Gross Error Elimination Based on the Polynomial Least Square Method in Integrated Monitoring System of Subway

Zi-Yu Ma[1] Da-Wei Li[2] Fang-Wu Dong*[1]
[1]Faculty of electrical and rail traffic, ZheJiang Textile & Fashion College
[2] Ningbo City Rail Transportation Group Company Limited operating subsidiaries
NingBo city, ZheJiang, China, 315211, telp 086-574-863296228
Corresponding author, e-mail: Mazy1962@163.ccom, 15058291155@139.com, dongfw01@163.com

Abstract

The measurement data of parameter in the electrical equipment contains many noises in subway integrated monitoring system. To eliminate the impact of gross error in the measurement data, a polynomial least square curve fitting algorithm is used in this paper. Based on the Rajda criterion, the algorithm gives the variance estimation of the noises, and then uses dynamic threshold to detect and replace the measurement data with gross error by statistical estimation. Finally, a data processing procedure has been presented to deal with the gross error. The practical application indicates that the proposed algorithm can effectively eliminate the gross error in many types of measurement signals so as to ensure the reliability of the monitoring system.

Keywords: Subway Integrated Monitoring, least square, dynamic threshold, error handling

1. Introduction

To monitor the work state of subway equipments, many electrical parameters such as voltage, current, and work frequency are measured and transmitted to controlling computer in real time. Nevertheless, there are many noises in the measurement datas of the parameters. For example, various interference sources exist in onsite monitoring. Detection devices also have chance to generate errors [1-2]. Besides, some signal disturbances cannot be avoided in the data collection process and remote transmission process. Therefore, gross measurement signals sometimes contain a certain amount of errors, although the probability of the occurrence of those gross errors is low. Because the amplitude of the error is relatively large, the untreated gross measurement datas with error cannot be directly inputted to the computer for data processing, which would lead to inaccurate input data, wrong processing results, or even misoperation of the devices [3-5]. Therefore, data processing of the measurements should be conducted to eliminate the error and ensure comprehensive monitoring system be in good work condition. Based on the statistical theory and calculation method, this paper uses a polynomial least square method to automatically eliminate the gross errors in the measurements. The method is then effectively applied in electrical parameters acquisition and data processing in a subway station.

2. Gross Error Processing
2.1. Gross Error Processing Method

When the equipments in the integrated monitoring system of subway are in normal work condition, the voltage, current, and other dynamic physical parameters are usually continuous with normal distributions respect to time. According to the normal distribution of error theory, any changes of a physical parameter are continuous with no jump [6]. In a certain period of time, a set of data can be obtained by using polynomial least square curve fitting on a continuous recording of a measured parameter [7-8]. Dynamic mean value of the parameter can then be got by averaging the above two sets of data. According to the Rajda criterion, i.e. the small

probability does not exist [9-10], a proper threshold is set for the gross error, which replaces the obviously wrong measurement datas with corresponding statistical values [11-12].

2.2 Mean Value of Dynamic Noise

In comprehensive monitoring system, the system measurement signal can be assumed as a sequence of recorded data with respect to the time ((T_i, Y_i), i=1, 2…n). When step length of data sampling is set as T, the ith data point is generated in time instant T_i:

$$T_i = \Delta T \cdot (i - 1) \tag{1}$$

The mth least square fitting polynomial is:

$$PY_i = \sum_{k=0}^{m-1} c_k T_i^k \tag{2}$$

The deviation on each data point is:

$$|\varepsilon_i| = |Y_i - PY_i| \tag{3}$$

To minimize the summation of the square of the deviation, polynomial least square curve is used to fit the data,

$$S = \sum_{i=1}^{n} (Y_i - PY_i)^2 = \sum_{i=1}^{n} \left[Y_i - \sum_{i=1}^{n} c_k T_i^k \right]^2 \tag{4}$$

When (4) reaches its minimaal value, multiple function value conditions indicate:

$$\frac{\partial S}{\partial C_k} = 2 \sum_{i=1}^{n} \left[Y_i - \sum_{k=1}^{n} c_k T_i^k \right] T_i^k \tag{5}$$

