Application of Principal Component Analysis Algorithm in Abnormal Diagnosis of Electricity Bill Reading and Receiving Data

Weihua Zhao*

{14413707@qq.com}

State Grid Henan Electric Power Company Marketing Service Center , Zhengzhou,450051 China

Abstract. In order to understand the abnormal diagnosis of electricity bill reading and receiving data, an application research based on principal component analysis algorithm in the abnormal diagnosis of electricity bill reading and receiving data is put forward. In this paper, firstly, an abnormal diagnosis method of electricity bill reading and receiving data based on principal component analysis algorithm is proposed. The collected data of electricity bill copying and checking is taken as the research sample data, and it is denoised and preprocessed by wavelet transform. On this basis, the correlation coefficient matrix is used to solve the principal components of data, and the number of principal components of data contribution rate is determined. Secondly, the principal component analysis expression is used to establish a data anomaly diagnosis model, and the data anomaly diagnosis of electricity bill copying and receiving is realized through the model. Finally, in order to verify the comprehensive effectiveness of the proposed method based on the principal component analysis algorithm, a simulation experiment was carried out on Win7 system, CPU i55600U@2.6GHz and memory 16GB3200MHz. The experimental results show that by denoising the data, the accuracy of the abnormal diagnosis result of the proposed method is effectively improved and the diagnosis delay is obviously reduced.

Keywords: Wavelet transform; Principal component analysis algorithm; Electricity fee collection data; Abnormal diagnosis.

1 Introduction

Checking and collecting electricity charge is a comprehensive work, which mainly includes electricity charge meter reading, electricity charge accounting and electricity charge collection. Although intelligent meters are becoming more and more popular, some places still need to carry out the work of copying and checking. However, in the work of copying and collecting, we can find that there are still some problems, which may lead to related risks, which is unfavorable for power supply enterprises. Therefore, it is necessary to reduce or avoid these risks by optimizing and improving the work of copying and checking, so that the work of copying and checking electricity charges can achieve the expected ideal effect. It is a very important project to diagnose the abnormal data of electricity bill reading and checking. With the rapid development of science and technology, the reading and checking staff in electric power enterprises should not only improve their professional level, but also make full use of modern science and technology to improve their literacy. In recent years, relevant experts have

MSIEID 2023, December 08-10, Guangzhou, People's Republic of China

DOI 10.4108/eai.8-12-2023.2344734

also strengthened their research in this field. Some scholars have set smart meters as the information hub between users and power grids, and provided users with relevant information such as electricity consumption habits and load characteristics through power grids for data abnormality diagnosis. Others analyze the characteristics of wind power, obtain the probability power curve through Copula function, and establish a data anomaly diagnosis model combined with the time series characteristics of abnormal data. Although the above two methods have achieved satisfactory research results at this stage, due to the failure to denoise the data, the accuracy of abnormal diagnosis of electricity bill reading and receiving data is reduced and the diagnosis delay is increased. Therefore, this paper proposes an abnormal diagnosis method of electricity bill reading and receiving data based on principal component analysis algorithm. The simulation results show that the proposed method can not only comprehensively enhance the accuracy of abnormal diagnosis of electricity bill reading and receiving data, but also reduce the diagnosis delay[1-2].As shown in Figure 1:

