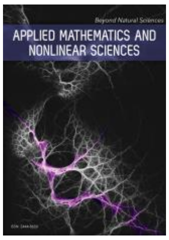
sciendo Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-18



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Research on the evaluation of equipment condition in assembly line workshop based on time series analysis

Meng Yu[1,†], Yan Zhao[1], Dongwen Wu[1], Maolin Pei[1], Tao Hu[1]

1. State Grid Jiangxi Power Supply Service Management Center, Nanchang, Jiangxi, 330000, China.

Submission Info

Communicated by Z. Sabir
Received June 10, 2024
Accepted September 16, 2024
Available online November 5, 2024

Abstract

Intelligent manufacturing requires the full use of modern information technology in traditional production scenarios, and assembly line workshops also need intelligent industrial upgrading in production management. To this end, this paper proposes a time series-based evaluation study of the state of equipment in an assembly line workshop. Through the theoretical analysis of the characteristics of assembly line workshop equipment state data, time series data mining technology is used to collect research and analysis data, which is then pre-processed. The differential autoregressive average model is used to construct the assembly line workshop equipment state evaluation model, and the simulation analysis of the assembly line workshop equipment state is carried out by combining the corresponding parameters with the research data. The data shows that the MAPE value of the ARIMA model for equipment failure is -0.0706%, which is more significant compared to the prediction effect of the traditional CNN model. In addition, the error range of the ARIMA model prediction of equipment state parameters of the assembly line workshop is 0.01~0.02. This study can accurately assess the equipment state of the assembly line workshop and lay the foundation for its intelligent transformation.

Keywords: ARIMA model; Time series; Assembly line workshop; Equipment condition assessment.
AMS 2010 codes: 15A60

†Corresponding author.
Email address: kaier1212@163.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2024-3033

1 Introduction

With the development of the economy, the manufacturing industry has become an important pillar of the national economy. The assembly line workshop equipment is a common industrial production equipment that first appeared in the field of processing manufacturing in order to improve the efficiency of the production process for the refinement of the division of labor, and its work is characterized by a single production unit focusing on a certain link in the production process, and to repeat this link under certain conditions [1-3]. The assembly line workshop outputs all products with the same process. Its production process is the completion of the process of dismantling, and each module to complete the process technology time is basically matched [4]. Assembly line workshop equipment is a combination of hardware and software, good design to ensure that the entire life cycle of the equipment can be expanded according to changes in demand, modify the production process, good stability, a single workstation to repeat the same production process, clear logic, low coupling between workstations, effectively avoiding complex production relationships and control conditions, high degree of automation, product processing, transfer, testing all automated to complete, saving labor, effectively reducing production costs [5-7]. In the process of equipment management, it is necessary to evaluate the state of the equipment and understand the implementation of equipment management and the direction of improvement, which is very helpful for the on-site production of the enterprise. Based on the time series analysis of assembly line workshop equipment, state evaluation reduces the time complexity of the network model, solves the problem of long-time dependence on time series data, and has significant advantages for the evaluation of long time series [8-9].

This paper analyzes the time series of the equipment state of the assembly line workshop in order to achieve the purpose of intelligent management of the assembly line workshop. The assembly line workshop is using time series data mining technology to preprocess and transform data on equipment state. In order to realize the accurate assessment of the equipment state of the assembly line workshop, it is proposed to use the differential autoregressive averaging model (ARIMA model) to construct the equipment state assessment model of the assembly line workshop and, at the same time, it is explained in detail the process of the implementation of the model and the principle of calculation. Combined with the collected data and ARIMA model, the fault state and equipment parameters of the assembly line workshop equipment are evaluated and analyzed, and finally, based on all the analytical results, the effectiveness of the model proposed in this paper in the evaluation of the state of the equipment in the assembly line workshop is confirmed.

2 Research on the condition evaluation of assembly line workshop equipment

2.1 Characterization of assembly line shop equipment condition data

2.1.1 Massiveness

Due to the large number of high-frequency sampling rate sensors arranged, the long time of acquisition makes the monitored data grow exponentially, and more intelligent processing methods are needed to adapt to the massive data processing work [10].

2.1.2 Non-authenticity of data

By electromagnetic radiation, sensor installation is not standardized, operational interference, etc., the data collected by the sensor does not fully reflect the real state of equipment operation. These inaccurate data will not only affect the reliability of the subsequent data analysis and processing, but also pose a hidden danger to the safety of production at the assembly line workshop. Analyzing the collected data requires a method that can adapt to the non-authenticity of the data.

2.1.3 High-dimensional low-value properties

The increasing automation of production in assembly line shops has made electromechanical equipment sensors a crucial component of production safety monitoring systems. The data collected by these sensors is heterogeneous. It is necessary to dig through the depth of the assembly line workshop equipment high-dimensional operation state data to automatically refine the change rule that embodies the change characteristics of the assembly line workshop equipment operation state in order to achieve the purpose of effective prevention and maintenance of the assembly line workshop electromechanical equipment.

2.2 Time series data mining

Data mining, also known as knowledge discovery in databases, is defined as the process of extracting information and knowledge implicitly contained in a large amount of incomplete, noisy, fuzzy, and random data that people do not know beforehand but are potentially useful. Time series data mining refers to data mining with time-dimensioned data as the main object. Not only does it query and traverse past data, but it also predicts future trends and behaviors to facilitate data analysis.

2.2.1 Basic steps

The basic steps of time series data mining mainly include data preprocessing, data transformation, mining modeling, and model evaluation. The detailed procedure is as follows:

1) Data preprocessing. The purpose of data preprocessing is to ensure the integrity of data items and improve the quality of data [11]. It mainly deals with abnormal data and empty data, deletes those abnormal data that affect the accuracy of the data mining results, and then utilizes the mean value, the plural, etc., to fill in the abnormal values. Of course, if the data is missing too much, it can be directly deleted from this column of data. Similarly, the empty data for some of the missing data is also filled in this way.

2) Data conversion. Data conversion is mainly to transform the preprocessed data in accordance with the data format of the model input [12].

3) Mining modeling. Mining modeling is an iterative process that constructs a mathematical model by analyzing the data characteristics and divides the data into two parts, one for model training to get the optimal model parameters and the other for model testing to get a model that meets the business requirements.

