

Research on Fault Diagnosis of ZPW-2000K Track Circuit Based on RS-BN Algorithm

Junwu LI, Guoning LI

Abstract: For problem of complex fault types and uncertain diagnostic features of the ZPW-2000K track circuit, traditional fault diagnosis mainly adopts manual diagnosis methods, which leads to low automatic diagnosis. This paper proposes a fault diagnosis method based on Rough Sets (RS) reduction model and Bayesian Network (BN) structure learning fusion. Firstly, data mining and feature extraction are performed on the fault data table, and expert knowledge is built into the prior knowledge base. Secondly, the K2 algorithm is used to train the fault feature quantity, and the BN model is built by combining the prior knowledge base. Then, a diagnostic decision table is established through the fault instance, and RS is used for attribute reduction, dimensionality reduction, and simplified model. The MLE algorithm is used again to learn the parameters to obtain the conditional probability table of the model, and the complete BN structure is established based on the RS-BN algorithm. Finally, the comparative analysis of the simplified model and the non-simplified model is carried out. Through the experimental simulation of the ZPW-2000K track circuit fault of a high-speed railway station, the accuracy and effectiveness of the diagnostic method are verified.

Keywords: attribute reduction; Bayesian network; fault diagnosis; rough set; ZPW-2000K track circuit

1 INTRODUCTION

ZPW-2000K non-insulated track circuit is based on the introduction and localization of French UM71 track circuit technology, combined with China's national conditions, and proposed a system with high security, high transmission and high reliability. It is important equipment for China's high-speed railway signal system, and it is also a key equipment to ensure the smooth and safe operation of railway high-speed trains and efficient transportation. Its main function is the open circuit inspection, occupancy inspection and data transmission of circuit. Its failure will seriously affect the transportation efficiency and lead to safety accidents. Therefore, accurate positioning and diagnostic fault are of great significance for the reliable operation of the train.

The ZPW-2000K track circuit has a complex system, high randomness and high uncertainty between fault causes and representations, which make fault diagnosis difficult. In recent years, fault diagnosis techniques such as expert systems, fault trees and neural networks have been continuously developed, and many scholars have applied this method to the fault diagnosis of track circuits. Literature [1] combined with decision tree and expert system, proposed a decision tree fault diagnosis system, but this method relies too much on expert experience and is easily misdiagnosed. In [2], the fault model of the ZPW-2000 track circuit and the fault phenomenon are explored, and the fault model of the system failure tree is established. However, it is extremely difficult to establish an accurate model. In [3], based on the working principle and fault characteristics of the track circuit, the FNN fault diagnosis model is established, but the neural network is easy to fall into local optimum.

The Bayesian Network (BN) is a hotspot in artificial intelligence research and has been successfully applied to many fault diagnoses [4-6]. Rough Sets (RS) is a classical theory for dealing with uncertainty. It can process and analyze incomplete data and has been widely used in many fields such as fault diagnosis, decision control, and pattern recognition [7-9]. Based on the advantages of BN and RS, this paper proposes a RS-BN ZPW-2000K track circuit fault diagnosis method. Firstly, data mining and

feature extraction are performed on faulty instances, and BN data learning and parameter learning are implemented in combination with expert experience. Secondly, RS theory is applied to attribute reduction, dimension reduction and denoising, to obtain minimum diagnostic rules and establish optimal BN. Finally, the BN model was verified and analyzed by the fault instance of the ZPW-2000K track circuit of a high-speed railway station.

1.1 ZPW-2000K Track Circuit System

The ZPW-2000K track circuit is specifically designed for passenger dedicated lines and high-speed rail systems. Its structural composition includes: indoor equipment and outdoor equipment. Indoor equipment includes: transmitter, receiver, loss redundant controller and lightning protection analog network disk. Outdoor equipment includes: tuning matching unit, compensation capacitor, equipment connection line, air core coil and air turbulence transformer. Its system structure [10] is shown in Fig. 1.