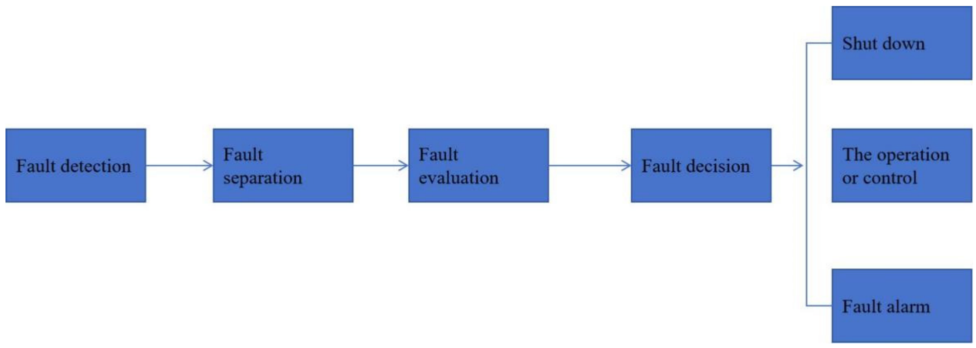

Figure. 1. Schematic diagram of fault diagnosis process

2 Common risk analysis of electricity charge copying and checking in power supply enterprises

As far as the current reality is concerned, the risks existing in the work of copying and collecting electricity fees are mainly in the following aspects, which requires managers to form a clear understanding. First, the risks in meter reading. In electricity meter reading, the main risk is the inaccurate meter reading data. According to the actual work, such problems have indeed occurred. If the meter reading is inaccurate, it will not accurately reflect the actual electricity consumption of users, which will lead to the inaccurate collection of electricity charges in the current period. There are many reasons for inaccurate electricity meter reading, such as staff copying errors, or there are problems such as stealing electricity. The influence of these factors will have a real impact on the accuracy of meter reading. Second, accounting risks. Meter reading is only the primary task of electricity charge copying and checking, and after meter reading, electricity charge accounting is needed. There are also some risks in electricity accounting. For example, there is a lack of thinking about the accuracy of meter reading data, which is often simply based on meter reading data, and some obvious problems are not paid attention to. At the same time, there may be accounting errors in electricity accounting, which will bring trouble to electricity collection. Third, the risk of fees. Electricity charge is the last link in the work of copying and checking, but in practice, the risk in this aspect is the greatest, that is, customers default on electricity charges. In the actual charging work, some customers have maliciously defaulted on electricity charges, and repeated collection has failed, which has caused direct damage to the economic interests of power supply enterprises[3-4].

3 Diagnostic method for abnormal data of electricity bill copying and checking

3.1 Denoising Pretreatment of Electricity Charge Copy Data

In the process of collecting electricity bill reading data, many data have spikes or abrupt changes, and the noise is not balanced white noise. For this kind of signal processing, sampling traditional Fourier transform can not achieve denoising. Because Fourier analysis completely analyzes the signal in frequency domain, it is impossible to obtain the change of the signal in any time period, and a random mutation of the signal on the time axis will affect the whole signal spectrum. Wavelet analysis can not only analyze the signal in time domain and frequency domain at the same time, but also accurately distinguish the abrupt part of the signal from the noise, and effectively eliminate the noise.

For function φ (t)=L2R, if Fourier function is performed, the following admissibility conditions need to be met:

$$\int_R = \frac{\Psi(\omega)^2}{\omega} d\omega \quad (1)$$

In the above formula, φ (t) represents the basic wavelet or wavelet generating function, and the wavelet generating function φ (t) is expanded or translated; At the same time, the scale factor is a; If the translation factor is b and the translated function is $\varphi_{(a)(b)}$ (t), then:

$$\varphi_{(a,b)}(t) = a^{\frac{1}{2}}(\frac{t-b}{a}), a > 0, d \in R \quad (2)$$

In the above formula, $\varphi_{(a)(b)}$ (t) represents the wavelet basis function that depends on a and b.

In the process of practical application, useful signals can be expressed as low-frequency signals or partially stable signals, while noise can be expressed as high-frequency signals, so the following processing is needed in the process of noise reduction: firstly, the initial signals are decomposed by wavelet; Secondly, the high frequency coefficients of wavelet decomposition are quantized by threshold value and other forms; Finally, the signal is reconstructed to achieve the purpose of noise reduction. Set the one-dimensional signal of noise as:

$$s(i) = f(i) + \sigma e(i) \quad (3)$$

In the above formula, f(i) stands for true signal; e(i) stands for noise; s(i) represents a signal containing noise. Usually, the one-dimensional noise reduction process can be divided into the following steps:

(1) Signal wavelet decomposition:

Select wavelet decomposition and determine the decomposition level N at the same time; Then the high-frequency coefficients of all levels are quantized[5].

(2) threshold quantization processing of high frequency coefficients:

Select an appropriate threshold to quantize the high frequency coefficients of 1~N slave layers.

(3) Wavelet reconstruction:

Wavelet reconstruction is realized through the coefficients of the N-th layer decomposed by wavelet and the high-frequency coefficients of the first to N layers after quantization.

In the above steps, the key is the selection of threshold and quantization, which will directly affect the quality of the signal. The threshold selection method mainly includes the following forms: unbiased likelihood estimation through stein, heuristic and so on.

In the data curve, the threshold of non-reflection point is the empirical value of its loss, which is mainly determined by different connection technologies and processes.

