

Method of detecting power equipment data based on LSTM-LOF

Zhicheng Ma^a, Jinxiong Zhao^a, Hong Zhao^a, Guangyuan Zheng^b, Baohui Wang^{b,*}

^a State Grid Gansu Electric Power Research Institute, Lanzhou 730070, China; ^b College of Software, Beihang University, 100191, China.

ABSTRACT

In the production work of the State Grid, the power equipment is often in a high and low temperature environment. The timely detection of the fault of the power equipment is an important part of the stable development of the State Grid. This paper proposes a method to detect the quality of power equipment based on the measurement data of power equipment at room temperature. Based on the LSTM model, the working data of power equipment is predicted in the form of serialized time nodes during the experiment. Then, the feature data is extracted and dimensionality reduction according to the operating environment, and finally the anomaly detection is carried out according to the local outlier factor (LOF) algorithm. The experimental results show that the prediction effect of the model prediction diagnosis results is ideal, and the LSTM-LOF combined with the model can be used to assist engineers in the quality diagnosis of power equipment.

Keywords: Long short-term memory, power equipment anomaly detection, feature extraction, local outlier factor algorithm

1. INTRODUCTION

As an important backbone enterprise related to national energy security and the lifeblood of national economy, State Grid Corporation undertakes the basic mission of providing safe and sustainable power supply for economic and social development. The power equipment plays an important role in the national grid. Generally speaking, to find the stability of power equipment, it is generally necessary to conduct artificial detection, measure its data, and then judge its stability according to its data. Three main issues need to be considered during data detection: estimation of current health status, prediction of future status and failure time, and determination of the impact of failure on system performance¹. Anomaly detection specifically refers to the early detection of potential product problems².

At present, most countries are encouraging the development of artificial intelligence and big data technology. In the direction of power equipment fault prediction, a large number of researches on the application of big data have been carried out and achieved fruitful results³⁻⁷. Compared with the traditional empirical knowledge and basic mathematical model for power equipment quality prediction, some scholars have applied neural network to data prediction of power equipment⁸⁻⁹. Based on the method¹⁰, this paper proposes a power equipment data prediction detection method, and then carries out the envelope of machine learning algorithm. It can fully excavate the rule knowledge between data, which not only overcomes the large judgment deviation of traditional experience and the lack of rules in the existing knowledge base, but also avoids the data accuracy, high efficiency and excellent accuracy which are difficult to achieve with the traditional mathematical model due to the complex structure of power equipment. It has reference significance.

*Baohui Wang: wangbh@buaa.edu.cn
Zhicheng Ma: 2633116578@qq.com.
Jinxiong Zhao: 1304518920@qq.com.
Hong Zhao: 2268107554@qq.com.
Guangyuan Zheng: zhengguangyuan@buaa.edu.cn

2. LSTM DEEP LEARNING MODEL

In neural network prediction, for RNN, only the output of the last node can be accepted, and the weight of the whole layer network is shared, so gradient disappearance or gradient explosion is easy to occur. LSTM¹¹ model is a deep neural network model, which can carry out linear and nonlinear fitting of data features, carry out deep extraction of data features, and dig potential rules. It usually shows better prediction performance than traditional models in prediction tasks.

2.1 LSTM Network structure

The LSTM network layer is composed of cell units connected. The unit structure is shown in Fig.1,

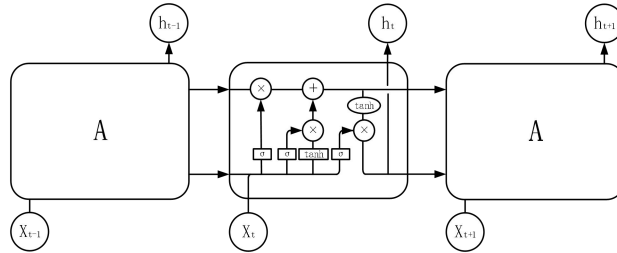


Figure 1. LSTM unit structure

It can be seen that the repeated structure A of LSTM has four layers. The main line running through these units is line 2 above. The main structure of LSTM consists of three gates that control the increase and decrease of cell information, namely the forget gate, the input gate and the output gate.