After inner production simplification is introduced in (5), polynomial (6) and (7) can be obtained:

$$\begin{bmatrix} n & \sum_{i=1}^{n} t_i & \dots & \sum_{i=1}^{n} X_i^m \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} T_i^2 & \dots & \sum_{i=1}^{n} T_i^{m+1} \\ \sum_{i=1}^{n} T_i^m & \sum_{i=1}^{n} T_i^{m-1} & \dots & \sum_{i=1}^{n} T_i^{2m} \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ \dots \\ c_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} T_i Y_i \\ \sum_{i=1}^{n} T_i^m Y_i \end{bmatrix} \tag{6}$$

$$A = \begin{bmatrix} c_1 \\ \dots \\ c_m \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} T_i & \dots & \sum_{i=1}^{n} T_i^m \\ \sum_{i=1}^{n} T_i & \sum_{i=1}^{n} T_i^2 & \dots & \sum_{i=1}^{n} T_i^{m+1} \\ \sum_{i=1}^{n} T_i^m & \sum_{i=1}^{n} T_i^{m+1} & \dots & \sum_{i=1}^{n} T_i^{2m} \end{bmatrix} \cdot \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} T_i Y_i \\ \sum_{i=1}^{n} T_i^m Y_i \end{bmatrix} \tag{7}$$

The fifted value of data points (T_i, PY_i)(i=1,2......n) is then obtained. The dynamic noise value is

$$\varepsilon_x = \sqrt{\frac{\sum\limits_{i=1}^{n}(Y_i - PY_i)^2}{n-1}} \tag{8}$$

2.3 Replacement of the gross error

According to the Rajda criterion, there is 97% confidence of the data when its absolute error is beyond 3ε. [13] Because the measurement in the integrated monitoring system may contain fault information, and thus it need to be retained for further fault identification. The gross error threshold is set as $5\hat{\varepsilon}_x$ with corresponding confidence being 99.99%[14] , i.e.:

$$|Y_i - PY_i| > 5\hat{\varepsilon}_x \tag{9}$$

Data satisfying (9) is the gross error, which is needed to eliminate and then to replace. Considering the continuous multi-point gross error processing, data in i+1, i+2 point can be determined. The replacement can be obtained according to formula (10) and (11) interpolation:

When two nearby points are need to interpolated, i.e. Y_i, Y_{i+1} are both gross errors, the replacement value are:

$$Y_i = \frac{Y_{i-1} - Y_{i-2} + Y_{i+2}}{3} \tag{10}$$

When a single point is need to replace, i.e. Y_i is a gross error, the replacement value is:

$$Y_i = \frac{Y_{i-1} + Y_{i-2} + Y_{i+1} + Y_{i+2}}{4} \tag{11}$$

3. Gross Error Processing
3.1. Gross Error Processing Flowchart

The flowchart of the gross error processing is shown in Figure 1. Detailed algorithm is as follows:

Step 1: the initial value in the reference section is set as the first value of the whole gross error data estimation $\hat{\varepsilon}^2$.

Step 2: estimate the state value of the next point. Compare the estimation and the real value, if the difference$|Y_i - PY_i|$ is less than $3\varepsilon_x$, this estimation is considered as reasonable. Otherwise, the point is thought as gross error and need to replace. Since the signals in monitoring system are all continuous and dynamic, the prediction can reply on the value in the last point. Because the error of the prediction gradually becomes larger when several consecutive gross errors appear, threshold should be set to be larger, such as $5\varepsilon_x$, to avoid misjudgment.

Step 3: update reference section. When the reference section moves backward by one point, its first value is removed and replaced by an adjacent value in the reference section (in this manner, the measurement is judged as gross error, and then replaced with its prediction). If the data processing is completed, switch to Step4, otherwise, go to Step 3.

