Anomaly Recognition of Non-fault Switch Action Curve Based on Image Moments and NBM

Wen Hao[1], Xie Guangxiang[2], Mao Junyun[3]

[1]Wuhan Metro Operation Co. Ltd. , Wuhan, China

[2]Wuhan Metro Operation Co. Ltd. , Wuhan, China

[3]Wuhan Railway Vocational College of Technology, Wuhan, China

Keywords: Switch action curve; Abnormal recognition; Image moments; NBM

Abstract: In order to alleviate the shortcomings of the existing methods in selecting the distinguishing features of the switch action curve, reduce the impact of limited training samples, and achieve accurate multi-task classification, a new method for anomaly recognition of switch non-fault action curve is proposed. Firstly, the characteristic graph of switch action curve is defined, which is divided into five parts according to action time-sequence, and the recognition feature set is constructed by calculating the zero to second moments of switch action curve. Then, five classifiers are trained for recognition based on Naive Bayesian Model (NBM). Experiments show that the method can adapt to small-scale sample training, effectively identify the different problems existing in any action time-sequence of switch, and achieve a more comprehensive switch fault early warning.

1. Research Background

As we all know, turnouts are the most basic and important equipment for rail transit. Once the turnout breaks down, it may endanger the traffic safety, so it is of great significance to carry out fault early warning analysis [1]. The commonly implemented warning of turnout failure is mainly based on manual analysis. By reading the non-faulty turnout action curve, it is observed whether there is abnormal fluctuation or obvious change trend. Manual analysis mainly has the disadvantages of relying too much on subjective judgment, time-consuming and laborious.

With the rise of subject areas such as signal processing, pattern recognition, and artificial intelligence, people began to study the intelligent diagnosis algorithm for ballast fault based on mathematical models such as grey relational analysis, support vector machine and artificial neural network [2][3][4]. This will alleviate the limitations of manual analysis. In recent years, many scholars have begun to further study algorithms that can achieve early warning [5] [6], in order to eliminate the fault in the bud. Regardless of fault diagnosis or early warning analysis, most of the algorithms are based on the identification of the switch action curve [7]. The idea is to obtain a large number of samples for feature extraction. Classification identification is then achieved based on statistical analysis or machine learning to obtain “reference standard data” for association comparison. However, there are still some shortcomings in these methods:

(1) At present, the mainstream method for anomaly identification is to divide the turnout action curve into 3-5 different time sequences according to the action sequence and characteristics, and then calculate the identification characteristics based on each time sequence unit for association and classification. This kind of method can accurately judge whether the switch action curve of each time sequence is normal or not, but it is not easy to carefully distinguish which kind of problem the abnormal time sequence of switch action curve belongs to. Taking the “conversion” timing as an example, there are usually different types of abnormalities such as significant fluctuations, trend anomalies, and continuous shock pulses. The anomaly of “significant fluctuations” should also be distinguished in the early, middle or late stages, which is difficult for many existing algorithms to identify effectively.

(2) Existing algorithms usually regard the switch action curve as a time series and calculate its

  DOI: 10.25236/scmc.2019.004

time series characteristics to identify and classify [[2][3][6]. There are two main drawbacks: (1) time series features can reflect abnormalities. However, it cannot be used to distinguish the specific time period in which an abnormality occurs. (2) The time series characteristics have strong sensitivity to numerical changes. The normal jitter of the ballast curve in a certain range may affect the judgment result.

(3) From the actual situation, there is a possibility that a fault or an abnormality may occur when the switch is continuously operated for hundreds of times, so there is a problem that the proportion of positive and negative samples is unbalanced. The number of negative samples is very limited and it is not easy to meet the requirements of general machine learning algorithm classifier training.

In this paper, the problem of abnormal identification of non-faulty action curves is taken as the research direction. A new identification method is proposed for the three shortcomings of the above analysis.

2. Identification Feature Selection and Classifier Training

2.1 Definition and Calculation of Moment Characteristics of Turnout Action Curve

Taking the power curve as an example, Fig. 1 shows the power curfor:

$$M_{pq} = \sum_{i=1}^{m} \sum_{j=1}^{n} i^p j^q f(i, j) \qquad (1)$$

Among them, i versus j Representing the horizontal and vertical coordinates of the pixel, $f(i, j)$ The pixel at the coordinate point of the image (i, j) value. Obviously, here $f(i, j)$ The value is 0 or 1. Further, the zero-order $p = 1 \& q = 0$ moment, the first-order moment and the second-order moment of the image are used to construct the image area, the center of gravity and the directional feature, thereby using the classification and recognition of the ballast action curve. The specific characteristics are calculated and their meanings are as follows:

a. Area characteristics. will $p = q = 0$ Substituting (1) to calculate the zero moment to obtain the area feature M_{00}, that is, the sum of the pixel valuesin the characteristic diagram of the switch action curve. When any abnormality occurs in the ballistic action curve, its area characteristics will also change abnormally. Therefore, the significance of calculating the area feature is mainly used to identify whether the ballast ac $p = 1 \& q = 1$ Calculate the first moment M_{11} And then separately $p = 2 \& q = 0$, $p = 0 \& q = 2$ Calculate the second moment M_{20} with M_{02}, calculating direction characteristics θ.

