A Methodology for Accuracy Assessment of Information Quality in the Processing Production System

(Completed Research Paper)

Zhiyong Gao[1], Jianmin Gao[1], Rongxi Wang[1, 2*], Yanjie Liang[1], Hongquan Jiang[1]
[1]State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, China
[2] Collaborative Innovation Center of High-End Manufacturing Equipment, Xi'an Jiaotong University, Xi'an, China
rongxiwang@163.com

Abstract: High reliability and accuracy of monitoring information is an essential prerequisite for the efficient and safe operation of system. In this paper, a method is proposed to identify the information accuracy in process production system based on collision detection of coupling relationship. The proposed method focuses on the monitoring data of distributed control system (DCS) and considers the strong coupling and complexity characteristics of the state information in process production system. The DCCA method is used to identify the coupling between the monitoring variables and establish the coupling network model between the monitored variables. The MF-DXA is used to extract the coupling multi-fractal features to reflect the changes in the coupling relationship. The Mahalanobis-Taguchi system is used to select the effective feature variables and Mahalanobis distance is used to detect the collisions of coupling relationship. The empirical research results indicate that the proposed method is feasible for information quality assessment.

Keywords: information quality, collision detection of coupling relationship, DCCA, MF-DXA, Mahalanobis-Taguchi system, process production system

INTRODUCTION

The production system of the process industry can be viewed as a distributed and complex manufacturing system consisting of large power equipment, electrical instruments, chemical reactors and control system. The process units are connected as a highly-coupled production system through power, fluid, pressure and control signals. The production system has the characteristics of tight coupling, highly relevant and complex. The distributed control system (DCS) is widely used in the production system of the process industry for monitoring the production status. The monitoring information generated by the DCS is the major basis for the control and scheduling of production. However, the data collected by the DCS have inaccuracy problem due to the sensor failure, vibration, temperature changes and environmental pollution. Inaccuracy data can bring safety production serious risks. The preliminary statistic show that, in the process industry of china, the number of accidents caused by the information quality annually is no less than 100 and the direct economic losses is no less than RMB 100 million. In order to eliminate and avoid such security exposure, at present, most of the process industry mainly adopt itinerant detective, additional instrument for critical monitoring point and routine calibration for instrument. However, the effect are still not perfect due to the lack of effective assessment method for information quality. Therefore, information quality in the process industry system is an ever-present problem that has negative effect on production safety. Thus, research on effective evaluation and monitoring methods for information quality in the process industry will be significance for the solution of information quality control in distributed and complex manufacturing systems.

As a research hotspot, information quality is widely studied in various fields. Fields like management information (Wang et al. 1995), radar system, accounting information, information of web (Neely et al. 2014), and medical information (Amjad et al. 2014) have been the most popular information quality research areas, where the total data quality management developed by MIT (Huang et al. 1999) has the most impact. Batini (Batini et al. 2009) presented a review about the dimension, evaluation and the methods of IQ research from the standpoint of information system. Most of the research on information quality in different fields has focused on definition, dimension and evaluation indicators, which has not only promoted the development of the information quality research but has also provided the research foundation for information quality control and evaluation. Lee (Lee et al. 2002) proposed the AIMQ in questionnaire survey that includes the IQ model and the analysis technique for measuring the information quality. The method of data life cycle takes the data collection, data organization, data representation and data application as a loop and has different measurement index for each process (Batini et al. 2009). The HIQM can identify the key point of IQ in the business process and detect the failure caused by IQ using the embedded blocks of IQ and conformance checking (Arazy and Kopak 2011; Cappiello et al. 2006). The IP-MAP can analyze the production process of IQ and the control element, and can improve the process intuitively by associating with dimension indices of IQ. Other researchers have applied the traditional control method of product quality to information quality control (Reiter et al. 2011; Han Xiaohong et al. 2009). The control matrices and the grey fuzzy theory have also been used for information quality evaluation and control (Zheng Hua et al. 2012;Del Pilar and MacKinnon. 2011; Lirong Song 2009; Mattias and Carl. 2009). At the same time, inaccurate information identification and false information detection have been studied by various scholars at home and abroad. Gu (Gu et al. 2015) proposed a detection method to detect the false data injection attacks by tracking the dynamics of measurement variations, and used the Kullback-Leibler distance (KLD) to calculate the distance between two probability distributions derived from the measurement variations. When false data are injected into the power systems, the probability distributions of the measurement variations will deviate from the historical data resulting to a larger KLD. Niu (Niu and Lu 2015) investigated the strategies for the Bayesian estimator to detect the false information based on the least square criterion. Matthew (Brenneman and Morton 2009) used the Tracy-Widom distribution to determine the false alarm rate of information theoretic methods. KH (Law and Kwok 2005) modelled the normal alarm patterns of intrusion detection systems and detected anomaly from incoming alarm streams using the k-nearest-neighbor (KNN) classifier for intrusion detection in computer security to reduce the number of false alarms. However, the confirmation of parameter k in KNN classifier is very difficult. Cho (Cho and Qu 2013) proposed a false alarm detection and recovery technique that adopts a trust mechanism for each node of wireless sensor networks to measure the neighbors' trustworthiness so that the node sends data packets only to the trustworthy neighbors. A false alarm occurs when a good node is considered as untrustworthy, but how to determine the node's trustworthiness has not been mentioned. The above-mentioned research on information quality has provided qualitative and quantitative methods for information quality control, evaluation and false information detection. However, most of the research has mainly focused on the management information quality and due to the characteristic of highly coupled and complex restriction between each factors of productions, the major achievements so far cannot be directly applied to the manufacturing information evaluation. Therefore, a new method should be proposed for the accuracy appraisement of manufacturing information.

