

Article

An Online Classification Method for Fault Diagnosis of Railway Turnouts

Dongxiu Ou ¹, Yuqing Ji ¹ , Lei Zhang ^{2,*}  and Hu Liu ³

¹ Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Shanghai 201804, China; ou.dongxiu@tongji.edu.cn (D.O.); jiyuqing@tongji.edu.cn (Y.J.)

² School of Transportation Engineering, Tongji University, No.4800 Caoan Road, Shanghai 201804, China

³ School of Rail Transit, Shanghai Institute of Technology, No.100 Haiquan Road, Shanghai 201418, China; liuhu@sit.edu.cn

* Correspondence: reizhg@tongji.edu.cn

Received: 17 July 2020; Accepted: 14 August 2020; Published: 17 August 2020



Abstract: Railway turnout system is a key infrastructure to railway safety and efficiency. However, it is prone to failure in the field. Therefore, many railway departments have adopted a monitoring system to monitor the operation status of turnouts. With monitoring data collected, many researchers have proposed different fault-diagnosis methods. However, many of the existing methods cannot realize real-time updating or deal with new fault types. This paper—based on imbalanced data—proposes a Bayes-based online turnout fault-diagnosis method, which realizes incremental learning and scalable fault recognition. First, the basic conceptions of the turnout system are introduced. Next, the feature extraction and processing of the imbalanced monitoring data are introduced. Then, an online diagnosis method based on Bayesian incremental learning and scalable fault recognition is proposed, followed by the experiment with filed data from Guangzhou Railway. The results show that the scalable fault-recognition method can reach an accuracy of 99.11%, and the training time of the Bayesian incremental learning model reduces 29.97% without decreasing the accuracy, which demonstrates the high accuracy, adaptability and efficiency of the proposed model, of great significance for labor-saving, timely maintenance and further, safety and efficiency of railway transportation.

Keywords: turnout system; fault diagnosis; incremental learning; scalable recognition

1. Introduction

Due to its characteristics of comfort, speed, energy-saving and environmental protection, the high-speed railway has an increasing ratio of carrying passengers and cargo, which puts increasing requirements on the safety of the railway system. According to the 2019 China Railway Statistical Bulletin [1], in 2019, the number of railway passengers sent across the country was 3.66 billion in China, an increase of 8.4% over the previous year. Moreover, in Europe, passenger and cargo transportation in railways has been, respectively doubled and tripled in the past decade [2] according to the report of the European Commission.

The turnout, which is usually installed between two or more strands of tracks, is responsible for switching the train direction, consequently being a key part of the whole signal system in railway infrastructure [3–5].

However, due to its complicated mechanical and electrical structure, exposure to the outdoor environment and the need to be frequently pulled, the turnout is more prone to failure. Based on the statistics given by China Railway Jinan Group Co, Ltd. in 2016 [6] and China Railway Guangzhou Group Co, Ltd. in 2018 [7], the number of turnout failures, respectively accounts for approximately 36%

and 55% of the total railway signal infrastructure failures. It is obvious that turnout system damage will increase a lot along with the increase in railway transportation demand.

In addition, the cost of turnout maintenance is also very high. For example in England, the maintenance cost of turnout is about 3.4 million pounds per 1000 km of railway [2]. Moreover, according to the report of the International Railway Union, the annual maintenance cost of the turnout system accounts for about 30% of the entire maintenance budget [8]. Moreover, in the United States, the cost is about ten times that of ordinary tracks [9]. Therefore, research on the turnout fault-diagnosis method is very important in not only improving railway reliability and safety, but also in reducing maintenance costs.

The main fault-diagnosis methods have been proposed can be divided into three categories, they are expert system-based, model-based and feature-based.

Xue et al. [10–13] designed and implemented a fault diagnosis and analysis system for railway signal equipment by establishing an expert knowledge base of railway signal equipment. This method overcomes the overreliance on the mathematical model and can effectively identify the faults. However, the application of expert system needs to obtain sufficient and comprehensive fault diagnosis expert knowledge and experience, which has certain difficulty. In addition, it is also difficult to generalize and summarize the experience of experts into the base [14].

Eker, O. F. et al. [15,16] established the mathematical model for turnouts to achieve fault recognition and prediction of turnout operating state. This method does not require a large number of samples, however, in practice, it is very difficult to establish an accurate mathematical model for complex equipment.

For the feature-based diagnosis method, Witczak, Marcin et al. [17] proposed a turnout fault-diagnosis method based on a neural network. They determined the input parameters and designed a variable threshold to adapt to new faults. Silmon, JA et al. [18] used qualitative trend analysis to extract the data trend of normal and fault samples and established a rule matching mechanism for different fault types.

However, these methods are more costly to deal with new types of faults. Therefore, incremental learning is introduced in this paper. In other engineering areas, some researchers have proposed different classification methods (fault-diagnosis method) based on incremental learning. Generally, these methods can be divided into two different kinds: hierarchy-based methods and improved models [19].

For hierarchy-based methods, a particular learning hierarchy is designed for incremental learning. Nong Ye et al. [20] discussed scalable and incremental learning of a non-hierarchical cluster structure from training data. This cluster structure serves as a function to map the attribute values of new data to the target class of these data, that is, classify new data. Jun Wu [21] proposed a new online semantic classification framework, in which two sets of optimized classification models, local and global, are online trained by sufficiently exploiting both local and global statistic characteristics of videos.

For improved methods, an original model is modified to adapt to the incremental learning scene. Max Kochurov et al. [22] presented an improved Bayesian method to update the posterior approximation of each new data so as to reduce the cost of training a deep learning network. David A [23] proposed an incremental learning target tracking method based on principal component analysis, which effectively adapted to the change of target appearance. Marko Ristin et al. [24] introduced two improved random forest (RF) algorithms, one based on k-means and the other based on SVM, which were used to avoid retraining RFs from scratch when merging new classes. Stefan et al. [25] proposed an incremental learning SVM method, which inputs the data into the algorithm in several batches and produces initial results after each step of training. In addition, many other neural network-based methods have also been proposed [26,27].

Some researchers have also proposed other different incremental methods for classification problems [28–30]. However, many of these methods cannot be combined well to the field experience of workers (especially for the turnout system), and also some of these methods cannot update itself,

which may reduce the efficiency of the model when new samples or fault types appear since the training samples are quite limited from the very beginning.

Our research group has been long engaged in the safety of railway turnout system, from turnout fault analysis [31] and simulation fault data generating [32], to fault-diagnosis method [33–35], based on which, this paper proposed a Bayes-based online turnout fault-diagnosis method with high accuracy, adaptability and efficiency, which can realize incremental learning and scalable fault recognition, i.e., the model can update itself and deal with new turnout fault types, making it more applicable to the fieldwork and eventually, of great significance for labor-saving, timely maintenance and further, safety and efficiency of railway transportation.

The remainder of this paper is organized as follows: Section 2 introduces the basic conceptions of the turnout system, including its structure, fault types and monitoring data. The methods of data processing and modeling are explained in Section 3, including feature processing, imbalanced data preprocessing, incremental learning method and scalable fault recognition. Section 4 presents a numeric experiment using monitoring data from Guangzhou Railway, followed by conclusions in Section 5. The research framework of this paper is as shown in Figure 1.

