Improved Bayesian network-based for fault diagnosis of air conditioner system

Jiaqi Xu, Qiang Wang*, Juan Zhou, Haiting Zhou, and Jiayan Chen

College of Quality and Safety Engineering, China Jiliang University, Hangzhou, PR China

Received: 26 February 2023 / Accepted: 15 June 2023

Abstract. To solve the problem of fault prediction and diagnosis of household air conditioning, an improved Bayesian network (BN) fault diagnosis model is proposed. Firstly, the orthogonal defect classification (ODC) is used to make statistics and analysis of air conditioning fault data, and the structure of BN fault diagnosis model is built based on the analysis results. Then, genetic algorithm (GA) is used to optimize the conditional probability of network nodes and determine the network parameters. Finally, the cooling and heating failure data of household air conditioning were taken as an example to diagnose. Compared with the traditional BN model, the accuracy of fault diagnosis increases from 81.13% to 92.83%, which verifies the effectiveness of the model.

Keywords: Household air conditioning systems / fault diagnosis / Bayesian network / orthogonal defect classification / genetic algorithm

1 Introduction

Air conditioner has been an essential electrical appliance in residential and commercial buildings, playing a role in regulating the temperature. Among the 7523 complaints about air conditioners in 2021, product quality and safety issues accounted for 25.6%, statistics of China Consumer Association [1]. However, due to the complexity of the air conditioning system, it has a variety of failures, and the causes of failure are all different. Therefore, efficient and accurate diagnosis of faults in the air conditioning system is essential.

Recently, researchers have been dedicated to exploring various fault detection and diagnosis methods for air conditioning systems. These methods can be roughly divided into three categories as model-based, rules-based, and data-driven [2]. Theoretically, the model-based method [3,4] is the best one, because this method constructs mathematic models of complex systems based on various identification methods, physical principles, etc. The fault diagnosis can be accomplished by analyzing the residuals generated by the model. But in practice, because of the difficulty in obtaining accurate mathematic model, model-based method is one of the least commonly used methods in fault diagnosis research. The rules-based methods [5–7] detect and diagnose faults according to the characteristics and rules of system, which obtained from practical experience and expert knowledge. However, the rules-based approach lacks generality because different systems need to develop different rules. Moreover, when this method deals with complex systems, the diagnostic efficiency will be reduced. The data-driven methods are the most commonly used methods for fault diagnosis of air conditioning systems, such as support vector machines (SVM) [8,9], principal components analysis (PCA) [10,11], neural networks [12–14], etc. These methods realize fault diagnosis of the system by analyzing historical data and operating conditions. Nevertheless, it is difficult to obtain complete fault data in practice, leading to various uncertainties in the fault diagnosis process. Due to the complexity of the system and existing uncertainties, sometimes the results are not reliable.

Bayesian network (BN) is a typical data-driven method with powerful capabilities to cope with various uncertainties and have been widely used for fault diagnosis [15]. Based on BN theory, Hu et al. [16] proposed an intelligent fault diagnosis network for variable refrigerant flow (VRF) air conditioning systems, which can make full use of knowledge information and more accurately describe the relationship between faults and variables. Verbert et al. [17] defined different BN fault diagnosis models in consideration of various operating modes of heating, ventilation, air conditioning (HVAC) systems to ensure the reliability and timeliness of diagnosis. Yang et al. [18] developed an automatic air handling units (AHU) fault...
diagnosis method, which effectively diagnosed ten typical faults in fans, dampers, pipes, filters, and sensors. Debashis Dey et al. [19] combined the rule-based method with BN and realized the fault diagnosis of AHU by simulating expert thinking with probability. Although BN has been applied to fault diagnosis of HVAC, chiller plants, AHU, VRF and other air conditioning systems, there are few researches for household air conditioning systems. Therefore, in this study a BN for household air conditioning systems will be constructed to diagnose system faults. Two significant elements of BN are its structure and probability, which are mostly determined by experiential knowledge lacking objectivity. In order to take full advantage of the ability of BN to handle uncertain information in fault diagnosis, several methods have been combined with BN to reduce the subjectivity in fault diagnosis. For instance, rough set theory [20], varying coefficient transfer learning [21], PCA [22], monotonicity constraints [23]. Most of these methods are done by extending the original data and then determining the structure and parameters. However, when the system is very complex, the rationality and effectiveness of data expansion will be affected. Fault diagnosis of household air conditioning systems using BN is not a doodle. Although it is feasible to transform data from qualitative to quantitative descriptions through expert knowledge in the absence of data, the results obtained are subjective and incomplete. Therefore, constructing an objective and scientific BN fault diagnosis model remains a great challenge.