In this paper, aiming at realizing the accuracy assessment of IQ in the process production system, a new method based on coupling relationship collision is proposed. Since the process production system monitoring information is nonlinear, non-stationary and long range correlated, the detrended cross-correlation analysis (DCCA) is used to construct the coupling network model, the multi-fractal detrended cross-correlation analysis (MF-DXA) is used to analyze the coupling between the non-stationary information series, and Mahalanobis-Taguchi system (MTS) is proposed to realize the abnormal information identification based on coupling relationship collision.

The outline of this paper is as follows: Section 2 introduces the judgement of coupling relationship and the construction of coupling network; Section 3 describes the coupling feature extraction method based on MF-DXA; Section 4 presents the evaluation method of information quality accuracy based on the information conflict detection. A case study is presented in Section 5 followed by the conclusion in Section 6.

COUPLING NETWORK MODELING

Podobnik (Podobnik et al. 2009; Podobnik and Stanley 2008) presented detrended cross-correlation analysis (DCCA) for analyzing the long-range cross-correlation between two non-stationary time series that is now widely used for coupling relationship. The DCCA procedure is as follows:

Step 1. Consider two time series $\{y_i\}$ and $\{y_i'\}$ with equal length N and calculate new time series for each original time series as the following equation.

$$\begin{cases} R_k = \sum_{i=1}^{k} y_i \\ R_k' = \sum_{i=1}^{k} y_i' \end{cases}, k = 1, 2, 3, ..., N \tag{1}$$

Step 2. Separate the new time series into N-n non-overlapping boxes with each box containing s=n+1 data points, where subscript value of each box is from i to i+n.

Step 3. Define the local trends $\{\tilde{R}_k\}$ and $\{\tilde{R}_k'\}$ for each box, which can be obtained by a linear least squares fits. Then determine the covariance of each box as follow:

$$f_{dcca}^2(n,i) = (N-n)^{-1} \sum_{k=i}^{i+n} (R_k - \tilde{R}_k)(R_k' - \tilde{R}_k') \tag{2}$$

Step 4. Calculate the covariance of all boxes as follow:

$$F_{dcca}^2(n) = (N-n)^{-1} \sum_{i=1}^{N-n} f_{dcca}^2(n,i) \tag{3}$$

With the change of n, $F_{dcca}^2(n)$ is also changed. Figure 1 shows relation curve of $\log_2 F_{dcca}^2(n)$ and $\log_2 n$, where the slope is the DCCA index.

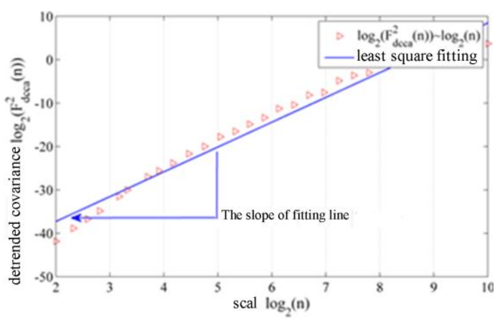

Figure 1 Detrended covariance function changing curve with scale

Goodness of coupling is used to measure the existence theorem of coupling, and DCCA index is used to express the coupling feature. The calculation is as follow:

Step 1. Define the DCCA index as the line slope to express the coupling feature of two time series, which represents the coupling strengths of two series.

Step 2. Define the goodness of coupling as the predicate condition to prove existence theorem of coupling. Calculate the goodness of coupling as the following equation:

$$r^2 = 1 - \frac{\sum(\log_2 F_i - \log_2 F_{fit})^2}{\sum(\log_2 F_i - \log_2 \bar{F})^2 + \sum(\log_2 F_i - \log_2 F_{fit})^2} \tag{4}$$

where $log_2 F_{fit}$ is the fitted value, expressly when the goodness of coupling r is greater than or equal to 0.9, the two time series are long-range cross-correlated, otherwise there is an unsteady coupling or no coupling between the time series.