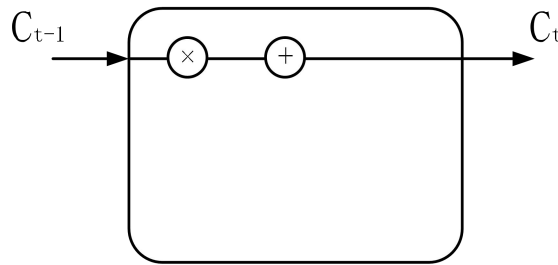


Figure 2. cell state

Forget Gate

The forget gate is responsible for the reduction of information in cells, controlling what information should be forgotten.

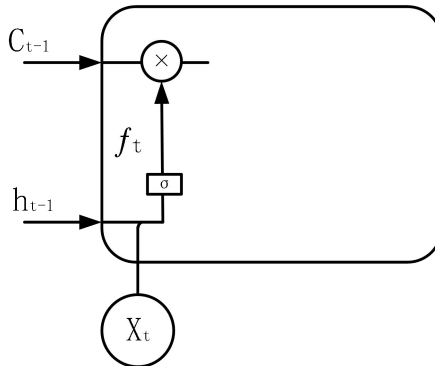


Figure 3. Forget Gate

The forget gate formula is $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$. After connecting h_{t-1} and x_t it multiplies a weight W_f and adds bias b_f . The weight and deviation are the parameters that the network model needs to learn. If the size of hidden state (i.e., size of hidden Layer of neurons) is h_{size} , then W_f is $h_{size} * h_{size}$. The value of h_{size} is set manually.

Input Gate

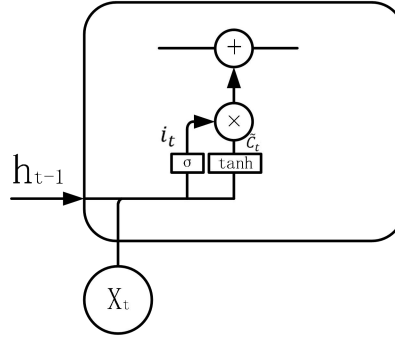


Figure 4. Input Gate

The input gate determines what information is saved in the cell and what information is added to the cell. The formula for the gate is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (1)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2)$$

Input Gate (a Sigmoid function layer) and \tanh layer, the two neural network layers will learn their parameters like the forgetting gate before.

Output gate

The output gate determines what information the cell needs to output. The output gate formulas are as follows:

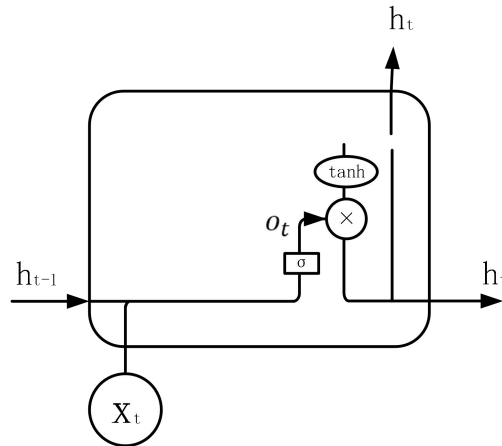


Figure 5. Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

$$h_t = o_t * \tanh(C_t) \quad (4)$$

Output gate also has its own weight parameters to learn.

The whole one-layer neural network is composed of several LSTM units, which send the original sequence (x_1, x_2, \dots, x_n) into the network, through the LSTM network, and finally expand into the required dimensions from the full connection layer.

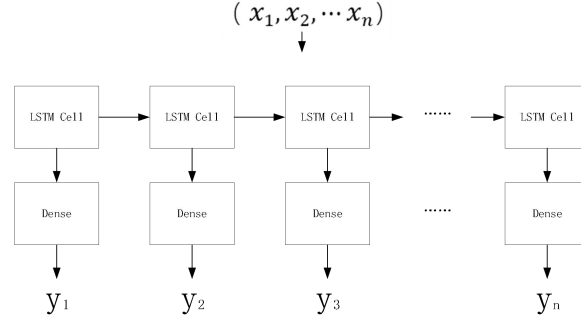


Figure 6. LSTM network structure

2.2 Evaluation

In this paper, MAE (mean absolute error) is selected as the evaluation model performance index, which represents the average value of the absolute error between the observed value and the real value. The calculation formula is:

$$MAE(X, h) = \frac{1}{m} \sum_{i=1}^m |h(x^{(i)}) - y^{(i)}| \quad (5)$$

3. DATA DIMENSION REDUCTION AND DETECTION MODEL

PCA algorithm is used to reduce the dimension of the predicted data, and the processed data is convenient for later detection, and then anomaly detection is performed according to local outlier factor (LOF) algorithm.

