A New Fault Diagnosis Method of Flexible OSC Clamp

Xingjun Tian¹, Qi Yao¹, Yufeng Yao¹, Tao Zhang¹, Yuntao Lei¹ and Jiu Li¹*

¹School of Electrical and Electronic Engineering, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, 050043, China
*Corresponding author’s e-mail:513106501@qq.com

Abstract. The capacity and safety of train transportation mostly depend on the working state of the flexible overhead contact system (OCS) clamp in the traction power supply system. The notable feature of the clamp loosening problem is its temperature change. Due to the influence of long-distance measurement, environmental factors, the traditional signal processing methods for temperature measurement have basically failed. The feature analysis method combined with empirical clustering and time series analysis can effectively determine the thermal health status of bare conductors in the overhead contact line and realize the diagnosis of the thermal state. In order to eliminate the much interference in measurement process, the data is pre-processed based on the point density criterion, and then the clustering method of big data is used to calculate the relative proximity of the sampled data, the normal data and the historical fault data. And the time-series similarity analysis method is used to obtain the predicted fault development time of the hidden equipment. Experiments show that the method can accurately identify the thermal health of bare conductors, the concept is clear and the algorithm is simple.

1. Introduction

The OCS clamp in railway power supply system is easy to become slightly loose, which is the early fault behavior of clamp connection state, so the early identification of clamp loosening is considered as an important measure to improve the reliability of traction power supply system [1]. With the continuous advancement of smart grid construction, many scholars have applied various intelligent algorithms to the field of fault diagnosis and achieved good results [2]. Accurate analysis of the thermal state of the bare conductor is of great significance to the safe operation of the railway system and to improve economic efficiency.

At present, the detection of bare conductors mainly relies on infrared non-contact temperature measurement and contact temperature measurement of PT100 temperature sensor and other modules, and only determines the operating range of the bare conductor thermal state by analyzing the temperature. However, it is extremely difficult to eliminate early failures and it is difficult to analyze hidden dangers. In the existing research, it has been proposed to use the monitoring resistance change to achieve the purpose of detecting the health of the bare conductor, and some scholars have proposed the idea of using thermal imaging analysis [3]. However, there is no algorithm support. The former has complex measurement variables, and the latter has too high detection cost, which is difficult to implement under the actual working conditions of the railway OCS.

In this paper, the idea of big data clustering is applied to bare conductor detection, combined with time series analysis method, a set of OCS bare conductor thermal state evaluation system is established. On the basis of measuring the temperature change of the naked conductor, a comprehensive evaluation of the current performance and hidden danger of the naked conductor is completed. The model firstly performs data preprocessing based on the point density criterion, marks
and eliminates noise points to complete the data; secondly, the sampling data points are selected to calculate the relative proximity between them and the historical fault data clusters; finally, according to the results of relative proximity calculation, the state of the equipment is marked and classified.

2. Device status based on big data clustering

The bare conductor fault is mainly caused by the increase of resistance due to looseness, which leads to heating. The collection of bare conductor fault history data samples will contain the collection of bare conductor thermal state data under bare conductor fault. All historical data points that fault a bare conductor are called a cluster of \( C \), Based on big data clustering and time series analysis, the idea of the thermal state evaluation system for the bare conductor of the OCS was adopted to calculate the mean reciprocal distance between the nearest data points \( x \) and \( y \), Thus, the proximity \( l(x, k) \) of the data point \( x \) to the data cluster \( C \) can be obtained.

\[
l(x, k) = \left( \frac{d(x, y)}{k} \right)^{-1}
\]

\( C \) is the data point \( x \) which the set of \( k \) nearest neighbors in cluster \( C_j \); \( y \) is the nearest neighbor of \( x \) in cluster \( C_j \); \( d(x, y) \) is the Euclidean distance between \( x \) and \( y \), which can be calculated by equation (2).

\[
d(x, y) = \sqrt{\sum_{i=1}^{m} \lambda_i (x_i - y_i)^2}
\]

\( m \) is the data dimension; \( x_i \) is the coordinate of the data point \( x \) on the dimension \( i \); \( y_i \) is the coordinate of the data point \( y \) on the dimension \( i \); \( \lambda_i \) is the normalization coefficient on the dimension \( i \), obtained in the data preprocessing stage.