To overcome this challenge, we combined expert knowledge with a small amount of data and achieved the construction of a BN fault diagnosis model in the absence of data by introducing orthogonal defect classification (ODC) and genetic algorithm (GA). We apply this model to fault diagnosis of household air conditioning (1172) statistical data from of a middle-scale city in China to verify the model.

2 Method

2.1 Working principle of air conditioning

Split wall-mounted type air conditioners are the most used type of air conditioners for household, composed of two parts: indoor and outdoor units, which are connected by a power line, a control line, a return pipe and a liquid supply pipe [24]. After connecting the circuit and pipeline, the refrigerant steam of low pressure and low temperature is compressed by the compressor in the outdoor unit and converted into superheated steam of high pressure and high temperature. Then, the generated superheated steam is discharged to the condenser, and the fan in the condenser plays the role of cooling and heat dissipation, so that the superheated steam becomes wet steam co-existing with steam and liquid, and becomes condensed liquid [25]. Finally, the condensed liquid is throttled and reduced by capillary tube to flow into evaporator. In the evaporator, the condensed liquid is vaporized into steam, so that the air around the evaporator is cooled, so as to achieve the effect of cooling. As for the realization of heating function, it is only needed to change the flow direction of the refrigerant through the electromagnetic four-way valve [26]. The working principle of the air conditioner is shown in Figure 1.

2.2 Fault diagnosis model based on improved BN

2.2.1 Structure of BN fault diagnosis model

BN is a directed acyclic graph based on probabilistic inference, which consist of nodes, directed arcs and conditional probabilities of nodes [27]. To construct the BN fault diagnosis model, firstly, we need to identify the variables and select the network nodes. Then the network structure is constructed according to the relationship between the nodes, and the conditional probability table of the network is determined, finally the model is completed. The flow chart of BN fault diagnosis model construction as Figure 2.

(1) Variable identification based on ODC

As the first step to construct BN fault diagnosis model, the quality of variable identification will significantly affect the diagnosis performance. Commonly, variables are identified by asking experts, but when dealing with complex systems such as air conditioning systems, relying only on the experience and knowledge of experts may suffer from incomplete identification and inefficient. ODC is a defect classification method proposed by IBM, which combines the statistical model of defect data and the causal analysis [28]. The basic idea of ODC is to carry out multidimensional analysis of defects by extracting and quantifying key attributes in defects, so that each defect can be uniquely classified into a class in each dimension, avoiding the problem of fuzzy and inaccurate classification. Thus, ODC is introduced into the variable identification of BN fault diagnosis model to improve the efficiency of identification while ensuring the comprehensiveness of identification.
An air conditioning fault statistics and analysis software is developed according to ODC, and the software interface is shown in Figure 3. Through statistical analysis of limited fault data, four key orthogonal defect attributes of event (e.g., system failure), type (e.g., loss of parts), source (e.g., refrigeration system), and object (e.g., condenser) were extracted. These attributes are divided into two sections: found defect and repaired defect. The event attribute belongs to the found defect section and are submitted by inspectors. The section of repaired defect includes the three attributes of type, source, object, which are submitted by repair personnel, as show in Figure 4.

(2) Structure construction

Under the foundation of variable identification, the structure of the BN fault diagnosis model is established, which can qualitatively illustrate the relationship between the variables [29]. At present, the structure of the BN fault diagnosis model for air conditioning is mostly divided into three layers consisting of fault layer, fault symptom layer and additional information. For the sake of more accurate diagnosis and location of air conditioning faults, a four-layer BN fault diagnosis model was constructed by using four orthogonal attributes of air conditioning, namely, event, source, object and type as show in Figure 5.

We consider the event attribute as the top layer of the BN fault diagnosis model, namely the fault layer. Since it belongs to the section of found defect and indicates possible faults in the air conditioning system, its attribute values are considered as leaf nodes. The source attribute belongs to the repair part, which indicates the origin of the air conditioning fault. Therefore, it is considered as the second layer, the fault source layer, and its attribute value is considered as the parent of the leaf node. The object attribute is a more specific description of the source attribute in the orthogonal attribute, so it is regarded as the third layer, that is, the fault object layer. Consequently, the attribute value of the source attribute is a child node of the attribute value of the object attribute. The last layer is the fault type layer, corresponding to the type attribute, which indicates the defect type of the faulty object and the attribute value is the root node of the BN fault diagnosis model.

