TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 13, No. 1, January 2015, pp. 101 ~ 105
DOI: 10.11591/telkomnika.v13i1.6695

An Abnormal Signal Diagnostic Emendation Technique Research

Tiejun Cao
School of Information Science and Engineering, Hunan International Economics University,
Changsha, China, postcode: 410205
email: matlab_bysj@126.com

Abstract

This paper introduced a sort of abnormity data method. We start with analysing an example first and carry through theory deducing once more and list this method arithmetic step and application field finally. With the development information and control technology, we need more and more important and imminent to diagnose abnormity data and eliminate them. The quality of measuring data is improved availably by the application of data emendation technique.

***Keywords**: abnormal signal, diagnosis, process error, algorithm, information, control*

1. Introduction

From the information theory, control theory point of view, the production process can be regarded as an information flow. Production process with a large amount of process information, reflect changes in the status of crafts and equipment, reflecting the various aspects of the interaction and association implies the regularity of production optimization. Process information is mainly reflected in a large number of state data, it is process control and process optimization basis. However, data collection and flow of the process, often associated with gross error data, such as sensors, switches, and recorder of the failure, artificial logs or random data entry errors. No matter what kind of data processing, data are required to reflect the objective reality as possible, accurate, reliable and complete. If the number of true false count, the results not only does not make sense, but also cause errors due to false information in decision-making and control. So, whether it is manual processing, or computer processing, data must be identified, the first sentence of its authenticity, namely, fault diagnosis and error correction. In view of the theory in this area more and more attention to workers and the actual producers, the paper where the fault error by regression analysis, an abnormal diagnostic methods data, and describes the application of the method.

2. Regression Analysis of Data from the Exception Instance about Interference

Regression analysis is an effective and practical method of mass modeling. Identification of outliers in meta-data approach is that people often draw two yuan scatter plot, then the human eye to observe the discrete state, but in the multi-dimensional data, using observation methods can not identify abnormal data. Some people say that regression analysis can be predicted and actual values of the absolute difference (error) to determine. This conclusion may not be reasonable. Consider a few examples, the data in Table 1.

Between y and x is assumed a linear relationship. The least square method with regression, a straight line L1: $y = 0.06833 - 0.08146x$ (see Figure 1)

The return value points y_i and residuals r_i are listed in Table 1, the return of the remaining standard deviation $\sigma_1 = 1.55$. As the largest absolute value of residual $| r_{max} | = 2.09$, residual standard deviation does not exceed twice Therefore, in accordance with usual practice, the data can not believe to be outliers, but if you look carefully you will find the data in Figure 1 except for the first six points, other points in general in a straight line. Therefore, point 6 is likely

Received September 16, 2014; Revised November 12, 2014; Accepted December 8, 2014

to be an abnormal point, if the first 6 removed, with the other five points with a straight line, then get L2 in the Figure 1: $y = -1.87333 - 0.97767x$

Table 1. The original data and regression results

N	x	y	L1		L2	
			y	r	y	r
1	-4	2. 48	0.39	2.09	2.04	0.44
2	-3	0. 73	0.31	0.42	1.06	-0.33
3	-2	-0.04	0.23	-0.27	0.08	-0.12
4	-1	-1.44	0.15	-1.59	-0.9	-0.54
5	0	-1.32	0.07	-1.39	-1.87	0.55
6	10	0	-0.75	0.75	-11.64	(11.64)
Residual standard deviation			σ_1=1.55		σ_2=0.55	
$\lvert r_{max}\rvert/\sigma$			1.35		1.00	

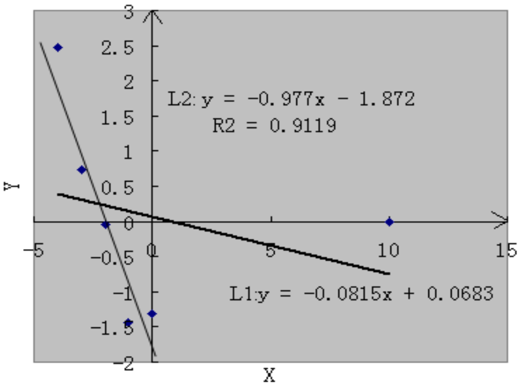

Figure 1. Experimental data and regression results

Far from a straight line L2 and L1, and L2 and the vast majority of points are very close; and L1 is a lot of points with a larger gap. From this example, least-squares regression can be seen in at least two weaknesses:

1) The individual data points may be a great influence on the regression results, its performance is the point of no return and participate in this point to participate in return, the larger difference between the results obtained, point 6 above, is this point.

2) anomalies in the data value, not necessarily with the size of the residuals found out, in the example anomaly of the first six points of residuals is quite small (0.75), while the largest residuals in point 1 (r_1 = 2.09) are not outliers; if because of its large and suspect it is the residual outliers, even weed out, it will be worse than the L_1 regression results. Thus, in the least-squares regression using the residuals to determine the size anomaly is unreliable, even misleading.

