Fault Diagnosis Technology of Railway Signal Equipment based on Improved FP-Growth Algorithm

Yueqin Yang
Urban Rail Transit College, Jilin Railway Technology College, Jilin, 132299, China

***Abstract*—The rapid development of computer information technology has made various fault diagnosis and detection technologies emerge in an endless stream. As one of the main transportation vehicles, the detection efficiency of fault diagnosis of railway signal equipment has important practical significance for maintaining the overall safe operation of railways. On the basis of the traditional FP-Growth algorithm, improve the TF-IDF algorithm to realize the weight discretization of text features, and realize the improvement of the FP-Growth algorithm by adjusting the adaptive confidence and support. The FP-Growth algorithm will be improved. FP-Growth algorithm is used for performance tests and applications. The results show that the minimum running time-saving of the proposed algorithm is 1500ms, and the average accuracy of P@N exceeds 85%, which is higher than that of the FP-Growth algorithm (81.4%) and VSM algorithm (82.1%). The PR curve of the improved algorithm is closer to the upper right, which effectively ensures the processing of correlated data, and the overall average precision performance under the influence of positive and negative signal-to-noise ratio values exceeds 95%. And the signal curve generated by the algorithm. The error range of the data under the four fault types of track circuit, turnout, signal, and connecting line floats between 1% and 5%. The improved FP-Growth algorithm can effectively analyze railway fault types and data. Perform analysis and data processing to minimize diagnostic errors.**

Keywords—FP-growth algorithm; gailway; signal equipment; fault diagnosis; relevance; frequent itemsets; vector space model

I. Introduction

The railway has always been in the backbone position in the modern comprehensive transportation system. It is an important infrastructure and the main means of transportation for the public in China, and also an important industrial sector to promote the national economy and social development. With the rapid development of China's railways in recent years, the safety requirements for railway transportation are becoming increasingly strict. With the in-depth development and application of information technology, the era of big data has come [1]. In the aspect of railway, the important premise to ensure traffic safety is to ensure the normal operation of railway signal equipment. Under the railway big data application platform, using data mining technology to process fault text data is one of the current research hotspots. The fault text data is extracted and mined, and the final results can guide maintenance personnel to process the fault equipment. As important technical equipment, railway signal equipment plays a key role in organizing and directing train operations, ensuring running safety and transportation efficiency. However, the types and causes of railway signal faults are complex, and traditional detection methods are difficult to detect railway signals. Once the signal equipment fails, it is easy to cause a major safety accident, which will seriously affect the maintenance efficiency and endanger the life safety of the staff. The rapid development of information technology provides a variety of diagnostic methods for the detection of railway signals. Different methods are used to diagnose railway faults, such as the application of neural networks, fuzzy logic fault diagnosis, expert diagnosis systems, etc. Data analysis of fault signals is one of the key contents of detection. Common text data association analysis algorithms include Apriori Algorithms and Frequent Pattern Growth (FP-Growth) algorithms. The correlation algorithm can provide a practical basis for fault diagnosis based on ensuring the correlation analysis of signal detection data. The FP Growth algorithm is often used in data mining processing because of its fewer traversal times and the ability to compress the data, but its application effect will be affected by the data type [2]. Therefore, it is studied to improve the traditional FP Growth algorithm and discretize the weight value of the text feature based on considering the type and characteristics of the railway fault causes, to ensure the objectivity and standardization of data processing, and adaptively set the confidence and support to reduce the non-objectivity of manually set parameters, to detect and analyze the railway signal fault problems with the improved FP Growth algorithm. The research ensures that the improved FP Growth algorithm can better analyze and process the fault information by discussing the status quo of relevant research methods and designing experimental methods, and verifies the effectiveness of the proposed algorithm with the help of simulation experiments and algorithm comparison, in order to provide new data diagnosis and treatment means for future railway fault maintenance and operation maintenance, so as to ensure the normal operation of railway construction.