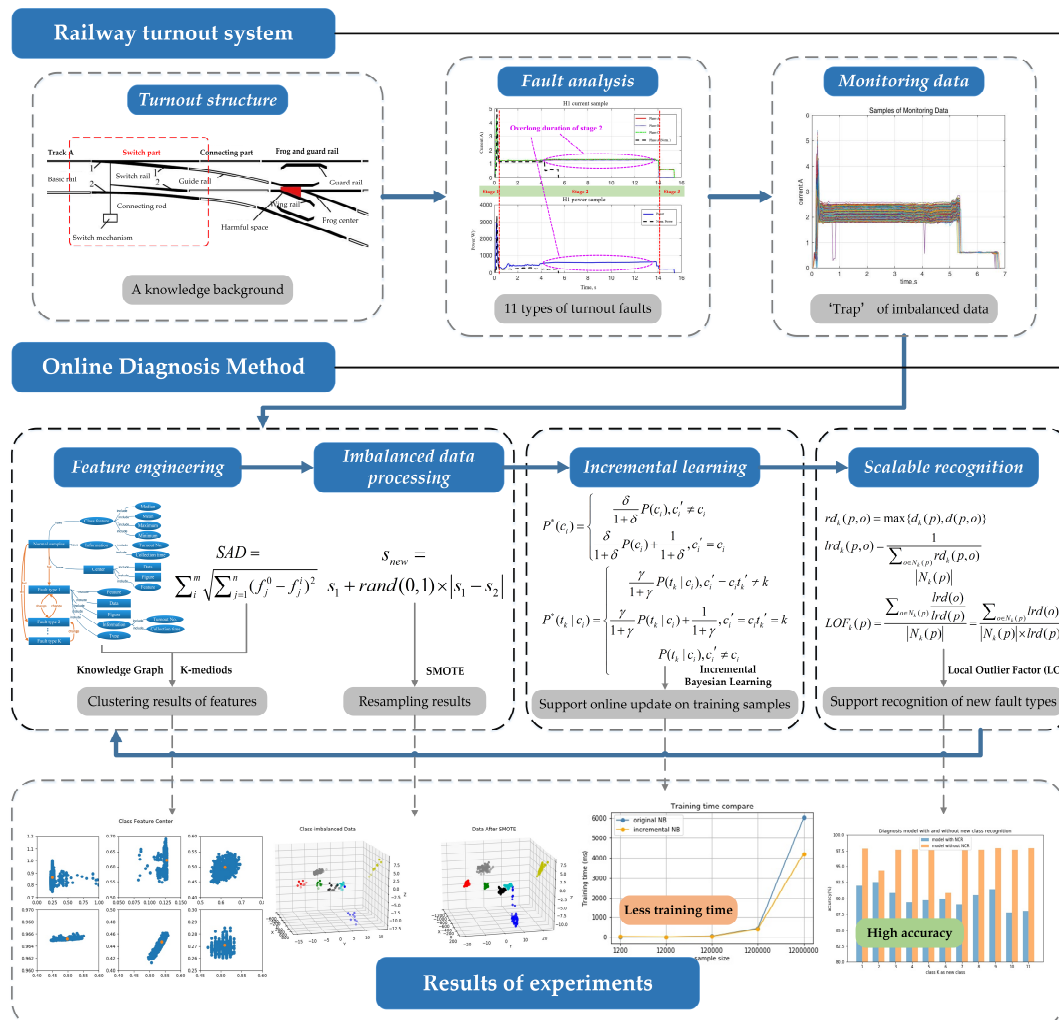


Figure 1. Research framework of this study.

2. Railway Turnout System

2.1. Turnout System

Turnout is a kind of railway track that can be used to change the direction of the route by moving the rails before the trains pass. The whole turnout system includes an indoor control circuit and an outdoor mechanical structure. Figure 2 [36] shows the basic structure of a single turnout system. As shown in Figure 2, the railway turnout system is mainly composed of three parts, among which, the switch part, where massive monitoring data can be obtained by the MMS (microcomputer monitoring system), plays an important role in switching the train direction.

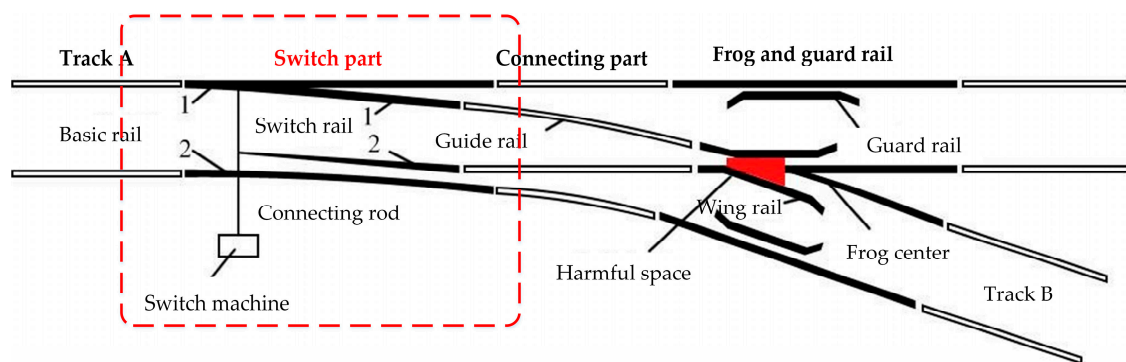


Figure 2. Railway turnout system. The basic structure of a single turnout system and the location of mechanical components.

The switch part, as Figure 2 shows, consists of two basic rails (basic rail 1 and basic rail 2), two switch rails (switch rail 1 and switch rail 2) and a switch machine [37]. Among them, the switch rails are connected with the switch machine through the connecting rod to realize the corresponding operation of the switch machine into horizontal conversion, locking, indication and other operations of the switch rails.

2.2. Fault Type Analysis

Due to the complicated mechanical and electrical structure, exposure to the outdoor environment and the need to frequently pull it, the turnout is more prone to failure. According to our and other scholars' previous research [33,38–40] and combined with the field data, the common types of fault in the turnout system are shown in Table 1 [33].

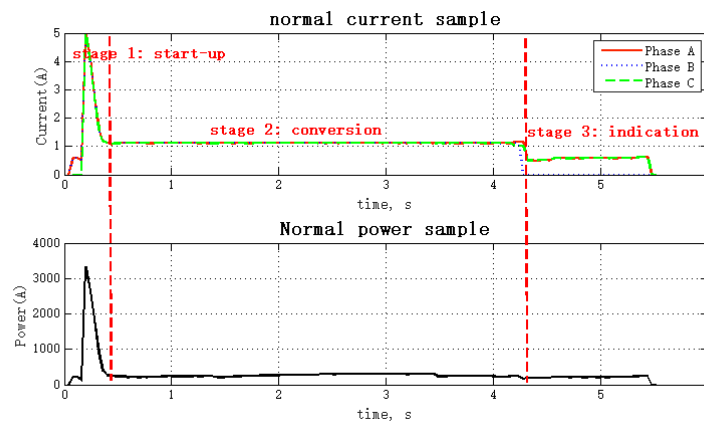
Table 1. Turnout fault types.