2 THEORIES AND ALGORITHMS

2.1 Bayesian Network

BN combines graph theory and probability theory for uncertain reasoning and data analysis. $BN = (G, P)$ consists of two parts: (1) directed acyclic graph $G = (I, E)$, where $I = \{A_1, A_2, A_3, \dots, A_n\}$ is the set of nodes, $E = \{E_1, E_2, E_3, \dots, E_n\}$ is the set of edges, and the directed edges reflect the inter-node dependencies. (2) The conditional probability table (CPT) represents the prior probability of each node and describes the probability distribution of the nodes. Assuming that the BN nodes are $A_1, A_2, A_3, \dots, A_n$, its joint probability distribution is as shown in Eq. (1).

$$p = \prod_{i=1}^n -p[A_i | \pi(A_i)] \quad (1)$$

The search-scoring method is a common method of BN structure learning. Accurate network structure can be obtained through data learning. Assuming that the sample contains n variables, the number of networks that exist through structure learning is as shown in Eq. (2).

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \frac{n!}{i!(n-i)!} 2^{i(n-1)} f(n-i) \quad (2)$$

The complexity of the BN structure will grow exponentially following the number of nodes n , and the BN structure learning is also considered to be the NP-Hard problem [11]. This paper uses the classical K2 algorithm, as shown in Eqs. (3) and (4).

$$P(D | Bs) = \prod_{i=1}^n score(i, pa_i) \quad (3)$$

$$score(i, pa_i) = \prod_{j=1}^{q_i} \left[\frac{\Gamma(\partial_{ij})}{\Gamma(\partial_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\partial_{ijk} + N_{ijk})}{\Gamma(\partial_{ijk})} \right] \quad (4)$$

where Bs represents the network structure and D represents the data.

The parameter learning of BN is relatively mature, so the Maximum Likelihood Estimation (MLE) is used to study the parameters of BN.