II. Related Works

The construction and operation of the railway system and the information transmission of the command and dispatch of the trains are all centered on the signal equipment. With the increase of traffic volume, it is very important to strengthen the fault detection of the signal equipment to ensure the normal development of railway safety operation and maintenance. Different scholars have put forward different views on the detection and processing of railway signals. For example, Fang et al. used FP growth algorithm to carry out secondary equipment defect monitoring, and used frame algorithm to establish an exception model [3]. Jiang et al. used FP growth algorithm for association analysis and data mining, but found that this algorithm still has problems of low query efficiency and frequent item set traversal [4]. Yang and other scholars used the text mining intelligent fault classification model of signal equipment to realize the feature extraction and transformation of text information, and combined the characteristics of different classifiers. The ensemble study of learning classification was carried out under the experimental results. The experimental results showed that the model has high accuracy and recall [5]. Yuan and his team used the biased classification algorithm to divide the data flow of railway fault signals, and divided the data into line segments. The data block matrix is constructed to realize the integrated calculation of data consumption, and the experiments have verified that the algorithm can effectively reduce the memory consumption and greatly improve the data operation efficiency [6]. Han combines the hash table technology with the original FP growth algorithm, and conducts database experiments on it. It is found that the improved algorithm effectively shortens the running time, and its advantages gradually become prominent with the decrease of support [7]. Scholars such as Jiyang QI fully consider other external and objective factors when carrying out fault diagnosis, not only using a single indicator such as fault symptoms as the judgment standard, but also using the Weibull distribution model to calculate the failure rate based on, and carry out example analysis, the experimental results prove that the proposed algorithm can greatly improve the accuracy of fault diagnosis, and effectively reduce the interference of multi-information fusion on fault diagnosis [8]. Due to the diversity and complexity of information storage and application, word segmentation algorithms and topic models are used to extract features from information data, and then use support vector machines to diagnose equipment faults. It is found that the algorithm can provide algorithm guidance value for railway operation and maintenance [9]. Zhang scholars use association rules to achieve fault diagnosis of railway signals, and use TF-IDF algorithm to achieve fault feature extraction. The average diagnostic rate of this algorithm is 11.44% higher than that of Bayesian network [10]. Scholars such as Cheng integrate particle swarm optimization and minimum entropy to improve back - rolling the product method is used to detect the faults of rolling bearings, and the auxiliary coordinate transformation is used to optimize the algorithm. In the experimental simulation, it is found that the algorithm has good deconvolution performance, which effectively improves the diagnosis accuracy [11]. Zheng et al. designed constraints and screened data features to diagnose UAV flight faults. The experimental results show that the rule base of the improved algorithm has good scale growth and good application effect [12]. Ding scholars added the tail attribute to the traditional FP Growth algorithm to achieve tire data analysis. The experimental results show that the improved algorithm can effectively find the quality influencing factors and data feature mining [13]. Li et al. The fault feature vector acquisition and neural network improvement are carried out with the particle swarm optimization algorithm. The improved algorithm can show a short training time under the fault signal data of different states and its diagnosis accuracy rate is above 95%, which has a good application potential [14]. Jin and his team of scholars the power grid fault diagnosis based on layer data fusion is used to construct the model to process the data of different information sources, and the restoration path algorithm based on the fault rules is used to make the optimal judgment of the power supply restoration path. The results show that the method has strong practicability and effective the accuracy of power grid fault diagnosis has been improved [15]. At the same time, Cho and other scholars have achieved a safe and efficient train by reviewing and analyzing the control unit and current relay interface of the automatic train protection system, and detecting the fault type after analyzing the cause of the fault [16].

Fuzzy logic method, artificial neural network method and other methods will be interfered by artificial thinking to a certain extent when detecting railway signals and the application of FP-growth algorithm to railway fault detection has attracted the attention of many scholars. Fang et al. used FP growth algorithm to conduct secondary equipment defect monitoring, and used Hadoop framework and Mapreduce algorithm to establish an exception model. The results show that the algorithm can effectively find equipment abnormal defects and provide a data basis for diagnosis [17]. In order to reduce the disadvantage that feature extraction of fault data mostly depends on manual work, Jiangquan scholars use convolutional neural network to perform two-dimensional transformation and feature extraction of signals, and the algorithm proposed has good application timeliness under different cost and load conditions [18]. Scholars such as Syakur MA use the FP-Growth algorithm to understand the clustering results and classification of sales data and customer data, and use It can better explore the correlation between the subject and object of commodity consumption, and combine it with the frequent pattern tree to determine the item set. The test results show that the algorithm can effectively mine the data relationship, improve the average processing time of the data, and achieve better smooth operation efficiency. Guarantee [19]. Scholars such as Ariestya WW use frequent patterns and a priori algorithms to process selection, transformation and evaluation of data, and then conduct marketing strategy analysis to analyze the degree of consumer demand. The experimental results show that this method is used in data processing. It has good application performance and effectively provides a supporting basis for the formulation of consumer marketing strategies [20].

The railway fault signal detection can analyze its fault signal according to the classification idea. Some scholars use text data mining, biased classification algorithms and frequent patterns to identify and analyze the fault signal. Considering the complexity of the fault signal and the multi-source of the data, some scholars use wavelet decomposition, particle swarm optimization algorithm and convolution neural network to transform the features, and they have achieved certain results. Some research results of FP growth algorithm application, defect detection and data classification also show that it can be used in better classification and recognition. Based on this research idea, the text feature data with discrete weight is proposed as the input value of FP Growth algorithm, and the algorithm is improved through adaptive setting, effectively taking into account the characteristics and changes of railway signal fault data.

III. FAULT DIAGNOSIS ANALYSIS OF RAILWAY SIGNAL EQUIPMENT BASED ON IMPROVED FP-GROWTH ALGORITHM

A. *Application of Improved FP-Growth Algorithm in Signal Fault Diagnosis*

The continuous increase of railway lines and the continuous updating of signal equipment make the fault diagnosis records in the operation and maintenance of the current railway system mostly exist in the form of text, and the traditional railway text data are mostly analyzed from the aspects of safety management, and are not based on the type of fault diagnosis. It is difficult to improve the efficiency and accuracy of fault diagnosis by analyzing the characteristics and causes of faults. It is difficult to improve the efficiency and accuracy of fault diagnosis. The phenomenon of multiple meanings, the difference of signal fault text records and the existence of redundant information make the diagnosis of railway signal faults difficult. There are problems such as untimely and inaccurate problems, which further increase the difficulty of railway signal equipment diagnosis. Therefore, on this basis, the research is based on text data to carry out feature extraction and data mining to improve the accuracy of fault diagnosis, that is, to establish an association rule based the fault detection of the fault signal can realize the diagnosis and real-time analysis of the fault type and cause from different aspects. First, the fault signal text is segmented and feature extracted to ensure the vectorization of its data. The Vector Space Model (VSM) will the feature item is used as the smallest unit of the text, and the corresponding feature weight is given to it, and the mathematical expression of the model is shown in the formula [21].

$$V(d_i) = (t_1, w_{i1}; t_2, w_{i2}; t_3, w_{i3}; \ldots; t_n, w_{in}) \tag{1}$$