Fault Label	Fault Type
H1 ¹	Mechanical jam
H2	Improper position of the slide chair
H3	Abnormal impedance in the switch circuit
H4	Bad contact in the switch circuit
H5	Abnormal open-phase protection device in the indicating circuit
H6	Abnormal impedance in the indicating circuit
F1 ²	The electric relay in the start circuit fails to switch
F2	Supply interruption
F3	Open-phase protection device fault
F4	Fail to lock
F5	Indicating rod block in the gap

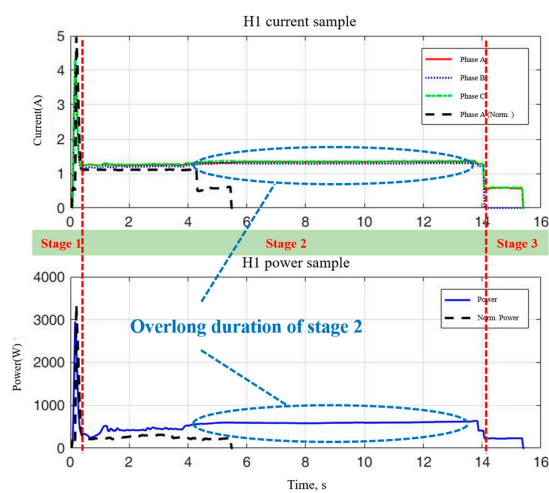
¹ H—hidden dangers, showing that the turnout system is in an unhealthy situation, which may further cause major faults; ² F—major faults that are more serious.

The detailed analysis of monitoring data of the turnout system under normal and abnormal working conditions is as shown in Figure 3. Figure 3a is the normal current and power sample, showing the three stages of a single movement of the turnout, i.e., startup (Stage 1), conversion (Stage 2) and indication (Stage 3).

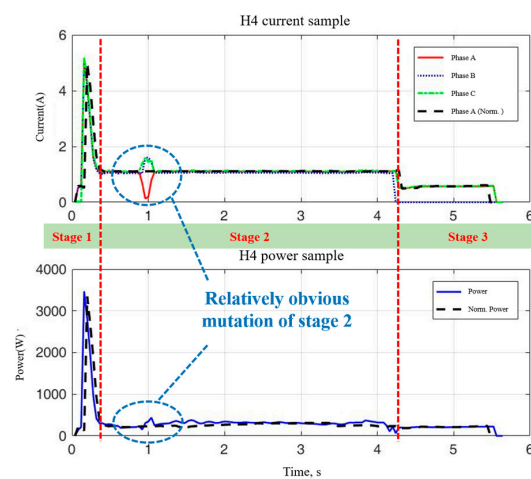
- Stage 1 is the starting of the switch machine and unlocking of the turnout, which needs to overcome much resistance;
- Stage 2 is the process of the movement of the switch rail to the other basic rail, as well as the locking of the turnout;
- Stage 3 is responsible for indicating the result of the conversion using a low-power indicating circuit.



(a)



(b)



(c)

Figure 3. Cont.

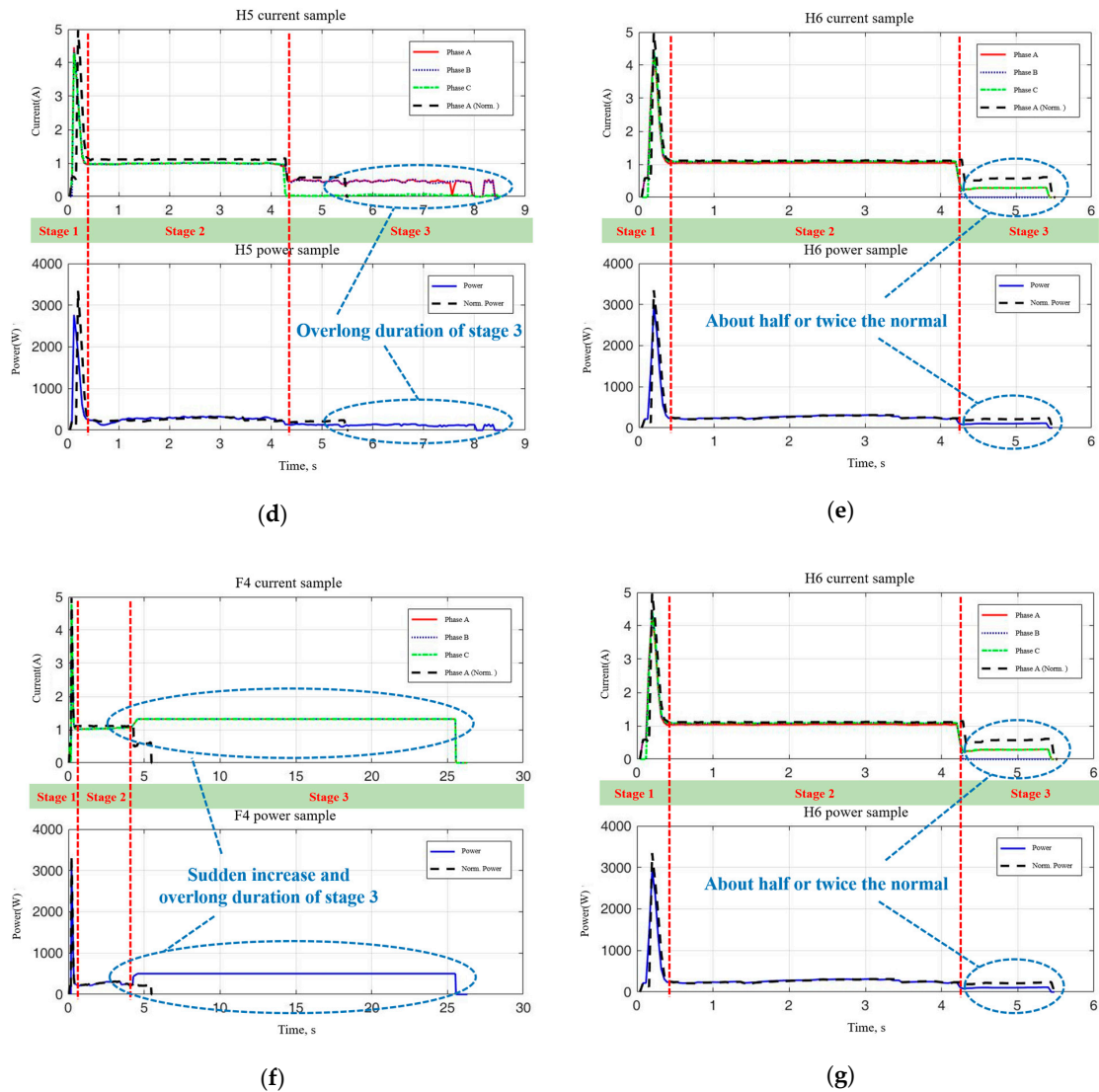


Figure 3. Current and power data of normal and fault types of turnout system. (a) Normal sample of current and power curve; (b) sample of fault H1; (c) sample of fault H4; (d) sample of fault H5; (e) sample of fault H6; (f) sample of fault F4; (g) sample of fault F5.

Figure 3b–g shows the samples of some main fault types, i.e., H1, H4, H5, H6, F4 and F5, which are relatively more frequently occur in the fieldwork. For example, the sample of fault H1 is as shown in Figure 3b, where the duration of Stage 2 is about 14 s, which is too long compared to the normal of about 4 s. As shown in Table 1, this fault is usually caused by mechanical jam, which means there may be a foreign body between the switch rail and the basic rail or there is a jam between the locking gear and the locking block.

2.3. Imbalanced Monitoring Data

Railway turnout system is a key infrastructure to railway safety and efficiency; therefore, many railway departments have adopted MMS to monitor turnouts, providing massive data every day. Among all kinds of monitoring data, the current and power data, which are representative and significant indicators of the turnout working condition, are used in this research.