3.1 PCA

Principal component analysis (PCA) is a common linear dimension reduction method. The main way is to project the information in the high-dimensional space to the low-dimensional space through linear projection, and make the projected information retain the maximum original information and reduce the dimension of data.

Assuming that the original data set has m samples and n features, the original data set can be expressed as the following $m \times n$ order matrix, shown as Eq.6.

$$X_{m \times n} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \quad (6)$$

Since the dimensional influence between features in the data set needs to be eliminated before calculating the matrix covariance, standardization needs to be carried out first, as shown in Eq.7.

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{Var}(x_j)}}, i = 1, 2, \dots, m, j = 1, 2, \dots, n \quad (7)$$

In Eq.8,

$$\bar{x}_j = \frac{1}{m} \sum_{i=1}^m x_{ij}, \text{Var}(x_j) = \frac{1}{m-1} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2 \quad (8)$$

Use the equation $|\lambda - R| = 0$ to calculate the characteristic root of the matrix in Eq.7, and then use the characteristic root $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ to calculate the orthogonal unitized eigenvectors e_1, e_2, \dots, e_n .

In the case of keeping the total variance of the data set unchanged, the contribution rate of the i th principal component is $\lambda_i / \sum_{i=1}^n \lambda_i$ ($i = 1, 2, \dots, n$), and the cumulative contribution rate of the first q principal components after reduction is:

$$\text{Sum}(q) = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^n \lambda_i} \quad (9)$$

After the original data matrix is processed by principal component analysis to reduce the dimensionality, the relationship between the principal components $C_1, PC_2, \dots, PC_s (s \leq n)$ and the original data feature x_i is:

$$\begin{cases} PC_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ PC_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ PC_s = a_{s1}x_1 + a_{s2}x_2 + \dots + a_{sn}x_n \end{cases} \quad (10)$$

In Eq.10, a_{ij} and PC_i are not related to each other, and $a_{i1}^2 + a_{i2}^2 + \dots + a_{in}^2 = 1$.

3.2 Local outlier factor

Local Outlier factor (LOF) is a density-based outlier detection method.

In sample set D, the distance between a sample point o and its nearest neighbor far from its kth is defined as the kth distance $d_k(o)$ of point o, and the Kth distance neighborhood $N_k(o)$ of point o is defined as all sample points whose distance from point o is not more than $d_k(o)$, as Eq.11:

$$N_k(o) = \{o' | o' \in D, d(o, o') \leq d_k(o)\} \quad (11)$$

In Eq.11, $d(o, o')$ is the distance from point o to point o' , if $d(o, o') < d_k(o)$, then the reachable distance $d_{reach}(o, o')$ of sample point o is defined as $d_k(o)$, else for $d(o, o')$, As shown in Eq.12:

$$d_{reach}(o, o') = \max\{d(o, o'), d_k(o)\} \quad (12)$$

Where, the nearest neighbor number k needs to be selected according to the actual sample size and sample distribution. However, sometimes sample distribution will affect the selection of outliers, so LOF uses locally accessible density to represent the density of sample point o, and defines the locally accessible density of point o as $lrd_k(o)$:

$$lrd_k(o) = \frac{|N_k(o)|}{\sum_{o' \in N_k(o)} d_{reach}(o, o')} \quad (13)$$

$|N_k(o)|$ is the number of sample points contained in the kth distance neighborhood of point o.

Define the average value of the ratio of the local reachable density of point o to the local reachable density of all points in the kth distance neighborhood $N_k(o)$ as the local outlier $LOF_k(o)$, as shown in Eq.14:

$$LOF_k(o) = \frac{\sum_{o' \in N_k(o)} \frac{lrd_k(o')}{lrd_k(o)}}{|N_k(o)|} \quad (14)$$

According to Eq.14, the lower the local reachable density of sample point o is compared with that of its kth nearest neighbor, the higher the value of $LOF_k(o)$ is, the greater the possibility that point o is an outlier.

4. THE SIMULATION TEST AND RESULT ANALYSIS

Since the failure of power equipment occurs in high and low temperature environment, high and low temperature data is used as the detection standard. Figure 7 is the flow block diagram of the abnormal detection method of power equipment proposed in this paper. The whole process block diagram is divided into data prediction module and envelope detection module, in which the data prediction module mainly completes the data prediction. The envelope detection module completes the feature extraction of the prediction test data and the data dimension reduction processing. Finally, a data set that can be input into the local outlier factor model is formed to complete the learning and classification of input features.