EXTRACTION OF MULTIFRACTAL FEATURE

Wei-xin Zhou made further efforts to augment the DCCA with multi-scale analysis, and presented the multi-fractal detrended cross-correlation analysis (MF-DXA) (Zhou 2008). In this paper, the MF-DXA is used to achieve coupling feature extraction. The MF-DXA procedure is as follows:

Step 1. Consider the two time series $\{x_k\}$ and $\{y_k\}$ with equal lengths (N). Convert each time series into a new time series by using the following equation.

$$\begin{cases} X(i) \equiv \sum_{k=1}^{i} [x_k - \bar{x}] \\ Y(i) \equiv \sum_{k=1}^{i} [y_k - \bar{y}] \end{cases} \tag{5}$$

where $\bar{x}$ and $\bar{y}$ are the mean values of original time series.

Step 2. Separate the new time series into $N_s \equiv [N/S]$ non-overlapping boxes with each box containing S data points. Since the record length N does not need be a multiple of the considered time scale S, a shorter part at the end of the new time series will remain in most cases. In order to disregard this part of the record, the same procedure is repeated from the end to the start of the record. Thus, $2N_s$ boxes are obtained altogether.

Step 3. Calculate the local trend of each box v by least-squares fitting method, and the detrended covariance of each box is obtained as following equation:

$$f_v(s) \equiv \frac{1}{s} \sum_{k=1}^{s} \left\{ X[(v-1)s+k] - \tilde{x}_v(k) \right\} \times \left\{ Y[(v-1)s+k] - \tilde{y}_v(k) \right\} \tag{6}$$

where $v = 1, 2, \cdots, 2N_s$, $\tilde{x}_v(k)$ and $\tilde{y}_v(k)$ are the fitted values. The fitting polynomial could be squares, cubic or higher.

Step 4. Average all the detrended covariances and calculate the qth order detrended covariance. Here q is the order of the moment that can take any real value.

If $q \neq 0$ then

$$F_{xy}(q,s) \equiv \left\{ \frac{1}{2N_s \sum_{v=1}^{2N_s} [f_v(s)]^{\frac{q}{2}}} \right\}^{1/q} \tag{7}$$

If $q \equiv 0$ then

$$F_{xy}(0,s) \equiv \exp[\frac{1}{2N_s} \sum_{v=1}^{2N_s} \ln f_v(s)] \tag{8}$$

Step 5. Repeat Steps 2, 3, and 4.for different q and s. One can observe that $F_{xy}(q,s)$ increases as s increases.

Step 6. In the double logarithm coordinate log-log, calculate the detrended fluctuation function $F_{xy}(q,s)$ for different q and s. If the two original time series are cross-correlated, then

$$F_{xy}(q,s) \sim s^{h_{xy}(q)} \tag{9}$$

where $h_{xy}(q)$ is the scaling exponent whose value depends on q. It is found that the coupling relationship shows a multi-fractal behavior. Furthermore, for a positive q value, $h_{xy}(q)$ describes the scaling behavior of the segments with large fluctuations. On the contrary, for a negative q value, $h_{xy}(q)$ describes the scaling behavior of the segments with small fluctuations.

For the purpose of extracting the key coupling features, the usage scope of singularity spectrum is enlarged through Legendre transform.

$$\begin{cases} \alpha = h_{xy}(q) + q h'_{xy}(q) \\ f(\alpha) = q[\alpha - h_{xy}(q)] + 1 \end{cases} \tag{10}$$

Where α is the Hölder exponent and $f(\alpha)$ denotes the dimension characterized by α using (10). One can directly relate α α and $f(\alpha)$ to $h_{xy}(q)$. Finally, the feature set $(\Delta f, \Delta\alpha, a_{max}, a_{min})$ can be used to describe the singularity spectrum.

INFORMATION QUALITY ASSESSMENT

Mahalanobis Taguchi system (MTS) considers the relevance between the feature variables. Mahalanobis space (MS) is established through mass normal feature samples, and selection of key features is done with the help of orthogonal arrays (OA) and signal-to-noise ratio (SNR) even when only a handful of abnormal data is presented. The MTS is designed to deal with non-typical and non-equilibrium dataset, which is especially applicable to analyze the monitoring information in the process industry. Owing to space reasons, the authors mainly focus on Mahalanobis distance and its modification.