More than 2000 pieces of current samples of a railway station for about five months are shown in Figure 4, where only about 3 abnormal samples can be found, and the rest are normal, which indicates that there is a huge imbalance between the normal and abnormal samples, let alone between the normal

and 11 types of fault samples. If provided with these data for training, the classification model is prone to be caught in a ‘trap’ of imbalanced data.

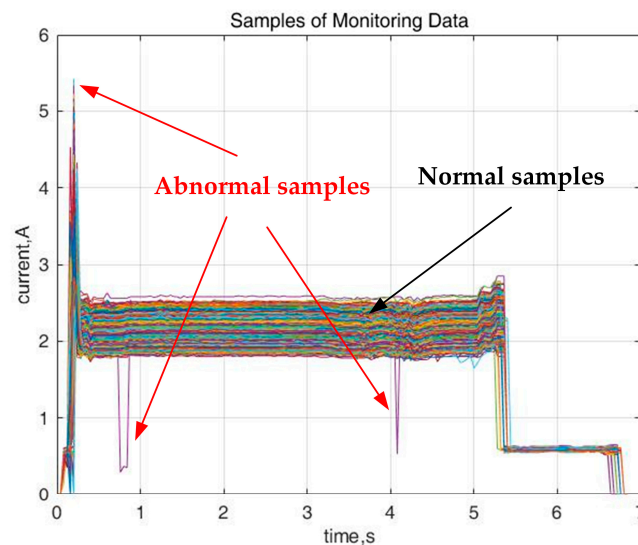


Figure 4. The cumulative current samples from a certain station.

Take an extreme example, suppose there are 1000 samples in total, of which there is only one abnormal sample, then even when the classifier judges all samples as normal, the accuracy can still reach 99.9%, however, missing the most important abnormal sample. Therefore, preprocessing of the imbalanced data are of great significance to improve the performance of the diagnosis model.

3. Online Learning and Diagnosis Method

3.1. Knowledge Graph-Based Feature Engineering

Feature engineering, which, to a large extent, determines the limit of machine learning, is of great importance to obtaining valuable input features for the online diagnosis model of this paper. Therefore, the first step is feature extraction from the original data. The feature extraction method for monitoring data of railway turnouts has been introduced in our previous research [33], where 21 features (e.g., maximum value, mean value, fluctuation factor, peak factor, etc.) are stored by the knowledge graph in this paper.

The knowledge graph is a theoretical method to visualize the core structure or knowledge framework of disciplines or research fields, where the representation of the knowledge graph may vary from one another. In this paper, the entity-relationship model is used to show the structure of the knowledge graph of the turnout monitoring data, where the basic elements are entities (squares), attributes (ellipses) and relationships (lines) as shown in Figure 5.

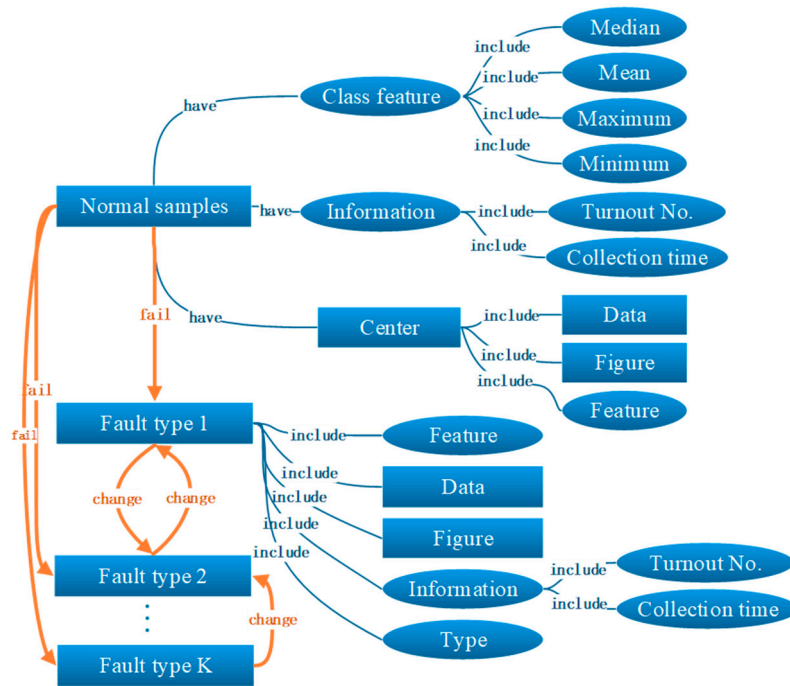


Figure 5. Structure of feature-based knowledge graph.

In reality, turnout systems are well maintained, and the fault samples are quite rare. According to the data from Guangzhou Railway, the abnormal samples account for less than 0.1% of the total. Therefore, the knowledge graph of normal samples and abnormal samples should be arranged in different ways.

As shown in Figure 5, for normal conditions, it is not necessary to put every single sample into the knowledge graph. Instead, the changing trend of sample features is the most important characteristic. Therefore, the knowledge graph records the center, i.e., the most representative sample of all the normal samples. Other normal samples are only recorded as feature ranges, e.g., mean, maximum, etc. For abnormal conditions, the samples are divided into different fault types. In each fault type, all samples will be recorded into the knowledge graph, including features, figures, data and information.

As mentioned above, it is very important to choose the center of normal samples and record it into the knowledge graph. After the feature set is formed, the next step is to choose the most representative sample, i.e., the center. To solve this problem, the K-medoids method (an adaption of K-means algorithm) [35], which requires the clustering center to be one of all the samples rather than creating a new clustering center, is applied in this paper. The steps of the K-medoids algorithm are as follows:

- The input feature set is $F_{m \times n}$ (n is the feature dimension) and choose any row as the original center;
- Define the cluster evaluation variable sum of absolute differences (SAD):

$$SAD = \sum_i^m \sqrt{\sum_{j=1}^n (f_j^0 - f_j^i)^2} \quad (1)$$

- Choose any sample except F^0 as F^k and calculate the corresponding SAD;
- If the SAD of F^k is smaller than the SAD of F^0 , update the F^k to be the new center;
- When all the samples have been searched, the SAD of the current center will reach the minimum and the iteration is finished. Otherwise, return to the third step).

3.2. Imbalanced Data Preprocessing

Before incremental learning modeling, the imbalanced problem of the dataset needs to be solved. In reality, since fault samples are quite rare, which may influence the classification accuracy. The class imbalance problem means that there is a class in the data set whose sample number is far more or less than others, which often leads to the failure of some machine learning models. In this study, the SMOTE (synthetic minority oversampling technique) [41], an improved method based on random oversampling, is applied to deal with the imbalance data.

SMOTE produces new samples for the minority class by selecting k-nearest neighbors and add random numbers to them. The detailed process is as follows:

- For each sample s_1 in the minority class, the Euclidean distance is taken as the standard to calculate its distance to all samples in the class so as to find its k-nearest neighbor;
- A sampling ratio N is set according to the sample imbalance ratio; For each minority sample s_1 , several samples are randomly selected from its k-nearest neighbors, assuming that the chosen nearest neighbor is s_2 ;
- For each chosen neighbor s_2 , a new sample s_{new} is constructed according to the following formula with the original sample s_1 :

$$s_{new} = s_1 + rand(0, 1) \times |s_1 - s_2| \quad (2)$$

3.3. Bayesian Incremental Learning

The naïve Bayesian (NB) classification model is a pattern recognition method based on the Bayesian principle which combines the class prior probability and the conditional probability of each feature attribute to calculate the posterior probability.