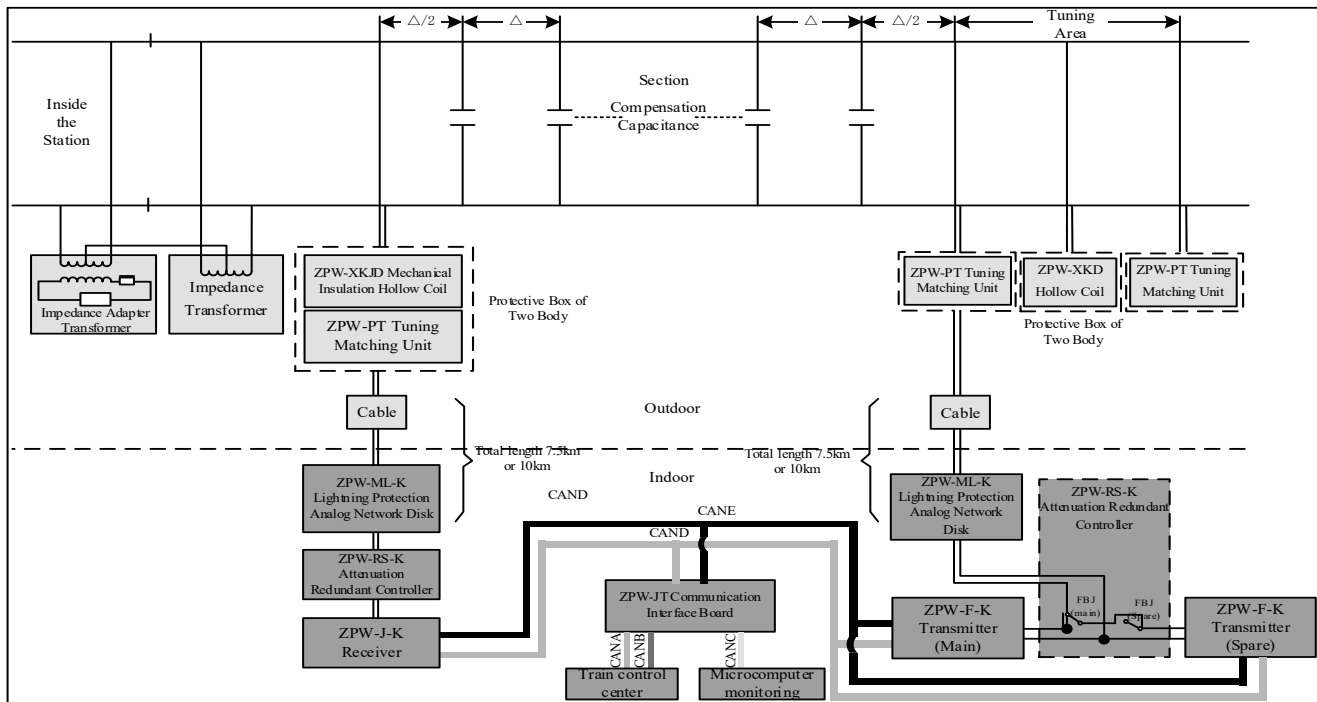


Figure 1 System structure of ZPW-2000K

2.2 Rough Sets

RS is a classical mathematical theory that deals with fuzzy and inaccurate problems [12]. Its attribute reduction can eliminate redundant information, simplify conditional attributes, and generate minimal decision rules without changing decision-making ability. The theory consists of the following three parts:

- (1) Knowledge expression system: $S = \{U, R, V, f\}$ is an ordered quaternion, where domain $U = \{X_1, X_2, \dots, X_n\}$ is a collection of sample points, $R = C \cup D$ is a set of conditional attributes and a set of decision attributes.
- (2) Unresolvable relationship: the type division relationship between objects in U . Attribute subset $B \subseteq R$, if objects $X_i, X_j \in U, \forall r \in B$, if and only if $f(X_i, r) = f(X_j, r)$, X_i and X_j are indistinguishable, denoted as $Ind(B)$.
- (3) Decision-making: it embodies the diagnostic rules of the sample from conditional attributes to decision attributes. Let $S = \{U, R, V, f\}$, $a \in R$, if $Ind(R - \{a\}) = Ind(R)$, it indicates that a is unnecessary in R , otherwise it is a necessary attribute.

3 FAULT DIAGNOSIS OF ZPW-2000K TRACK CIRCUIT BASED ON RS-BN ALGORITHM

The fault data table used in this paper is derived from the track circuit monitoring warning information and the fault repair form filled out by the maintenance personnel. The fault data table is recorded in natural language and has no rules. Usually, the computer cannot process the data, so data mining and unified coding of the fault data are required.

3.1 Establish a Diagnostic Knowledge Base

The establishment of the fault diagnosis knowledge base is traditionally based on expert experience. It is usually recorded in natural language and script, subjectively biased and not easily accessible. When a certain type of failure occurs, there is no experience to follow. In this paper, data mining and feature extraction are performed on the fault data table, and the diagnostic knowledge base is built by experts prior. Through data mining, the potential implicit relationship between fault points is found, which enriches the prior knowledge and reduces the dependence on expert knowledge. The establishment process is shown in Fig. 2.

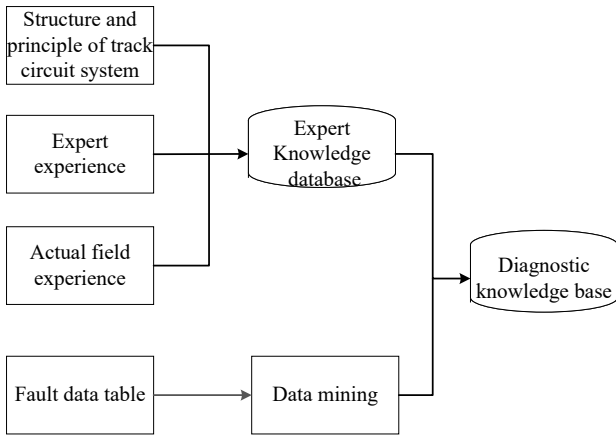


Figure 2 The process for establishing a diagnostic knowledge base

3.2 Determining the Diagnostic Model Node of BN

In this paper, the fault information is divided into three layers: the fault trigger layer, the fault mode layer and the fault feature layer, which represent the fault cause, the fault occurrence module and the fault representation phenomenon. Its hierarchical relationship is shown in Fig. 3.