(1) Standard Mahalanobis distance and generalize Mahalanobis distance

Assume Y is the sample with size m×n, where m and n are the number of features and the number of the samples, respectively. Then the standard Mahalanobis distance between x and y is as follow:

$$MD = (x-z)^T C^{-1} (x-z) \tag{11}$$

where z is the mean vector of feature with size n×1, z=Y*u/m. C is the covariance of Y.

$$C = \frac{1}{m-1}(Y - uz^T)(Y - uz^T)^T \tag{12}$$

In order to apply the MTS, the sample data is the feature set extracted by the MF-DXA. Mahalanobis distance is defined as the distance between jth sample and the sample population and is given as follow:

$$MD_j = \frac{1}{k} Z'_{ij} C^{-1} Z_{ij} \tag{13}$$

If multicollinearity exists, the covariance matrix C becomes singular matrix, i.e., C-1 is impossible to be calculate. For both Gram-Schmidt method and adjoin matrix method of MTS, the multicollinearity is solved by improving the Mahalanobis distance function. Here, the M-P generalized inverse matrix is used instead of the traditional inverse matrix. C+ is the M-P generalized inverse matrix of covariance matrix and the generalize Mahalanobis distance is as follow:

$$MD_j = \frac{1}{k} Z'_{ij} C^{+} Z_{ij} \tag{14}$$

where i=1,2,3,…,k corresponds to feature variables; j=1,2,3,…n corresponds to samples; n is the number of data samples; k is the number of feature variables; $Z_{ij} = (Z_{1j}, Z_{2j}, Z_{3j}, ..., Z_{kj})$ is the standardized feature variable; $Z_{ij} = {(x_{ij} - m_j)}/{S_i}$ is the element of feature variables; x_{ij} is the ith feature value of jth sample; m_i is the mean value of ith feature variable; S_i is the standard deviation of ith feature variable; and C^{-1} is the inverse matrix of covariance matrix.

(2) Orthogonal experiment design

The orthogonal experiment design is used to assess the importance of feature variables based on data analysis. The orthogonal table $L_n(t^c)$ is well-designed to reduce the experiment times where L denotes the orthogonal table, n is the experiment times, t is the number of level and c is the number of features. In this study, the feature variables are the factors in two levels' orthogonal table. The feature variable is used when the level is 1 and discarded when the level is 2.

(3) Signal-to-noise ratio

The SNR is calculated for each row in the orthogonal table using the following equation. Wwhere both normal generalized Mahalanobis and abnormal generalized Mahalanobis distances have been obtained using (14). A larger value of SNR is better.

$$SNR = -10 * \log[\frac{1}{t} \sum_{i=1}^{t} (\frac{1}{D_i^2})] \tag{15}$$

In order to select the feasible features, the gain of each feature variable is calculated using (16). According to the rule that the SNR needs to be larger, the features with larger gain values are retained.

$$Gain = (Avg.SNR)_{Level1} - (Avg.SNR)_{Level2} \tag{16}$$

CASE STUDY

The Construction of Coupling Network Model

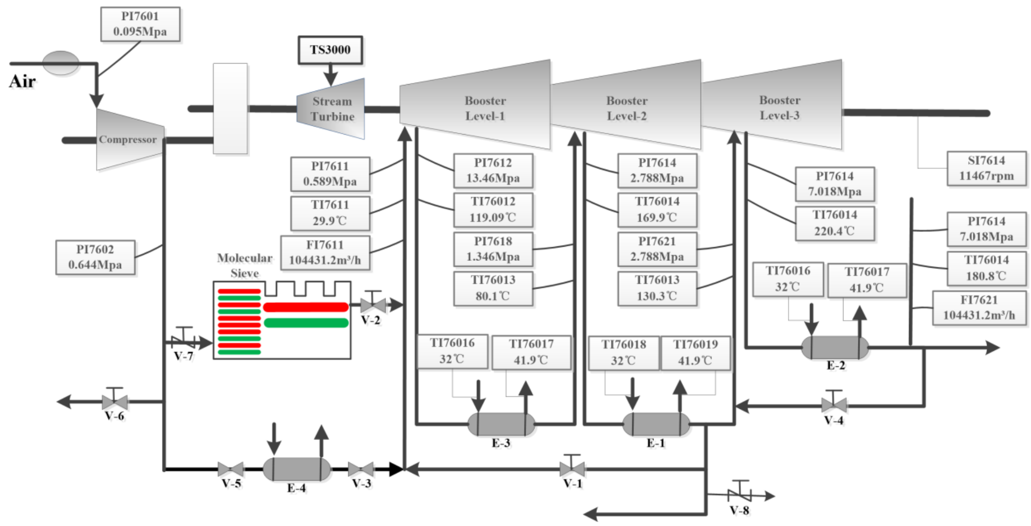

Figure 2 Schematic of the compressor system

With the compressor system in chemical enterprise, for example, 16 monitored variables in the compressor system are selected to construct the coupling relationship model. Figure 2 shows the components of the compressor system. Table 1 describes the meanings of each monitored variable. The data of each monitored variable collected at a continuous sampling rate of 1/10 Hz for 12 days is used with a total number of samples of 17280.