Presume that the feature set T consists of $\{t_1, t_2, \dots, t_m\}$ and the m feature attributes are independent of each other. The class variable is represented as $(c_1, c_2, \dots, c_n)$. A single sample is represented as $\langle T, c \rangle$.

According to the Bayesian principle, when the prior probability $P(c_i)$, the conditional probability $P(T|c_i)$ and marginal probability is known, the posterior probability $P(c_1|T)$, $P(c_2|T)$, $\dots$, $P(c_n|T)$ corresponding to n classes can be calculated. The largest posterior probability indicates the class of this sample:

$$P(c_i|T) = \frac{P(c_i) \times P(T|c_i)}{P(T)} \quad (3)$$

The Laplace estimation is applied to get the prior probability and conditional probability of each feature attribute:

$$P(c_i) = \frac{1 + |D_{c_i}|}{|C| + |D|} \quad (4)$$

$$P(t_k|c_i) = \frac{1 + |D_{t_k|c_i}|}{|B_k| + |D_{c_i}|} \quad (5)$$

where $|D_{c_i}|$ is the number of the sample of class c_i in the training data, $|C|$ denotes the class number, $|D|$ denotes the number of training samples, $|D_{t_k|c_i}|$ is the number of training samples which have the attribute t_k and belongs to the class c_i and $|B_k|$ is the number of training samples which have the feature attribute k . Since the features are independent of each other, the conditional probability can be calculated as:

$$P(T|c_i) = \prod_{k=1}^m P(t_k|c_i) \quad (6)$$

naïve Bayesian classifier is of strong mathematical logic and high classification accuracy. However, the traditional naïve Bayesian classifier has two weaknesses. First, the classification accuracy of the classifier is closely related to the size and integrity of the training data set, which, generally, is difficult

to guarantee. Second, when it is applied to turnout fault diagnosis, with the sample size gradually increasing, for each additional batch of sample data, the quasi-prior probability and conditional probability of each characteristic attribute must be recalculated, which is very time consuming and cannot meet the real-time requirement of online fault diagnosis.

However, the combination of the Bayesian algorithm and incremental learning can solve these problems well while maintaining its inherent strong logic and high diagnostic accuracy. When a fault occurs, the newly added sample is input into the Bayesian classification model to get the diagnosis result. If the result is reliable, then this new sample can be used to update the model and ultimately helps to increase the sample data sets and accuracy of the model.

For the new sample, when meeting the judgment conditions, it can be used to update the model, i.e., to update the class prior probability and characteristic properties of conditional probability, which are defined as follows:

$$P^*(c_i) = \begin{cases} \frac{\delta}{1+\delta} P(c_i), c_i' \neq c_i \\ \frac{\delta}{1+\delta} P(c_i) + \frac{1}{1+\delta}, c_i' = c_i \end{cases} \quad (7)$$

$$P^*(t_k|c_i) = \begin{cases} \frac{\gamma}{1+\gamma} P(t_k|c_i), c_i' = c_i, t_k' \neq k \\ \frac{\gamma}{1+\gamma} P(t_k|c_i) + \frac{1}{1+\gamma}, c_i' = c_i, t_k' = k \\ P(t_k|c_i), c_i' \neq c_i \end{cases} \quad (8)$$

where $P^*(c_i)$ is the updated class prior probability and $P^*(t_k|c_i)$ is the updated conditional probability. To be specific, $\delta = |C| + |D|$ and $\gamma = |B_k| + |D_{c_i}|$.

As for the judgment conditions, the posterior probability of the fault label c_i' satisfies the following condition:

$$P(c_i'|T) \geq \alpha(1 - P(c_i'|T)) \quad (9)$$

The overall incremental fault diagnosis process is as shown in Figure 6.

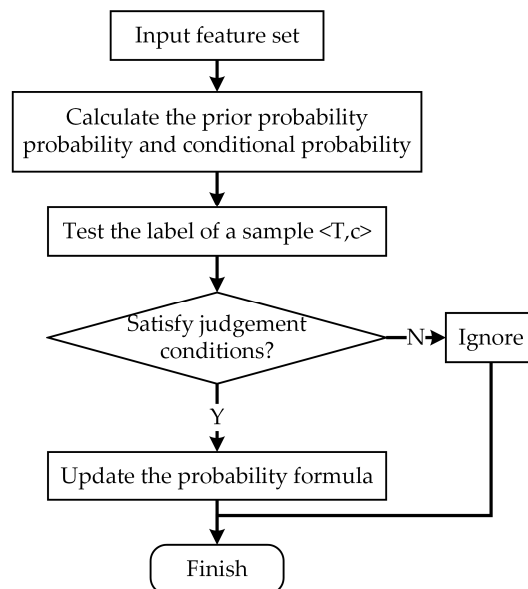


Figure 6. Flow chart of Bayesian incremental learning.

3.4. Scalable Fault Recognition

Though the incremental learning method was widely used in different fields and it can reduce the training cost, there is a key problem that cannot be ignored is the recognition of new classes. Since the types of faults may increase along with the mechanical changes of the turnouts, the new fault types are

very normal in reality. Therefore, to find a method to recognize new faults and realize the automatic update of the knowledge graph is very important for the field application.

The scalable fault recognition problem (SFR) can be abstracted into a novelty detection problem. The model of novelty detection will fit a rough boundary to define the contour of the initial samples, i.e., the normal samples. When a new sample appears, it will be judged whether it is within the boundary.

Local outlier factor (LOF) [42], an efficient novelty detection method, aims to detect those abnormal data that are quite different from the characteristic attributes of normal data. One of the assumptions of LOF is that the normal data must be clustered, which puts forward certain requirements on the original data set. The turnout data samples, which were clustered according to the mentioned methods of data processing, basically meet this requirement.

The principle of LOF is to calculate the density of a certain point and to judge whether the point is in the sparse region according to the density of the data cluster of the dataset itself. A point in the sparse region is recognized as a novelty point. Due to its "local" characteristics, LOF can handle clusters with different densities correctly. The local outlier factor $LOF_k(p)$ is used to measure the novelty degree of point p :

$$rd_k(p, o) = \max\{d_k(p), d(p, o)\}, \quad (10)$$

$$lrd_k(p, o) = \frac{1}{\frac{\sum_{o \in N_k(p)} rd_k(p, o)}{|N_k(p)|}}, \quad (11)$$

$$LOF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{lrd(o)}{lrd(p)}}{|N_k(p)|} = \frac{\sum_{o \in N_k(p)} lrd(o)}{|N_k(p)| \times lrd(p)}, \quad (12)$$

where $d(p, o)$ is the distance between point p and o ;

$d_k(p)$ is the distance between point p and the k_{th} nearest neighbor of p ;

$rd_k(p, o)$ is the reachability distance between point p and o ;

$lrd_k(p, o)$ is the local reachability density. It measures the density of point p and its k neighbors.

The larger the reachable distance, the smaller the density;

$LOF_k(p)$ is the local outlier factor. It defines a concept of relative density, which can deal with heterogeneous data and make outlier detection more accurate.