Step 4: the data after the replacement in the last round is re-estimated $\hat{\varepsilon}^2$ to judge whether there are new gross error is remove in this round of data. If not, the computation is completed. Otherwise, go to Step2. This process is actually a loop of replacement of the gross errors. The larger errors are firstly replaced and then followed by smaller errors. The whole replacement can generally be finished when the loop is run for 3 times.
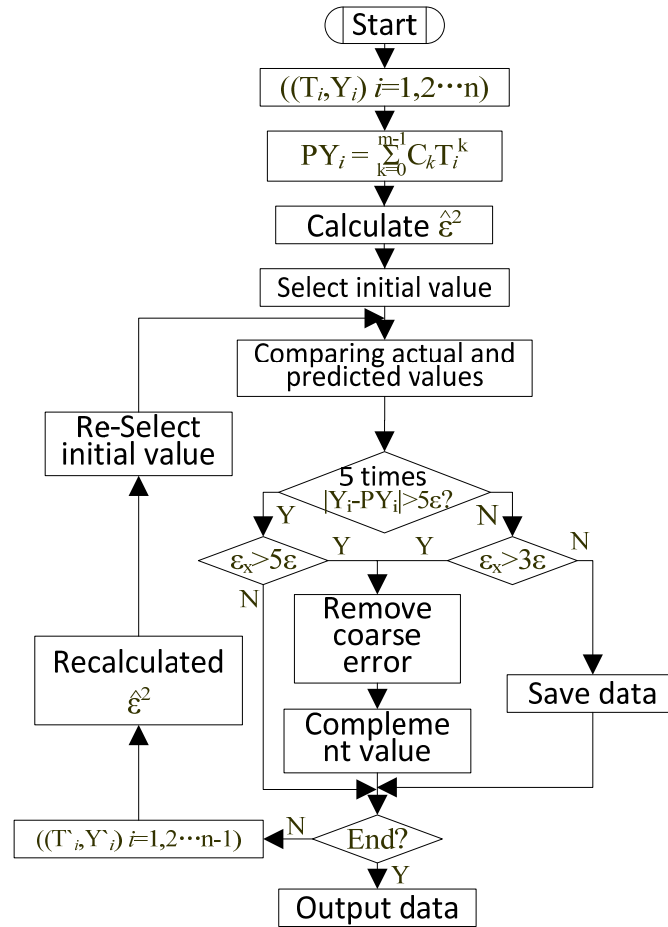
Start

$((T_i, Y_i)\ i=1,2\cdots n)$

$PY_i = \sum_{k=0}^{m-1} C_k T_i^k$

Calculate $\hat{\varepsilon}^2$

Select initial value

Comparing actual and predicted values

5 times $|Y_i - PY_i| > 5\varepsilon$?

Re-Select initial value

$\varepsilon_x > 5\varepsilon$

$\varepsilon_x > 3\varepsilon$

Remove coarse error

Recalculated $\hat{\varepsilon}^2$

Complement value

Save data

$((T`_i, Y`_i)\ i=1,2\cdots n-1)$

End?

Output data

Figure 1. Flowchart of gross error processing

4. Practical Application
4.1. Gross Error Data Processing

A data serie of a motor working current is shown in Table 1 and Figure 2. The initial ε_x is 0.13 from calculation. From Figure 2, at least three measurement datas are obviously wrong and need to replace.

Table 1. Measurement data of the current for a working motor (Unit: A)

