

CONSTRUCTION OF AN ELECTROMECHANICAL AUTOMATION AND FAULT DIAGNOSIS SYSTEM USING DATA-DRIVEN TECHNOLOGY IN THE CONTEXT OF BIG DATA

Lulu Yuan

College of Mechanical and Electrical Automation, Henan Polytechnic Vocational and Technical College, Nanyang, 473000, Henan, China

Email: yuanlulu_ny@163.com

Abstract: The sensor faults in electromechanical equipment are diagnosed and predicted to accelerate the industrial automation and stable operation of electromechanical equipment. Firstly, several basic fault prediction methods are introduced, and the data of electromechanical equipment during production can be well diagnosed and processed through data-driven technology. Secondly, the principal component analysis (PCA) method is introduced, and accordingly, the optimized dynamic principal component analysis (DPCA) method and several common sensor faults are proposed. Thirdly, the specific steps of the PCA and DPCA methods are discussed in sensor faults diagnosis, respectively. The PCA method can study the multivariate nonlinear data concurrently. On this basis, the optimized DPCA method can analyse the dynamic features of the real-time data collected by sensors. Finally, the fault type that produces a fixed deviation is studied. Then, the experiment of the draw-wire displacement sensor is simulated through the optimized PCA and DPCA. The square prediction error (SPE) and prediction-error rate are compared with the error. The results show that both methods can predict the fault of the sensor. The results of the T^2 statistics obtained through the PCA show that the sensor fault can be effectively predicted, but the prediction-error rate and fault-omission rate are also high. In the DPCA algorithm, the SPE statistics show that the DPCA algorithm can predict the fault with a small prediction-error rate and fault-omission rate, which is far lower than the PCA algorithm. Therefore, dynamic factors analysis can improve sensor fault prediction, which is of great significance to the acceleration of industrial automation and equipment intelligence.

Keywords: Data-Driven; Electromechanical Equipment; Sensor; Automation; Principal Component Analysis.

1. Introduction

In recent years, with the continuous improvement of China's productivity, the national economy, science, and technology have been fully developed. The Internet and computers have been popularized throughout the country. The production means and technological levels of industries have upgraded, and the traditional mechanical equipment is being transformed into digitalization, automation, and intelligence system, thus entering a new era of mechatronics. At the same time, due to industrial expansion, the production technology, production process, and production equipment are becoming more and more complicated and precise. Vast data are being generated through industrial production, including process parameters and equipment parameters [1,2]. Most of these data contain information, such as production process and equipment operation. Based on these data, a fully

automated fault diagnosis system must be constructed for electromechanical equipment to optimize the production process and realize industrial automation [3,4].

Under the era of big data, data collection and utilization should be strengthened, and these data should be mined for equipment information to realize goals in the China Manufacturing 2025 plan, that is, improving the automation, intelligence, and stability of electromechanical equipment. Consequently, a fault prediction and diagnosis system can be established for electromechanical equipment [5]. Sensor is one of the core components of electromechanical equipment [6] and can transform different forms of data through data collection and signal detection. If a sensor fails during operation, data will not be able to be transferred timely and accurately, this may affect the mechanical and electrical equipment or even brings risks to the production process [7]. Data-driven

technology can build a fault prediction and diagnosis system for sensors, improving the automation degree and operation stability of the electromechanical system. Meanwhile, the sensors' working environments are very complex, and the fault signals are not single linear data but dynamic nonlinear data [8–10], so a single variable method cannot predict and diagnose these faults.

To accurately predict these failures and ensure the safety and stability of electromechanical equipment, many Chinese and foreign scholars have conducted a multivariate statistical analysis based on data-driven technology. A thorough understanding and appropriate utilization of data-driven technology are inevitable for the

development of the new era, which has important theoretical and practical significance.

2. Sensor-fault Prediction System for Electromechanical Equipment

2.1 Fault Prediction Technology

Generally, the fault of the sensor can be predicted through estimation of the operation state of the system in the future time through a measurable feature signal. Probability-based approach, model-based approach, and data-driven approach are commonly used [11]. The advantages and disadvantages of the three methods are shown in figure 1.