According to the definition of the local outlier factor, $LOF_k(p) \approx 1$ means the local density of p is similar to its k neighbors; $LOF_k(p) < 1$ means p is in a high-density area, indicating it is a normal point; If $LOF_k(p) \gg 1$, then the point p is far away from the normal data cluster, which means it is very likely to a novelty point.

Since the original LOF can only deal with binary classification problems, the final output should consider all the binary classification models, inspired by the multiple classification models. If all the binary classification models output -1 , i.e., this sample is not similar to any known classes, then it is identified as a new class sample.

When the sample size is big enough, a clustering method is applied, and the unidentified samples can be labeled more easily. Finally, the densest N clusters among these unidentified samples are selected as the candidate sample set of the new classes. If the density of this sample set exceeds the specified threshold, then it is determined as the representative sample set of the new class for further knowledge graph update.

4. Experiment Result

4.1. Clustering and Resampling

In this paper, the experimental data are from several stations of Guangzhou Railway from 2016 to 2018. As mentioned before, the class center is the most representative sample of a class and one of the key components of the knowledge graph. Taking the normal class as an example, the dimension of features is 21, under which, the feature clustering result is abstract and hard to understand. For a

clearer display, the clustering results are divided into several pairs, e.g., feature 2 (span of Stage 1) & feature 3 (maximum of Stage 1), feature 6 (span of Stage 2) & feature 9 (mean of Stage 2), feature 11 (standard deviation of Stage 2) & feature 12 (peak factor of Stage 2), etc., as shown in Figure 7, where the orange point stands for the center, and the values of all features have been normalized.

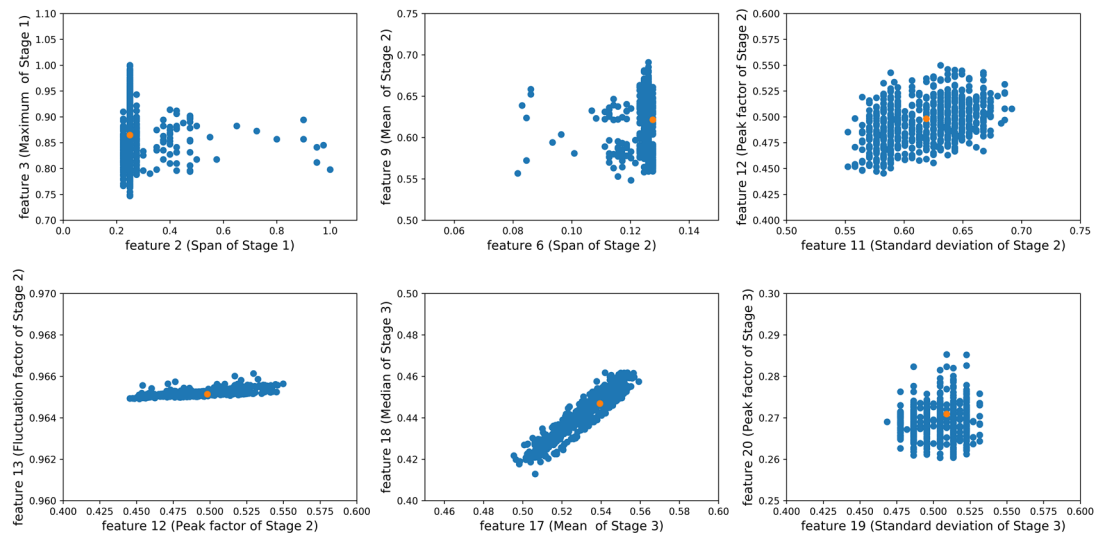


Figure 7. Class center of features.

Before using the feature set to train the diagnosis model, the class-imbalanced data should be preprocessed with the SMOTE resampling method. In this paper, the original data set consists of 1000 normal samples and 10 samples of each 11 fault types, and the final data set includes 12,000 samples in total (each class with 1000 samples).

To show it more clearly, an example data set is organized with 100 normal samples, and each of 6 different fault types is of 10 samples, i.e., the normal sample size to the abnormal is 10:1. Moreover, an LDA feature reduction method is applied to reduce the feature dimension to 3, which can be easily shown in the picture. Figure 8 shows the original example data set and the resampled data set after SMOTE. The fault samples are multiplied according to the original imbalance ratio. After SMOTE, the sample sizes of all classes (including the normal class) are the same. This method cannot only be used in data preprocessing, but also in generating fault data manually.

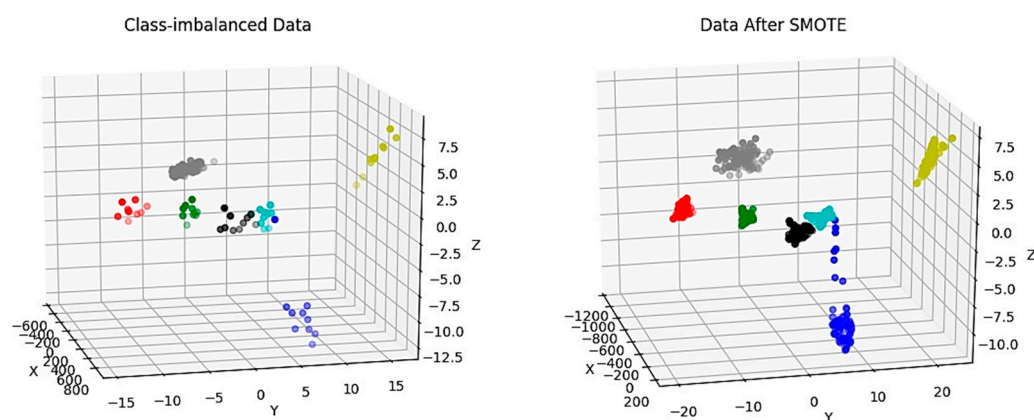


Figure 8. Date of original and resampled.

4.2. Incremental and Scalable Fault Diagnosis Model

With the input data set prepared, it comes to the next step, the scalable fault recognition. Different parameters can be set to get the scalable fault recognition accuracy of different models to determine the optimal parameter value. The main parameters of the LOF model are “n_neighbors” (n), which is the number of neighbors and “contamination” (c), the percentage of outliers in the test sample. With the method of grid search, the parameter choosing result is shown in Table 2. Therefore, the optimal values for the parameters should be $n = 40$ and $c = 0.01$.

Table 2. Local outlier factor (LOF) parameter choosing result (MS).

$c \backslash n$	25	30	35	40
0.0001	95.35	95.35	93.76	98.13
0.001	95.35	95.35	93.76	98.13
0.01	95.38	95.83	96.01	99.11

For the experiment, only the samples not identified as a new class can be input into the diagnosis model. For a single fault type, which is regarded as a new class, the diagnosis models with and without scalable fault recognition are compared. As shown in Figure 9, the blue bars represent the accuracy of the original model and the yellow bars represent the accuracy of the model with scalable fault recognition. To 11 fault types, the accuracies improve to different degrees and the average accuracy improves about 6.8%. Moreover, the total accuracy of scalable fault recognition is 99.11%. When the training data increasing over time, the accuracy will increase as well.