The specific implementation steps are as follows:

Step 1: Preprocess the original room temperature data, input it into the LSTM model, and output the high and low temperature environmental data of power equipment products.

Step 2: Conduct data preprocessing on the data obtained in Step 1. Firstly, conduct data standardization operation, and then use principal component analysis to complete the data dimension reduction operation.

Step 3: The pre-processed data is taken as the data set, and the resampling method is adopted to sample the data set, and n new sub-data sets are obtained. The n sub-data sets are taken as the input to train n local outlier factor models, and n different base learners are obtained. The learning results of n base learners are averaged through the integration decision. The final classification results are obtained.

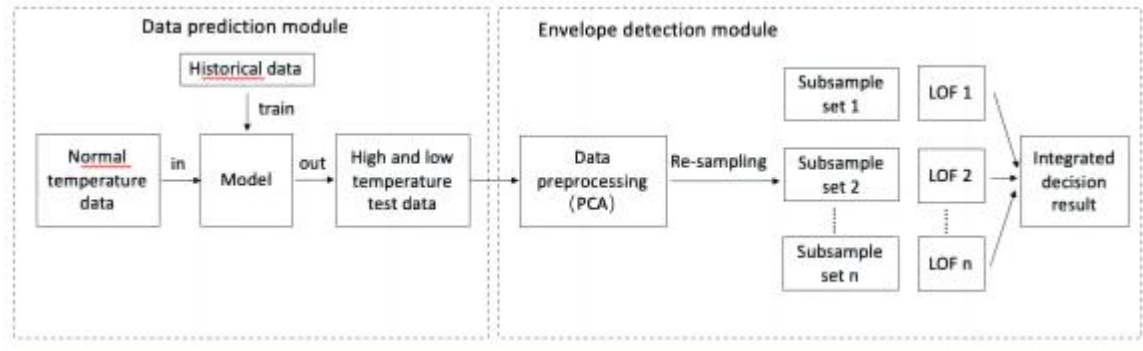


Figure 7. Flow chart of anomaly detection method

4.1 Model evaluation indicators

Anomaly detection is essentially a classification problem, the all samples are divided into two categories: abnormal state and normal state, the test using confusion matrix in table 1 for test results were described,

Table 1. Confusion matrix

	Predicted to be abnormal	Predicted to be normal
Actually abnormal	True Positive(TP)	False Negative(FN)
Actually normal	False Positive(FP)	True Negative(TN)

This experiment uses Precision (P) and Recall (R) as the algorithm evaluation indicators, and the calculation formula is as follows:

$$P = \frac{TP}{TP+FP} \quad (15)$$

$$R = \frac{TP}{TP+FN} \quad (16)$$

P represents the proportion of predicted anomalies and the number of actual or abnormal samples to the total number of predicted anomalies. The larger the value, the better the anomaly detection effect of the algorithm. R represents the number of predicted anomalies and actual or abnormal samples as a percentage of the total number of actual anomalies. The larger the value, the better the anomaly detection effect of the algorithm.

4.2 Data introduction

The experimental data are transformer data collected by Gansu Power Company of State Grid. Transformer products use load, local discharge, dielectric loss of insulating oil, polarization index, volume resistivity, sugar aldehyde content, ground current of iron core, insulation resistance of iron core, DC resistance of winding as detection characteristics. The high temperature of transformer products for environmental testing is 62 °C and the low temperature is -26 °C . The characteristic values of transformer products are recorded every 20 minutes. There are 50 kinds of sampling transformer products. In the test process of a transformer product, the monitoring data of its characteristic load is shown in Figure 8. The residual characteristics of the product are similar to the tendency of the load. It can be seen that the variation trend of product eigenvalue is highly correlated with the ambient temperature. Therefore, in this example, 9 characteristics of transformer products are analyzed as key parameters.

