors in this algorithm, we first need to solve the priority problem between data nodes. As long as the corresponding priority sequence follows the corresponding sequence A, sequence B, and sequence C, the corresponding three sequences determine the priority sequence of the data nodes in turn. The node priority of corresponding different data shows the priority of the data node compared with other data in the network. The corresponding data priority network structure schematic diagram is shown in Figure 2. It can be seen from Figure 2 that the priority relationship between corresponding data is mainly the hierarchical relationship between data nodes. In the actual priority arrangement, it is necessary to calculate the sum of mutual information between nodes with the same priority among different data and other data nodes of the system.

Based on the solution of the above priority problem, the genetic algorithm in its Bayesian network is improved. The corresponding improved processing level mainly includes four levels, and the corresponding optimization framework is shown in Figure 3. At the level of data parameter coding, the traditional data parameter coding method is abandoned, the Bayesian network structure is represented by the matrix corresponding to $m \times m$, and the corresponding adjacency matrix in the Bayesian network structure is represented by matrix $h$. In the actual genetic operation and corresponding coding optimization process, the adjacency matrix in the Bayesian network is directly regarded as an individual in the system. The corresponding operation of the adjacency matrix is regarded as the operation of the system individual. Finally, the Bayesian network architecture is constructed based on the adjacency matrix; in the corresponding Bayesian network initialization population, the node mutual information formula based on information theory is optimized and improved. For the data edge in the independent mapping, the corresponding node is retained and removed by determining the corresponding obstacle set. Based on this operation step, the corresponding candidate Bayesian network structure can be further obtained; based on the fitness function of detection and repair the abnormal network structure is detected, repaired, and removed in the process of population or data iteration and the fitness of the remaining structure or population is detected. The detection tool used is the fitness function. Based on this, the better the individual, the higher the corresponding fitness. In this study, the selection of fitness function is mainly based on the scoring function; the control parameter optimization of genetic algorithm is mainly to optimize the parameters of Bayesian core algorithm genetic algorithm at this level. The main optimization parameters include the identified uncertain population size, the crossover rate, and mutation rate of the uncertain population, and the termination threshold of the algorithm is proposed in this paper, in which the corresponding population crossover rate. The rate of variation is mainly processed and analyzed based on adaptive function.

In order to further improve the accuracy of the above Bayesian network structure in identifying uncertain information in metrology and improve the performance of the whole algorithm, this study also optimizes its priority relationship, which is also the further optimization of the core genetic algorithm of Bayesian network in this study. Based on the original priority confirmation, this study reestablishes the logical relationship between different data nodes and defines it as causality. Based on this causality, the corresponding node priority judgment standard is set as shown in

![Figure 1: Schematic diagram of Bayesian network architecture.](image-url)
formula (1). The corresponding $m$ in the formula represents the corresponding data sample code, and the corresponding $n$ represents the total number of samples of the whole data node. The corresponding $N$ represents the corresponding data random variable. It can be further seen from the judgment formula that the interference factors in the judgment of causality are deleted in this study, and the causality is calculated and processed based on the corresponding conditional probability. At the same time, the relative causality of nodes in the three dimensions of $A$, $B$, and $C$ is fully considered in the formula, and the causal effects caused by their corresponding causality are weighted.

$$\text{xy}(m - n) = \ln \left[ \frac{P(y_{i+1})}{P(n_{i-1})} \right]. \quad (1)$$

Based on this, it can be further concluded that the decision steps of optimizing the priority sequence of data nodes are as follows:

(1) Step 1. Triggers from a specific node in the specified Bayesian network structure to gradually calculate the causal effect between the node and other nodes in the system.

(2) Step 2. Substitute the above corresponding causal effect into the judgment formula for judgment processing. When the corresponding node priority order is processed, the corresponding priority judgment rules are as follows: when $\text{xy}(m - n) > \text{xy}(n - m)$, the priority of the corresponding node $x$ increases by 1; otherwise, the priority of the corresponding node $y$ increases by 1.