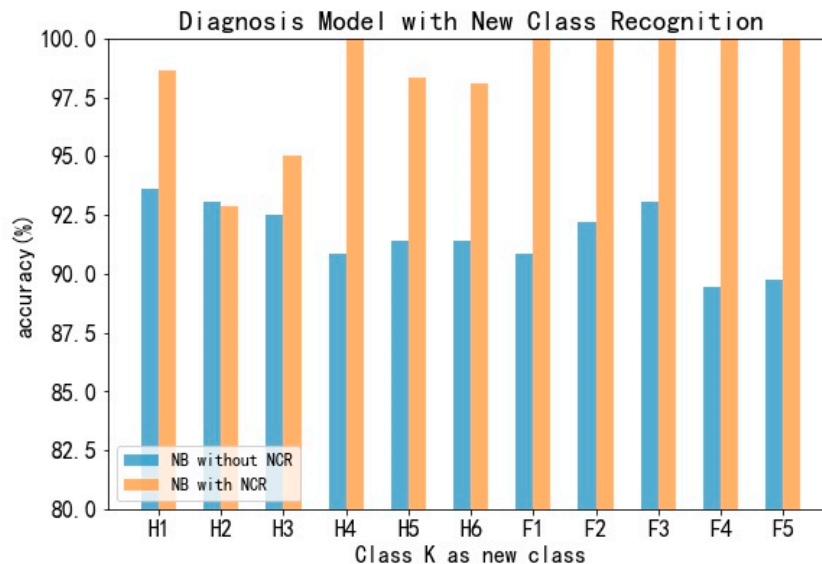


Figure 9. Result of scalable fault recognition.

After leaving out the new class samples, the incremental Bayesian model and naïve Bayesian model are compared both in training time and accuracy. The incremental NB model divides the whole data set into 10 batches and trains them separately. As shown in Table 3 and Figure 10a, it can be easily found that as the sample size increases, the incremental NB takes less training time than the original NB. When the sample size reaches 12 million, the training time of incremental NB decreases by about 30%. For the training time, the larger the training data set is, the greater the gap between the two models will be. At the same time, it can be found in Table 4 and Figure 10b that the accuracy of incremental NB is always above the original NB, with an average improvement of 0.32%. This is mostly

caused by the step-by-step training model of incremental NB, which can choose the most valuable samples for training.

Table 3. Training time comparison (MS).

Sample size	1200	12,000	120,000	1,200,000
NB	6.62	10.27	59.84	476.52
Incremental NB	38.48	19.53	51.32	436.72

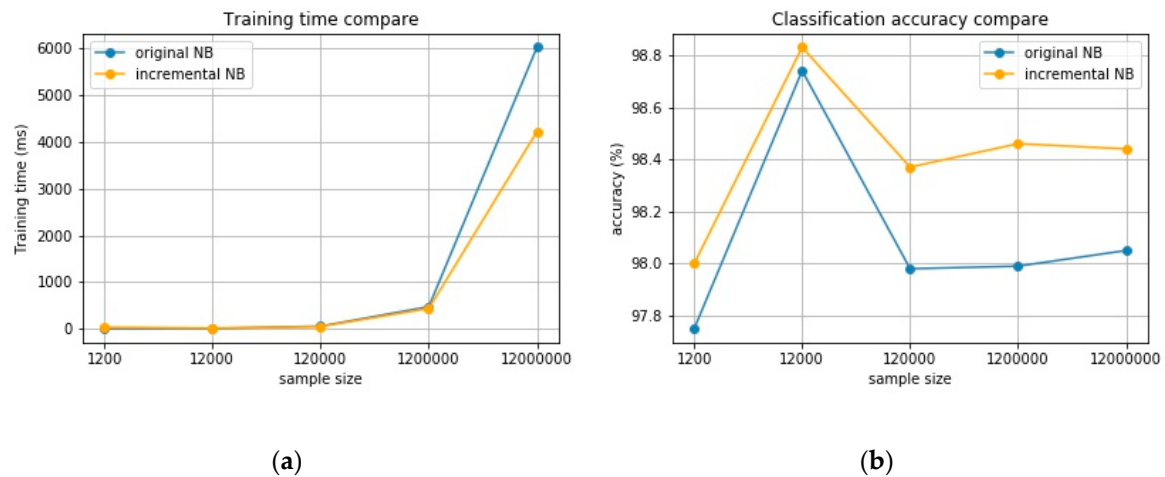


Figure 10. Comparison of original naïve Bayesian (NB) and incremental NB. (a) Comparison of training time; (b) comparison of classification accuracy.

Table 4. Classification accuracy comparison (%).

Sample size	1200	12,000	120,000	1,200,000
NB	97.75	98.74	97.98	97.99
Incremental NB	98.00	98.83	98.37	98.46

5. Conclusions

Based on imbalanced monitoring data, this paper proposes a solution for railway turnout fault diagnosis, which can realize incremental learning and scalable fault recognition. First, this paper introduced the basic conceptions of the railway turnout system. Next are the knowledge graph-based feature engineering and processing of the imbalanced monitoring data. Then, an online learning and diagnosis method based on Bayesian incremental learning and scalable fault recognition is proposed. According to the results of the experiment, with data from Guangzhou Railway, the scalable recognition method can reach an accuracy of over 99%. Moreover, compared to the non-incremental model, the incremental learning model saves about 30% of training time when the sample size reaches 12 million—with a slight accuracy improvement as well. In addition, it should be noted that the proposed solution is applicable to most railway turnout systems, except for those whose switch machines are of three-phase AC type.

Future research directions of this paper may include studying a more intelligent method to improve the accuracy of scalable fault recognition, and further improvement of the knowledge graph structure is also required.