4) Model Evaluation. After the model is built, it must be tested and evaluated to determine its value. The accuracy obtained from the test set is only meaningful for the data used to build the model, but in practical applications, this valid model is not necessarily the correct model, which is directly due to the assumptions implicit in the model-building process. Therefore, the models obtained from the data in different scenarios are not the same, and the test models are best applied in practice and verified by the real data collected so that the models obtained are the most in line with the current business requirements.

2.2.2 Application embodiment

Time series data mining technology is widely used in various fields and has even become a hot spot in the software research and development market, such as communications, finance, manufacturing and other industries that use time series data mining as the underlying technical support to fulfill business requirements [13]. The most typical example is the equipment failure warning, which utilizes the historical operation data of the equipment and predicts the data in the future period, thus reflecting the change of the equipment state and achieving the effect of a failure warning. In addition, in the stock market, the time series data of stocks not only reflect the stock price information and trends in the historical period but also imply the trend information in the future period so as to evaluate the investment value of stocks.

2.3 ARIMA-based equipment condition assessment model

Data mining for the characteristics of time series data in the context of this topic, the ARIMA model is used for data prediction and assessment of the state information of the relevant parameters of the key components of the assembly line workshop equipment. The construction process of the one-dimensional ARIMA time series prediction model for equipment condition assessment. The differential autoregressive averaging model's basic principles are explained in the first step, and in the second step, the time series data information is selected for analysis. To smooth the time series data, the data smoothness test is used in the third step to obtain the minimum differential order required. In the fourth step, the model identification stage refers to the initial validation of the time series data after differential processing to obtain the model parameters. In the fifth step, the parameter estimation stage is to obtain optimal parameter values suitable for time series data through data calculation and analysis. In the sixth step, the model adaptability test refers to the way of analyzing the data by white noise calculation, comparing whether the observed values of the past data are consistent with the predicted values of the model. Below is a detailed description of the operation process.

2.3.1 Differential Autoregressive Average Model (ARIMA)

Differential autoregressive sliding average (ARIMA) model mainly includes moving average process (MA), autoregressive process (AR), autoregressive moving average process (ARMA) and ARIMA process, and its biggest feature is to transform the non-stationary series into a stationary series to realize the prediction of the short-term future value by the difference method.The essence of the ARIMA model prediction is to predict future trends through the autocorrelation of the time series. To predict the future trends, $ARIMA(p,d,q)$ the mathematical description of the model formula is shown in (1):

$$y_t = u + \sum_{i=1}^{p} \Phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{1}$$

Where y_t is the time series obtained after d degrees of difference; u denotes constant: ε_t denotes white noise with t always expectation of 0 and variance of σ^2; Φ_i and Θ_j are the coefficients for each term: p is the coefficient of the autoregressive term, and q is the coefficient of the moving average term. ARIMA is a linear model, which has a strong ability to fit the effect in time series with strong linear trend, but often has poor fitting effect ability in time series with nonlinear trend.

The steps of ARIMA model construction are: (1) The time series to be predicted are smoothed, and if the series does not meet the smooth condition, it needs to be made to meet the smoothness time series condition by the operation of differencing. (2) After meeting the smoothing conditions of the time series through the autocorrelation coefficient (ACF) and partial correlation coefficient (PACF) to initially determine the model of the order p and q. (3) Estimation of the parameters of the model, as well as parameter significance test and residual stochasticity test, to diagnose the model to determine the reasonableness of the model. (4) The constructed model is utilized to forecast the time series.

The condition for using the ARIMA model is that the time series as the object to be forecasted is a zero-mean smooth time series, i.e., there is no obvious upward or downward trend in the curve characteristics of the smooth stochastic series. However, for non-zero-mean non-smooth stochastic time series for prediction, the time series needs to be pre-smoothed before being converted into a time series to meet the ARIMA model prediction conditions. In the model construction step (1), zero-meanization for non-zero-mean stochastic time series requires that each item of the original data to be observed be subtracted from the mean of the series to obtain a zero-mean stochastic time series. This zero-mean stochastic time series is shown in equation (2):

$$X_t = Y_t - \overline{Y} \tag{2}$$

The zero-mean non-stationary random time series needs to be made into a zero-mean stationary time series through the operation of differencing. If the time series still does not satisfy the smooth condition after the first-order difference, the difference operation can be carried out again through several iterations until the time series satisfies the smooth condition. The mathematical description of the process is shown in equation (3), where d means after d times of differencing. That is:

$$\begin{gathered} \Delta y_t = X_t - X_{t-1} \\ \Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \\ \cdots \\ \Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1} \end{gathered} \tag{3}$$

The determination criteria for selecting parameters p,q based on trailing or truncated tails are shown in Table 1. The determination of ARIMA model parameters p,q in step (2) is determined by determining whether the PACF and ACF are trailing or truncated, where trailing means that the values of the PACF and ACF decay geometrically or oscillatorily to converge to 0, and truncated means that the values of the PACF and ACF decay to 0 in a discontinuous manner.

Table 1. Determination criterion of model

ρ_k	Φ_{kk}	Model selection
Trailing	p Step tail	$AR(p)$
q Step tail	Trailing	$MA(q)$
Trailing	Trailing	$ARMA(p,q)$

After step (2) completes the determination of the model, step (3) needs to carry out the estimation of p,q parameters to portray the appropriateness of the model, which usually adopts the method of determining the order of the information criterion function. The most commonly used are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), the main idea of the two is to calculate the value corresponding to each order to determine the degree of model appropriateness, and the smaller the value of the AIC and BIC results corresponding to the p,q parameter as the best model parameters. The formulas for AIC and BIC are shown in equations (4) and (5), where L is the maximum likelihood value under the model, n is the amount of predicted data, and k is the number of variables in the model. Namely:

$$AIC = -2\ln(L) + 2k \tag{4}$$

$$BIC = -2\ln(L) + \ln(n) \times k \tag{5}$$

After determining the parameters of the ARIMA model p,d,q, a residual test is also needed. The residual refers to the difference between the original data and the data fitted by the model, if the residual sequence conforms to the stochastic normal distribution and is not autocorrelated, it means that the effective linear part of the original data has been completely extracted into the ARIMA model. In this paper, the Durbin-Watson autocorrelation test is used to test the residuals. When the DW value is significantly close to 0 or 4, the residuals exist autocorrelation, indicating that the ARIMA model needs to be re-selected; when the DW value is close to 2, the residuals do not exist autocorrelation, indicating that the ARIMA model complies with the test standard, and it can be used to make the model prediction. In summary, the value of the ARIMA model p,d,q can be calculated, and finally, through the model parameter test, the optimal model is obtained to forecast the target time series.