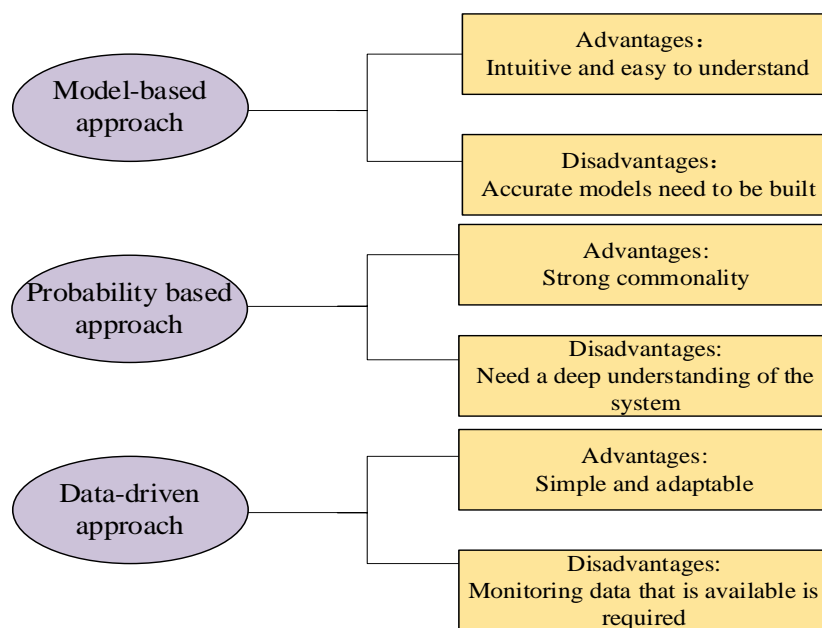


Figure 1: Three fault prediction approaches

The probability-based approach includes Bayesian probability theory, Weibull model, and probability density function of residual life. The probability density function of residual life is calculated through statistical data, and the probability density will change with time. Model-based prediction approach includes autoregressive moving average technology, physical model method, and Kalman particle filter technology. The mathematical relationship between the operation status of each component and the degree of damage can be fitted through a specific physical model of the equipment system, which can predict the service life and the cause of the failure [12].

The above two approaches may be feasible theoretically, but the fault data of the sensor in the real production process are not single linear relationships, and a precise mathematical model cannot be built to predict faults. Therefore, the electromechanical system can only be evaluated through historical data, which is the most essential

idea of the data-driven method. The equipment operation information and production process parameters can be extracted based on historical data and the data-driven model. Finally, the data-driven approach can analyse and process real-time data and diagnose sensor equipment, predicting the service life and fault. The data-driven approach usually contains signal processing, multivariate statistical methods, and neural network construction. Among them, multivariate statistical methods are more suitable for handling complex and unpredictable signals in an actual working environment than other methods [13].

2.2 Data-driven Fault Diagnosis

The data-driven approach overcomes the shortcomings of the signal processing method and has some intelligence. Its main idea is to grasp knowledge and data, and it is based on the data processing technology, especially, the experience

diagnosis knowledge which can be fully used. Meanwhile, dialectical thinking and mathematical logic operation are combined, and signal processing and data processing are unified, realizing both the reasoning process and algorithm process. Finally, the faults can be diagnosed through the data-mining operation of system rules. Fault prediction and diagnosis based on data-driven technology is one of the hottest research directions in the field of electromechanical automation, especially, in the field of complex nonlinear models, with many influencing factors, such as satellites. With the continuous development of the Internet and big data processing technology, data is growing at an unprecedented speed. Data mining methods can make full use of these data to generate useful rules or fault diagnosis models for knowledge. This can solve the bottleneck of knowledge acquisition in traditional modelling and mining methods and improve the effectiveness of fault diagnosis. Fault diagnosis based on data mining has the following advantages: 1. Very rich data sources can be easily obtained, such as system parameter data, historical data, and experience data, from which knowledge can be mined. 2. Data mining and fault diagnosis can complement each other. Fault diagnosis knowledge can select data mining methods and analyse the data. Conversely, the rules obtained through data mining can improve the diagnosis knowledge base and make up for the manual summarized experience knowledge. 3. Data mining can easily solve the uncertainty and imprecision of knowledge description. Engineers cannot accurately solve incomplete and imprecise description problems. The data mining method does not need to understand the complex mechanism of the system, so it can better describe the credibility of knowledge, increasing the channels to obtain knowledge [14].

2.3 Multivariate Statistical Methods

The multivariate statistical methods can reduce the dimension of complex data for simplified analysis. The main methods include principal component analysis (PCA), factor analysis (FA), and canonical variate analysis (CVA). These methods are mature and can solve many problems in practical work, so they have become the most popular data-driven technology. Here, the multivariate statistical methods are chosen as the basic experimental method. Various sensors in the electromechanical equipment system will generate voluminous data every day, and these data can hardly be mathematically modelled. However, based on the regularity between these process data, the multivariate statistical method can reduce their dimension, simplify these data, and analyse the data [15].

2.4 Principal Component Analysis (PCA)

Under the era of big data, data can be easily acquired, utilized, and stored. However, traditional data analysis methods cost a lot of time, and the rules of data are very difficult to find. Consequently, researchers have proposed a new and efficient analysis method-PCA [16].

PCA is the major algorithm of multivariate statistical theory, and it can construct a new vector space based on the original data. The new vectors are obtained through dimension reduction of the original data, containing the main features of the original data. Although the data are compressed, no data information has been lost, which is the key to the PCA. Therefore, PCA can effectively solve the data redundancy, high data coupling, and difficult calculation. Although data accuracy is slightly affected by data compression, the basic information of data is almost unchanged, so the PCA becomes the core of fault diagnosis and prediction technology.

The PCA can obtain new variables through a linear combination of the old variables. These new variables have a reduced dimension, containing most of the information in the old variables. The old and new variables are independent of each other [17]. The specific operation process is shown below.