In formula (1), is t_n the first n feature item w_{in} of the text, and is the feature weight corresponding to the d_i feature item t_n. The weighted algorithm TF-IDF (Term Frequency - Inverse Document Frequency) algorithm is often used in data mining, and its core idea is to evaluate the importance of a word to a certain document in the corpus, its calculation formula is shown in the formula.

$$W_{t.d} = tf_{t.d} * idf_{t.d} \tag{2}$$

formula (2), is $W_{t.d}$ the weight value of the $tf_{t.d}$ feature value in the text, t and $idf_{t.d}$ is the word frequency and the inverse document frequency, respectively. The calculation formulas of the word frequency and the inverse document frequency are shown in formula (3) and formula (4).

$$tf_{t.d} = n_{t.d} / \sum k_{nt.d} \tag{3}$$

$$idf_{t.d} = \log\left(N / N_{t.d} + 1\right) \tag{4}$$

equations (3) and (4), $n_{t.d}$ is the number of times the feature value appears in the $\sum k_{nt.d}$ text, is the sum of the number of feature values in the text, is the total number N of texts, and is the number $N_{t.d}$ of texts with feature values in the text. The individual differences of text data and the appearance of synonyms will increase the difficulty and accuracy of fault diagnosis. The research adjusts the traditional Term Frequency-Inverse Document Frequency (TF-IDF) algorithm to obtain the formula (5)).

$$W_{t.d}' = tf_{t.d} * idf_{t.d} = tf_{t.d} * \log\left(N / N_{t'.d} + 1\right) \tag{5}$$

In the formula, is $W_{t.d}'$ the final weight of the feature item, $tf_{t.d}$ and t_n is $idf_{t.d}$ the word frequency and inverse document frequency of the feature item, $N_{t'.d} = N_{tn.d} + N_{tm.d}$, respectively, and is the sum of the number of texts t_n containing the feature item t_n, t_m. At the same time, the fault feature weight is discretized and divided into three categories according to its importance., and the value of 0-2 indicates the low, general and high importance. Fig. 1 is a schematic diagram of the level division under the weight discretization.

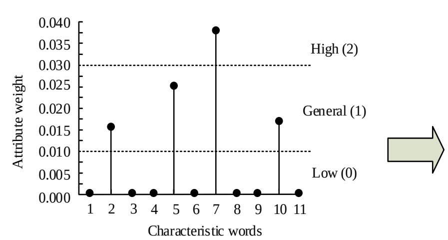
When marking the keywords extracted by the algorithm, it is evaluated by means of the hit rate (Recall), the accuracy rate (Precision) and the comprehensive evaluation (F measure).

$$\mathrm{Re}\,call = \frac{\left|Y_{real} \cap X_{re}\right|}{Y_{real}} * 100\% \tag{6}$$

$$\mathrm{Pr}\,ecision = \frac{\left|Y_{real} \cap X_{re}\right|}{Y_{real}} * 100\% \tag{7}$$

$$F - measure = \frac{2 * \mathrm{Re}\,call * \mathrm{Pr}\,ecision}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision} * 100\% \tag{8}$$

In formulas (6)-(8), it Y_{real} represents the keyword set obtained by the measurement algorithm, which is Y_{real} the keyword set under manual marking. To diagnose and detect railway signal faults, it is possible to determine the fault feature layer, fault type layer and the fault cause layer realizes the subdivision and classification of fault information, and then provides a reference for subsequent countermeasures and improvements. For a variety of railway fault signal information, the text information data is often analyzed by means of association analysis. The value correlation or association pattern between itemsets is found in a large number of databases, emphasizing the interdependence between data, and then helping decision making and judging the feasibility of data. The key lies in the decomposition of association tasks and the formulation of association rules. The running framework of association rules is shown in Fig. 2.



Characteristic words	Attribute weight	Weight dispersion result
1	0	0
2	0.015	1
3	0	0
4	0	0
5	0.025	1
6	0	0
7	0.038	2
8	0	0
9	0	0
10	0.017	1
11	0	0

Fig. 1. Schematic diagram of grade division under weight discretization.

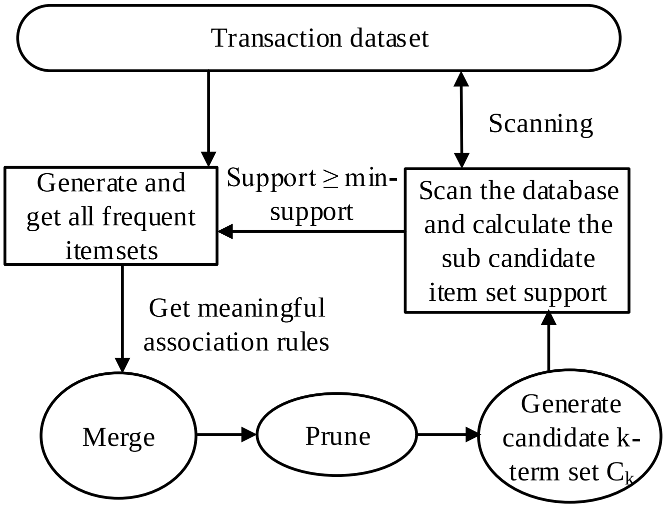


Fig. 2. Schematic diagram of association rule operation framework.