2.3.2 Selection of time series data samples

This paper uses factory assembly line workshop equipment for each dimension 2020-2022 3 years of continuous equipment health status data. It is a month as a time interval for the data statistics, equipment, and each dimension health status data samples, as shown in Table 2. Table 2 describes the assembly line workshop equipment each dimension state sample data, each dimension of the experimental data research and analysis process is the same, followed by the number of failures after inspection X6 as an example of research and analysis.

Table 2. Data samples of each dimension of the equipment

Workshop equipment dimensions	First month	Second month	Third month	……
Temperature test rate X1	0.89	0.67	0.58	……
Angle qualification X2	0.87	0.77	0.47	……
Wireless number pass rate X3	0.48	0.76	0.87	……
……	……	……	……	……
The number of failures after the inspection X6	4.96	3.98	7.26	……
……	……	……	……	……

2.3.3 Examining data and processing data

The first step in the process of predictive analysis of the health status of equipment in a single-dimensional assembly line workshop is the need to ensure that the data sequence maintains a certain degree of smoothness, and if the sample data appear to be unstable, the difference method is used to transform the non-smooth data sequence into a smooth data sequence, so as to further analyze and study the data. The distribution of the sample data series of the number of faults after inspection is shown in Figure 1. The total number of faults after inspection time series samples appears to have obvious mutations, so the sequence is called an unsteady time series. For this phenomenon of unsteady time series, it is necessary to use the difference method to process the data and then convert it into a smooth time series. Therefore, in this subsection, for the unsteady time series data, first-order differencing is used to try to process the data, and if the results of the time series still show that it is unsteady, then continue to use second-order differencing to process the data until it becomes a smooth time series.

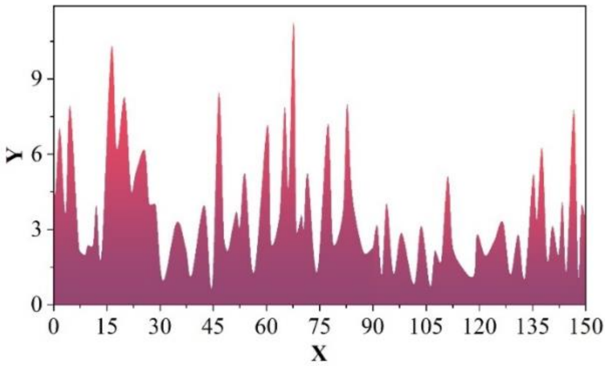


Figure 1. The number of faults in the inspection is according to the sequence distribution

The first-order difference and second-order difference time series are shown in Figure 2. Unsteady time series - the number of faults after inspection after the first-order difference (red), and then after the second-order difference (blue) processing, the sequence from the unsteady time series into a smooth time series, basically in the vicinity of 0 up and down uniform fluctuations. The method of testing the time series is generally commonly used using the observation method, but it also needs to be used with the rigorous statistical test method ADF test again. The test results of the significance of the test is (True, 4.67, 0.0001) due to the $p \ll 0.01$, so after second-order difference processing, data transformed into a smooth time series.

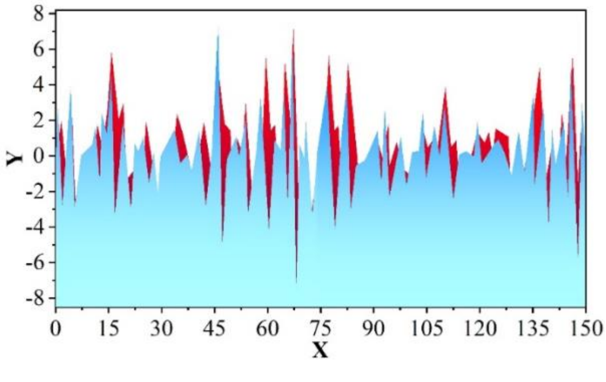
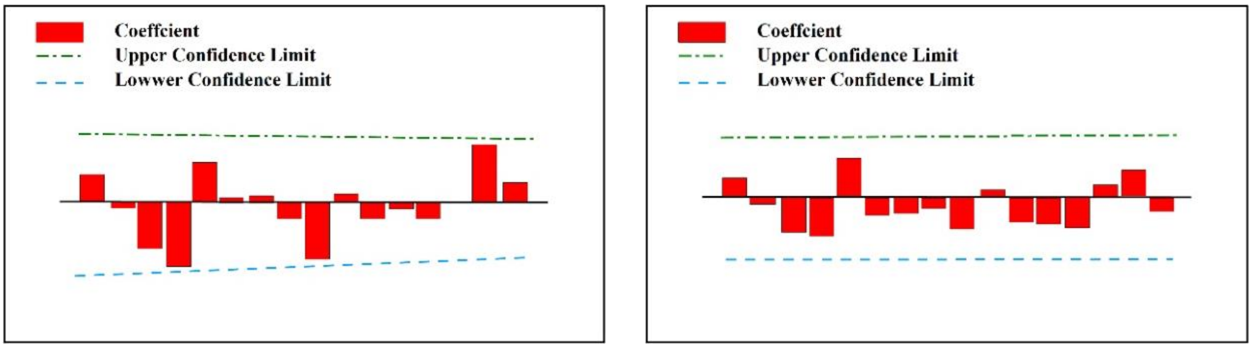


Figure 2. First order difference and second order difference time sequence

2.3.4 Model identification

Model identification is mainly through identifying the truncated and trailing characteristics of the autocorrelation function and the truncated and trailing characteristics of the partial correlation function of a given series, so as to determine what kind of model the series is. According to the theory of autocorrelation analysis, if the time series is an AR model (autoregressive model), its autocorrelation coefficient shows trailing characteristics after p step, and the partial correlation coefficient shows truncated characteristics; if the time series is a MA model (moving average model), its autocorrelation coefficient shows truncated characteristics after q steps, and the partial correlation coefficient shows trailing characteristics; if the time series is a $ARMA(p,q)$ model, then its autocorrelation coefficient and partial correlation coefficient show trailing characteristics. After the experimental data processing, the autocorrelation and partial correlation characteristics are obtained, and the autocorrelation and partial correlation coefficient characteristics of the ARIMA model are shown in Fig. 3, in which Figs. (a)~(b) shows the autocorrelation coefficient and partial correlation coefficient, respectively.