Firstly, the training observation data X containing m observation variables is collected, and each variable contains n observation values. These data are expressed as an $n \times m$ dimensional data matrix, as shown in equation (10).

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} \quad (1)$$

In equation (1), matrix X can be decomposed into the sum of the outer product of m vectors, as shown in equation (2).

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_m p_m^T \quad (2)$$

In equation (2), $t_j \in R^n$ is the score vector, $p_j \in R^m$ is the load vector, and the score vector of X is also called the principal element (namely, the implicit variable). When t_i is extracted, the maximum value of variance should be taken, and each score vector is orthogonal to each other. That is, when $i \neq j$, the length of each load vector satisfying $t_i^T t_j = 0$ is 1 and orthogonal to each other, so equations (3) and (4) can be obtained.

$$p_i^T \times p_j = 0 \quad (3)$$

$$p_j^T \times p_i \neq 0 \quad (4)$$

If equation (2) is multiplied by p_i on both sides, then equation (5) can be obtained.

$$Xp_i = t_1 p_1^T p_i + t_2 p_2^T p_i + \dots + t_m p_m^T p_i \quad (5)$$

When equations (3) (4) are introduced into equation (5), equation (6) can be obtained.

$$t_i = Xp_i \quad (6)$$

Thus, each score vector is the projection of the matrix X on the load vector corresponding to the score vector. The length of the vector t_i is proportional to the coverage of the data matrix X projected to the p_i direction, and the length of the vector t_i is also proportional to the standard deviation of the data matrix X projected to the p_i direction.

The detailed steps to solve the PCA are shown as the following.

(1) X is normalized to prevent over-coupling, and then the sample covariance matrix S is obtained, as in equation (7).

$$S = \frac{1}{n-1} X^T X \quad (7)$$

(2) Eigenvalue of decomposition matrix S can be expressed as in equation (8).

$$S = V \Lambda V^T \quad (8)$$

Equation (8) shows the correlation of covariance matrix, and Λ denotes the diagonal matrix, which contains nonnegative real eigenvalues $(\lambda_1, \lambda_2 \dots \lambda_n)$. V represents the orthogonal matrix, where $V^T V = I$, and I stands for a unit matrix.

$$\lambda_i = \frac{1}{n-1} t_i^T t_i \quad (9)$$

In equation (9), λ_i denotes the sample variance of the i th principal component. If the variance of the i th principal component is accumulated more than a set threshold, the i th principal component is regarded as a comprehensive unit. Then, the i th principal component is transformed into an i -dimensional vector from the original m -dimensional vector, where $i \leq m$.

(3) Calculation of score matrix T

To obtain an optimal data variation with lesser random noise influence on PCA, the load vectors of d maximum eigenvalues are retained, and the column $P \in R^{m \times d}$ of the load matrix is selected to correspond to the load vector associated with the previous d eigenvalues. Then, the score matrix $T = XP$ will be included in the projection from X to the low dimensional space. For the new sample vector x , the

score vector of PCA is $t = P^T x$, the estimation vector is $\hat{x} = PP^T x$, and the residual vector is $e = (I - PP^T)x$. The T^2 statistics determine the measurement of variation in the PCA model. T^2 represents the sum of normalized squared scores.

$$T^2 = t^T \Lambda^{-1} t = x^T P \Lambda^{-1} P^T x \quad (10)$$

In equation (10), Λ refers to an eigenvalue diagonal matrix related to the primitive element, and the control upper limit of T^2 is obtained through F distribution.

$$T_{lim}^2 = \frac{p(n-1)}{n-p} F(p, n-1, \alpha) \quad (11)$$

In equation (11), $F(p, n-1, \alpha)$ represents the critical point of the upper limit of 100 α % of the F distribution with degrees of freedom of p and $n-p$. Besides, the measurement of PCA residual can be detected through square prediction error (SPE). SPE is defined as in equation (12).

$$SPE = e^T e = x^T (I - PP^T) x \quad (12)$$

When the sensor fails, some sample data will appear outside the normal range of the PCA diagram and SPE, and SPE or T^2 statistics will also be outside the boundary of the confidence interval.

2.5 Dynamic PCA

PCA can analyse multidimensional variables, which are all static variables, while the output of the sensor is dynamic variables. Hence, an improved analysis algorithm-dynamic PCA (DPCA) is proposed, which combines dynamic time series with PCA algorithm. Through time window removal, the dynamic time series group is promoted to the analytical range of PCA, so that the original static data is transformed into real-time dynamic data, and the PCA algorithm can analyse the system operation state accurately [18].

Given any X observation data, the augmented matrix with l delays can be expressed as in equation (13).

$$X(l) = \begin{bmatrix} x_t^T & x_{t-1}^T & \dots & x_{t-l}^T \\ x_{t-1}^T & x_{t-2}^T & \dots & x_{t-l-1}^T \\ \dots & \dots & \dots & \dots \\ x_{t+l-n}^T & x_{t+l-n-1}^T & \dots & x_{t-n}^T \end{bmatrix} \quad (13)$$

In equation (13), x_t^T represents the m -dimensional observation data at time t , and l stands for the length of time delay.