Table 1 Part of monitoring variable of the compressor unit

Monitoring sites	Description of monitoring variable	units	Monitoring sites	Description of monitoring variable	units
A_API7601	Inlet pressure of air compressor	kPa	A_API7621	Inlet pressure of booster level-3	MPa
A_API7602	Exhaust pressure air compressor	MPa	A_API7622	Exhaust pressure of booster level-3	MPa
A_AFI7601	Inlet flux of air compressor	Nm3/h	A_APDI7601	Pressure difference at throat of air compressor	kPa
A_API7611	Inlet pressure of booster level-1	MPa	A_RZI7634	Axial displacement of booster	mm
A_ATI7611	Inlet temperature of booster level-1	℃	A_ATI7642	Thrust pad temperature of booster	℃
A_AFI7611	Inlet flux of booster	Nm3/h	A_ATI7615	Exhaust temperature of booster level-2	℃
A_API7612	Exhaust pressure of booster level-1	MPa	A_ATI7622	Exhaust temperature of booster level-3	℃
A_API7614	Exhaust pressure of booster level-2	MPa	A_RZI7630	Axial displacement of air compressor	mm

The process to construct the coupling model is as follow:

Step 1. Use the DCCA proposed in Section2 to analyze the coupling between any two time series in table 1. Calculate the DCCA index and the goodness of coupling by equations (3) and (4). If the goodness of coupling is greater than 0.9, the coupling between two time series does exist. Establish the connection between the two time series by the DCCA index.

Step 2. Repeat Step 1 for all monitoring variables in table 1 and obtain the goodness of coupling and DCCA indices of all monitoring variables.

Step 3.Eestablish the connection between any two series by the result of step 2, which forms the simplified network model as shown in Figure 3.

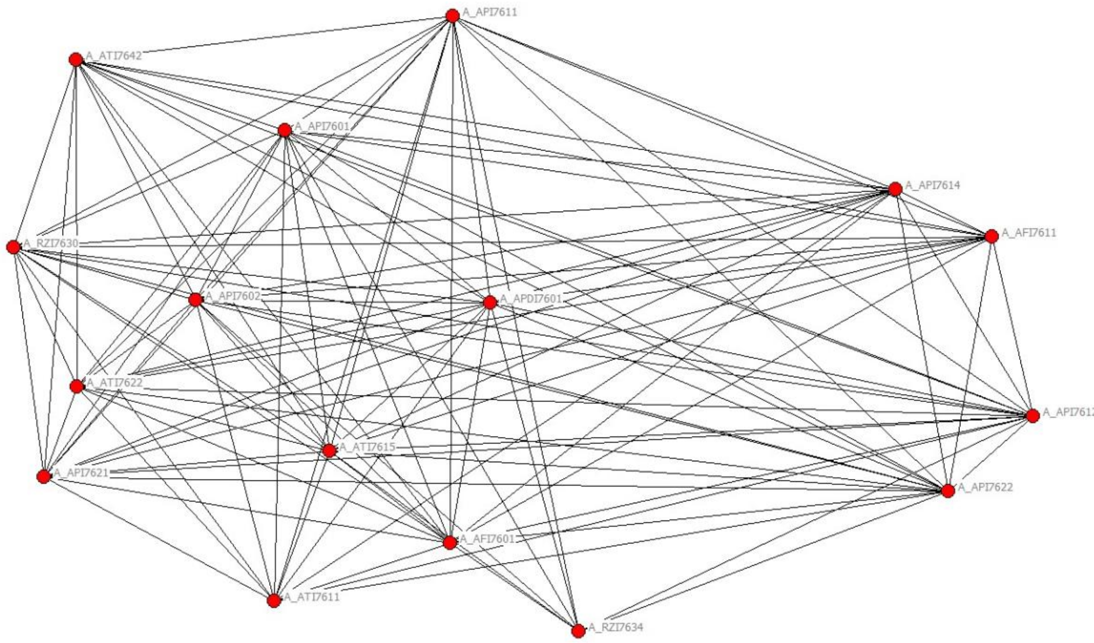

Figure 3 The network model of monitoring variables in compressor system

Coupling Multi-fractal Feature Extraction

Four monitoring variables (API7622, API7612, API7614, AFI7611) of coupling network model as shown in Figure 3 are selected. What they indicate is specified in Table 1. For each monitoring variable, the data sample interval is 1 minute and the sample length is 1440. Calculate the coupling feature set $(\Delta f, \Delta\alpha, a_{max}, a_{min})$ between any two monitoring variables through the method proposed in Section 3. Figure 4 shows the curve of coupling feature between API7622 and other three monitoring variables.