When thresholding signal f, there are mainly two methods: soft and hard thresholding. The hard threshold is to compare the absolute value of signal transformation with the threshold, and the point less than or equal to the threshold becomes zero, while the point greater than the threshold remains unchanged. It is found that the denoising effect of soft threshold is obviously better than that of hard threshold.

3.2 Based on the principal component analysis algorithm, the establishment of abnormal diagnosis model of electricity bill reading and receiving data.

Principal component analysis has obvious advantages in data preprocessing, especially in the application of dealing with complex data. It mainly combines it with other multivariate technologies. Firstly, the random signals in the subsystem are compressed, and a unified statistical model of large-scale system dimension reduction-—PCA model is constructed. Then, the statistical characteristic parameters of PCA model output in the system are extracted, and the specific location and nature of the fault are further determined by artificial intelligence technology through the analysis results, and all the variables are effectively retained in PCA model[6-7].

Principal component analysis mainly explains the original variable information through the correlation between the initial variables and uses a few linear combinations of the initial variables, so as to effectively reduce the dimension of the data. Usually, there is the following relationship between the principal variables obtained by principal component analysis and the initial variables, namely:

(1) Linear combination of initial variables is formed by different principal components;

(2) The number of initial variables is obviously higher than the number of principal components;

(3) Principal component effectively retains most information of initial variables;

(4) Different principal components are not interrelated.

Principal component analysis (PCA) is carried out on the data after wavelet transform. The principal component analysis algorithm is used to represent the data in the high-dimensional variable space as much as possible under the condition of ensuring the minimum loss of information in the electricity bill reading and receiving data. The number of principal components is a very important parameter in the PCA model. When the number of principal components used is too small, it will lead to the loss of information in variables and the error of the model will increase; When more principal components are used, the measurement noise in the data will be introduced too much, which will enhance the calculation of analysis and the complexity of diagnosis, so it is very important to determine the number of principal components.

Set the observed data of electricity meter reading and collection collected under normal conditions as X, which contains a total of m observation variables and n observation values, using m × The data matrix of n represents these data, namely:

$$X = \begin{bmatrix} x_{11}, & x_{12}, & \cdots, & x_{1m} \\ x_{21}, & x_{22}, & \cdots, & x_{2m}, \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1}, & x_{n2}, & \cdots, & x_{nm} \end{bmatrix} \quad (4)$$

Where the matrix X can be decomposed into the following form:

$$X = \begin{cases} t_1 p_1^T + t_2 p_2^T + \cdots t_m p_m^T \\ \sum_{i=1}^{m} t_i p_i^T \end{cases} \quad (5)$$

Set the variance of X as: = COV (X), where the eigenvalues can be expressed as λ1, λ2,..., λm, and λ 1 λ 2... λ m, the decomposition theorem of the binding matrix can obtain:

$$\sum = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_{\mathrm{m}} \end{bmatrix} \quad (6)$$

Solve the principal components of the electricity bill charging data through the correlation coefficient matrix, namely:

$$\hat{X} = \begin{cases} \sum_{i=1}^{l} t_i p_i^T \\ \hat{T}\hat{P}^T \end{cases} \quad (7)$$

The number of hosts used to clarify the contribution rate of the data is determined mainly through the cumulative contribution rate of the variance. At the same time, the principal element analysis expression is used to establish the data abnormality diagnosis model, and realize the data abnormality diagnosis through the model, namely:

$$J(P_1) = \frac{1}{m}\sum_{i=1}^{m}\left\| x_1 - P_1 P_l^T x_i \right\|^2 \tag{8}$$

To sum up, the abnormal diagnosis of electricity bill copying and collecting data has been completed.

4 Simulation experiment analysis

In order to verify the comprehensive effectiveness of the proposed method based on principal component analysis (PCA), a simulation experiment was carried out on Win7 system, CPU i55600U@2.6GHz and memory 16GB3200MHz[8]. Because the proposed method uses wavelet transform to denoise the collected data of electricity bill copying and receiving in the early stage of research, the denoising effect of the proposed method is given in Table 1.