The specific process of fault prediction based on DPCA can be summed up as the following.

(1) The training sample database is formed through fault-free data collection.

(2) According to $r_{new}(l) = r(1) - \sum_{(l-1)}(l-i+1)r_{new}(i)$, time delay l can be calculated, obtaining the time series of dynamic variables.

(3) The augmented matrix $X(l)$ is built with dynamic data.

(4) The augmented matrix $X(l)$ is standardized to obtain \bar{x} .

(5) The covariance matrix is calculated according to equation $S = \frac{1}{n-1} \bar{X}^T \bar{X}$.

(6) The principal component variance vector $[\lambda_1, \lambda_2 \dots \lambda_m]^T$ and its load matrix $[p_1, p_2 \dots, p_m]$ are calculated;

(7) According to equation $CPV = \frac{\sum_{i=1}^A \lambda_i}{\sum_{i=1}^m \lambda_i}$, the principal number is determined.

(8) The DPCA model is constructed to determine the principal component subspace and residual subspace.

(9) SPE_{lim} and T^2 statistics T^2_{lim} are calculated.

(10) The augmented matrix is constructed and standardized through sample data with faults.

(11) SPE and statical value of T^2 are calculated.

(12) The control limits of SPE and T^2 are compared with the calculated statistics. If the statistics are not within the control limits, it indicates that there is a fault. Comparatively, if the statistics are within the control limits, there is no fault [19].

In summary, the DPCA fault diagnosis process is shown in figure 2.

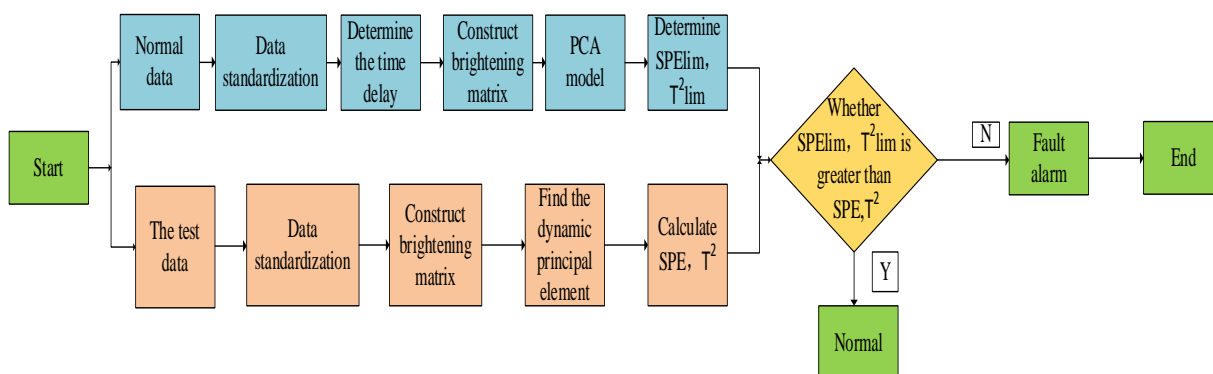


Figure 2: Flowchart of DPCA fault diagnosis

2.6 Fault Prediction and Diagnosis Process for Sensors of Electromechanical Equipment

A transducer/sensor is a signal detection and collection device and can detect signals, collect signals, and transform signals into other forms, such as image data and electronic data according to specific rules. Thus, information can be transmitted, processed, stored, displayed, recorded, and

controlled. Most electromechanical equipment contains three main structures: software system, mechanical system, and sensor. Sensors can connect software system and mechanical system and converts the collected data into output data through data processing, completing data storage, data transmission, and data control.

The mechanism of the electromechanical system is shown in figure 3.

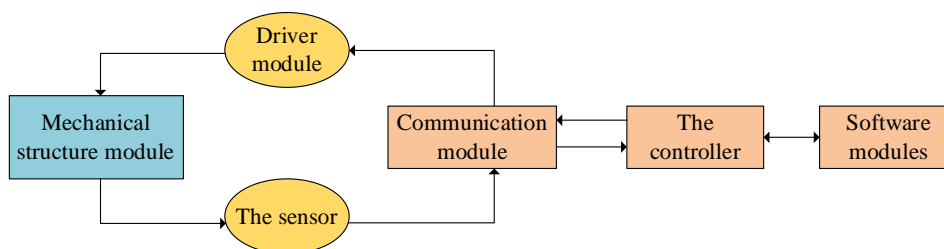


Figure3: Structural of electromechanical equipment

To ensure the normal operation of electromechanical equipment, sensors must collect and process signals timely, accurately, and efficiently. Generally, the fault prediction process of large electromechanical equipment can be divided

into four stages: real-time data acquisition, data processing, working state identification, and prediction and diagnosis.

The process of sensor fault detection is shown in figure 4.