Association rules mainly realize the processing of itemsets by scanning the data set and determining the support degree, which is $A \rightarrow B$ an implication of the form, which A, B is an itemset containing one or more items, $A \cap B = \Phi$, which is the antecedent A, B of the association rule. and consequent. The constraints of association rules are the minimum values of support and confidence, and their mathematical expressions are shown in equations (9) and (10).

$$\sup(A) = count(A \subseteq T)/|D| \tag{9}$$

Equation (9) is the calculation formula of the support degree, which is expressed as the ratio of the frequency of a certain item set appearing in the transaction to the corresponding transaction number, where D, A is the corresponding transaction database and item set, and $\sup_{-} count$ is the corresponding support degree count. Confidence is expressed as the ratio of the probability that two related rules appear at the same time to the probability that a single rule appears $confidence(A \Rightarrow B)$, and its mathematical expression is Equation (10).

$$confidence(A \Rightarrow B) = \sup_{-} count(A \cup B)/\sup_{-} count(A) \tag{10}$$

Frequent Pattern-Growth (FP-Growth), as a typical algorithm for solving association problems, is different from the Apriori algorithm that scans each potential frequent itemset, which only scans the database twice and does not generate candidate sets. In the case of compressing the database containing frequent itemsets into a frequent pattern tree (Frequent Pattern-Tree, FP-tree), the frequent itemsets are obtained. The FP-Growth algorithm reduces the processing of itemsets to a certain extent. In the transaction database processing, the sorted frequent itemsets are used to reconstruct the database, then the root node of the tree is created, and the transactions are compressed into the FP-Tree branch in the form of merge prefix to realize the frequent itemsets. Fig. 3 shows the construction process of the algorithm pattern number.

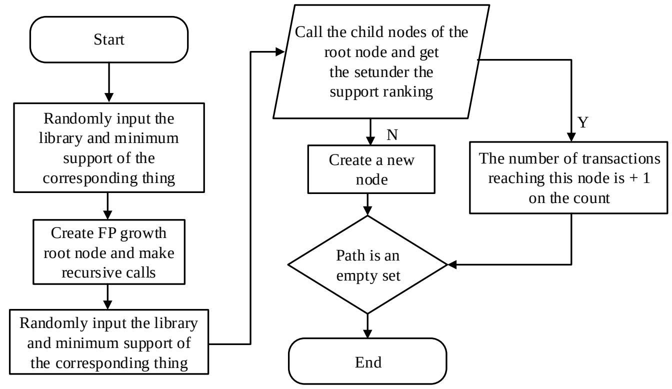


Fig. 3. Construction process of algorithm pattern number.

However, the traditional FP-Growth algorithm will artificially set the minimum support degree and the minimum confidence degree when dealing with related transactions, which makes the selection of algorithm parameters to a large extent subjective and arbitrary, and it is difficult to ensure the authenticity of the algorithm and the operation. Feasibility. Also, the ambiguity of determining the size of frequent itemsets will also cause the pattern tree to consume more memory during the traversal process, thereby reducing the efficiency of data mining. Therefore, the research is based on this to eliminate the defects of artificially setting parameters., based on the characteristics of the data itself, the minimum support and confidence are set. The improved FP-Growth algorithm first sorts the support number of the transaction set in descending order, and builds a sequence table corresponding to "order-value", in which There is a monotonically decreasing relationship trend between the sequence value and the ordinal value. The polynomial curve fitting function is set by the numbers corresponding to the ordinal-value table, and its mathematical expression is shown in formula (11).

$$y = f(x) = \sum_{i=0}^{t} m_i * x^i \tag{11}$$

In formula (11), y, x is the serial number value and the sequence value, t is the support number, i is the item set, and m is the upper limit of the interval satisfied by the function. Perform the quadratic derivative of the obtained function to obtain the second-order derivative function, as shown in formula (12). Show:

$$y' = f''(x) = \sum_{i=2}^{t} i(i-1) * m_i * x^{i-2} \tag{12}$$

When the sequence value of the second derivative function of the fitting curve appears for the first time in the interval, $y' = f''(x) = 0$ the sequence number value under the point is rounded down, and the rounded result is used as the algorithm parameter. Then set according to the adaptive support degree Threshold, get deleted items less than support degree, merge them to get sub-database, build its TP-Tree with sub-database, at this time, the data sub-database realizes the association of frequent itemsets under the recursive call tree structure Mining, but also to ensure that the obtained results are strong association rules.

B. Design of Railway Fault Signal Diagnosis System

The detection and diagnosis of railway signals based on the improved FP-Growth algorithm is to mine the rules in the fault storage knowledge base, and the reasoning flow chart is shown in Fig. 4.

By judging the fault characteristics of railway signals and saving them in a comprehensive database, and judging whether there are countermeasures for the type of problem, if so, the fault diagnosis results will be output directly, if not, the information characteristics will be screened in the historical fault database. Extraction, reasoning and correlation judgment, and consider a variety of factual basis in the process of fault type and cause analysis, and directly seek manual diagnosis when it is difficult to obtain a clear diagnosis result. This fault diagnosis process can effectively realize the processing of information data, detection, judgment, etc. effectively ensure the efficiency and accuracy of detection. The diagnosis of railway fault signals is the basis for staff to carry out equipment maintenance. The design of railway signal equipment fault maintenance system can make the diagnosis results intuitive, so as to carry out corresponding diagnosis countermeasures Research. On the basis of improving the FP-Growth algorithm, the research proposes a multi-level signal diagnosis structure including fault diagnosis, fault data management, post-maintenance management and user management. The picture shows the architecture diagram of the railway signal equipment fault repair prototype system.