The internal data point of \( C \) is taken as the data point to be tested, and is substituted into equation (1) for calculation, and the degree of proximity of the data point with respect to the cluster is obtained, which is called intra-cluster proximity. The intra-cluster proximity of all internal data points is averaged to obtain the base proximity \( l_{ref} \) of \( C \).

For the data point \( x \) to be calculated: the greater its proximity to the cluster, the closer it is to the cluster, which is more likely to fail. The degree of proximity of data point \( x \) and cluster \( C \) is represented by the relative proximity \( l_c(x, k) \).

\[
l_c(x, k) = \frac{l(x, k)}{l_{ref}}
\]

3. Trend prediction based on time series similarity analysis

The change of the fault temperature of the naked conductor has time sequence characteristics. For the naked conductor with latent fault risk, the speed and trend of the transition from the current state to the fault state can be judged based on the similarity analysis of time series.

3.1 Determining the time series of data to be tested

When the relative proximity between the data point \( x \) of the device to be evaluated and the bare conductor fault data cluster \( C \) is in the interval \([0.6, 0.8]\), the data points \( x \) are taken at the plurality of data points before the acquisition time as the time series \( X \) of the data to be tested. The steps are as follows:

- Set the length \( N \) of the sequence \( X \);
- Determine the last data point time stamp \( t_{SN} \): Set the acquisition time of the data point \( x \) to \( t_{SN} \);
- Calculate the relative proximity: set the maximum search time \( t_{max} \) of the sequence \( X \). Calculating the relative proximity \( l_c(x(t), k) \) of the data point \( x(t) \) in the range of \( t \in [t_{fb}, t_{max}, t_{max}] \) and the fault data cluster \( C \) according to equation (3);
- Determine the initial data point time stamp \( t_{x1} \): Search from the \( t_{xN} \) to the first data point \( x(t) \) with relative proximity \( l_0(x(t),k) = 0.5 \delta_c \), and \( \delta_c \) is the search allowable error. If there is a data point that meets the requirements, set the corresponding data acquisition time to \( t_{x1} \). If there is no data point that meets the requirements, set \( t_{xN} \) to \( X_1 \);

- Get the time series: take \( D \) as the sampling interval, extract \( d_{xk} = \Delta t_0/(N-2) \) data points in chronological order from the beginning, and get the final time series \( X=\{x(x_1), x(x_2), \ldots, x(x_N)\} \).

### 3.2 Determining the fault history data time series
Taking the data point \( y \) adjacent to a data point \( x \) in the bare conductor fault cluster \( C \) as an example, the time series \( Y \) before the data point \( y \) is selected as the comparison sequence. The process is basically consistent with the time series \( X \) of obtaining the data to be tested, and only the first and last data points are determined to have different time stamping methods, and the difference is as follows.

- Determine the initial data point time stamp \( t_{y1} \): The acquisition time of the fault data point \( y \) is \( t_{yb} \), and the data point \( y \) is also recorded as \( y(t_{yb}) \). Set the maximum search time \( t_{ymax} \) for sequence \( Y \). Calculate the relative proximity of the data \( y(t) \) in the range \( t\in[t_{yb} - t_{ymax}, t_{ymax}] \) and the fault data cluster \( C \). \( l_0(y(t),k) \) searches for the data point satisfying \( l_0(y(t),k) = l_0(y(t_{y1}),k) \pm \delta_{y1} \) from the time \( t_{yb} \), and \( \delta_{y1} \) is the initial search allowable error. If a data point that meets the requirements is found before the time \( t_{ymax} \), the corresponding acquisition time is set to \( t_{y1} \). Otherwise, end all steps and consider the search to fail.