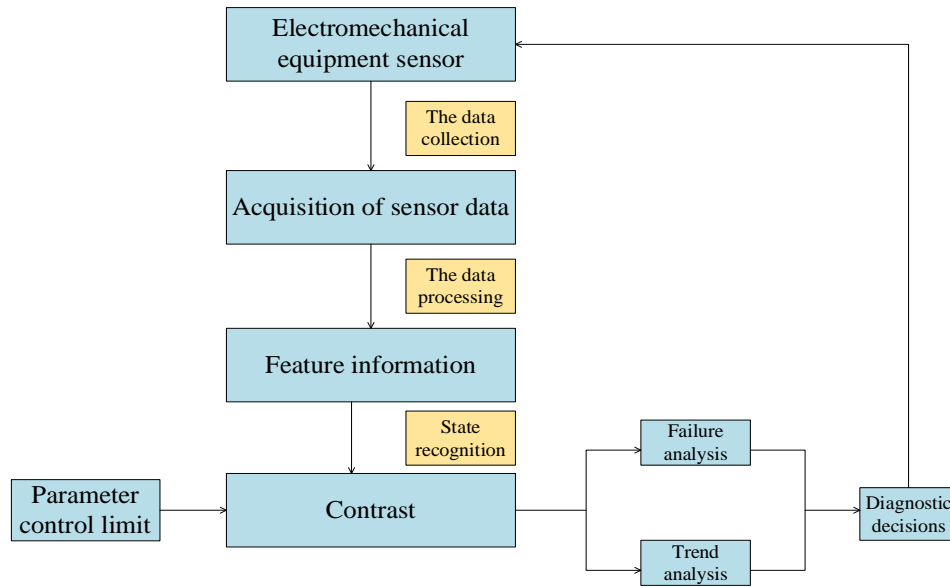


Figure 4: Flowchart of sensor fault prediction

The first step is real-time data acquisition. Here, the sensor data is obtained through the construction of the roll position automatic adjustment system. Real-time data acquisition is the premise of sensor analysis and fault prediction. Only the real data collected under actual working conditions can be used in subsequent analysis and diagnosis [20].

The second step is feature signal extraction (data processing). This process can obtain the features of the sensor running state and then classify and extract the collected signal. The original data with noise cannot be used but denoised to obtain a single real data signal. These denoised single real data can be regarded as the feature signal of the sensor.

The third step is state identification. The statistical recognition method, fuzzy recognition method, and neural network recognition method can compare the processed data with the feature data.

The fourth step is data prediction and diagnosis. The results are analysed and compared, and if the sensor fails, the fault is immediately reported. Then, the fault is treated according to the cause, and if no failure occurs, further prediction can judge the service life of the sensor.

2.7 Fault Diagnosis Experiment of Draw-wire Displacement Sensor based on PCA

Sensor faults mainly include fixed deviation, complete failure, drift deviation, and reduced accuracy. Here, the main cause of the fault is deviation. The difference between the feedback value and the real value of the sensor with a fixed deviation fault is constant. The feedback value of the sensor with complete failure is a constant. The difference between the feedback value of the sensor

with drift deviation and the real value is always changing [21].

In general, the sensor-measured data includes the real value, system deviation, and random deviation, which can be expressed as in equation (14).

$$x_t = x_t^* + f_t + v_x \quad (14)$$

In equation (14), x_t denotes the feedback at time t.

x_t^* represents the real value of t time.

f_t stands for the system deviation at time t.

v_x refers to the random deviation caused by free noise at time t.

Here, the deviation fault is analysed. The deviation fault is caused by the harsh environment of the sensor. As the sensor's service time prolongs, there will be a fixed deviation between the real value and the feedback value, as shown in equation (15).

$$f_t = C \quad (15)$$

Most draw-wire displacement sensors work under electromagnetic interference, low temperature, high humidity, or many pollutants. Therefore, the failure rate is relatively high, with a lower average service life than general sensors. In the case of a failure, great losses will be caused to the electromechanical system and the production process. Therefore, an accurate and effective fault diagnosis and prediction method is of great significance to the draw-wire sensors. Through the observation of the common features of the draw-wire sensor, a fault prediction method based on PCA is designed and optimized. In particular, the dynamic data-driven approach is added. The comparison of the T² statistics and SPE statistics with the initial

threshold can determine the sensor's faulty state [22].

Here, the experiment is simulated on the automatic control system of the rolling mill. Five monitoring sensors in a rolling mill are selected and analysed. The historical data of 200 fault-free states are collected as the offline training database, and the model is constructed based on this database. The PCA model is established through the sample data under normal working conditions. The first 50 sets of sample data of fault data set are normal data, and the fault is introduced from the 51st set of data.

The principal component contribution rate is 85%, and the control limit is 95%.

3. Simulation Experiment on Fault Prediction of Draw-wire Sensor

3.1 Comparison of Data Deviation of the Draw-wire Sensor

Figure 5 illustrates the fixed deviation and drift deviation between the feedback value and the real value of the draw-wire sensor on fault occurrence.