(a) ARIMA ACF (b) ARIMA PACF

Figure 3. Self-correlation and partial correlation coefficient of ARIMA model

2.3.5 Parametric valuation

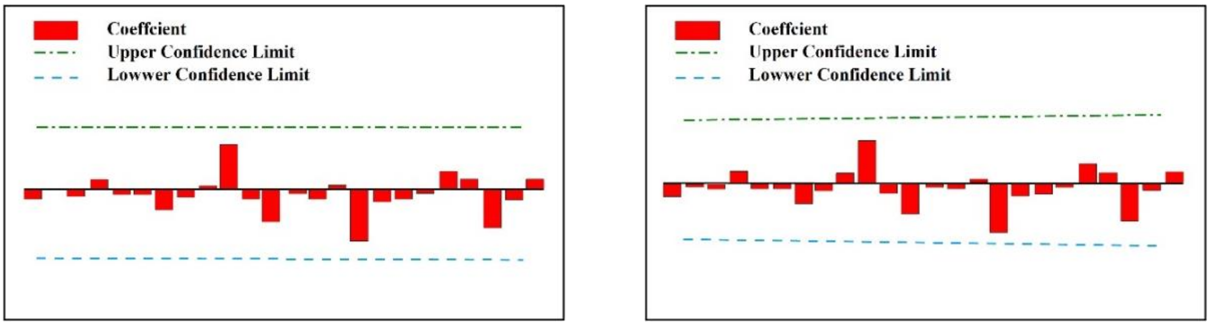
The corresponding optimal parameter values of the ARIMA prediction model are shown in Table 3. In the process of determining the parameter valuation, firstly, the Python tool is used to make appropriate adjustments to the p,q parameter, and then the adjustments are gradually increased between the range of 0 and 5, and the method of substituting the order into the calculation formula is used to analyze and draw relevant conclusions through the study and analysis of the data permutations and combinations. When $p=0$ and $q=1$, the corresponding Sig value of the prediction model is the largest, and the BIC (Bayesian Information Criterion) value is the smallest, based on the knowledge related to the BIC theory, the best prediction model is obtained as ARMA (0, 1).

Table 3. The ARMA prediction model corresponds to the optimal parameter value

Parameter (p,q)	Significance(Sig.)	R^2	BIC
(0,1)	0.514	0.809	18.453

2.3.6 Reasonableness test

The residual sequences of the experimentally obtained unidimensional optimal prediction models are examined and analyzed for the study using white noise. The rationality test is shown in Fig. 4, where (a)~(b) are the autocorrelation and bias correlation of the residual sequence, respectively. From Fig. 4, it can be seen that the autocorrelation of the residuals is not obvious, and it is determined that the residuals are white noise sequences, so it is then deduced that the constructed ARIMA model of the health state of the equipment in the assembly line workshop is compliant with the requirements.



(a) ARIMA residual ACF (b) ARIMA residual PACF

Figure 4. Rationality test

3 Simulation analysis of equipment state based on ARIMA model

3.1 Failure assessment of assembly line shop equipment

By predicting the condition monitoring data of the equipment in the assembly line workshop, we can know the future operation trend of the equipment in the assembly line workshop and then make a judgment on whether the equipment fails or not. In this paper, the following two methods for making a judgment on the equipment will fail, and the above-constructed model for verification and analysis will be used.

3.1.1 Threshold determination

Statistics of historical monitoring data for the state of the equipment failure when the state is analyzed to determine the failure threshold when the predicted value exceeds the threshold, it is determined to be a failure, assembly line workshop equipment viewing threshold determination statistics shown in Figure 5. Statistics 188 times assembly line workshop equipment historical fault data corresponding to the surface temperature can be found when the temperature reaches 140 °C, the number of failures appeared to increase steeply, that is, 85.3% of the faults appear after 140 °C, so 140 °C is set as the fault threshold when the predicted value exceeds the threshold value is judged to be about to fail.

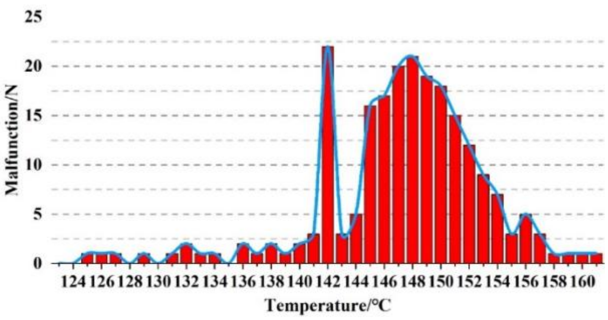


Figure 5. The pipeline workshop equipment view threshold value decision

3.1.2 Determination of variance

Statistical historical monitoring data for the assembly line workshop equipment failure when the state of the analysis of the difference between the predicted value and the actual value, the variance of the difference as a parameter to characterize the failure. The comparison between the actual value and the predicted value is shown in Fig. 6, and the statistics of the variance range corresponding to the number of failures is shown in Fig. 7. Combining Fig. 6 and Fig. 7, it can be seen that when a fault is about to occur, the monitored real data fluctuates greatly from the predicted data, which can be used as a sign to characterize a fault that is about to occur. The 100 times the variance value is counted with the corresponding number of times the fault occurs, and the variance threshold is set to 5, i.e., when the variance is greater than 5, it is determined that a fault is about to occur.