$$\theta = \frac{1}{2} \arctan\left(\frac{2b}{a-c} \right) \qquad (3)$$

among them, $\mathrm{a} = \frac{M_{20}}{M_{00}} - x_c{}^2, \mathrm{b} = \frac{M_{11}}{M_{00}} - x_c y_c, \mathrm{c} = \frac{M_{02}}{M_{00}} - y_c{}^2$.

The meaning of the calculation of directional features is mainly used to distinguish between significant fluctuations and abnormal trend changes, and their location.

2.2 NBM-based Classifier Training

This paper adopts a naive Bayesian model that is outstanding for small-scale data processing and robust to missing valuesand isolated noise points. Error! Reference source not found. (Navie Bayesian Model, NBM) training classifiers alleviate the problem of negative sample missing mentioned above, and can achieve multi-classification tasks and incremental training, and have certain robustness to the fluctuation of the ballast action curve in the normal range.

Firstly, the process of naive Bayesian classification of the ballast action curve based on the identification features of this paper is given:

① Set $x=\{M_{00}, x_c, y_c, \theta\}$ Identify a feature set for a curve to be classified, with a category set $C=\{y_1, y_2, ..., y_n\}$

② Separately calculated $P(y_1|x), P(y_2|x), ..., P(y_n|x)$; , there is $x \in y_k$.

The key to classification is seeking $P(y_1|x), P(y_2|x), ..., P(y_n|x)$ .Naive Bayesian theory assumes that each feature attribute is independent, so according to Bayes' theorem, the following derivation can be made:

$$P(y_i|x)=\frac{p(x|y_i)p(y_i)}{p(x)}, i=1,2,...,n \quad (4)$$

In the above formula, the denominator is regarded as a constant for all classifications, and since each characteristic attribute is conditionally independent, further molecules are:

$$P(x|y_i)P(y_i)=P(y_i)\prod_{j=1}^{4}P(x(j)|y_i)=$$

$$P(M_{00}|y_i)P(x_c|y_i)P(y_c|y_i)P(\theta|y_i)P(y_i) \quad (5)$$

The feature attribute in this paper is a continuous variable, so Gaussian distribution is used to characterize the conditional probability distribution of each feature attribute.

$$P\left(X=x(j)|Y=y_i\right)=\frac{1}{\sqrt{2\pi}\sigma_{ij}}e^{\frac{\left(x(j)-\mu_{ij}\right)^2}{2\sigma_{ij}^2}} \quad (6)$$

Where parameters μ_{ij} , σ_{ij}^2 Representing i In the category j Statistical average of the characteristics and statistical variance.

In summary, the essence of the nbm training classifier based on the proposed algorithm is to obtain the statistical mean and variance of different identification features under different conditions of the known training sample set.

In order to enhance the classification effect, consider the time division of the ballast action curve, and train the nbm classifier for each sequence. Fig. 3 shows the structure of the training nbm classifier.

In the Fig., the sequence 1, the sequence 2, the sequence 3, the sequence 4, and the sequence 5 are sequentially started, unlocked, converted, locked, and hooked to indicate five stages. The key to classifier training is not the total number of samples, but rather the types of abnormal samples with different timings.

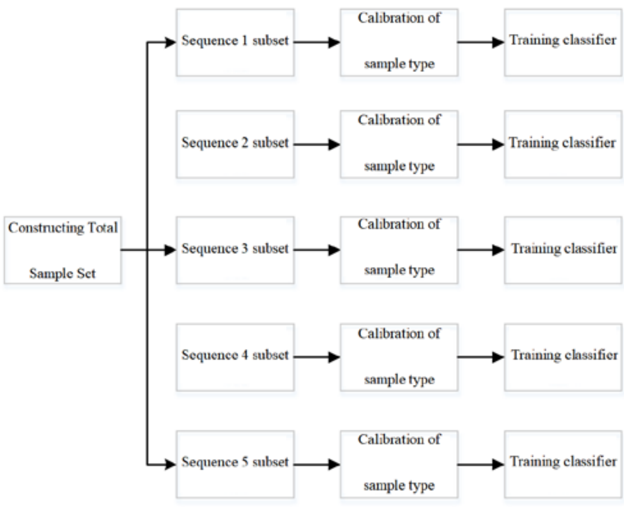

Fig. 3 Schematic diagram of the overall structure of the training nbm classifier

3. Turnout Motion Curve Anomaly Identification Process

Taking the power curve as an example, the specific steps of the anomaly identification method of the ballast action curve are given.