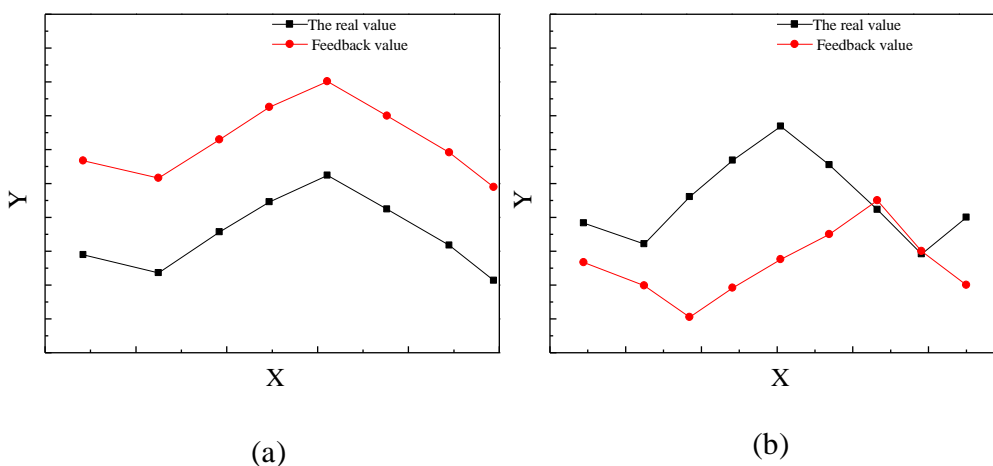


Figure 5: The deviation between real value and feedback value of sensors (a) fixed deviation (b) drift deviation

Figure 5 displays that the difference between the feedback value of the sensor with a fixed deviation and the real value is constant.

The PCA method can simplify these data. To facilitate the data processing, five sensors are numbered, respectively. Number 1 denotes the upper-right sensor of K12, number 2 represents the

lower-left sensor of K12, number 3 denotes the upper-left sensor of K12, number 4 denotes is the lower-right sensor of K12, and number 5 represents the upper\lower sensor of K12. Then, the standard deviation and mean value of the sample vectors of the five sensors are calculated, as shown in figure 6.

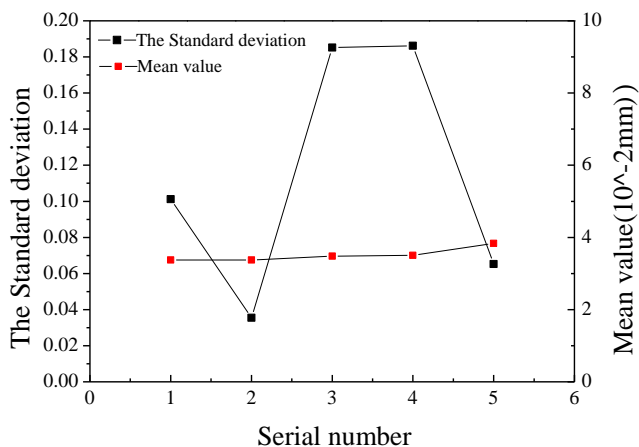


Figure 6: Mean value and standard deviation of sensors

Figure 6 demonstrates that the dispersion degree of the sample data collected by the five sensors is large, and there is a great difference between the sample data of different groups.

Therefore, the results can compare with subsequent test results, and the reliability of the test results is high.

3.2 Simulation Experiment of Fault Prediction based on PCA

The main output data of the draw-wire sensor is the rate change before and after the displacement. The sensor feedback changes with the rate change, through which sensor fault can be predicted.