- Determine the end data point time stamp \( t_{yN} \): Perform a second search in the \( t_{yj} \) to \( t_{yb} \) interval, and search for the first data point backward from \( t_{y0} \), so that \( l_0(y(t),k) = l_0(y(t_{yN}),k) \pm \delta_{y2}, \delta_{y2} \) is the secondary search allowable error. If you find a data point that meets the requirements before time \( t_{yb} \), set the corresponding acquisition time to \( t_{y} \). Otherwise, end all steps and consider the search to fail.

### 3.3 Similarity calculation of time series
The distance between two time series can reflect their similarity characteristics. The smaller the distance is, the higher the similarity of the two time series will be. Commonly used time series distances include Euclidean distance \([4]\), edit distance \([5]\), dynamic time bending distance, and so on \([6]\). Among them, the edit distance between two strings sequence is used to calculate conversion steps needed to edit. Euclidean distance is the subscript of the corresponding two time series data points, the sum of the distance between corresponding computing dynamic time warping distance allows synchronization point, its essence is by finding the minimal path between the two time series to calculate the distance between them. Since the time series data element in this paper is the temperature information of the naked conductor, and it is necessary to prevent the phenomenon that the distance evaluation result is too large due to the dislocation of some data, the dynamic time bending distance is adopted to measure the similarity of the two time series. The time series \( X=\{x(t_1),x(t_2), \ldots, x(t_N)\} \) and the time series \( Y=\{y(t_1),y(t_2), \ldots, y(t_N)\} \) Dynamic time bending distance can be obtained by recursive calculation \((4)\).

\[
D_{dtw} = \begin{cases} 
0 & X = \{\emptyset\} \text{and } Y = \{\emptyset\} \\
\infty & X \neq \{\emptyset\} \text{or } Y = \{\emptyset\} \\
d(x(t_1), y(t_1)) + D_{dtw}(X, Y) & X = \{\emptyset\} \text{and } Y \neq \{\emptyset\} 
\end{cases} 
\]

\[
D = \min[D_{dtw}(X,R(Y)), D_{dtw}(R(X), Y), D_{dtw}(R(X), R(Y))] 
\]

\[
d(x(t_1), y(t_1)) = |x(t_1) - y(t_1)|; \quad R(X) = \{x(t_2), x(t_3), \ldots, x(t_N)\}; \quad R(Y) = \{y(t_2), y(t_3), \ldots, y(t_N)\}. 
\]

In order to conveniently measure the distance between different length sequences, the dynamic time bending distance calculation results are normalized, and the normalized distance finally used to evaluate the sequence similarity is

\[
D(X,Y) = D_{dtw}(X,Y) 
\]
4. Comprehensive evaluation of bare conductor status

4.1 State evaluation method flow
The comprehensive evaluation method of bare conductor state is divided into three stages, the flow of which is shown in Figure 1.

Figure 1. Method flow chart

- First, data preprocessing is performed based on the point density criterion, and the noise points are marked and eliminated, and the data is complemented.
- Next, select the sampled data point \( x \) and calculate its relative proximity to the historical fault data cluster.
- Finally, based on the relative proximity calculation results, the device status is marked and classified, and is divided into three parts:
  - The relative proximity of the data point to the fault data cluster is greater than 0.8, which indicates that the equipment is at a high risk of failure, and the bare conductor is recorded as a fault;
  - When the relative proximity is in the interval \([0.6, 0.8]\), it indicates that the bare conductor is at a higher risk, and the bare conductor state is marked as a latent fault. The time series analysis method is used to further judge the speed and trend of the device transition to the fault state;
  - When the relative proximity of the data points and all data clusters are less than 0.6, the device status is marked as healthy and the health score of the device is calculated.

4.2 Computing device health score
When the relative proximity calculation result is less than 0.6, the health score of the computing device is used as a comprehensive evaluation index of the device state. For determining the bare conductor fault \( j \), the known data point \( x \) and its proximity \( l_c(x, k) \) can be converted into a bare conductor with a fault-related health score. Set 100 to full score, the expression of the device's health score \( F(t) \) is

\[
F(t) = 100 \left[ 1 - \max(1, l_c(x, k)) \right] \quad (6)
\]

5. Case analysis

5.1 Lab Environment
The schematic diagram of experimental set-up was shown in Figure 2, and the surface temperature of the wire clamp group was measured every once in a while with an infrared thermometer. See Figure 3 for the experimental site.