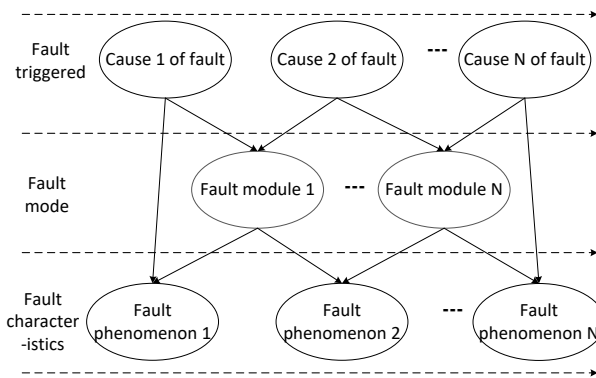


Figure 3 Fault node causal hierarchy

The fault node hierarchy table of the ZPW-2000K track circuit is established according to the causal hierarchical relationship between the diagnostic knowledge base and the faulty node, as shown in Tab. 1, Tab. 2, and Tab. 3.

Table 1 Fault mode node

Number	Definition	Number	Definition
M1	Indoor sending end	M4	Outdoor receiving end
M2	Indoor receiving end	M5	Small rail end
M3	Outdoor sending end		

Table 2 Fault initiating node

Number	Definition	Number	Definition
R1	24KV power off	R13	Compensation capacitor failure
R2	Transmitter failure	R14	Receive lead wire breakage
R3	Receiver failure	R15	Send lead wire disconnection
R4	Send level line error	R16	Receive tuning failure
R5	Receive level line error	R17	Send tuning failure
R6	Transmit path failure	R18	Receive SPT disconnection
R7	Receive path failure	R19	Receive SPT short circuit

Table 2 Fault initiating node (continuation)

Number	Definition	Number	Definition
R8	Send cable disconnection	R20	Send SPT disconnection
R9	Send cable short circuit	R21	Send SPT short circuit
R10	Receive cable disconnection	R22	impedance transformer fault
R11	Receive cable short circuit	R23	Track adjustment error
R12	Attenuator failure		

Table 3 Fault feature node

Number	Definition	Number	Definition
S1	Power output voltage failure	S10	Low rail receiving voltage
S2	Low send of power output voltage	S11	Low out voltage of small rail
S3	Low receive of power output voltage	S12	Low in voltage of small rail
S4	Low send voltage of the device	S13	Voltage overrun of small rail
S5	Low receive voltage of the device	S14	Red light belt
S6	Low cable send voltage	S15	Train degraded operation
S7	Low cable receive voltage	S16	System work interruption
S8	Low out voltage of main rail	S17	System transfer interruption
S9	Low transmission voltage of rail surface	S18	System works abnormally

3.3 Establish a BN Model for a Prior Diagnostic Knowledge Base

Based on the prior diagnosis knowledge base, the fault node causal hierarchy of Fig. 3, and the fault node information of Tabs. 1 to 3, a BN diagnostic model of the ZPW-2000K track circuit is established based on the diagnostic knowledge base. The faulty node adopts discrete coding, and its state value takes values: 1-occurred, 0-did not occur. The model was built using Matlab's BNT toolbox, as shown in Fig. 4.

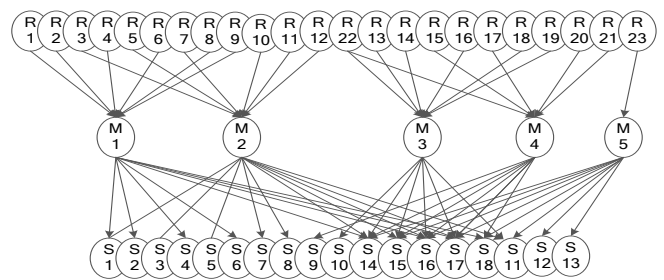


Figure 4 BN model based on prior diagnosis knowledge base

3.4 Establish a BN Model Based on K2 Learning Algorithm

The K2 algorithm is a local search algorithm for data optimization. It combines the hill climbing search algorithm and the Bayesian scoring index to optimize the network model with high accuracy and excellent search efficiency. Because of the established fault data table, the information is relatively complete, so the K2 algorithm is used to mine the potential causal relationship between the fault points. Take the M1-indoor sender as an example, and use the Lean_Struct_K2() function to model, as shown in Fig. 5.

$$I(Q, P) = H(Q) - H\left(\frac{Q}{P}\right) = \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (7)$$

Calculate attribute combination dependencies and use the combination of minimum dependent values as the best attribute group. The optimal decision diagnosis rule is obtained based on the optimal attribute group, and an optimal diagnosis decision table is established. The condition attribute reduction in Tab. 4 is: $\{S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S14, S15, S17\}$, and the dimension of the fault feature points is reduced to 14, which reduces the complexity of the model.