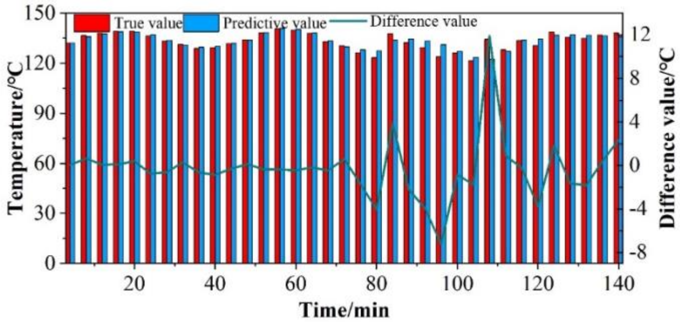


Figure 6. The real value is compared to the predicted value

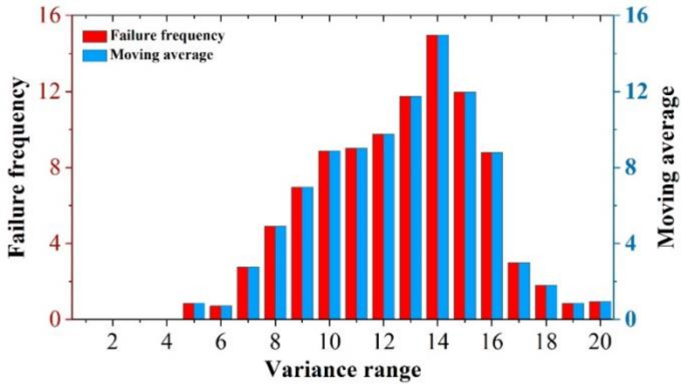


Figure 7. The variance range corresponds to the number of failures

3.1.3 Model validation analysis

The model predicts the surface temperature of the assembly line workshop equipment from the monitoring data of the surface temperature of the assembly line workshop equipment over some time, and the original monitoring data of the assembly line workshop equipment is shown in Fig. 8. It can be seen that the temperature of the assembly line workshop equipment varies within the range of 120°C to 144°C over 140 minutes, and the integrated average temperature value is 135.35°C.

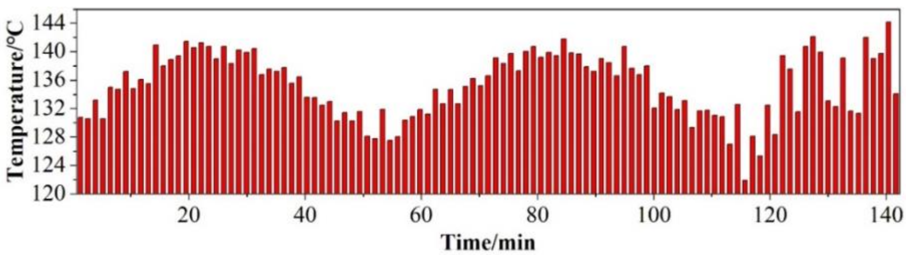


Figure 8. Production line workshop equipment raw monitoring data

To determine the smoothness of the monitored equipment status data changes should be analyzed from a period of time to analyze the specific representative, set a time period m, time interval $[t-m,t]$ within the k sampling points, recorded as D_i, $i=1,2,3\cdots,k(k>1)$, then calculate the average change in the value of the data is:

$$\overline{D}_m=\frac{|D_k-D_{k-1}|+|D_{k-1}-D_{k-2}|+\cdots+|D_2-D_1|}{k} \tag{6}$$

Assuming a time period n, $n\geq m>0$, and a time interval $[t-n,t]$, within which there are p sampling points, denoted as D_j, $j=1,2,3\cdots,p$, the average value of the data in the computed interval is:

$$\overline{D}_n=\frac{D_1+D_2+\cdots+D_p}{p} \tag{7}$$

Then the smoothness of the data is calculated:

$$D(m,n)=\frac{\overline{D}_m}{\overline{D}_n} \tag{8}$$

Set the base time interval for data acquisition to T_{unit}, i.e., the data adoption interval set in the algorithm must be an integer multiple of this base time. Set the maximum data sampling interval as $T_{\max}$ and the minimum data sampling interval as $T_{\min}$. Introduce the life cycle influence factor δ to obtain the adaptive sampling interval ΔT, which can be obtained:

$$\Delta T=\left[\frac{T_{\max}}{T_{unit}}\cdot\frac{D(m,n)}{\delta}\right]T_{unit} \tag{9}$$

The adaptive interval collection points are shown in Fig. 9. Next, the raw monitoring data are analyzed in terms of life cycle stages and the corresponding life cycle impact factors are obtained, and then the smoothness of the data is combined with the calculation of the adaptive data collection intervals to find the corresponding sampling points, which are a total of 39 sampling points.

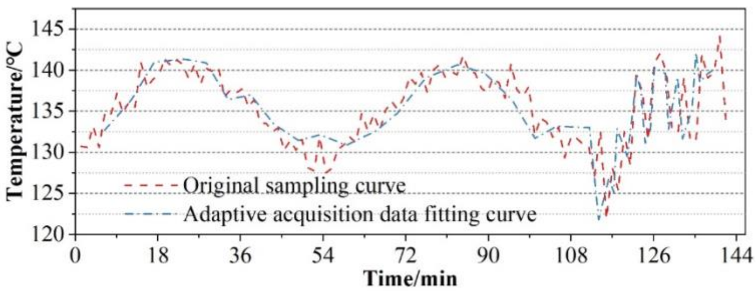


Figure 9. Adaptive interval gathering point

On the basis of 39 sampling points, the ARIMA-based assembly line workshop equipment condition assessment model is compared and analyzed with the traditional CNN-based assembly line workshop equipment condition assessment model, and the results of the model comparison analysis are shown in Table 4. The detailed data analysis is shown below:

Table 4. Model comparison analysis results

Time point	True value/℃	CNN Predictive value/℃	CNN Error value/℃	ARIMA Predictive value/℃	ARIMA Error value/℃
9	132.71	133.1931	-0.4831	132.9286	-0.2186
15	136.3809	136.3063	0.0746	136.3149	0.066
20	141.0075	140.093	0.9145	140.4087	0.5988
23	141.4192	141.2661	0.1531	141.5773	-0.1581
30	140.9641	141.0633	-0.0992	140.9111	0.053
34	136.5066	137.9296	-1.423	137.3981	-0.8915
40	136.9383	137.2536	-0.3153	136.6588	0.2795
44	133.5799	135.0541	-1.4742	134.4682	-0.8883
50	131.5192	132.8898	-1.3706	132.0823	-0.5631
55	132.2285	132.9185	-0.69	132.424	-0.1955
61	130.8654	132.0557	-1.1903	131.3927	-0.5273
66	132.517	133.0534	-0.5364	132.7215	-0.2045
72	134.8628	134.8776	-0.0148	134.9302	-0.0674
76	139.1747	138.4874	0.6873	138.7953	0.3794
86	140.807	140.4323	0.3747	140.0818	0.7252
90	139.8281	139.995	-0.1669	139.7296	0.0985
97	136.4203	137.3031	-0.8828	136.8138	-0.3935
102	132.71	133.1931	-0.4831	132.9286	-0.2186
106	136.3809	136.3063	0.0746	136.3149	0.066
114	141.0075	140.093	0.9145	140.4087	0.5988
116	141.4192	141.2661	0.1531	141.5773	-0.1581
119	140.9641	141.0633	-0.0992	140.9111	0.053
117	136.5066	137.9296	-1.423	137.3981	-0.8915
118	136.9383	137.2536	-0.3153	136.6588	0.2795
120	133.5799	135.0541	-1.4742	134.4682	-0.8883
123	131.5192	132.8898	-1.3706	132.0823	-0.5631
124	132.2285	132.9185	-0.69	132.424	-0.1955
125	130.8654	132.0557	-1.1903	131.3927	-0.5273
126	132.517	133.0534	-0.5364	132.7215	-0.2045
127	134.8628	134.8776	-0.0148	134.9302	-0.0674
130	139.1747	138.4874	0.6873	138.7953	0.3794
132	140.807	140.4323	0.3747	140.0818	0.7252
131	139.8281	139.995	-0.1669	139.7296	0.0985
133	136.4203	137.3031	-0.8828	136.8138	-0.3935
135	132.71	133.1931	-0.4831	132.9286	-0.2186
137	136.3809	136.3063	0.0746	136.3149	0.066
138	141.0075	140.093	0.9145	140.4087	0.5988
139	141.4192	141.2661	0.1531	141.5773	-0.1581
140	140.9641	141.0633	-0.0992	140.9111	0.053

The error rate of the predicted value compared to the true value was calculated using the mean MAPE of the absolute values of the relative percentage errors, calculated as:

$$MAPE = \frac{\sum_{i=1}^{N} \frac{\left|P_i - O_i\right| \times 100}{P_i}}{N} \tag{10}$$

Where, P_i is the true value, O_i is the predicted value, and N is the number of sampling points, then the MAPE value of CNN prediction model is calculated to be -0.241%, while the MAPE value of ARIMA prediction model is -0.0706%, i.e., the RIMA prediction model has a higher priority in the assessment of the equipment condition of the assembly line workshop.

The difference between the true value and the predicted value is calculated and the variance corresponding to the difference is calculated and the results of the comparison of the variance of the true value and the predicted value are shown in Table 5. It is found that the variance 0.196607, although it does not exceed the upper limit set threshold of 140 ℃, but it exceeds the set variance threshold, so it is determined that the monitored point of the monitored equipment is about to fail.

Table 5. The actual value is compared to the predicted value variance

Time point	True value/℃	Predictive value/℃	Error value/℃	Variance value/℃
9	132.71	132.9286	-0.2186	
15	136.3809	136.3149	0.066	
20	141.0075	140.4087	0.5988	
23	141.4192	141.5773	-0.1581	
30	140.9641	140.9111	0.053	
34	136.5066	137.3981	-0.8915	
40	136.9383	136.6588	0.2795	
44	133.5799	134.4682	-0.8883	
50	131.5192	132.0823	-0.5631	
55	132.2285	132.424	-0.1955	
61	130.8654	131.3927	-0.5273	
66	132.517	132.7215	-0.2045	
72	134.8628	134.9302	-0.0674	
76	139.1747	138.7953	0.3794	
86	140.807	140.0818	0.7252	
90	139.8281	139.7296	0.0985	
97	136.4203	136.8138	-0.3935	
102	132.71	132.9286	-0.2186	
106	136.3809	136.3149	0.066	
114	141.0075	140.4087	0.5988	0.196607
116	141.4192	141.5773	-0.1581	
119	140.9641	140.9111	0.053	
117	136.5066	137.3981	-0.8915	
118	136.9383	136.6588	0.2795	
120	133.5799	134.4682	-0.8883	
123	131.5192	132.0823	-0.5631	
124	132.2285	132.424	-0.1955	
125	130.8654	131.3927	-0.5273	
126	132.517	132.7215	-0.2045	
127	134.8628	134.9302	-0.0674	
130	139.1747	138.7953	0.3794	
132	140.807	140.0818	0.7252	
131	139.8281	139.7296	0.0985	
133	136.4203	136.8138	-0.3935	
135	132.71	132.9286	-0.2186	
137	136.3809	136.3149	0.066	
138	141.0075	140.4087	0.5988	
139	141.4192	141.5773	-0.1581	
140	140.9641	140.9111	0.053	

3.2 Prediction of assembly line workshop equipment state parameters

3.2.1 Prediction sets

The original time series of vibration amplitude is shown in Fig. 10, obtaining the historical state parameter data, randomly intercepting the bearing vibration data of the key components in the assembly line workshop equipment for some time as the experimental data of the time series, and a total of 140 data points are intercepted, and considering that the prediction model is more suitable for the short-term prediction, the first 125 are selected as the training set, and the last 15 data are selected as the prediction set.

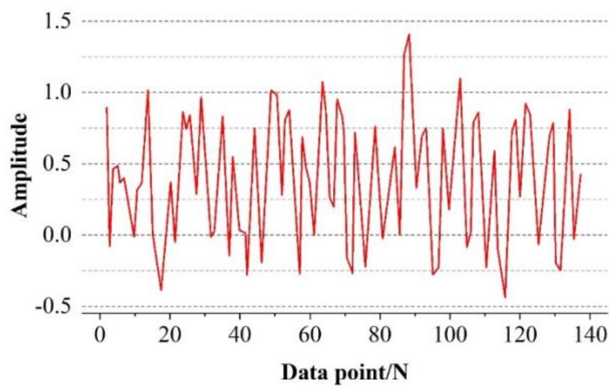


Figure 10. The sequence of the original time of the vibration amplitude

3.2.2 Determination of smooth series

The results of the ADF test for the original time series are shown in Table 6. It has been observed that the vibration amplitude time series is a smooth series. For further verification of its smoothness, it is tested using the unit root (ADF) test, and the calculated p value is 0.00047, which is much less than the critical value of 0.05 and the t-statistic is less than the critical value at any confidence level, so the series is smooth.