Reasons for these problems, mainly due to least-squares method to minimize the sum of squared residuals, it is equally treat all by requiring that arrogance may be close to the regression line every point, when the data contains outliers, due to the minimum squares it is "equal" to the regression results influenced by it, but can not give more accurate regression equation. In this regression equation to get on base out of the residual nature is not reliable. In addition, the regression of observed values for the dependent variable from the variables x and y of two parts, while the y component of the residual is the difference between its return value, the x-component which does not fully reflect some of the factors. Below we analyze theoretically.

3. Theoretical Analysis of Generated Gross Error

Linear model:

$$y_i = x_i{}'\beta + \varepsilon_i \qquad i = 1,2,\cdots,n$$

In which x 'is the transpose of x, (xi,yi) is observed data, $x_i \in R^n, y_i \in R^1, \beta = (\beta_1, \beta_2, \cdots, \beta_m)'$ is regression coefficients to be estimated, ε_i is

$$y_i = (y_{i1}, y_{i2}, \cdots, y_{in})',$$

Random error. $x_i = (x_{i1}, x_{i2}, \cdots, x_{in})' = (x_{ij})_{n\times m}$

Obtained by the method of least squares regression, results are:

$$\hat{\beta} = (x'x)^{-1}x'y \tag{1}$$

$$\hat{y} = x\hat{\beta} = (x'x)^{-1}x'y = Hy \tag{2}$$

Among: $H = x(x'x)^{-1}x = [x_i'(x'x)^{-1}x_j]_{n\times n} = (h_{ij})_{n\times n}$ (3)

H is the column vector x is a linear subspace spanned by the projection matrix; also known as the "hat matrix" (Hat matrix), the first point of the residual L is:

$$r_l = y_l - \hat{y}_l = y_l - \sum_{j=1}^{n} h_{lj}y_j = (1-h_{ll})y_l - \sum_{j\neq l} h_{lj}y_j$$

If the observed value of y, yi is an abnormal value, assuming it's normal is y_i*,And y_i by y_i* with gross error ⊿y_i , that is, y_i = y_i* + ⊿y_i,, if there is no gross error, the normal situation is the i-th residual:

$$r_i^* = (1-h_{ii})y_i^* - \sum_{j\neq i} h_{ij}y_j$$

But $r_i = (1-h_{ii})(y_i^* + \Delta y_i) - \sum_{j\neq i} h_{ij}y_j$

As the gross error to change the i-th residual, then:

$$r_i - r_i^* = (1-h_{ii})\Delta y_i$$

Thus, while the gross error ⊿y_i is generally relatively large, but if h_{ii} close to 1 (note hii is the projection of the diagonal matrix elements, therefore, 0 ≤ h_{ii} ≤ 1). The corresponding normal r_i* and r_i difference is not large. Therefore, in the i-th data on the gross error, and from that point of the residuals ri not shown, due to gross errors ⊿y_i first k (k ≠ i) point residuals changes are:

$$r_k - r_k^* = (1-h_{ki})\Delta y_i \qquad (k \neq i)$$

Therefore, if h_{ki} relatively large, but the gross errors in the k reflected points on the residuals, which means that, with the residual value of r_i to determine the size of the corresponding point value of y_i is not abnormal is not reliable, because residual change not only the y component of gross errors, but also to the x component, (h_{ij} is completely determined by the x).

From the analysis of data on the return of the i-th effect size close to the corresponding h_{ii}, h_{ii} is large, this effect may be large; generally referred to h_{ii} big point of "leverage points" (Leverage point) or the potential impact points. Generally h_{ii} considered the best in 0.2 the following as well, if possible, should be at least 0.5 or less. In the example above h_{66} = 0.936 is very close to 1, so this is a great impact on the return, but also in the x data, it does stray far.

4. Error Diagnosis Algorithm

Based on the above discussion and analysis, if we capture to n m-dimensional data set group $x = (x_1, x_2, ..., x_n)' = (x_{ij})_{n \times m}$, to diagnose whether the fault which error data, then the following algorithm steps:

STEP1: calculate x'x

STEP2: find the inverse matrix $(x'x)^{-1}$ of x'x

STEP3: In addition to abnormal values to determine a cut boundary value F (0 <F <1);

STEP4: calculation $\alpha_i = x_i'(x'x)^{-1}x_i$, i = 1,2, ..., n, If α_i<F, then xi is the normal number of points; if $\alpha_i \geq F$, then xi outliers.

STEP5: If one parent of the new data set z, then the calculation $\alpha=z'(x'x)^{-1}z$,,With the rules distinguish STEP4 z is abnormal data or normal data.

In order to improve the computing accuracy of the data, the the original data set X can be normalized to center.