Step1: Obtain a sequence of non-fault turnout power curves to be identified (acquired at a specific frequency from 0 seconds until the action is completely ended).

Step2: The collected curve sequence is constructed into a fixed size (determined by the number of collected points and the power value range), and is divided into 5 steps according to the starting, unlocking, converting, locking, and communicating. Fixed-size parts.

Step 3: Calculate the identification feature set by calculating the zero to second moment of the image corresponding to each time series feature map according to the content of Section 2.1.

Step4: Use the trained 5 NBM classifiers to classify the five timing features separately.

Step 5: Combine the classification results of each time series feature to obtain the final abnormal recognition result of the turnout power curve.

4. Experimental Results and Analysis

In order to verify the effectiveness of the proposed method, the non-faulty power curve actually generated by any reentry channel in a Wuhan metro line is classified and the training and identification results of time series 3 (conversion) are selected for display and analysis.

Firstly, a sample set containing 100 curve samples is constructed by means of framed photographing or manual acquisition, and the sample subsets of the sample curve are manually typed, which are divided into normal, pre-significant fluctuations, medium-term significant fluctuations, late significant fluctuations, and trends. Abnormal rise, there are six types of continuous shock pulses, which are recorded as type 1, type 2, …, type 6. The 100 calibrated samples are then divided into 80 training samples and 20 test samples, and the classifier 3 is trained according to the method herein based on 80 training samples. Finally, 20 test samples were combined with 80 training samples for identification verification. The experimental results are shown in Table 1.

Table 1 Identification experiment results for the sample curve timing 3

Types of	Number of samples	Recognition rate	False rate
1	24	100%	1.39%
2	13	92.31%	0%
3	18	94.44%	1.22%
4	16	93.75%	2.38%
5	19	89.47%	1.23%
6	10	100%	0%

Wherein, the recognition rate indicates the probability that all samples of the type are correctly identified as the calibration type, and the false positive rate indicates the probability that other types of samples are recognized as the type.

The results of this experiment can be seen: (1) The recognition rate of each type has basically reached more than 90% (type 5 is close to 90%), and the number of misidentified samples is no more than two, and the overall recognition effect is better. (2) The recognition rate of type 1 and type 6 reaches 100%, which indicates that the method has a better recognition effect on the curve with normal curves and abnormal features. (3) The recognition rate of type 5 is relatively low. Two samples are misidentified as type 1 and type 4, respectively. The probability of type 4 being misunderstood is the highest, indicating that the recognition feature is weaker than other types.

5. Conclusion

In this paper, based on image moment and nbm, the non-fault action curve anomaly recognition method is used to construct the recognition feature for classification, and five nbm classifiers are trained to identify various time series curves. Experimental results show that This method can not only distinguish the normality of the ballast action curve, but also can effectively identify which type of abnormal curve belongs to, and provide the basis for the warning of ballast failure. How to further improve the recognition rate of all types is the next research goal of this paper.

References

[1] Wang Guang. Prediction of railway turnout fault based on text data [d]. Beijing Jiaotong University, 2017.

[2] Zhao Linhai,Lu Qiao. A Method for Fault Diagnosis of Turnout Based on Grey Correlation [J]. Journal of the China Railway Society, 2014, 36(2):48-50.

[3] Yang Xiaorui, Wang An, An Bangjun. Research on Turnout Fault Diagnosis Based on Switching Traction and Pso-svm[J]. Computer Measurement & Control, 2017, 25(7): 55-58.

[4] Li Yamei,Wei Wenjun. Study on the fault diagnosis system of turnout based on fuzzy neural network [J]. Railway Computer Application, 2012, 21(1):35-39.

[5] Xu Qingyang. Drainage fault diagnosis and health status prediction [d]. Beijing Jiaotong University, 2018.

[6] Zhong Zhiwang, Chen Jianyi, Tang Tao, et al. The svdd-based fault detection and health assessment method [j]. Journal of Southwest Jiaotong University, 2018, 53(4): 842-849.

[7] Zhang Xing. Research and implementation of fault current curve identification method based on similarity[d]. Lanzhou Jiaotong University, 2016

[8] Xia Ting. Moment-based image analysis method [d]. Southeast University, 2007.

[9] Chen Yating, Lu Lingyun. A tcp adaptive discriminator based on naive Bayesian classification in large-scale hybrid networks[J]. Journal of Communications, 2018,39(s1):189-194.