Table 6. Original time sequence ADF test results

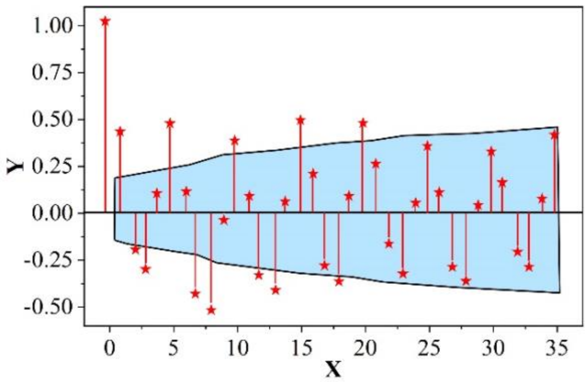
Inspection item	Test result
Test Statistic Value	-4.12374
P-Value	0.00047
Critical Value (1%)	-3.24351
Critical Value (5%)	-2.44218
Critical Value (10%)	-2.27001

3.2.3 Model ordering

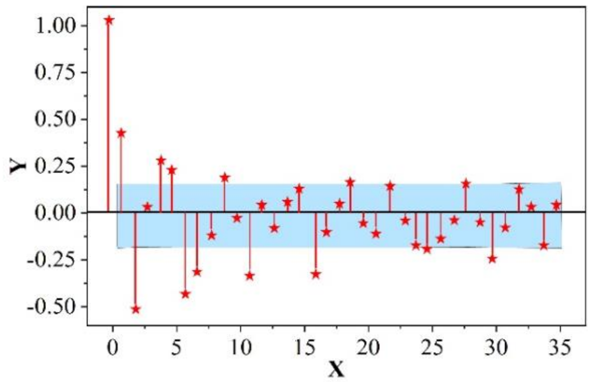
The time series autocorrelation and partial autocorrelation are shown in Figure 11, and the autocorrelation and partial autocorrelation plots are plotted to determine the appropriate type of model. The autocorrelation plot and partial autocorrelation function are found to be trailing, so the ARIMA model is chosen. In determining the value of p and p, it can be determined by observing the autocorrelation and partial autocorrelation plots, but it is prone to large errors, so the choice was made to take the BIC (Bayesian Information Criterion) to determine the order of the model. It can be expressed as:

$$BIC = K \ln N - 2\ln(L) \tag{11}$$

Where, K is the number of model parameters, L is the likelihood function, N is the number of samples, when the BIC is minimized is the model works better, the BIC is calculated by the values of parameters p and p, and the model order is determined as ARIMA (2, 0, 6).



(a) Autocorrelation



(b) Partial Autocorrelation

Figure 11. The time series is independent of correlation

3.2.4 Residual white noise test

The white noise test is performed on the residuals of the predicted values of the ARIMA (2,0,6) model, the residual time series is shown in Figure 12, the autocorrelation and partial autocorrelation of the residual series are plotted, and the white noise test is performed, the autocorrelation plot of the residual series and the partial autocorrelation is shown in Figure 13, and it is found that p-value in the order of 1 to 35 is much larger than the critical value of 0.05, and it is considered that the residuals are a white noise series. Therefore, the ARIMA (2.0,6) model can be fitted and predicted.

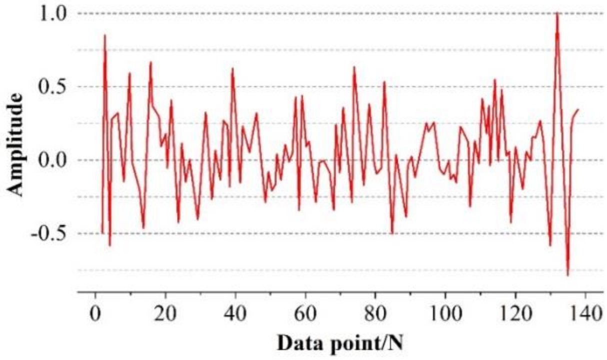
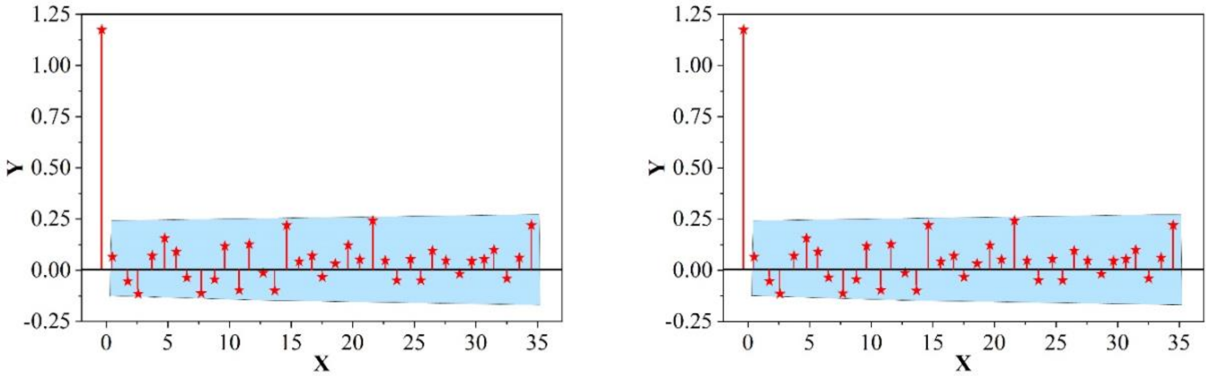


Figure 12. Residual time sequence



(a) Autocorrelation (b) Partial Autocorrelation

Figure 13. The residual differential sequence is based on the correlation diagram

3.2.5 Fitting and prediction

ARIMA model fitting and prediction results are shown in Figure 14. The ARIMA (2,0,6) model can be fitted and predicted, where the blue line represents the true value and the red line represents the fitted and predicted values. As can be seen from Figure 14, the fitted value is closer to the true value, the pre-fitting results are more accurate, the basic trend is consistent with the true value, the prediction results roughly predict the trend, and the difference between the actual results is small, the specific error value range of 0.01~0.02, indicating that the ARIMA model is able to accurately predict the state parameters of the equipment in the assembly line workshop.