As shown in Fig. 5, the architecture diagram of the fault repair prototype system includes the presentation layer, the business logic layer and the data access layer to realize the input and storage of the fault data, the logical interaction of the data business information and the human-computer interaction of the personnel related to the fault information., and then realize the cause and effect of fault data and interactive processing. The most important one is the business logic layer. The fault word database is established through the fault signal record, and then the knowledge base and data reasoning are used to diagnose the strong correlation. The fault causes of the rules are identified, and this is used as a reference for the maintenance plan to judge whether it needs manual modification. Continuously improve the fault case database of railway signal equipment and improve the overall fault diagnosis efficiency.

At the same time, when combining association rule mining and manual identification and diagnosis results, in order to avoid the omission of important fault causes, $P@N$ is used as the evaluation index, and its meaning is expressed as the accuracy rate of the retrieved fault cause as the real cause, and its calculation formula as shown in formula (13).

$$P@N = \sum_{j=1}^{n} I(j)/n \tag{13}$$

In formula (13), n is the set number of output fault causes, $I(j)$ and is the ratio between the input value of the test set conforming to the fault cause under the diagnosis algorithm and the actual cause of the fault, if the diagnosed cause is consistent with the actual cause, then $P@N$ The value of 1.

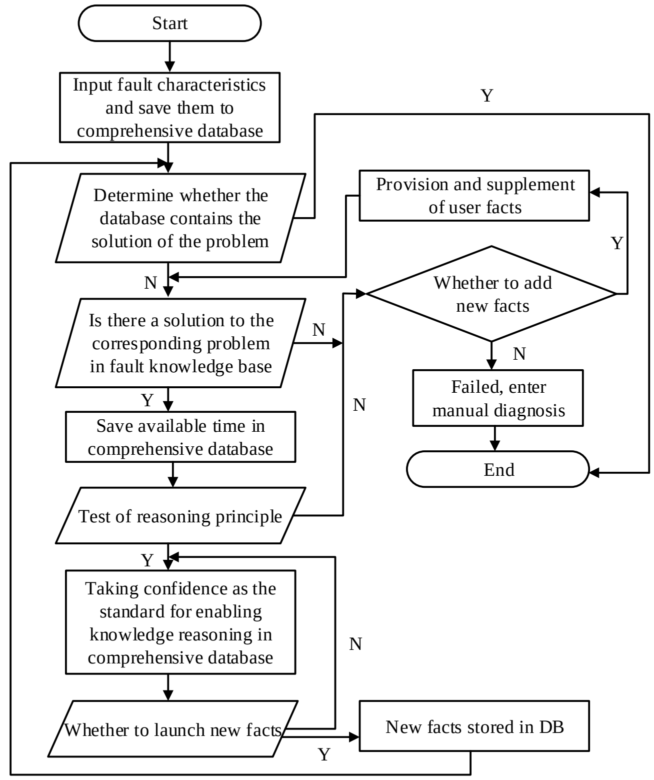


Fig. 4. Schematic diagram of fault diagnosis process.

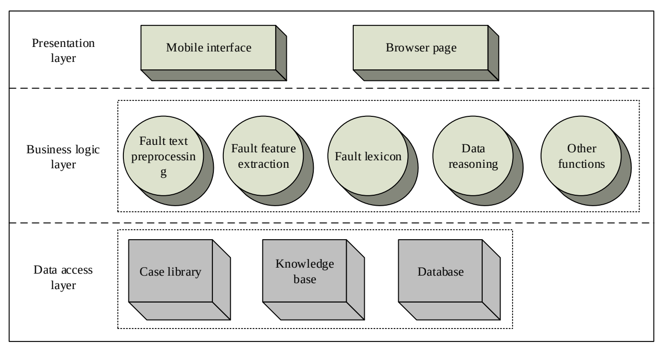


Fig. 5. Architecture diagram of railway signal equipment fault maintenance prototype system.

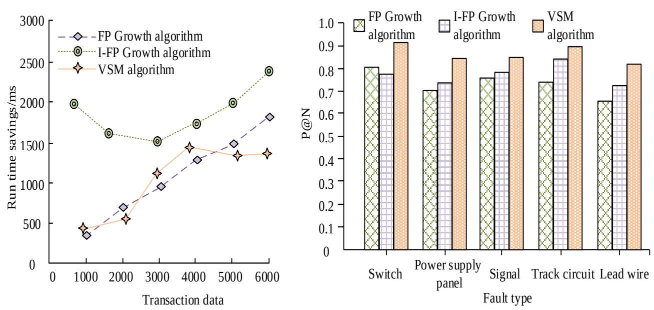
IV. Application Analysis of Railway Signal Equipment Fault Diagnosis under the Improved FP-Growth Algorithm

The development momentum of railway construction projects has shown a good trend with the enhancement of China's economic strength. As one of the main means of transportation in China, the safety monitoring awareness of railways increases with the development of safety accidents. Research is improving FP- With the support of the Growth algorithm, the fault signal system is designed and analyzed experimentally. First, the experimental development environment is set up. The fault detection application environment includes two parts: browser and mobile terminal. The main parameter setting conditions are shown in Table I.