5. Conclusion

This method can be used to diagnose and remove abnormal data, use data to improve the reliability, availability and effectiveness.

Troubleshooting: If our signal detection means well, we are faced with a demand for the part (such as industrial process control) in normal and abnormal data are collected and must have, on the section through the algorithm to determine the failure α_i corresponding interval, put $(x'x)^{-1}$ into the register has been quite good, on-line information and data for the detection of Z, calculated$\alpha=z'(x'x)^{-1}z$; According to the size of α determine the severity of failure, such as equipment performance, the implementation of the online real-time quantitative analysis and fault alarm.

Planning and Statistics Administration: plant project management, statistical reporting and decision-making should be based on incoming and outgoing materials and energy devices correct measurements. However, they are with a random error, and even some gross error data also make managers are not able to grasp the true economic benefits of the factory. Diagnosis and correction using the data flow and temperature measurements, in order to obtain reliable data management.

Signal tracking process: the use of online diagnosis and correction techniques to analyze process data, so tracking the operational status of equipment and devices, trends and interference, identify the error or equipment failure situation.

Process control and optimization: a process simulation, advanced optimization and diagnostic calibration software used in combination to provide a reliable process optimization. In process control applications, the implementation of diagnostic correction software automatically, real time access to process data, and average time to do the calculation and correction process to obtain consistent data. Can also be corrected and the data input process simulation program with the latest economic data, computing, simulation and optimization of operating parameters to achieve optimal economic results. After optimization of process control parameter values recommended for the new set point, which is the online feedback control system. Optimization can be offline, online, and it can constitute a control system optimizer and online closed-loop system.

References

[1] Zhao Wei, Sun Jiang. Praised; a new dynamic process data correction; Control Theory and Applications; 1999; 04 33-38.

[2] Gang Rong Wang Xi. Measuring instruments of a single node gross error identification method; *Chemical Engineering*. 2000; 51(1): 41-45.

[3] Rae put Bing-Zhen, Chen Ming. *Data coordination and gross error detection of synchronized robust estimation method*. Tsinghua University (Natural Science); 2000; 40(2): 25-30.

[4] Bo Li, Bing-Zhen, Chen Yong Ming put the East section. Fault detection based on time redundancy error correction coefficient; Tsinghua University; 2000; 40(10): 27-32.

[5] Bo Li, Bing-Zhen, Chen Yong Ming. East section; fault detection based on time redundancy error correction coefficient; Tsinghua University; 2000; 40(10): 33-35.

[6] YUAN Yong-root*. Chemical process data correction techniques from theoretical research to practical conversion*. China Institute of Process Systems Engineering Systems Engineering Committee of the first academic conference proceedings; Jinan. 1991; 10: 110-114.

[7] Xu Chao, Pan Zhaohong. *Chemical consistency of the data correction process; China Institute of Process Systems Engineering Systems Engineering Committee* of the first academic conference proceedings; Jinan, 1991; 10: 118-120.

[8] Chen Jie. *Mathematical Modeling rapid detection of chemical process data in the theoretical basis of gross errors*; chemical process in, the Third Annual Meeting Proceedings. 1990; 120-124

[9] Chen Jie. C*hemical process data in the rapid detection of gross errors*. Mathematical simulation of chemical process third Annual Meeting Proceedings. 1990; 125-129

[10] Li-li LIU the Li-wei, XIONG, Yan-ye. The abnormal signal detection method based on the fat tail distribution. *Journal of Nanjing University (Natural Science).* 2011; (1).

[11] XIA Ya-qin, CUI Xiao-yan, LI Jun-zhi, CHEN Wei-sheng, LIU Cheng-yan. The study of the earthquake the previous abnormal acoustic signals. Beijing University of Technology. 2011; (3).

[12] Zhang Xiang, Zhang Jianqi, Qin Hanlin, Liu Delian. Anomaly detection using multi-resolution decomposition of hyperspectral images. *Infrared and Laser Engineering.* 2011; (3).

[13] Liu Zhanfeng, Si Jingping, Liang Hongbo. based on vibration signal analysis engine cylinder wall clearance abnormal diagnosis. S*mall internal combustion engine and motorcycle.* 2011; (2).

[14] YAN Ji-hong, WANG Wei LU Lei. Automatically update the fault diagnosis model based on artificial immune algorithm. *Computer Integrated Manufacturing Systems*. 2011; (4).

[15] HU Su-yun, E Jia-qiang, GONG Jin-ke. Diesel cold start the process of fuzzy logic inference unusual diagnosis. *Journal of Hunan University (Natural Science).* 2011; (11).

[16] WU Dinghai, ZHANG Peilin, REN Guoquan, XU Chao, FAN Hongbo, based on the Bayes classifier of a hypersphere and the diesel engine anomaly detection applications. *Journal of Mechanical Engineering*. 2011; (6).