(3) Step 3. Perform algorithmic traversal processing on all nodes in the whole system until the causality corresponding to all nodes in the system is calculated. Based on this, the priority vector of the corresponding nodes is arranged in ascending order to obtain the corresponding priority order.

Based on the above analysis, this study introduces the priority sequence processing with causality into the traditional Bayesian genetic algorithm and improves and optimizes the four key levels of the corresponding genetic algorithm, so as to form a hybrid uncertain information recognition algorithm. The corresponding algorithm architecture is shown in Figure 4. From the figure, it can be further concluded that the steps of the optimization algorithm in this study are as follows:

(1) Step 1. Perform priority processing on the corresponding data population to be processed, and perform algorithm traversal processing on all nodes in the whole system until the causality corresponding to all nodes in the system is calculated. Based on this, the corresponding nodes are arranged in ascending order to obtain the corresponding priority order.

(2) Step 2. Perform iterative analysis based on the processed individual data nodes.

(3) Step 3. Form the initial population based on the algorithm proposed in this paper.

(4) Step 4. Encode the population to form the corresponding adjacency matrix, regard the operation of the corresponding adjacency matrix as the operation of the system individual, and finally construct the
Bayesian network architecture based on the adjacency matrix.

(5) Step 5. For the data edge in the independent mapping, the corresponding node is retained and removed by determining the corresponding obstacle set. Based on this operation step, the corresponding candidate Bayesian network structure can be further obtained.

(6) Step 6. For the entire Bayesian network system, detect, repair, and remove the abnormal network structure in the process of population or data iteration.

(7) Step 7. Determine the fitness of different data nodes in the system based on the fitness function and select the data nodes with high fitness.

(8) Step 8. Repeat the above process until the optimal Bayesian network structure is found.

Based on the above-improved algorithm, the disadvantages of falling into local optimization and premature convergence of the whole Bayesian network can be further avoided in theory. At the same time, it can further improve the accuracy of the whole Bayesian network in identifying uncertain information.

4. Experiment and Analysis

Based on the above theoretical analysis of the algorithm, this section will experiment based on the convolutional neural network keywords in a large data knowledge base. In the experiment, the algorithm proposed in this study is compared with the Bayesian network under the traditional genetic algorithm. The corresponding experimental environment is consistent, and the corresponding experimental design is shown in Figure 5. The corresponding data codes in Figure 5 represent different keywords and the frequency of corresponding keywords.

According to the above network learning results of convolutional neural networks, network learning processing is carried out based on the algorithm in this study. Based on the improved node priority algorithm and causal effect processing, the corresponding algorithm mainly determines the corresponding semantics and whether the corresponding keywords in the keyword data set correspond. The corresponding learning results are shown in Figure 5. The corresponding data codes in Figure 5 represent different keywords and the frequency of corresponding keywords.

In order to verify the accuracy of this algorithm in identifying uncertain information, this study makes a fuzzy
search for the keyword convolutional neural network based on the same database. At the same time, the compared algorithm is the traditional Bayesian network based on the genetic algorithm. The corresponding accuracy and the corresponding algorithm standard deviation are shown in Figures 7 and 8. It can be seen from the figure that the corresponding algorithm proposed in this study is significantly higher than the traditional algorithm in accuracy and standard deviation, which fully reflects the obvious advantages of the improved genetic algorithm and the priority sequence based on causality.

In order to further verify the accuracy of the corresponding algorithm and the fluctuation of standard deviation under a large sample size, three rounds of new database samples are added to the database of the original experiment, and comparative experiments are carried out based on the new data samples. The experimental results are shown in Figures 9 and 10. It can be seen from the experimental results in Figure 9 that with the continuous increase of database samples, the accuracy of the corresponding algorithm has achieved the effect of continuous improvement. Compared with the traditional algorithm, the accuracy of the corresponding algorithm proposed in this study improves faster, and the overall increase in accuracy of the traditional algorithm decreases with the increase of samples; it can be seen from Figure 10 that the
corresponding standard deviation index does not improve significantly with the increase of database samples, but the algorithm proposed in this study still has obvious advantages over the traditional algorithm.