2.2.2 Parameter learning of BN fault diagnosis model

Since BN is a combination of graph theory and probability theory, parameter learning also has a significant impact on the accuracy of BN fault diagnosis model. The parameter
learning of BN is carried out on the basis of the constructed BN structure to learn the conditional probability table (CPT) of each node. Currently, there are three commonly used parameter learning methods: Expectation Maximization, Maximum Likelihood Estimation (MLE) and Bayesian estimation. However, for these methods, it is hardly possible to gather sufficient data because of the complexity of air conditioning system. GA is an optimization algorithm that simulates the process of natural system evolution according to the law of biological evolution to obtain the optimal solution of the problem, which can effectively solve complex and multivariate nonlinear problems [30]. Thus, the traditional GA is introduced into the parameter learning of BN, so that the network parameters can be searched in the search range given by the experts even in the absence of data, optimizing the CPT while reducing the influence of the subjective factors of the experts on the results.

It is of great significance to construct an appropriate fitness function for GA, which determines the efficiency and accuracy of parameter learning. Therefore, we adopt the maximum likelihood function as the fitness function of GA. Firstly, according to the obtained structure of the BN air conditioning fault diagnosis model G and network node parameters θ, the corresponding parameter set \( D = \{X_1, X_2, ..., X_m\} \) is constructed. The conditional probability distribution of all \( X \) in \( D \) is the same and independent of each other when the parameter a is given, so the likelihood function can be obtained as shown by equation (1) [29].

\[
L(\theta|D) = P(D|\theta) = \prod_{i=1}^{m} P(X_i|\theta).
\]

The parameter \( \theta \) log-likelihood function was obtained by logarithmic equation (1), as shown by equation (2).

\[
l(\theta|D) = logL(\theta|D) = \sum_{i=1}^{m} logP(X_i|\theta).
\]

Let the characteristic function of the sample be \( m_{ijk} = Y(i, j, k : X_i) \), where \( i \) is the number of nodes in the BN, \( j \) is the number of combinations of parent nodes, and \( k \) is the value of the current node. \( m_{ijk} = 1 \) only when the node value is \( k \) and the number of parent nodes is \( j \), and \( m_{ijk} = 0 \) under other conditions. Then the likelihood function is shown in equation (3).

\[
l(\theta|D) = \sum_{i=1}^{m} \log P(X_i|\theta) = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{m_j} Y(i, j, k : X_i) \log(\theta_{ijk}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{m_j} m_{ijk} \log(\theta_{ijk})\]  

Thus, the fitness function can be obtained as shown by equation (4)

\[
F(x) = \max(\log[−l(\theta|D)]).
\]

After completing the construction of the fitness function, the CPT of the network nodes is optimally searched using GA. The process of adopting GA for parameter learning are as follows.

- The initial population of CPT is randomly created based on the search range given by the expert.
- The fitness value of each individual in the population is calculated by equation (4).
- Three genetic operations, selection, crossover and mutation, are carried out according to the size of the fitness value.
- Determine whether the termination condition is met and output the optimal CPT.

The optimal CPT obtained by GA is combined with the network structure of BN fault diagnosis model, and a complete BN fault diagnosis model is obtained, and the fault diagnosis analysis of air conditioning is carried out.

3 Results and discussion

3.1 BN fault diagnosis model for air conditioning

The fault data of 1172 household air conditioning systems in a middle-scale city were collected by the air conditioner fault statistical analysis software. The obtained fault information was analyzed in multiple dimensions based on four attributes: event, source, object, and type to complete the variable identification of the household air conditioning systems, and the identification results are shown in Table 1. According to the results of variable identification and the relationship between the four attributes, the obtained attribute values are regarded as network nodes, and the structure of BN fault diagnosis model for household air conditioning can be constructed, as shown in Fig. 6.

After constructing the structure of the BN fault diagnosis model, the CPT of the network nodes need to be determined and parameter learning is performed based on the obtained network structure. At first, the prior probabilities of each state of root nodes T1-T4 can be obtained by statistically collecting the fault data of household air conditioning systems, as shown in Table 2. The N and Y in the table indicate the state of the node, the former indicates that the current node event has not occurred and the latter indicates that the node event has occurred.