(4) Establish a BN diagnostic model based on RS-BN algorithm

According to the optimal decision diagnosis rule and the information fusion BN model, a BN fault diagnosis model based on RS-BN algorithm is established, as shown in Fig. 8.

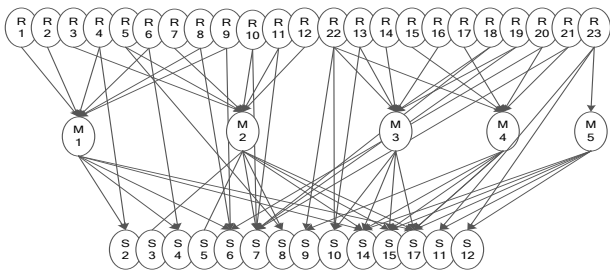
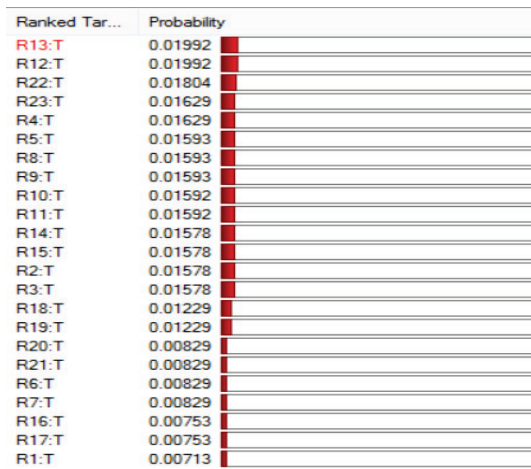
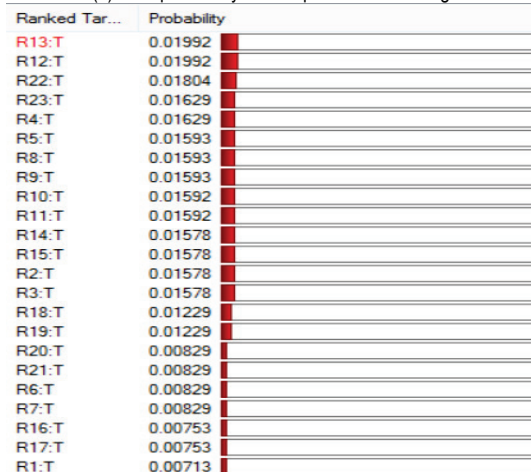


Figure 8 BN model based on RS-BN algorithm



(a) Prior probability of BN1 parameter learning



(b) Prior probability of BN2 parameter learning

Figure 9 Comparison of parameter learning node probabilities

3.7 Determine the Parameter Model of BN

The accuracy of parameter learning depends on the accuracy of the build model. After establishing the optimal BN model, it is necessary to determine the prior probability of each faulty node and establish a conditional probability table (CPT) of the faulty node. In this paper, the MLE algorithm [13] is used to study the parameters of BN. The BN model of the information fusion in Fig. 6 is defined as BN1, and the BN model of the RS attribute reduction in Fig. 8 is defined as BN2, and the prior probability of the faulty node is learned by GenIe2.0 software. The results are shown in Fig. 9.

It can be seen from the comparison of (a) and (b) in Fig. 9 that, in the case where the number of samples is the same, the reduced BN2 model and the unreduced BN1 model have the same conditional probability for the fault nodes obtained by the MLE parameter learning. It shows that model reduction cannot only simplify the model, but also get the same prior probability.

4 INSTANCE VERIFICATION OF FAULT DIAGNOSIS

4.1 Instance Verification 1

Select a fault data from the fault instance as a diagnostic example of the BN model, as shown in Tab. 5.