**Figure 7:** Comparison column chart of accuracy of identifying an uncertain information level.

**Figure 8:** Standard deviation comparison column chart for identifying uncertain information level.

**Figure 9:** Comparison bar chart of accuracy of identifying uncertain information (increase in sample size).

As shown in Figures 11(a) and 11(b), it is a broken line diagram of the overall system performance comparison between the algorithm in this study and other traditional algorithms. It can be seen from the figure that the algorithm proposed in this study consumes less resources at the level of identifying uncertain information, and its corresponding algorithm has higher performance. At the same time, comparing Figures 11(a) and 11(b), it can be seen that with the continuous increase of the corresponding sample size, the performance advantage of the corresponding algorithm is more obvious, and the performance difference with the traditional algorithm is greater.

Based on the above experimental results, the advantages of this algorithm are basically verified. At the same time, it also provides an optimized algorithm for the application of Bayesian network in metrology, improves the recognition accuracy of Bayesian network for uncertain information, and improves the performance of the whole algorithm.

**5. Summary**

This study mainly analyzes the theoretical points of Bayesian network structure learning and comprehensively analyzes the current theoretical research status of Bayesian network based on the research status of Bayesian network application level. Based on the current application of Bayesian network in metrology, this study adds an adaptive genetic algorithm with causality to the original Bayesian structure, so as to optimize the operation strategy of its structure and quantitatively describe the order of the corresponding data nodes. The corresponding data nodes are arranged in the order of node priority, and the initial architecture of Bayesian network is initialized based on this. Finally, the network is initialized through information reciprocity and data score correction, so as to obtain the final learning results. In this study, the convolution neural network algorithm in a database is used for experimental verification. The experimental results show that the accuracy of the experimental...
results given by the Bayesian network structure proposed in this study is about 10% higher than the traditional accuracy, and its corresponding learning results basically cover the important algorithms, hypotheses, and verification of convolution neural network. From this level, the algorithm proposed in this study has obvious advantages in bibliometrics. This study will focus on the construction of Bayesian network in the case of multiple nodes when the corresponding data nodes increase. At the same time, it will further analyze and study the analysis of Bayesian network in complex cases and its recognition accuracy in metrology [21].

Data Availability
The figures used to support the findings of this study are included in the article.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Acknowledgments
The authors would like to show sincere thanks to those techniques who have contributed to this research. This work was not supported by any funds.