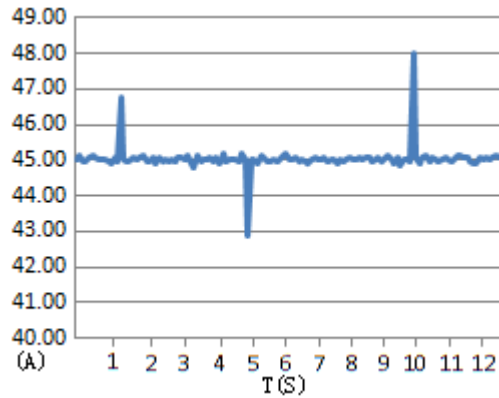| NO.	A	B	C	D	E	F	G	H	I	J
1	45.04	45.14	44.97	44.97	45.07	45.14	45.09	44.99	45.04	45.01
2	44.97	44.93	45.10	44.96	46.79	45.02	44.97	45.04	45.09	44.99
3	45.09	45.12	44.98	44.98	45.08	44.93	45.07	44.94	45.03	44.96
4	45.01	44.98	45.10	45.09	44.99	45.14	44.93	44.79	45.13	44.97
5	44.99	45.04	45.06	44.98	45.09	44.89	45.21	44.98	44.99	45.04
6	45.02	44.94	45.19	44.99	42.85	44.99	45.01	44.93	45.11	45.06
7	44.98	44.97	45.01	44.95	45.01	45.11	45.19	44.99	45.01	45.09
8	44.98	45.03	44.97	44.93	45.01	45.09	45.03	44.99	45.07	44.98
9	44.99	45.02	44.91	44.99	44.94	45.03	45.06	45.01	45.03	45.09
10	44.99	45.06	45.09	44.95	45.03	45.08	44.97	45.01	45.12	44.99
11	44.93	45.09	44.83	45.03	45.04	44.98	47.98	45.09	44.89	45.06
12	45.11	44.98	45.07	45.02	44.96	45.03	45.09	44.99	44.95	45.03
13	45.12	45.13	45.09	45.07	44.98	44.93	44.92	45.06	45.01	45.10
14	44.99	45.06	45.13	45.06						

Figure 2. Current measurement data of a working motor (Unit: A)

After the first loop of gross error elimination, errors are basically eliminated as shown in Figure 3. However, the figure contains some burrs. When several loops are completed, burrs are basically eliminated as shown in Figure 4.

4.2. Algorithm Improvement

Because the proposed algorithm is a kind of forward algorithm, it can only judge the data in the next point. Therefore, if the first data point is with gross error, this point cannot be replaced. In order to solve this problem, it's needed to judge whether the first data point contains gross error. In other words, if (9) is not satisfied, another starting point should be selected.
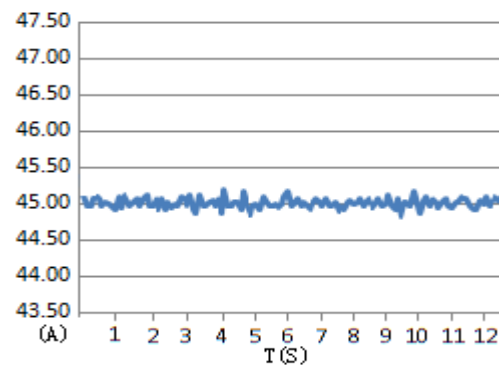

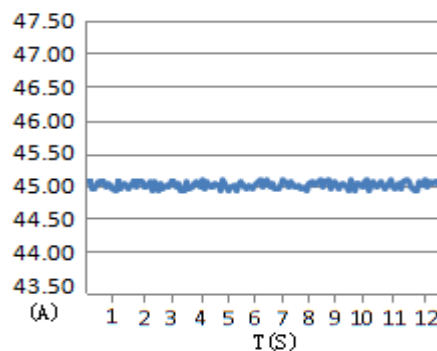Figure 3. Gross error elimination result after the first loop

Figure 4. Final gross error elimination result

When this algorithm is used to eliminate gross errors, data with large error have been removed in Figure 3, but there still exist many small burrs, which affect the smoothness of curve. This phenomenon is usually treated by using average method. Average method can effectively improve the smoothness of the data curve. However, it only uses the average value of the adjective measurements to replace the burr, and cannot fundamentally eliminate the error. Therefore, in practical application, bandpass filtering is used to further improve the data smoothness after the gross error being replaced.