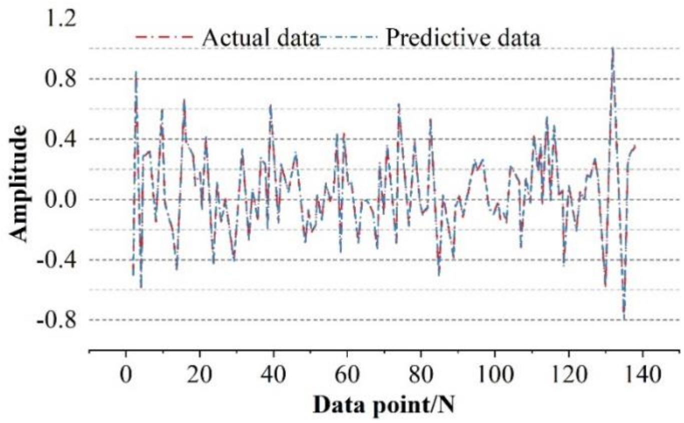


Figure 14. ARIMA model fitting and prediction results

4 Conclusion

As an important part of factory safety production, assembly line workshop equipment directly affects the sealing and safety in the process of factory operation, and the state assessment of assembly line workshop equipment has been attracting much attention. This paper summarizes the characteristics of state data for assembly line workshop equipment and obtains the initial data required for research and analysis with the help of time series data mining technology. Considering the complexity of the operation of the equipment in the assembly line workshop, it is proposed to use the ARIMA model to construct the state assessment model of the equipment in the assembly line workshop and evaluate and analyze the equipment in the assembly line workshop from the two aspects of the state parameters and failures. In the process of equipment failure assessment, the MAPE value of the ARIMA prediction model is -0.0706%, which is better than the CNN model. In the prediction of equipment state parameters, the ARIMA model accurately predicts the state parameters of the assembly line workshop equipment, and its error value ranges from 0.01 to 0.02. On the whole, the model constructed in this paper improves the technical guarantee for the safe and stable operation of the assembly line workshop equipment in the factory.

Funding:

[Fund projects] Research on Key Technologies of Metrological Verification Flexible Production for Advanced Measurement System (Project number: 52185223000H).

References

[1] Wang, C., & Wei, K. (2019, November). Construction of Intelligent Manufacturing Digital Workshop Ability Assessment Model for CPS. In 2019 2nd World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM) (pp. 627-630). IEEE.

[2] Zhang, H., Li, L., Li, L., Cai, W., Liu, J., & Sutherland, J. W. (2022). An integrated energy efficiency evaluation method for forging workshop based on IoT and data-driven. Journal of Manufacturing Systems, 65, 510-527.

[3] Wu, P., Qi, M., Gao, L., Zou, W., Miao, Q., & Liu, L. L. (2019, August). Workshop heterogeneous equipment information perception analysis system. In 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) (Vol. 1, pp. 3-7). IEEE.

[4] Liu, X., Yu, S., Li, Z., & Jiang, L. (2015, September). Developing a Performance Evaluation System Based on Key Performance Indicators for Discrete Manufacturing Workshop. In Conference on Engineering And Technology Management GCETM-2015 (p. 297).

[5] Wei, W., Liu, L., Yang, M., Li, J., & Wu, F. (2021, October). Predictive maintenance system for production line equipment based on digital twin and augmented reality. In International Workshop of Advanced Manufacturing and Automation (pp. 479-486). Singapore: Springer Singapore.

[6] Zhang, Y., Zhang, C., Yan, J., Yang, C., & Liu, Z. (2022). Rapid construction method of equipment model for discrete manufacturing digital twin workshop system. Robotics and Computer-Integrated Manufacturing, 75, 102309.

[7] Zhang, J., Ni, Z., Shi, G., Lai, J., Liu, J., & Wang, H. (2022, November). Health evaluation of equipment operation and maintenance of cigarette factory power workshop based on AHP-FCE model. In International Conference on Advanced Sensing and Smart Manufacturing (ASSM 2022) (Vol. 12351, pp. 293-299). SPIE.

[8] Shan, H. Y., Li, L. N., Yuan, Y., & Wang, C. (2018, December). Simulation and optimization of production line in Em-plant based assembly workshop. In 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 1381-1385). IEEE.

[9] Wang, Q., Ma, J., Jiang, Z., & Zhang, J. (2023). Research on the evaluation method of the operation status of digital workshop in discrete manufacturing industry. Production Engineering, 17(2), 247-261.

[10] Chen Ji lin,Huang Yanhao,Qiu Weijiang,Chen Xinglei,An Ning & Zhang Kaiyu. (2021). Research on feature extraction method of power grid online data based on big data. Journal of Physics: Conference Series(1),

[11] Wei Shi,Guangcong Xue,Xicheng Yin,Shaoyi He & Hongwei Wang. (2024). DRMM: A novel data mining-based emotion transfer detecting method for emotion prediction of social media. Journal of Information Science(3),590-606.

[12] V.K. Saxena & Shashank Pushkar. (2019). Managing privacy of sensitive attributes using fuzzy-based data transformation methods in privacy preserving data mining environment. Int. J. of Business Information Systems(2),249-264.

[13] Andreas P. Mentzelopoulos,Dixia Fan,Themistoklis P. Sapsis & Michael S. Triantafyllou. (2024). Variational autoencoders and transformers for multivariate time-series generative modeling and forecasting: Applications to vortex-induced vibrations. Ocean Engineering(P2),118639-118639.

About the Authors

Meng Yu (1983.10-), female, Han nationality, Nanjing, Jiangsu, master, State Grid Jiangxi Power Supply Service Management Center, senior economist, main research direction: research and application of power supply technology.

Yan Zhao (1986.08-), female, Shandong Tai’an, master, State Grid Jiangxi Power Supply Service Management Center, senior engineer, main research direction: electric energy metering verification technology research.

Dongwen Wu (1990.04-), male, Suizhou Hubei, master, State Grid Jiangxi Power Supply Service Management Center, engineer, main research direction: new electric energy measurement and detection technology research.

Maolin Pei (1980.10-), male, Han nationality, Fengcheng, Jiangxi, master, State Grid Jiangxi Power Supply Service Management Center, professor-level senior engineer, main research direction: new electric energy measurement and detection technology research.

Tao Hu (1990.09-), male, Han nationality, Ezhou Hubei, master, State Grid Jiangxi Power Supply Service Management Center, engineer, main research direction: new electric energy measurement and detection technology research.