Secondly, the initial search range of GA for other network node parameters is determined by expert empirical knowledge. Taking the conditional probability P (S3/O8, O9) between node S3 and node O8 and node O9 as an example, the initial search range is shown in the Table 3.

Finally, the likelihood value is calculated according to the constructed likelihood function, and GA is used to optimize the parameters of network nodes to obtain the
optimal CPT. Taking $P(S3 | O8, O9)$ between node S3 and node O8 and node O9 as an example, given a maximum genetic generation of 100, a crossover probability of 0.7, and a variation probability of 0.1. Optimization is performed according to the set GA parameters, and the CPT obtained is shown in Table 4. As can be seen from the table, O8 and O9 have about the same effect on S3. However, when both fail at the same time, S3 will definitely fail.

And the optimization curve is shown in Figure 7. The value of the likelihood reflects how well the data match the model parameters. A higher value of likelihood indicates that the parameter is closer to the real one. Obviously, with the increase of the number of iterations, the likelihood degree of the model slowly increases and gradually converges. At this time, there is a big gap between the parameters and the real network parameters. After 35 iterations, the likelihood value reaches the maximum, and the final learned parameter can be used as the optimal parameter of the model.

3.2 Fault diagnosis results and discussion

GeNiE is an open BN simulation software developed by the University of Pittsburgh, which can set the status of each node according to the actual demand and realize the fault

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute value</th>
<th>Symbol</th>
<th>Attribute</th>
<th>Attribute value</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event</td>
<td>Cooling and heating failure</td>
<td>E1</td>
<td>Object</td>
<td>Pipe fittings and valves</td>
<td>O1</td>
</tr>
<tr>
<td></td>
<td>Not working</td>
<td>E2</td>
<td></td>
<td>Condenser</td>
<td>O2</td>
</tr>
<tr>
<td></td>
<td>Water leakage</td>
<td>E3</td>
<td></td>
<td>Evaporator</td>
<td>O3</td>
</tr>
<tr>
<td></td>
<td>Abnormal sound</td>
<td>E4</td>
<td></td>
<td>Compressor</td>
<td>O4</td>
</tr>
<tr>
<td></td>
<td>Blade not working</td>
<td>E5</td>
<td></td>
<td>Sensor</td>
<td>O5</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>E6</td>
<td></td>
<td>Electric control box</td>
<td>O6</td>
</tr>
<tr>
<td>Source</td>
<td>Refrigeration system</td>
<td>S1</td>
<td></td>
<td>Motor</td>
<td>O7</td>
</tr>
<tr>
<td></td>
<td>Electric control system</td>
<td>S2</td>
<td></td>
<td>Louver</td>
<td>O8</td>
</tr>
<tr>
<td></td>
<td>Ventilation system</td>
<td>S3</td>
<td></td>
<td>Air duct and Components</td>
<td>O9</td>
</tr>
<tr>
<td></td>
<td>Box system</td>
<td>S4</td>
<td></td>
<td>Panel</td>
<td>O10</td>
</tr>
<tr>
<td></td>
<td>Accessories</td>
<td>S5</td>
<td></td>
<td>Box body</td>
<td>O11</td>
</tr>
<tr>
<td>Type</td>
<td>Loss of parts</td>
<td>T1</td>
<td></td>
<td>Support</td>
<td>O12</td>
</tr>
<tr>
<td></td>
<td>Missing parts</td>
<td>T2</td>
<td></td>
<td>Water tray</td>
<td>O13</td>
</tr>
<tr>
<td></td>
<td>Abnormal assembly</td>
<td>T3</td>
<td></td>
<td>Others</td>
<td>O14</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>T4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Node</th>
<th>N</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0.5623</td>
<td>0.4377</td>
</tr>
<tr>
<td>T2</td>
<td>0.8439</td>
<td>0.1561</td>
</tr>
<tr>
<td>T3</td>
<td>0.7039</td>
<td>0.2961</td>
</tr>
<tr>
<td>T4</td>
<td>0.8899</td>
<td>0.1101</td>
</tr>
</tbody>
</table>

| P (S3|O8, O9) | S3 = N | S3 = Y |
|---------|--------|--------|
| O8 = N, O9 = N | 1 | 0 |
| O8 = N, O9 = Y | 0.5758 | 0.4242 |
| O8 = Y, O9 = N | 0.4251 | 0.5749 |
| O8 = Y, O9 = Y | 0 | 1 |

Table 1. Variable identification results.