4.3. Analysis of Results and Discussion

In the gross error processing, according to Rajda criterion, $|Y_i\text{-}PY_i|>3\hat{\varepsilon}$ is set as the judgment condition. Because fault information may exist in the integrated monitoring system of subway, and the noises become large in the fault sites, $|Y_i\text{-}PY_i|>3\hat{\varepsilon}$ is not proper any more. If this criterion is not changed, some useful information will be ignored, which affects the reliability of the system. Therefore, In practical application, if the $5\hat{\varepsilon}$ successive measurements contain gross errors, $|Y_i\text{-}PY_i|>5\hat{\varepsilon}$ is set as a new criterion.

Through the analysis of the measurement in Figure 2, figure 3 and Figure 4 from an equipment operation integrated monitoring system of subway, the proposed gross error processing can completely meet the requirements of monitoring system, eliminate the phenomenon of the slow response and the misoperation of the comprehensive monitoring system. The maximum relative errors of all the parameters are controlled within 2%. Therefore, the polynomial least square curve fitting method with Rajda criterion shows good performance in detecting and replacing gross error. It has good practicality in not only eliminating the gross error, but also improving the reliability of the system.

5. Conclusion

This paper introduces the polynomial least square method into the data processing of the integrated monitoring system of subway. The measurement datas are firstly statistically analyzed, and then smoothed by removing the gross errors detected using Rajda criterion. The removed measurement datas are replaced by values obtained from statistical theory, which improves the accuracy of the measurements and the reliability of the system. The practical application shows that, the method has the characteristics of high reliability, strong practicability in monitoring equipment operation parameters for the metro integrated monitoring system.

References

[1] Ayong Hiendro. A Quantities Method of Induction Motor Under Unbalanced Voltage Conditions. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2010; 8(2): 73-80.
[2] ZHANG DQ, CHEN SC. Anovel kernelized fuzzy c-means algorithm with application in medical image segmentation. *Artificial In telligence in Medicine*. 2004; 32 (1): 37- 52.
[3] Didik R Santoso. A Simple Instrumentation System for Large Structure Vibration Monitoring. *TELKOMNIKA*. 2010; 8(3): 265-273.
[4] Lu Fengjie, Zhang Yongrui, Yang Gang. Study of the Gas Sensor Precision Based on Leastsquare Methed. *Electronic Science and Technology*. 2007; (6): 36-38.
[5] K Rajagopala, K Panduranga Vittal, Hemsingh Lunavath. Computation of Electric Field and Thermal Properties of 3-Phase Cable. T*ELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(2): 265-273.
[6] Zhang Liang, Zhang Xinping. Eliminating Intolerble Errors Contained in Flight Research Data. *Journal of Northwestern Polytechnical University*. 2011; (4): 637-638.
[7] XIONG Yan-yan, WU Xian-qiu. The Generalizing Application of Four Judging Criterions for Gross. *Physical Experiment of College of china*. 2010; 23(1): 66-67.
[8] Yang Li, Zhang Li, etc. An Arithmetic of Eliminating Faulty Value in On-line Monitoring. *High Voltage Apparatus*. 2000; (10): 3-4.
[9] Sun Tao, Zhang Hongjian. A Method of Outlier Detection and Correction Based on First Order Differential. *Chinese Journal of Scientific Instrument*. 2002; 23(2): 198-199.
[10] Pan Xudong Wang Guanglin Shao Dongxiang,etc. Measurement of Spool Valve Overlap Value of Servo Valve. *Journal of Vibration, Measurement & Diagnosis*. 2009: 29 (4): 292-297.
[11] FEI Yetai. *Error Theory and Data Processing*. Beijing: Mechanical Industry Press. 2002; 153-195.
[12] CHENG Maolin. The Least Squares Integral Method of Data Fitting Function. College Mathematics. 2006; (1): 70-73.

[13] NI Hui, LI Zhong, SONG Hongxing,etc. Moving Least Square Curve Fitting with Interpolation Conditions. *Journal of Zhejiang Sci-Tech University.* 2011; (1): 135-138.
[14] SHENG Wang, Liu Wenyan. Processing Dynamic Measuring Errors By Using Softwar. *Process Automation Instrumentation.* 2003; (11): 21-22.