Table 1. denoising effect of the method proposed in this paper

Test object number	Initial signal-to-noise ratio/(SNR/db)	Signal to noise ratio after denoising /(SNR/db)
initial	25.63	30.12
001	24.36	32.10
002	22.10	33.42
003	23.45	33.12
004	26.32	32.10
005	28.12	33.38

By analyzing the experimental data in Table 1, it can be seen that the proposed method can effectively improve the signal-to-noise ratio of the data, and at the same time effectively suppress the oscillation phenomenon of the data, and has obvious effect on abnormal data processing.

In order to verify the accuracy of the abnormal diagnosis results of electricity bill reading and receiving data, the average absolute error and the average error are set as test indicators. The lower the values of the two indicators, the higher the accuracy of the abnormal diagnosis results of electricity bill reading and receiving data, and the lower the accuracy of the abnormal diagnosis results of electricity bill reading and receiving data. In the experiment, two methods were selected as the comparison methods, and the accuracy comparison results of abnormal diagnosis results of electricity bill reading and receiving data were given in Table 2:

By analyzing the experimental data in Table 2, it can be seen that the proposed method effectively reduces the average absolute error and the average error of the abnormal diagnosis result of the electricity bill reading and receiving data, and makes the abnormal diagnosis result of the electricity bill reading and receiving data more accurate. This mainly depends on the initial stage of the proposed method, which uses wavelet transform to denoise the collected electricity bill reading data, effectively filters out the noise in the data, and comprehensively improves the accuracy of the data abnormality diagnosis results[9-10].

Table 2. Accuracy of abnormal diagnosis results of electricity bill reading and receiving data by the proposed method

Variable number	Average absolute error/(%)	Average error/(%)
01	0.0208	0.0188
02	0.0186	0.0166
03	0.0175	0.0173
04	0.0156	0.0165
05	0.0143	0.0173

5 Conclusion

In order to effectively improve the accuracy and reduce the delay of abnormal diagnosis of electricity bill reading and receiving data by traditional methods, combined with principal component analysis algorithm, a method of abnormal diagnosis of electricity bill reading and receiving data based on principal component analysis algorithm is proposed. The effectiveness and superiority of the proposed method are fully verified by specific simulation experiments.

References

[1] Huang, W. , Lu, X. , Hu, J. , & Ye, Q. . (2021). Research on monitoring and diagnosis technology of data anomaly in distribution network. IOP Conference Series Earth and Environmental Science, 6(3)2, 042008.

[2] Brito, L. C. , Susto, G. A. , Brito, J. N. , & Duarte, M. A. V. . (2022). An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery. Mechanical Systems and Signal Processing, 163(1), 108105.

[3] Zhang, Y. , Yan, W. , Hong, G. S. , Fuh, J. F. H. , Wang, D. , & Lin, X. , et al. (2022). Data fusion analysis in the powder-bed fusion am process monitoring by dempster-shafer evidence theory. Rapid prototyping journal4(5), 28.

[4] Wang, Y. , Bai, X. , Liu, C. , & Tan, J. . (2022). A multi-source data feature fusion and expert knowledge integration approach on lithium-ion battery anomaly detection. Journal of Electrochemical Energy Conversion and Storage36(2), 19.

[5] Gent, J. , Navaratnam, K. , & Sharp, A. . (2021). Prenatal screening and diagnosis in multiple pregnancies. Obstetrics, Gynaecology & Reproductive Medicine, 31(10), 288-290.

[6] N. S .(2021).Oil condition monitoring and predicting actions using an Artificial Intelligence technique: Principal Components Analysis algorithm.IOP Conference Series: Materials Science and Engineering(1), 1(5)6, 111096-.

[7] Gao, D. , Wu, Z. , Guo, J. , Xu, Y. , & Lu, H. . (2021). Thermal protection system damage diagnosis method using machine learning algorithm. Journal of Spacecraft and Rockets36(7), 1-11.

[8] Abdolalikazemi, Z. . (2021). Diagnosis of breast cancer using image processing with svm and knn. International Journal of Engineering and Technology, 13(2), 11.

[9] Shia, J. A. , & Auslander, D. M. . (2021). A time-series rank ordering control system data-driven fault detection approach for hvac systems in buildings. ASME Journal of Engineering for Sustainable Buildings and Cities, 2(1), 1-12.

[10] Fanli Z H R H H J .(2023).Suppress ambient temperature interference strategy based on SnO2 gas semiconductor sensor using dynamic temperature modulation mode and principal component analysis algorithm.Sensors and Actuators: B. Chemical, 1(6)7.