TABLE I. DEVELOPMENT ENVIRONMENT OF THE SYSTEM

Software type	Development environment		Version
Browser	***Operating environment***		***Windows7 64bit***
	Hardware environment	Hardware environment	i7-4790 CPU @ 3.60GHz 8G R
		Development tool	MyEclipse8.5
	Software environment	Database	MySQL (5.5.2)
		JDK	JDK1.8
		The server	Tomcat7.0
Mobile end	Operating system		Android4.0.4 or above
	Hardware environment		RAM 512M ROM 1G or above
	Development tool		Android Studio

The key to the detection of railway faults is to accurately judge the causes of the fault types. However, the levels and contents involved in railway operation and maintenance are relatively complex, and the original fault storage data has a certain amount of meaningless content. Research Using the VSM algorithm as the comparison algorithm, randomly select transaction data in the fault database for analysis, and ensure the consistency of the corresponding algorithm parameter settings when comparing the algorithms. The VSM algorithm uses the space vector of the feature items to represent the text information. The text similarity problem is transformed into a dimensional space vector problem. The data is counted and sorted on the algorithm running time and correlation accuracy, and the results are shown in Fig. 6.

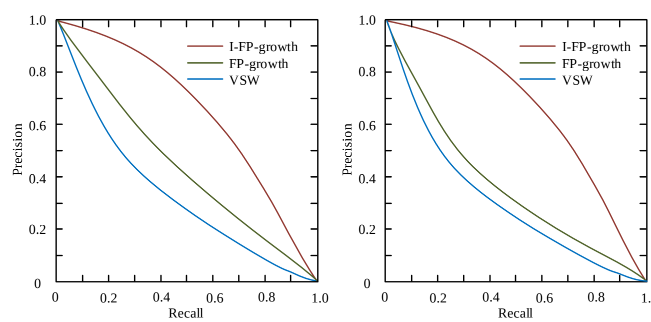


(a) Comparison results of running time of three algorithms

(b) Comparison results of accuracy of three algorithms under different fault types

Fig. 6. Comparison results of running time and fault correlation accuracy of three algorithms.

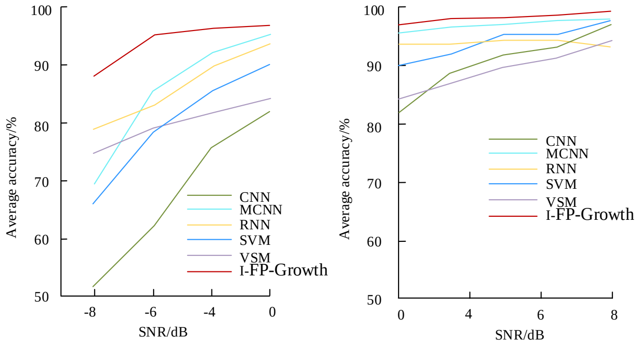
In Fig. 6, "I - FP Growth algorithm" represents the improved FP-Growth algorithm proposed by the study. It can be seen from Fig. 6(a) that with the increase of transaction data volume, the performance of the three algorithms. The running time is quite different. The specific performance is that the overall running time of the traditional FP-Growth algorithm shows an upward trend. The minimum and maximum values are 350ms and 1700ms respectively, while the improved FP-Growth algorithm is It shows a downward trend first and then upward trend, and takes 3000 data sets as the dividing point, the minimum and maximum values are 1500ms and 2400ms, respectively. The VSM algorithm shows a relatively obvious numerical fluctuation. The reason is that the improvement of FP-Growth The algorithm can reduce the frequent itemsets occupying memory by dividing the database into subsets, and avoid the redundant database under the condition tree generation of the original algorithm. The results in Fig. 6(b) show that the P@ N is higher than the other two algorithms, and its average accuracy exceeds 85%, which is about 5 times and 9 times that of the other two algorithms. The average accuracy of the traditional FP-Growth algorithm and the VSM algorithm are 81.4% and 82.1, respectively. %. The reason is that the improved FP-Growth algorithm can mine the correlation between the fault information by discretizing the weight feature words, which greatly improves the diagnostic accuracy of the algorithm. However, the traditional FP Growth algorithm and VSM algorithm lack of consideration of the potential information in the fault text data, and when encountering texts with similar contents but with synonyms, the calculated similarity will be far from the actual. Because the number of topics is artificially determined, there is a certain degree of experience, which ultimately leads to less fault features, and it is difficult to deeply mine the associated meaning of fault information, so it is difficult to maintain the accuracy of inferred information. The accuracy of the three algorithms is calculated, the result is shown in Fig. 7.



(a) Dataset A (b) Dataset B

Fig. 7. PR curves of three algorithms.

The research uses the Precision-Recall (PR) curve as a judging tool. The more the PR curve is convex to the upper right, the better the data classification effect is. The PR curves on A and dataset B are closer to the upper right, and their average accuracy reaches 87.15%, which is much higher than the 83.12% and 81.06% of the FP-Growth algorithm and the VSM algorithm. The reason is that the VSM algorithm and the FP-Growth algorithm. It is difficult to reduce the negative impact of objective factors and algorithm parameters on the results of data processing only with vector features and correlations. The improved FP-Growth algorithm can effectively take into account the characteristics of railway signal faults for correlation analysis, reducing redundant and redundant data. Then, the statistical results of the algorithm are organized, and the results are shown in Fig. 8.