Table 5 Fault data instance

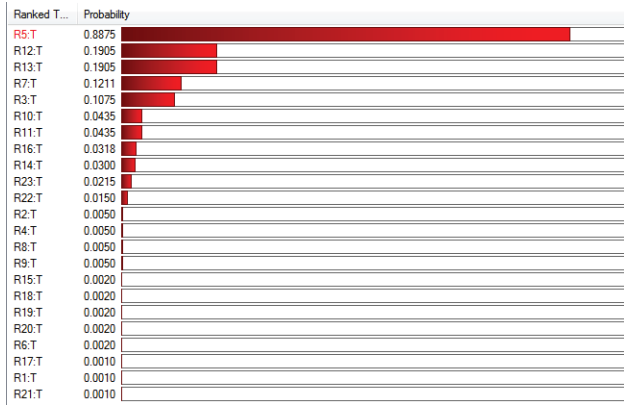
Number	Fault description	Fault mode	Fault cause	Maintenance recommendations
1	March 23, 2017, 17:40:18 View console display information : X1LQAG Red light belt failure, low out voltage of main rail, Normal driving	Indoor receiving end	Receive level adjustment line error	It is recommended to check the receiving channel: receiving level adjustment short line problem.

In the fault instance of Tab. 5, extract the fault feature set $\{\text{'red band fault'}$, $\text{'derailment out voltage'}\}$, and the fault feature set for the diagnostic models BN1, BN2: $T1 = \{0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0\}$,

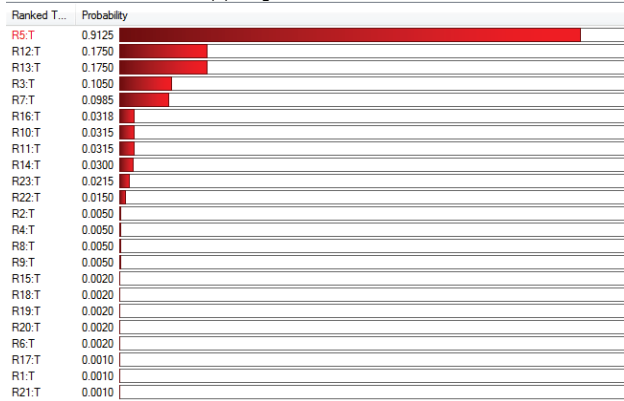
$T2 = \{0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0\}$. Add evidence $T1$, $T2$ to the inference engine and infer the probability of failure occurrence: $p(R = 1|T=1)$. The BN1 and BN2 models were diagnosed and inferred using the Linked Tree Algorithm (JT) of GenIe2.0 software. The diagnosis results are shown in Fig. 11.

It can be seen from the diagnosis results of a) and b) in Fig. 10 that under the known evidences $T1$ and $T2$, the maximum failure probability diagnosed by the BN1 model and the BN2 model is $R5$ (receive level line error). It is consistent with the actual cause of the failure and verifies the accuracy of the model. For the BN2 model, the posterior probability of $R5$ is 0.9125, which is significantly higher than the 0.8875 of the BN1 model, while the probability values of $R12$ and $R13$ are decreased, indicating that the BN2 model is reduced by the attribute of RS, which improves the fault

knowledge clarity and fault diagnosis ability better than BN1.



(a) Diagnosis result of BN1



(b) Diagnosis result of BN2

Figure 10 Comparison of model diagnosis reasoning results

4.2 Instance Verification 2

For the 100 pieces of verification data of the ZPW-2000K track circuit, the fault is extracted according to the failure mode ratio, and the extracted data is shown in Tab. 6.