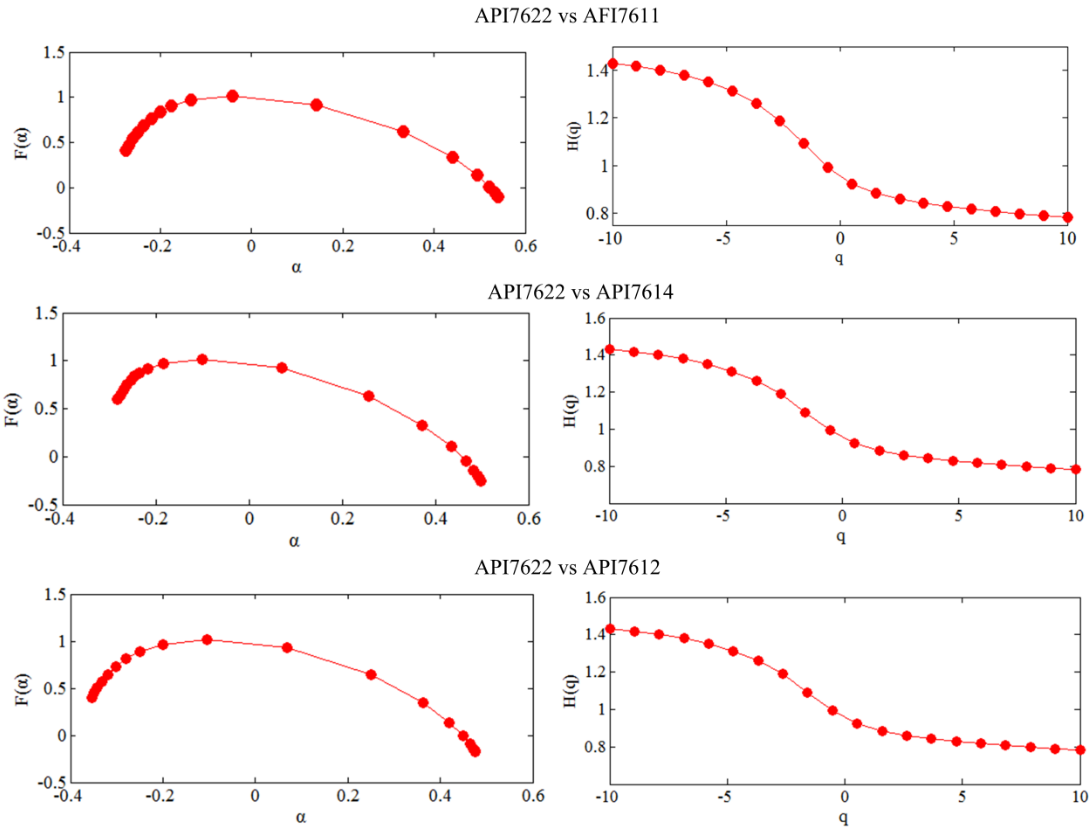

Figure 4 The curve of coupling feature of monitoring variable API7622

IQ Accuracy Assessment

The feature vector selected by Section 4 is the key index for the assessment of IQ. The Mahalanobis distance between the selected feature vectors can be obtained by equation (14). The accuracy of IQ can be evaluated using the Mahalanobis distance. Figure 5 shows the implementation steps.

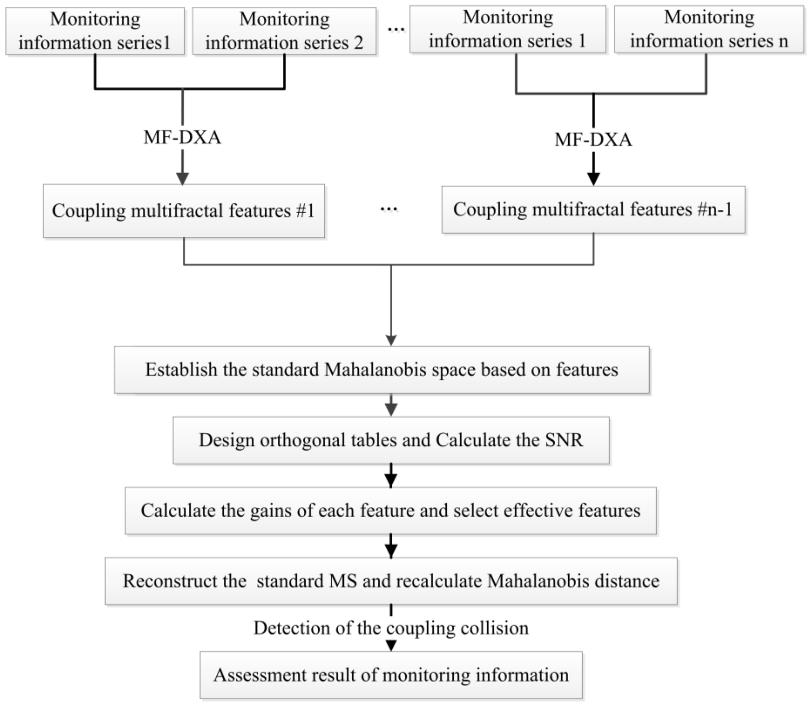

Figure 5 Procedure of assessment of IQ

The compressor system in chemical enterprise in Figure 2 of Section 5.1 is taken as the example to explain the produce of IQ evaluation. The abnormal monitoring information of A_API7622 appeared on December 21, 2013 and resulted to the accident. Several monitoring variables have larger fluctuation at the same time. Therefore, the related monitoring variables (API7622, API7612, API7614, AFI7611) are selected to validate the methods of IQ evaluation proposed in this paper.