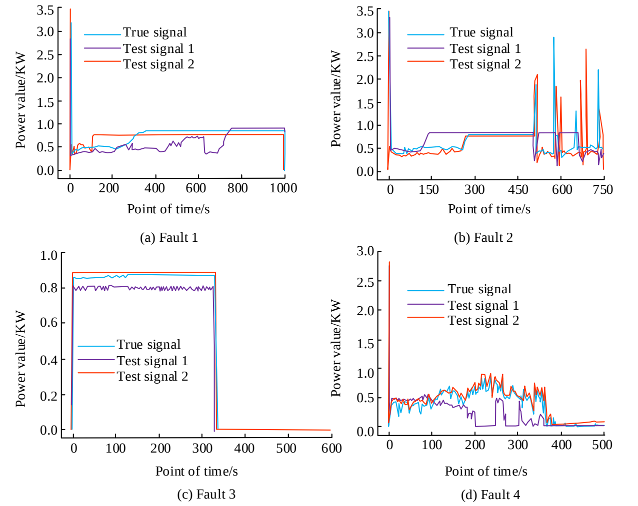


(a) Average precision results of different algorithms under positive signal-to-noise ratio

(b) Average precision results of different algorithms under negative signal-to-noise ratio

Fig. 8. Average precision results of different algorithms under signal-to-noise ratio.

Fig. 8 shows the average precision results of SNR under different algorithms. Fig. 8(a) shows that when the SNR is negative, the average precision results shown by various algorithms are quite different. When the ratio is -8 to -6, the average precision is from large to small as I- FP-Growth > RNN > VSM > MCNN > SVM > CNN, and when the signal-to-noise ratio is -6 to -0, the average precision is from large to the smaller is I- FP-Growth > MCNN > RNN > SVM > CNN. The average precision of the I- FP-Growth algorithm and the CNN algorithm shows a good upward trend, and the I-FP-Growth algorithm is better than the CNN algorithm. In Fig. 8(b), when the signal-to-noise ratio is positive, the overall average accuracy performance of the algorithms from good to poor is I -FP-Growth > MCNN > SVM > CNN > RNN > VSM, where I -FP-Growth algorithm The average precision shown is more than 95%, and the maximum value is 98.2%. The accuracy performance is second only to the I -FP-Growth algorithm for the MCNN algorithm, and the maximum value of the MCNN algorithm is 96.3%. As the ratio of signal power and noise power, the signal-to-noise ratio is one of the important indicators to characterize the signal quality. It may be difficult to express a good communication effect if signal power or noise power is used to represent the quality of signal processing. However, the reduction of signal-to-noise ratio will increase the bit error rate, which will greatly reduce the visual presentation effect of signal quality. When the SNR value is very low, it indicates that the working environment suffers from great noise interference, and the method in this paper still obtains satisfactory diagnostic accuracy in the task of adding very low SNR value noise signal. Although the noise causes a great difference in the data distribution of the training and test data sets in the target task, the method in this paper can still achieve effective fault diagnosis, which shows that the method in this paper has excellent robustness, can complete high-precision and stable fault diagnosis when the noise is seriously disturbed, and has good feature stability. The FP-Growth algorithm is less affected by the value of the signal-to-noise ratio, and is in an upward trend as a whole, and its average precision performance is better than other algorithms. The proposed algorithm is tested for signal generation and selected under different fault types. The signal data of the real curve is compared with it, and the results are shown in Fig. 9.



(a) Fault 1

(b) Fault 2

(c) Fault 3

(d) Fault 4

Fig. 9. Signal generation results under four fault types.

Fig. 9 represents the traditional FP-Growth algorithm, "Test signal 2" represents the improved FP-Growth algorithm, Fig. 9(a)-(d) are the track circuit fault, switch fault, signal fault and the connecting line is faulty. The results in Fig. 9(a) show that the test data 2 has a large difference between the real signal and the data before the time point of 300s, but with the increase of time, the magnitude of the power difference between the two is not More than 5%. However, the error between the signal curve graph of test data 1 and the real signal curve is large, and the fluctuation between the time point range (0-750) s is large. In Fig. 9(b), the test data the generated signal curve of 2 is continuous, and the time point at which the peak appears is basically the same as that of the real signal, and the trend is roughly the same. The signal curve of the improved algorithm in Fig. 9(c) is in the range of time points (0-380)s. The error range amplitude The value is less than 3%. The traditional FP-Growth algorithm curve has short and small fluctuations in a large range, and the error value compared with the real signal curve exceeds 6.8%. In Fig. 9(d), the curve of the improved FP-Growth algorithm is compared with the real signal. The fluctuation of the curve is basically the same, and the error margin value is less than 1%. However, the signal trend curve of the traditional FP-Growth algorithm has begun to deviate from the real signal curve after the time point is 100s, and the signal missing phenomenon occurs at the time point of 400s. The above results show that the improved FP-Growth algorithm can effectively detect and analyze signals of different fault types, and its accuracy and algorithm performance are better, which can effectively provide data support for fault diagnosis and maintenance countermeasures.

V. Conclusion

At present, the detection and diagnosis of railway faults mostly rely on manual operation. The complexity and redundancy of data types increase the difficulty and accuracy of fault analysis, and it is difficult to provide a guarantee for railway safety operation and maintenance. Based on the analysis and extraction, the improved FP-Growth algorithm is used to detect the fault signal data, and its performance test and application analysis are carried out. The results show that the improved FP-Growth algorithm takes 3000 data sets as the dividing point. The maximum and minimum time are 2400ms and 1500ms, respectively, which are significantly higher than the 350ms and 1700ms of the traditional FP-Growth algorithm, and its P@N accuracy is 5 times and 9 times that of the other two algorithms. The classification performance of data mining is better, and its average accuracy reaches 87.15%, which is much higher than 83.12% and 81.06% of the FP-Growth algorithm and VSM algorithm. The average accuracy of the Growth algorithm is better than other algorithms, with a maximum value of 98.2%. The error values between the signal generation results and the real signals under the four fault types are small, all lower than 5%. The above results show that the improved FP-Growth algorithm proposed by the study is less affected by the value of the signal-to-noise ratio, and its performance is better than other algorithms in terms of running time saving and accuracy, and can effectively diagnose fault data. The research only selects text data as the data source for analysis, which makes the fault problems it solves more limited and less comprehensive. For example, the data generated by microcomputer monitoring may also be an important source of fault problems. Strengthening the diagnosis and analysis of multiple data is one of the important contents to be studied in the future. At the same time, the research only uses the feature weight number as the source of fault analysis, without taking into account the post fault risk, Therefore, it is necessary to carry out combined weighting in the later stage to ensure the accuracy of fault information features.