Table 6 Fault verification data extraction

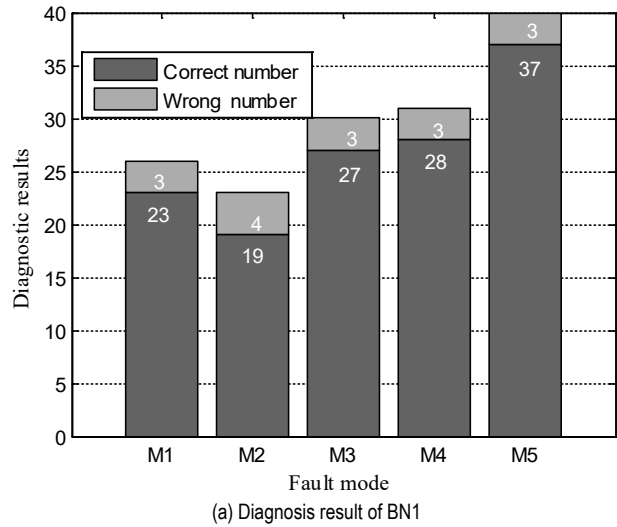
Number	M1	M2	M3	M4	M5
Fault module	Indoor send	Outdoor receive	Outdoor send	Outdoor receive	Small rail
Amount	26	23	30	31	40

The BN1 and BN2 models are diagnosed and inferred using the *j* tree_inf_engine () function in Full-BNT toolbox of BN. The comparison of the diagnostic results is shown in Fig. 12.

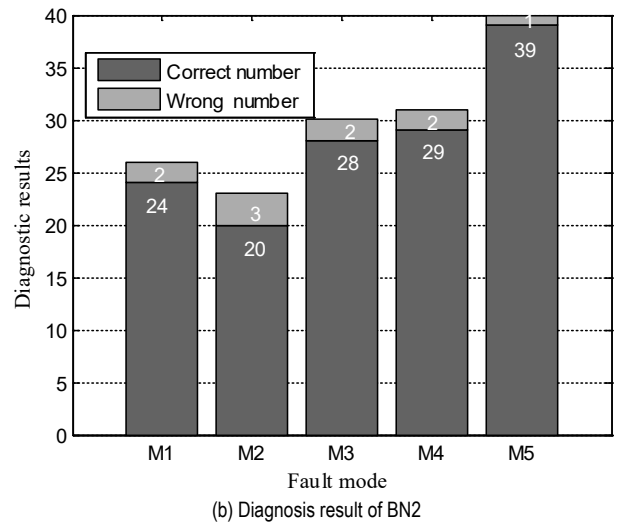
The comparisons of the fault diagnosis accuracy rates of the BN1 and BN2 model are shown in Tab. 7.

As can be seen from Fig. 11 and Tab. 7, the average diagnostic accuracy of the BN1 model is 89.33%, and that of BN2 is 93.33%. In particular, there are only 3 and 1 misdiagnosis data for M5, and the accuracy rate of fault diagnosis is 92.50% and 97.50%, which greatly verified the effectiveness of the two models. Although the diagnostic accuracy of M2 is relatively low, it has reached more than 82%. This is because the number of instances of M2 is 75. The lower number of instances leads to the diagnostic model of this module which is not very accurate. As the sample size continues to increase, the accuracy of fault diagnosis will continue to increase. The

diagnostic accuracy of the BN2 model is higher than that of BN1. This is because the RS attribute reduction eliminates model redundancy and unnecessary attributes, eliminates interference and noise, and makes the model diagnosis rules more accurate and clearer. At the same time, the model structure is simplified, the misdiagnosis rate is reduced, and the diagnostic accuracy is improved. The BN2 diagnostic model is more efficient and has high likelihood.



(a) Diagnosis result of BN1



(b) Diagnosis result of BN2

Figure 11 Comparison of inference diagnosis

Table 7 Fault diagnosis correct rate comparison

Fault mode	Diagnostic accuracy	
	BN1	BN2
M1	(23/26) 88.46%	(24/26) 92.30%
M2	(19/23) 82.61%	(20/23) 86.96%
M3	(27/30) 90.00%	(28/30) 93.33%
M4	(28/31) 90.32%	(29/31) 93.55%
M5	(37/40) 92.50%	(39/40) 97.50%
Average accuracy rate (%)	89.33%	93.33%

5 CONCLUSION

- (1) Through the full integration of ZPW-2000K system structure and expert experience, the fault prior diagnosis knowledge base and BN model structure are established. The potential internal hidden relationship between the fault points is mined by using the fault instance of the ZPW-2000K track circuit and the classic K2 algorithm learning BN structure.