![Schematic diagram of experimental set-up](image1)

![Experimental site](image2)

**Figure 2. Schematic diagram of experimental set-up**

**Figure 3. Experimental Show**

### 5.2 Experimental result

This paper use visual studio 2017 and write the implementation software based on C#. This software has the functions of temperature measurement, automatic recording of temperature measurement time, data recording, data calculation according to the above algorithm, output calculation results and related parameters, as well as failure alarm, hidden danger alarm and other related functions. In the engineering site, it can be switched to the timing temperature measurement mode and automatic calculation mode, and the score response of the hidden naked conductor is estimated to need to be replaced.

By the short circuit experiment: only in electricity 10 minutes we found wiring is lax bare conductor wire clip, program play appear alarm prompt, part of the clamp the measured temperature as shown in table 1. The time-temperature sequence diagram is shown in figure 4. Does not exclude the fault in the experimental process clamp to test, in the time series analysis found that there are other unqualified samples. This result is consistent with the background normalized distance response. The results are shown in table 2.

| Time series | clamp1 | clamp2 | clamp3 | clamp4 | clamp5 | clamp6 | clamp7 | clamp8 |
|-------------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1           | 17.4   | 17.4   | 17.4   | 17.4   | 18.6   | 19.9   | 39.2   |
| 2           | 17.3   | 17.3   | 17.3   | 17.3   | 17.3   | 23.5   | 34.1   |
| 3           | 18.5   | 18.51  | 17.2   | 18.2   | 17.1   | 18.4   | 25.0   | 54.6   |
| 4           | 18.4   | 18.4   | 18.4   | 18.4   | 19.4   | 19.3   | 61.1   |

According to the above algorithm, this paper write the fault time prediction function based on the relationship between the normalized distance value and a large number of sample fault records. Due to time problems, not enough data samples have been obtained, and the module is still in continuous update.
6. Conclusion
This paper proposes a method for evaluating the bare conductor state based on clustering and time series analysis. The following conclusions can be verified:

- Based on the point density criterion, the method preprocesses the data and eliminates the influence of noise data. Based on the clustering method, the state of the naked conductor is divided into three categories of healthy, latent faults and faults. Compared with the traditional fault diagnosis algorithm, the judgment accuracy of the latent faults and faults in the naked conductor is higher.
- The equipment health score evaluation index is set based on fault weighting, which can directly reflect the equipment health status and provide reference for the equipment operation and maintenance arrangement.

References
[1] XIN J B, KANG W, WENG X L, Chen T, XIE B, GUO X. A Transformer State Evaluation Method Based on Clustering and Time Series Analysis[J]. Power System Protection and Control, 2019, 47(03):64-70.
[2] ZHANG Y, PENG Y X. Study on the thermal state evaluation system of bare conductors in railway contact network and substation [J]. SHENHUA Technology, 2018, 16(07):83-86.
[3] Liang Z, XU M Z, HOU J P. Evaluation method of thermal state of outdoor terminals for ultra-high voltage cross-linked cables [J]. High voltage technology, 2017,43(02),682-688.
[4] MENG J, Zhang W. Volume measure in 2DPCA-based face recognition [J]. Pattern Recognition Letters, 2007, 28(10): 1203-1208.
[5] Production and Technology Department of State Grid Corporation of China. Guide for condition evaluation of oil-immersed power transformers (reactors): Q/GDW 169—2008[S]. Beijing China Electric Power Press,2008.
[6] RISTAD E S, YIANILOS P N. Learning string-edit distance [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(5): 522-532.
[7] ZHONG J Y, ZHANG Y, WEN F Z, et al. Fault diagnosis method for transmission system based on time series similarity matching[J]. Power System Automation, 2015, 39(6): 60-67.