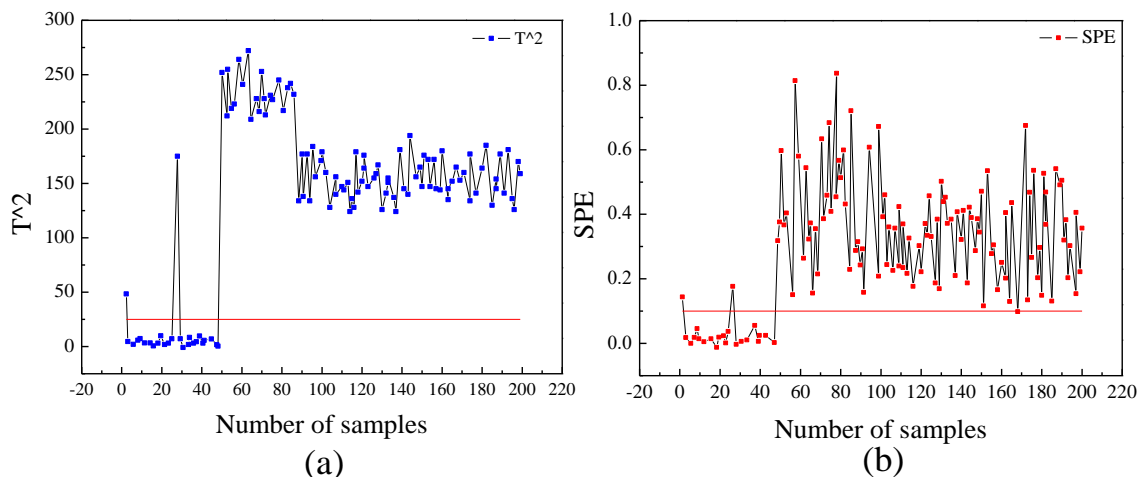


Figure 7: Statistical analysis of PCA fault prediction: (a) T^2 (b) SPE

The feature of deviation fault is that the difference between the feedback value and the real value is a constant value. In figure 7, the red baseline represents the threshold of the statistics, and the serial number of the sample data exceeding the threshold represents failure occurrence in the corresponding sensor. In the experiment, the first 50 samples are collected without fault, so they haven't exceeded the threshold. The T^2 statistics is 0.085, while that of SPE is 0.132.

The results show that the PCA method can predict the sensor fault in a long time and promptly in the automatic control system platform of the steel

The data of five sensors are collected, and the real-time data of 200 fault-free states are trained. The model is constructed based on these data.

The experimental prediction results obtained through the PCA method are shown in figure 7.

milling group, but the prediction-error rate and fault-omission rate are high.

3.3 Simulation Experiment of Fault Prediction based on DPCA

The DPCA method can simulate the fault prediction of the wire-draw displacement sensor with the same data as in the previous section, and the fault is introduced from the 51st set of sample data. The corresponding T^2 statistics and SPE statistics are shown in figure 8.

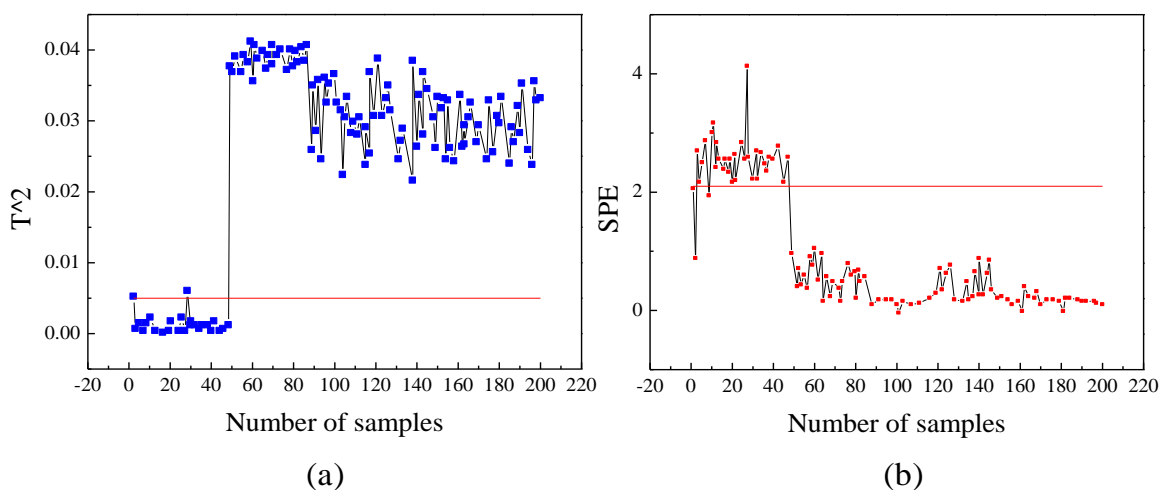


Figure 8: Statistical analysis of DPCA fault prediction: (a) T^2 (b) SPE

SPE statistics obtained through the DPCA method can accurately predict the fault of the sensor with a rare miss. In terms of prediction-error rate, the effect of SPE statistics is lower than that of T^2 . Comparison of the DPCA method and the PCA method suggests that although the prediction-error rate of T^2 statistics through the DPCA method is higher than that of the PCA method, the prediction-error rate of SPE statistics through the DPCA method is far lower than that of the PCA method. This indicates that the dynamic feature of real-time data is a very important factor in the fault prediction of the sensor. Consequently, the DPCA method containing time-delay can greatly improve the accuracy of the fault prediction of the draw-wire sensors.

4. Conclusion

Here, PCA and DPCA can predict the fault of draw-wire sensors. PCA is the most important and widely used method in data-driven technology. Compared with the traditional data analysis method, PCA can deal with multi-dimensional nonlinear data. To analyse the sensor feedback data with dynamic features, a DPCA method is further proposed based on the PCA. The experiment is simulated on the platform of the automatic adjustment system of the mill position of the steel unit. The results show that the optimized DPCA method can analyse and predict the fault of the sensor more effectively than the PCA method, and the prediction accuracy is higher.

The shortcomings are summed up as the following. 1. There are many types of sensor faults, while only the fixed deviation fault is studied here, and further research is needed for the diagnosis of other faults. 2. Although the proposed DPCA method can accurately predict the fault and reduce the false alarm ratio of the fault, the prediction space of this method is very limited, and the working conditions in actual production are complex and unpredictable. Therefore, new optimization methods should be explored through a combination of various systems.