VI. Findings

The research is supported by Provincial Project: Research and Practice of “223” Talent Training Model for Railway locomotive operation and maintenance professional (No. JGJX2022D721).

References

[1] E. Wang, D. Yang, C. Wang, et al., “Optimized Fault Detection Algorithm Aided by BDS Baseband Signal for Train Positioning,” Chinese Journal of Electronics, vol. 29(1), pp.34-40, 2020.

[2] N. Xu, X. Xiu, J. Liu, et al., “A novel scheme for multivariate statistical fault detection with application to the Tennessee Eastman process,” Mathematical Foundations of Computing, vol. 4(3), pp.167-184, 2021.

[3] X. Fang, W. Huang, D. Ye, et al., “Application of distributed parallel FP-growth algorithm in secondary equipment defect monitoring,” Power System Protection and Control, vol. 49(8), pp.160-167, 2021.

[4] X. Jiang, X. Xiao, W. Wen, et al., “Operation optimization of thermal power unit based on Parallel FP-Growth algorithm,” Guangxi Electric Power, vol. 42(1), pp.11-14,19, 2019.

[5] L. Yang, P. Li, R. Xue, et al., “Intelligent classification of faults of railway signal equipment based on imbalanced text data mining,” Tiedao Xuebao/Journal of the China Railway Society, vol. 40(2), pp.59-66, 2018.

[6] C. Yuan, P. Li, and Y. Zhang, “Parallel processing algorithm for railway signal fault diagnosis data based on cloud computing,” Future Generation Computer Systems, vol. 88, pp.279-283, 2018.

[7] C. doves Han, “Research and improvement of FP-Growth algorithm,” Journal of Longyan University, vol. 38(5), pp.8-13, 2020.

[8] Q. I. Jiyang, and L. Wang, “Fault diagnosis method of equipment based on multi-information fusion,” International Equipment Engineering and management, vol. 25(2), pp.77-97, 2020.

[9] Z. Zhong, T. Tang, and F. Wang, “Research on fault feature extraction and diagnosis of railway switches based on PLSA and SVM,” Tiedao Xuebao/Journal of the China Railway Society, vol. 40(7), pp.80-87, 2018.

[10] X. Zhang, “Research on fault diagnosis method of railway signal equipment based on association rules,” Railway Standard Design, vol. 66(4), pp.175-181, 2022.

[11] Y. Cheng, N. Zhou, W. Zhang, et al., “Application of an improved minimum entropy deconvolution method for railway rolling element bearing fault diagnosis,” Journal of Sound & Vibration, vol. 425:53-69, 2018.

[12] W. Zheng, Q. Li, and Y. Zhang, “Uav flight fault diagnosis based on improved FP-Growth algorithm,” Science & Technology Innovation and Application, vol. 2019(13), pp.16-19, 2019.

[13] D. Ding, Y. Yi, and M. Li, “Tire quality data analysis based on FP-growth algorithm,” China Mechanical Engineering, vol. 30(2), pp.244-251, 2019.

[14] Y. Li, and Z. Pan, “Analog circuit fault diagnosis methods based on rbf neural network,” Telecommunications and Radio Engineering, vol. 78(13), pp.1193-1201, 2019.

[15] M. Jin, H. Li, J. Xie, et al., “Rule based intelligent fault diagnosis and recovery power delivery algorithm,” Dianli Xitong Baohu yu Kongzhi/Power System Protection and Control, vol. 47(4), pp.176-181, 2019.

[16] B. H. Cho, and H. Lee, “Study on operation characteristics and failure causes of LED type railway traffic lights,” Journal of the Korean society for railway, vol. 21(8), pp.767-773, 2018.

[17] X. Fang, W. Huang, D. Ye, et al., “Application of distributed parallel fp-growth algorithm in secondary equipment defect monitoring,” Power System Protection and Control, vol. 49(8), pp.160-167, 2021.

[18] J. Zhang, S. Yi, G.U.O. Liang. et al., “A new bearing fault diagnosis method based on modified convolutional neural networks,” Chinese Journal of Aeronautics, vol. 33; 167(02), pp.54-62, 2020.

[19] M. A. Syakur, B. K. Khotimah, E. Rochman, et al., “Using K-means algorithm and FP-growth base on FP-tree structure for recommendation customer SME,” Journal of Theoretical and Applied Information Technology, vol. 96(4), pp.1102-1113, 2018.

[20] W. W. Ariestya, W, Supriyatin, and I. Astuti. “Marketing strategy for the determination of staple consumer products using fp-growth and apriori algorithm,” Jurnal Ilmiah Ekonomi Bisnis, vol. 24(3), pp.225-235, 2019.

[21] C. Tan, G. Ji, R. Liu, et al., “LTSA-LE: A local tangent space alignment label enhancement algorithm,” Tsinghua Science and Technology, 26(2), pp.135-145, 2020.