References
[1] V. Asvatourian, P. Leray, S. Michiels, and E. Lanoy, “Integrating expert’s knowledge constraint of time dependent exposures in structure learning for Bayesian networks,” Artificial Intelligence in Medicine, vol. 107, no. 4, Article ID 101874, 2020.
[2] M. Velikova, P. J. F. Lucas, M. Samulski, and N. Karssemeijer, “On the interplay of machine learning and background knowledge in image interpretation by Bayesian networks,” Artificial Intelligence in Medicine, vol. 57, no. 1, pp. 73–86, 2013.
[3] X. Xie, Y. Liu, Y. Luo, and Q. Du, “Surface water quality evaluation based on bayesian network,” Journal of Coastal Research, vol. 93, no. sp1, p. 54, 2019.
[4] Q.-Q. Meng, Q.-X. Zhu, H.-H. Gao, Y.-L. He, and Y. Xu, “A novel scoring function based on family transfer entropy for Bayesian networks learning and its application to industrial alarm systems,” Journal of Process Control, vol. 76, no. 2, pp. 122–132, 2019.
[5] E. N. Cinicioglu and P. P. Shenoy, “A new heuristic for learning Bayesian networks from limited datasets: a real-time recommendation system application with RFID systems in grocery stores,” Annals of Operations Research, vol. 244, no. 2, pp. 385–405, 2016.
[6] A. C. Constantinou, Y. Liu, K. Chobtham, and Z. Guo, “Large-scale empirical validation of Bayesian Network structure learning algorithms with noisy data,” International Journal of Approximate Reasoning, vol. 11, no. 3, p. 131, 2021.
[7] T. Talvitie, R. Eggeling, and M. Koivisto, “Learning Bayesian networks with local structure, mixed variables, and exact algorithms,” International Journal of Approximate Reasoning, vol. 115, no. 3, pp. 69–95, 2019.
[8] M. Bartlett and J. Cussens, “Integer Linear Programming for the Bayesian network structure learning problem,” Artificial Intelligence, vol. 244, no. 3, pp. 258–271, 2017.
[9] W. Yin, S. Garimalla, A. Moreno, M. R. Galinski, and M. P. Styczynski, “A tree-like Bayesian structure learning algorithm for small-sample datasets from complex biological model systems,” BMC Systems Biology, vol. 9, no. 1, p. 49, 2015.
[10] H. Dan, M. Wierzb, and B. Lerner, “Joint maximization of accuracy and information for learning the structure of a Bayesian network classifier,” Machine Learning, vol. 109, no. 5, pp. 11–21, 2020.
[11] F. Dondelinger, D. Husmeier, and S. Lébre, “Dynamic Bayesian networks in molecular plant science: inferring gene regulatory networks from multiple gene expression time series,” Euphytica, vol. 183, no. 3, pp. 361–377, 2012.
[12] M. Scanagatta, G. Corani, M. Zaffalon, J. Yoo, and U. Kang, “Efficient learning of bounded-treewidth Bayesian networks from complete and incomplete data sets,” International Journal of Approximate Reasoning, vol. 95, no. 3, pp. 152–166, 2018.
[13] M. Grzegorczyk and D. Husmeier, "Regularization of non-homogeneous dynamic Bayesian networks with global information-coupling based on hierarchical Bayesian models," *Machine Learning*, vol. 91, no. 1, pp. 105–154, 2013.

[14] Y. Xiang, "Direct causal structure extraction from pairwise interaction patterns in NAT modeling Bayesian networks," *International Journal of Approximate Reasoning*, vol. 105, no. 2, pp. 175–193, 2019.

[15] E. J. Hall, K. Um, M. A. Katsoulakis, and D. M. Tartakovsky, "Causality and Bayesian Network PDEs for multiscale representations of porous media," *Journal of Computational Physics*, vol. 394, no. 3, pp. 658–678, 2019.

[16] L. Schiatti, G. Nollo, G. Rossato, and L. Faes, "Extended Granger causality: a new tool to identify the structure of physiological networks," *Physiological Measurement*, vol. 36, no. 4, pp. 827–843, 2015.

[17] T. Qin, L. C. Tsoi, K. J. Sims, X. Lu, and W. J. Zheng, "Signaling network prediction by the Ontology Fingerprint enhanced Bayesian network," *BMC Systems Biology*, vol. 6, no. S3, p. S3, 2012.

[18] M. Dhamala, G. Rangarajan, and M. Ding, "Analyzing information flow in brain networks with nonparametric Granger causality," *NeuroImage*, vol. 41, no. 2, pp. 354–362, 2008.

[19] P. P. Rodrigues, D. Ferreira-Santos, A. Silva, J. Polónia, and I. Ribeiro-Vaz, "Causality assessment of adverse drug reaction reports using an expert-defined Bayesian network," *Artificial Intelligence in Medicine*, vol. 91, no. 9, pp. 12–22, 2018.

[20] C. Chen, X. Wang, H. Yu, N. Zhao, M. Wang, and H. Chen, "An enhanced comprehensive learning particle swarm optimizer with the elite-based dominance scheme," *Complexity*, vol. 202024 pages, Article ID 4968063, 2020.

[21] N. Shahid, I. H. Naqvi, and S. B. Qaisar, "One-class support vector machines: analysis of outlier detection for wireless sensor networks in harsh environments," *Artificial Intelligence Review*, vol. 43, no. 4, pp. 515–563, 2015.