For the evaluation, two types of monitored dataset are prepared: one is 20 days' normal dataset and the other is abnormal information, and data collected at a continuous sampling rate of 1/10 Hz. The joint conditions between the monitoring variables are shown in Figure 3. The procedure of IQ evaluation is as follows.

Step 1. Divide the normal and abnormal information into several segments each with an equal length of 2000. Then calculate the feature set $(\Delta f, \Delta\alpha, a_{max}, a_{min})$ for each segment by using the methods proposed in Section 3. Each monitoring site has $3\times4=12$ coupling features. The feature set $(\Delta f, \Delta\alpha, a_{max}, a_{min})$ of normal information is shown in Figure 4.

Step 2. Establish the normal standard Mahalanobis space using equations (3) – (4) and select the effective feature set through the produce described in Section 4.

Step 3. Calculate the Mahalanobis distance between the normal standard Mahalanobis space and the feature set selected by Step 2.

For the uniform description of coupling variance of monitored variables and detection of a collision, the coupling relationship change is supervised by sliding window with slip step of 50 and the generalized MD in every sliding window is calculated after feature selection. The analysis results are shown in Figure 6. Among the three graphs, Figure 6(a) shows the coupling relationship variance of A_API7622 before or after feature selection and the monitoring chart clearly manifest the advantages of applying the generalized MTS. Figure 6 (b), (c) and (d) are the monitoring charts of A_AFI7611, A_API7612 and A_API7614, respectively. The generalized MD of A_API7622 changes significantly when the sliding window number is larger than 25 indicating the change of coupling relationship, while the generalized MD of the rest three monitoring sites remains stable. The coupling relationship collision among the four monitoring sites occurs, which indicates that the monitoring information of A_API7622 is false.

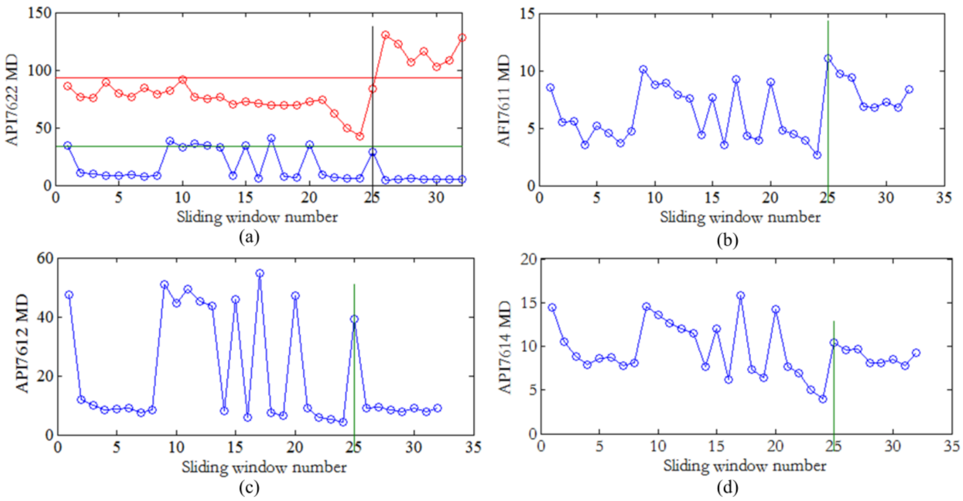

Figure 6 MD monitoring curves

CONCLUSION

In this paper, the IQ assess met methodology is developed for identifying and assessing the information accuracy in the process production system. This encompasses three major components: constructing the coupling network model, feature extraction of coupling multi-fractal and information accuracy assessment. The coupling network model describes the coupling relationship and the coupling strength between the monitoring variables. It lays a foundation for coupling feature extraction, information conflict detection and information assessment. The feature extraction of coupling multi-fractal can obtain effective features that are changed with changing of information accuracy. It is a prerequisite for information conflict detection. The information conflict detection and the information assessment provide the methods to detect the abnormal information in time. The key contribution of overall research stems from the mind of the inherent coupling and abnormal information can be detected by the change of coupling relationship. The assessment methodology as a whole provides a practical information quality assessment tool for process production system.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant by No. 51375375). The authors gratefully acknowledge the valuable cooperation of Weihe Coal Chemical Group in accomplishing this research project.

REFERENCES

Amjad, M. M. M., Rasta, K., Gerdes, M., and Fensli, R. 2014. "An Approach Towards Information Quality Management of Electronic Health Records," *7th International Conference on Health Informatics, HEALTHINF 2014 - Part of 7th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2014, March 3, 2014 - March 6, 2014*, Angers, Loire Valley, France: SciTePress, pp. 270-275.