- (2) The BN model established by the a priori diagnostic knowledge base and learned by the K2 algorithm is used for information fusion, and the advantages of the two are combined to establish a new BN model to further improve the accuracy of the diagnostic model.
- (3) We fully combine the advantages of BN and RS. RS theory is used to reduce the attributes of the initial decision table, reduce the dimension, eliminate redundant and non-nuclear attributes, reduce the model, and generate the simplest diagnostic rules to establish the best BN model structure.
- (4) Diagnostic analyses of the reduced model and the unreduced model are carried out by taking the actual fault of a high-speed railway station as an example. By comparison, the RS-BN fault diagnosis model proposed in this paper is compact in structure and efficient in diagnosis, with high reliability and high practical likelihood. It provides practical decision-making support for on-site electrical maintenance personnel to quickly and effectively diagnose track faults, which has practical significance for the development of fault diagnosis technology for ZPW-2000K track circuit.

6 REFERENCES

- [1] Wu, Z. P. (2016). Fault Diagnosis of ZPW-2000A Track Circuit Based on Decision Tree Algorithm. Southwest Jiaotong University.
- [2] Hu, L. Q. (2018). Fault Analysis and Countermeasure of ZPW-2000 Track Circuit Based on Fault Tree. Southwest Jiaotong University.
- [3] Huang, Z., Wei, X., & Liu, Z. (2012). Fault Diagnosis of Railway Track Circuits Using Fuzzy Neural Network. *Journal of the China Railway Society*, 34(11), 54-59.
- [4] Liang, X., Wang, H., Guo, J., et al. (2017). Bayesian Network Based Fault Diagnosis Method for On-board Equipment of Train Control System. *Journal of the China Railway Society*, 39(8), 93-100.
- [5] Jiang, Q., Huang, B., Ding, S. X., et al. (2016). Bayesian Fault Diagnosis with Asynchronous Measurements and Its Application in Networked Distributed Monitoring. *IEEE Transactions on Industrial Electronics*, 63(10), 6316-6324. <https://doi.org/10.1109/TIE.2016.2577545>
- [6] Liu, B., Chen, Y. (2015). *Auditing for Data Integrity and Reliability in Cloud Storage*. Springer New York. https://doi.org/10.1007/978-1-4939-2092-1_17
- [7] Wu, G., Yuan, H., Song, Z., et al. (2017). Fault Diagnosis for Power Transformer Based on Rough Set and Multi-class Support Vector Machine. *High Voltage Engineering*, 1(11), 202-208.
- [8] Wang, Y., Tang, J., & Zhao, Y. (2017). Prediction of risks in flight operations based on rough sets and support vector machine. *China Safety Science Journal*, 27(9), 158-163.
- [9] Lu, R., Li, X., & Chen, Y. (2017). Rough Set Bidding Decision Method for Railway Construction Project. *Journal of Transportation Systems Engineering and Information Technology*, 17(1), 14-19.
- [10] Li, W. T. (2017). *High-speed railway train operation control technology: ZPW-2000 series non-insulated track circuit system*. Beijing: China Railway Publishing House.
- [11] Ramazzotti, D., Nobile, M. S., Antoniotti, M., et al. (2018). Structural Learning of Probabilistic Graphical Models of Cumulative Phenomena. *International Conference on Computational Science*. https://doi.org/10.1007/978-3-319-93698-7_52
- [12] Ma, W., Huang, Y. Y., Li, H., et al. (2014). Image Segmentation Based on Rough Set and Differential Immune Fuzzy Clustering Algorithm. *Journal of Software*, 1(11), 2675-2689.
- [13] Balov, N. (2011). A Gaussian mixed model for learning discrete Bayesian networks. *Statistics & Probability Letters*, 81(2), 220-230. <https://doi.org/10.1016/j.spl.2010.11.009>

Contact information:

Junwu LI, Corresponding author
School of Automation & Electrical Engineering,
Lanzhou Jiaotong University,
Lanzhou 730070, China
E-mail: lijunwu_lzjtu@163.com

Guoning LI
School of Automation & Electrical Engineering,
Lanzhou Jiaotong University,
Lanzhou 730070, China