Overall, with the popularization of big data and artificial intelligence (AI) technologies, electromechanical equipment based on data-driven technology is getting mature, and the demand for sensor stability and accuracy is rising. A more accurate and effective prediction and diagnosis model for sensor faults still are ceaselessly being experimented with by domestic and international scholars. Meanwhile, more diversified optimization algorithms are needed to simplify the analysis of complex and diverse data and obtain valuable information from historical data, accelerating the process of industrial automation.

References

- [1] Huang Z., Tian L. (2018) Research on Fault Prediction and Health Management System of Intelligent Equipment. *Journal of Graphics*, 39(5), 855-861.
- [2] Xu Q., Ai Q., Du X, et al. (2020) Early Fault Detection of Electro-hydraulic Servo System of Ship Steering Gear Driven by Model and Data. *Journal of Shanghai Jiaotong University*, 54, 411(05), 15-28.
- [3] Lin Z., Wang S., An W, et al. (2020) Time-varying Reliability Evaluation and Fault Prediction of Data-driven Equipment. *Chemical Industry Progress*, 350(11), 44-49.
- [4] Wu X. [2019] Research Status and Method Analysis of Fault Prediction of Industrial Production Process Based on Data Driven. *Inverter World*, 34(6), 65-69.
- [5] Guo X. (2018) Application Research of Intelligent Device Fault Prediction and Health Management (PHM) Technology Based on Big Data. *Digital World*, 22(12), 10.
- [6] Ju J., Hu S., Shan Z, et al. (2017) Research on equipment health management driven by big data. *Journal of Ordnance Engineering*, 38(6), 73-75.
- [7] Cai G., Zhuang F., Liu Y, et al. (2020) Research on Fault Data-driven Mechanical Seal Life Prediction Method. *Equipment Management & Maintenance*, 463(01), 42-44.
- [8] Wang Q., Liu J., Wei B. (2020) Zhang Cheng. Research on Fault Identification Method of Data Driven Clustering Analysis. *Journal of Mechanical Engineering*, 56(18), 24-31.
- [9] Fan Y., Li T., Liu Z, et al. (2019) Data-driven Modeling and Application of Robot Shaft Temperature Prediction. *Automation Expo*, 18(10), 56-61.
- [10] Li Y., Xiao B., He W, et al. (2020) Research on Port Machinery Condition Monitoring Platform Driven by Big Data. *Port Handling*, 250(1), 4-8+51.
- [11] Duan W., Han X., Ma Li, et al. (2020) Running State Evaluation and Early Warning of Wind Turbine Gearbox Based on SCADA Data Driven. *Power Science and Engineering*, 36, 247(11), 5-10.
- [12] Ying W., Li P., Lei J, et al. (2017) A Data-Driven Architecture for Personalized QoE Management in 5G Wireless Network. *IEEE Wireless Communications*, 24(1), 102-110.
- [13] Stamatescu G., Stamatescu I., Arghira N, et al. (2019) Data-driven Modelling of Smart Building Ventilation Subsystem. *Journal of Sensors*, 19(4), 1-14.

- [14] Sanz-Herrera J A., J Mora-Macías., J Ayensa-Jiménez, et al. (2020) Data-Driven Computational Simulation in Bone Mechanics. *Annals of Biomedical Engineering*, 68(15), 299-304.
- [15] Niu Z., Reformat M., Tang W, et al. (2020) Electrical Equipment Identification Method with Synthetic Data Using Edge-oriented Generative Adversarial Network. *IEEE Access*, 39(99), 1-1.
- [16] Chen Y L., Michael H., Michael P. (2017) Data-driven coarse graining of large biomolecular structures. *PLOS ONE*, 12(8), 57.
- [17] Goncalves G., Batchvarov A., Liu Y, et al. (2020) Data-driven surrogate modelling and benchmarking for process equipment. 86(10), 72.
- [18] Ding S., Wang Z., Kong W, et al. (2019) Electrode regulating system modelling in electrical smelting furnace using recurrent neural network with attention mechanism. *Neurocomputing*, 359(24), 32-40.
- [19] Kearny C., Gerber A., Merwe A. (2017) Data-driven enterprise architecture and the TOGAF ADM phases[C]// 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). *IEEE*, 46(1), 9.
- [20] Ghorbanian V., Mohammadi M H., Lowther D. (2019) Design concepts of low-frequency electromagnetic devices based on a data-driven approach Vahid Ghorbanian, Mohammad Hossain Mohammadi and David Lowther. *Compel*, 38(5), 1374-1385.
- [21] Goodall P., Sharpe R., West A. (2019) A data-driven simulation to support remanufacturing operations. *Computers in Industry*, 105,48-60.
- [22] Mishra S., Glaws A., Cutler D, et al. (2020) Unified architecture for data-driven metadata tagging of building automation systems. *Automation in Construction*, 120(103), 411.