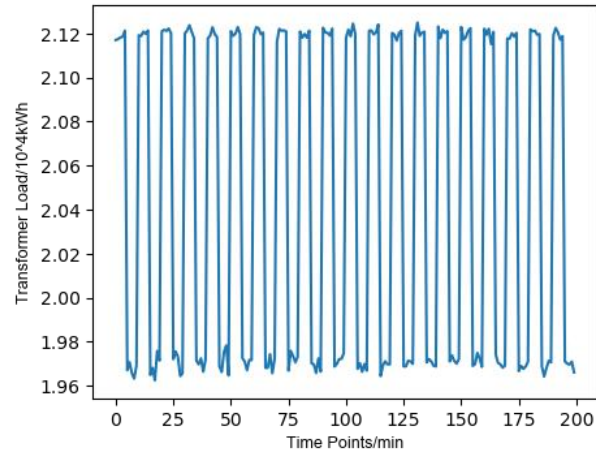


Figure 8. Trend diagram of transformer load high and low temperature environment

Transformer products will be tested at normal temperature, and the load curve of transformer products at normal temperature is shown in the figure 9 below:

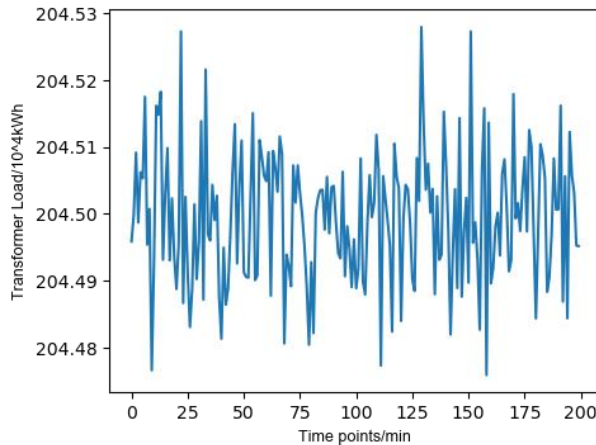


Figure 9. Ambient trend diagram of transformer load at normal temperature

4.3 Analysis of experimental process and results

Due to different data units, the contribution to fitting is also different, so it is necessary to standardize the data to the same order of magnitude, and accelerate the convergence of the model. The standardized formula is shown in Eq.17.

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (17)$$

The experiment completed the programming calculation in Python3.8 environment, relying on the deep learning framework keras and TensorFlow.

Data processing

The mean and variance of normal temperature data were introduced into the network as cell state and hidden State, and the temperature sequence of the test was introduced into the network to predict the feature sequence in the test environment.

Each individual product data are extracted, united into the same data length, after contrast found at around 200, most data length hence take 200 for unification of data length, will integrate the different products, the result was a 3 d data, one of the first dimension represents the product number, the length of the second dimension represents the data, The third

dimension represents the number of features. Use the first 40 products as a training set and the last 10 products as a validation set.

LSTM neural network selected the transfer function of Sigmoid function, the number of trained hidden nodes is 32, the optimizer uses Adam, the magnitude of network gradient update learning rate is 0.001, learning rate attenuation is 0.0002, the maximum number of network training is 20.

Effect display

Figure 10 shows the convergence of the loss function on the training set and verification set as the number of iterations increases. From the overall loss value, the loss value tends to converge to 0.09. After 15 training rounds, it began to converge.

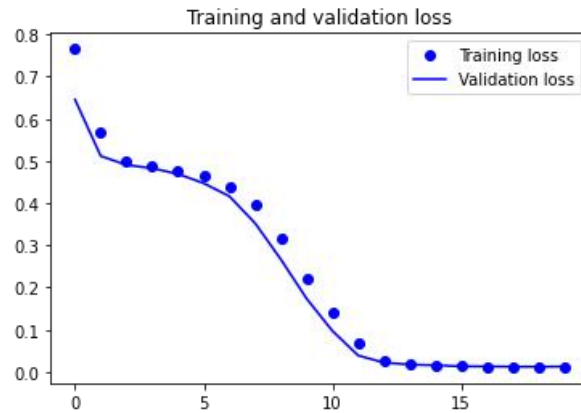


Figure 10. LSTM model loss convergence curve

The final model prediction is compared with the real data (taking characteristic current reserve coefficient Product 1 as an example), as shown in Figure 11. The same can be used for data prediction of abnormal products, as shown in Figure 12.