Author Contributions: Conceptualization, D.O., Y.J., L.Z. and H.L.; methodology, D.O., Y.J. and L.Z.; software, Y.J. and H.L.; validation, L.Z. and H.L.; formal analysis, Y.J., L.Z. and H.L.; resources, D.O., Y.J. and L.Z.; writing, D.O. and Y.J., L.Z. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China, Grant Numbers 2018YFB1201403 and 2016YFB1200401, and Research Program of Shanghai Science and Technology Committee, Grant Number 18DZ2202600.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 2019 Railway Statistical Bulletin; Ministry of Transport of the People's Republic of China: Beijing, China, 2020; pp. 1–8.
2. Atamuradov, V.; Camci, F.; Baskan, S.; Sevkli, M. Failure Diagnostics for Railway Point Machines Using Expert Systems. In Proceedings of the 2009 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics And Drives, Cargese, France, 31 August–3 September 2009; pp. 1–5.
3. Yang, Y. *Station Signal Control System*, 1st ed.; Southwest Jiaotong University Press: Chengdu, China, 2012.
4. Marquez, F.P.G.; Tercero, D.J.P.; Schmid, F. Unobserved component models applied to the assessment of wear in railway points: A case study. *Eur. J. Oper. Res.* **2007**, *176*, 1703–1712. [[CrossRef](#)]
5. Márquez, F.P.G.a.; Schmid, F.; Collado, J.C. A reliability centered approach to remote condition monitoring. A railway points case study. *Reliab. Eng. Syst. Saf.* **2003**, *80*, 33–40.
6. 2017 Annual Report; China Railway Jinan Group Co Ltd.: Jinan, China, 2018.
7. 2018 Annual Report; China Railway Guangzhou Group Co Ltd.: Guangzhou, China, 2019.
8. *Inspection of Switches and Crossings State of the Art Report Preliminary Report*; International Union of Railways (UIC): Paris, France, 2011.
9. Marquez, F.P.G.; Weston, P.; Roberts, C. Failure analysis and diagnostics for railway trackside equipment. *Eng. Fail. Anal.* **2007**, *14*, 1411–1426. [[CrossRef](#)]
10. Yanqing, X. Research on Realization Method of Turnout Equipment Fault Diagnosis Expert System. Master's Thesis, Beijing Jiaotong University, Beijing, China, 2012.
11. Min, S. Research on Fault Diagnosis Expert System of Railway Signal Equipment. Master's Thesis, Qingdao University, Qingdao, China, 2009.
12. Yang, Y. Research on A pplication of Neural Network and Expert System in Rail Switch Fault Diagnosis. Master's Thesis, Lanzhou Jiaotong University, Lanzhou, China, 2013.
13. Cao, H.; Yue, L.; Yang, F. Application of Expert System and Neural Network in Fault Diagnosis of Switch Control Circuit. *Heilongjiang Sci. Technol. Inf.* **2010**, *62*. (In Chinese) [[CrossRef](#)]
14. Youmin, H. Research on Fault Diagnosis Method of High-Speed Railway Turnouts. Master's Thesis, Beijing Jiaotong University, Beijing, China, 2014.
15. Eker, O.F.; Camci, F.; Guclu, A.; Yilboga, H.; Baskan, S. A Simple State-Based Prognostic Model for Railway Turnout Systems. *IEEE Trans. Ind. Electron.* **2012**, *58*, 1718–1726. [[CrossRef](#)]
16. Yilboga, H.; Eker, Ö.F.; Güçlü, A.; Camci, F. Failure prediction on railway turnouts using time delay neural networks. In Proceedings of the 2010 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Taranto, Italy, 6–8 September 2010; pp. 134–137.
17. Witczak, M.; Korbicz, J.; Mrugalski, M.; Patton, R.J. A GMDH neural network-based approach to robust fault diagnosis: Application to the DAMADICS benchmark problem. *Control Eng. Pract.* **2006**, *14*, 671–683. [[CrossRef](#)]
18. Silmon, J.A.; Roberts, C. Improving railway switch system reliability with innovative condition monitoring algorithms. *Proc. Inst. Mech. Eng. F J. Rail Rapid Transit* **2010**, *224*, 293–302. [[CrossRef](#)]
19. Diehl, C.P.; Cauwenberghs, G. SVM Incremental Learning, Adaptation and Optimization. In Proceedings of the International Joint Conference on Neural Networks, Portland, OR, USA, 20–24 July 2003; pp. 2685–2690.
20. Ye, N.; Li, X. A scalable, incremental learning algorithm for classification problems. *Comput. Ind. Eng.* **2002**, *43*, 677–692. [[CrossRef](#)]
21. Wu, J.; Hua, X.-S.; Zhang, H.-J.; Zhang, B. An online-optimized incremental learning framework for video semantic classification. In Proceedings of the 12th annual ACM international conference on Multimedia, New York, NY, USA, 10–16 October 2004; pp. 320–323. [[CrossRef](#)]
22. Kochurov, M.; Garipov, T.; Podoprikin, D.; Molchanov, D.; Ashukha, A.; Vetrov, D. Bayesian incremental learning for deep neural networks. *arXiv* **2018**, arXiv:1802.07329.

23. Ross, D.A.; Lim, J.; Lin, R.-S.; Yang, M.-H. Incremental learning for robust visual tracking. *Int. J. Comput. Vision* **2008**, *77*, 125–141. [CrossRef]
24. Ristin, M.; Guillaumin, M.; Gall, J.; Van Gool, L. Incremental learning of random forests for large-scale image classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 490–503. [CrossRef] [PubMed]
25. Ruping, S. Incremental learning with support vector machines. In Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 29 November–2 December 2001; pp. 641–642.
26. Pratama, M.; Anavatti, S.G.; Angelov, P.P.; Lughofer, E. PANFIS: A novel incremental learning machine. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *25*, 55–68. [CrossRef] [PubMed]
27. Xiao, T.; Zhang, J.; Yang, K.; Peng, Y.; Zhang, Z. Error-driven Incremental Learning in Deep Convolutional Neural Network for Large-scale Image Classification. In Proceedings of the 22nd ACM international conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 177–186. [CrossRef]
28. Kotsiantis, S. Increasing the accuracy of incremental naive Bayes classifier using instance based learning. *Int. J. Control Autom. Syst.* **2013**, *11*, 159–166. [CrossRef]
29. Gu, B.; Quan, X.; Gu, Y.; Sheng, V.S.; Zheng, G. Chunk incremental learning for cost-sensitive hinge loss support vector machine. *Pattern Recognit.* **2018**, *83*, 196–208. [CrossRef]
30. Xu, Y.; Shen, F.; Zhao, J. An incremental learning vector quantization algorithm for pattern classification. *Neural Comput. Appl.* **2012**, *21*, 1205–1215. [CrossRef]
31. Ou, D.; Xue, R. Railway Turnout Fault Analysis Based on Monitoring Data of Point Machines. In Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 26–28 October 2018; pp. 546–552.
32. Cui, K.; Tang, M.; Ou, D. Simulation Data Generating Algorithm for Railway Turnout Fault Diagnosis in Big Data Maintenance Management System. In *International Symposium for Intelligent Transportation and Smart City*; Springer: Singapore, 2019; pp. 155–166. [CrossRef]
33. Ou, D.; Xue, R.; Cui, K. A data-driven fault diagnosis method for railway turnouts. *Transp. Res. Rec.* **2019**, *2673*, 448–457. [CrossRef]
34. Ou, D.; TAng, M.; Xue, R.; Yao, H. Hybrid fault diagnosis of railway switches based on the segmentation of monitoring curves. *Eksploatacja i Niezawodn.–Maint. Reliab.* **2018**, *20*, 514–522. [CrossRef]
35. Ou, D.; Tang, M.; Xue, R.; Yao, H. A Segmented Similarity-Based Diagnostic Model for Railway Turnout System. Available online: <https://trid.trb.org/view/1572323> (accessed on 14 July 2020).
36. Jining Iron Machinery Equipment Co., Ltd. Turnout System. Available online: <http://www.juguuji.com/news/Data/354.html> (accessed on 14 July 2020).
37. Jin, G.; Yan, W.; Lifang, L. *Railway Signal Infrastructure*; Southwest Jiaotong University Press: Chengdu, China, 2008.
38. Ding, Z.; Guoning, L. Fault diagnosis of S700K switch machine based on improved WNN analyses power curve. *J. Railway Sci. Eng.* **2018**, *15*, 229–236. (In Chinese)
39. Ruifeng, W.; Wangbin, C. Research on Fault Diagnosis Method for S700K Switch Machine Based on Grey Neural Network. *J. Chin. Railway Soc.* **2016**, *38*, 68–72. (In Chinese)
40. Qingyang, X.; Zhongtian, L.; Huibing, Z. Method of Turnout Fault Diagnosis Based on Hidden Markov Model. *J. Chin. Railway Soc.* **2018**, *40*, 98–106. (In Chinese)
41. Post, C.; Rietz, C.; Büscher, W.; Müller, U. Using Sensor Data to Detect Lameness and Mastitis Treatment Events in Dairy Cows: A Comparison of Classification Models. *Sensors* **2020**, *20*, 3863. [CrossRef] [PubMed]
42. Zhao, S.; Li, W.; Cao, J. A user-adaptive algorithm for activity recognition based on k-means clustering, local outlier factor, and multivariate gaussian distribution. *Sensors* **2018**, *18*, 1850. [CrossRef] [PubMed]