Table 2. The prior probabilities of the root nodes.

Table 3. The initial search range of $P(S3|O8, O9)$.

Fig. 6. The structure of the BN air conditioning fault diagnosis model.

Fig. 7. The optimization curve of $P(S3|O8, O9)$.
diagnosis of the system [31]. Based on the obtained network structure and node parameters, the academic version of GeNIe is adopted to construct the simulation diagram of BN air conditioning fault diagnosis model, as shown in Figure 8.

According to the forward inference results of the model, E1, E2, E3, E4, E5, E6 with the probabilities of 9.15%, 1.62%, 1.32%, 6.58%, 0.76% and 6.94%, respectively. Among them, the probability of E1 is the highest, consequently, the fault diagnosis is carried out to find out the most likely causal chain. The node E1 is set to the occurrence state, namely, \( P(E1 = Y) = 100\%\). Then the inference is carried out to obtain the posterior probability of each node when E1 occurs, and the parent node with higher posterior probability is searched in reverse.

Through the fault diagnosis of E1, it can be found that the main source of the fault is the node S1, with a probability of 90.62%. The four parent nodes of S1 are O1, O2, O3, O4, of which the higher probability is node O2 and node O3 with 43.69% and 49.11%, respectively. Among the four fault types, the highest probability of T1 is 70.61%. Thus, the causal chain of E1 is \{T1-S1-O2-E1\} and \{T1-S1-O3-E1\}. The inference result of E1 is shown in Figure 9.

In order to further verify the effectiveness of BN fault diagnosis model of air conditioning under the conditions of limited, this model is compared with the traditional BN model. The inference results of the traditional BN are shown in Figure 10, and the comparison results of each model with the actual are shown in Figure 11. The evaluation accuracy \( A \) of each model is calculated by the...

![Fig. 7. The optimization curves of \( P(S3=N|O8=N, O9=Y) \).](image)

![Fig. 8. The BN air conditioning fault diagnosis model.](image)
following equation:

$$A = 1 - \left( \frac{1}{n} \sum_{i=1}^{n} \frac{|P(E_i) - P_a(E_i)|}{P_a(E_i)} \right),$$

(5)

where $P(E_i)$ is the probability of occurrence of node $E_i$ inferred by each model, $P_a(E_i)$ is the actual probability of occurrence of node $E_i$, and $n$ is the number of nodes $E_i$.

By comparison with the actual, the accuracy of the improved BN fault diagnosis model is 92.83%, which is higher than 81.13% of the traditional BN fault diagnosis model, indicating that the model has better diagnosis accuracy.

Fig. 9. The inference results of E1.

Fig. 10. The inference results of the traditional BN.

Fig. 11. The comparison results of each model.
4 Conclusion

This paper proposes an improved BN fault diagnosis model, which can diagnose the fault of household air conditioning. As for its BN structure, ODC is used to statistically analyze 1172 air conditioning fault data, and four key attributes of event, source, object, and type are extracted to achieve variable identification of household air conditioning systems. The BN structure was determined based on the identification results and the relationship between the attributes. For parameter learning, the prior probability of the root node is determined from the fault data and the CPT of the other nodes is given by the experts. Then GA is adopted to find the best in the given search range and determine the network parameters of other nodes. After completing the construction of the model, the BN simulation software GeNeI is used to calculate that the most likely failure of the household air conditioning system is the cooling and heating failure, with a probability of 9.17%. By diagnosing the fault, the main factor leading to the fault was the loss of parts of the condenser and evaporator in the refrigeration system. Compared with the conventional BN, the accuracy of the improved BN increased by 11.7%, thus verifying the reliability and effectiveness of the model.

Funding information

This work was supported by technical support project of State Administration for Market Regulation of China (No. 2021YJ024) and Zhejiang Special Support Program for High-Level Personnel Recruitment of China (No. 2019R52017).

References

24. Z. Li, Two-color illustration of air conditioner maintenance from starter to master (Chemical Industry Press, 2012)


29. L. Zhang, H. Guo, Introduction to Bayesian Networks (Science Press, 2006)


Cite this article as: Jiaqi Xu, Qiang Wang, Juan Zhou, Haiting Zhou, Jiayan Chen, Improved Bayesian network-based for fault diagnosis of air conditioner system, Int. J. Metrol. Qual. Eng. 14, 10 (2023)