Arazy, O., and Kopak, R. 2011. "On the Measurability of Information Quality," *Journal of the American Society for Information Science and Technology* (62:1), pp. 89-99.

Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. 2009. "Methodologies for Data Quality Assessment and Improvement," *Acm Computing Surveys* (41:3).

Brenneman, M. T., and Morton, Y. T. 2009. "False Alarm Rate Estimation for Information-Theoretic-Based Source Enumeration Methods," *EURASIP J. Adv. Signal Process* (2009), pp. 3-3.

Cappiello, C., Ficiaro, P., and Pernici, B. 2006. "Hiqm: A Methodology for Information Quality Monitoring, Measurement, and Improvement," *25th International Conference on Conceptual Modeling, ER 2006, November 6, 2006 - November 9, 2006*, Tucson, AZ, United states: Springer Verlag, pp. 339-351.

Cho, Y., and Qu, G. 2013. "Fader: False Alarm Detection and Recovery for Trust-Aware Routing in Wireless Sensor Networks," *2013 2nd IEEE International Conference on Connected Vehicles and Expo, ICCVE 2013, December 2, 2013 - December 6, 2013*, Las Vegas, NV, United states: IEEE Computer Society, pp. 647-648.

Del Pilar Angeles Maria, MacKinnon Lachlan M. Managing data quality of integrated data with known provenance[J]. *International Journal of Information Quality*, 2011, 2(3): 244-263.

Gu, C., Jirutitijaroen, P., and Motani, M. 2015. "Detecting False Data Injection Attacks in Ac State Estimation," *Ieee Transactions on Smart Grid* (6:5), pp. 2476-2483.

Huang, K.-T., Lee, Y. W., and Wang, R. Y. 1999. *Quality Information and Knowledge*. Prentice Hall PTR.

Han Xiaohong, Qin Xiansheng, Wang Keqin. The Process-oriented QFD Control Mode of Information Quality [J]. *Machine Design and Manufacturing Engineering*, 2009, 38(3): 14-18.

Law, K. H., and Kwok, L. F. 2005. "Ids False Alarm Filtering Using Knn Classifier," *5th International Workshop on Information Security Applications, WISA 2004, August 23, 2004 - August 25, 2004*, Jeju Island, Korea, Republic of: Springer Verlag, pp. 114-121.

Lee, Y. W., Strong, D. M., Kahn, B. K., and Wang, R. Y. 2002. "Aimq: A Methodology for Information Quality Assessment," *Information & Management* (40:2), pp. 133-146.

Lirong Song. Theoretic thought on the constraints of information quality under network information-sharing environment[C]. *2nd International Symposium on Information Science and Engineering, ISISE 2009*, p 290-293

Mattias Gustavsson, Carl Wänström. Assessing information quality in manufacturing planning and control processes[J]. International Journal of Quality & Reliability Management, 2009, 26(4):325 - 340.

Neely, M. P., Matheus, A., and Neely, W. 2014. "Data and Information Quality of the Web: Do Perceptions Predict Performance?," *International Journal of Information Quality* (3:4), pp. 340-360.

Niu, R., and Lu, J. 2015. "False Information Detection with Minimum Mean Squared Errors for Bayesian Estimation," *2015 49th Annual Conference on Information Sciences and Systems, CISS 2015, March 18, 2015 - March 20, 2015*, Baltimore, MD, United states: Institute of Electrical and Electronics Engineers Inc.

Podobnik, B., Grosse, I., Horvatić, D., Ilic, S., Ivanov, P. C., and Stanley, H. E. 2009. "Quantifying Cross-Correlations Using Local and Global Detrending Approaches," *The European Physical Journal B* (71:2), p. 243.

Podobnik, B., and Stanley, H. E. 2008. "Detrended Cross-Correlation Analysis: A New Method for Analyzing Two Nonstationary Time Series," *Physical Review Letters* (100:8).

Reiter, Michael, Breitenbücher Uwe1, Dustdar Schahram, et al. A Novel Framework for Monitoring and Analyzing Quality of Data in Simulation Workflows[C]. *Proceedings of 7th IEEE International Conference on eScience. eScience 2011*, Stockholm,Sweden.2011:105-112.

Wang, R. Y., Storey, V. C., and Firth, C. P. 1995. "A Framework for Analysis of Data Quality Research," *Ieee Transactions on Knowledge and Data Engineering* (7:4), pp. 623-640.

Zhou, W.-X. 2008. "Multifractal Detrended Cross-Correlation Analysis for Two Nonstationary Signals," *Physical Review E* (77:6).

Zheng Hua, Zhu Qinghua, Wu Kewen. Provenance management for data quality assessment [J]. *Journal of Software*, 2012, 7(8): 1905-1910.