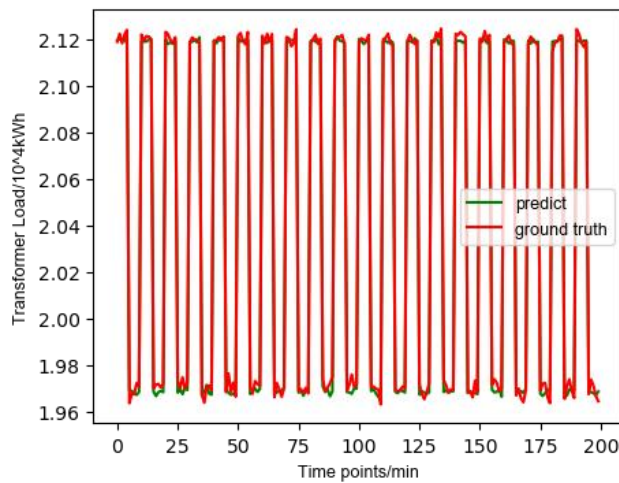


Figure 11. Data fitting comparison

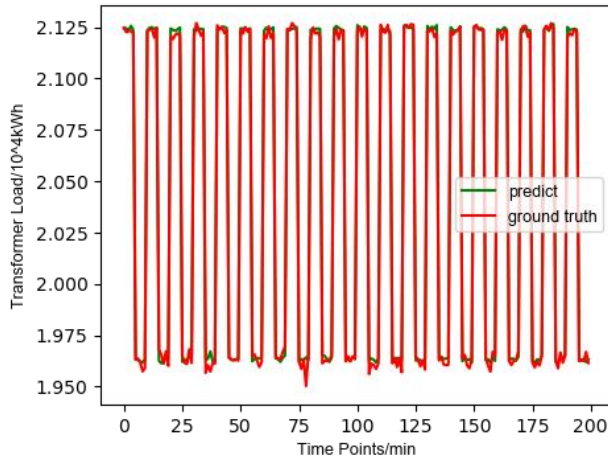


Figure 12. Abnormal product data prediction

It can be seen that the overall trend of the data is well fitted, which is also applicable to other parameters. It has certain practical significance and better guiding function for the power equipment experiment data prediction.

In order to reduce the redundancy of the features of the data set, speed up the training of the model, and reduce the amount of calculation of the anomaly detection algorithm, the principal component analysis of the features of the data set was carried out in this experiment. The contribution rate of each principal component is shown in the figure. It can be seen from Figure 13 that the cumulative contribution of the first two principal components is close to 50%. Therefore, in this paper, 9 transformer product features in the original data set are reduced to 2 principal component features, which can effectively reduce the actual operation of the subsequent LOF detection algorithm and improve the efficiency of transformer product abnormal detection.

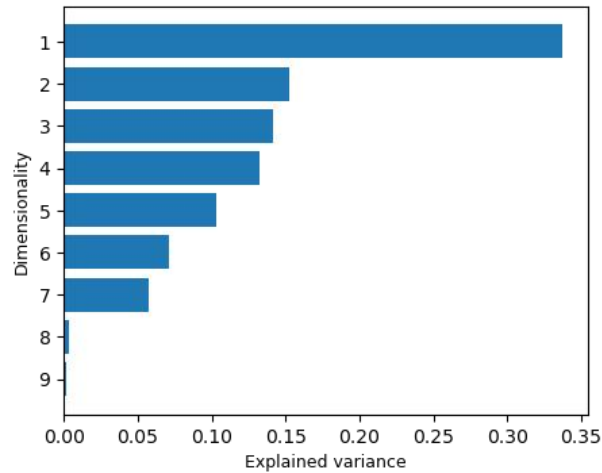


Figure 13. The principal component contribution rate

The two-dimensional data after dimensionality reduction are input into the detection model of LOF to achieve successful visualization of the envelope surface. The classification effect of the high-temperature data model is shown in the figure. The normal products are inside the cochromatic envelope, and the abnormal products are outside the red envelope.

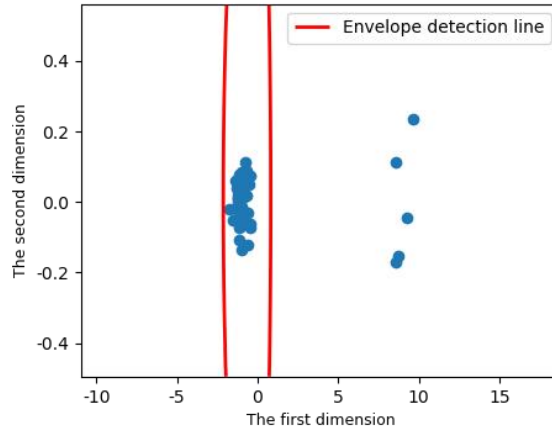


Figure 14. High temperature data envelope detection

Similarly, the predicted low temperature data can also be extracted, and then abnormality detection is carried out by LOF algorithm after dimensionality reduction. The envelope detection results are shown in Figure 15.

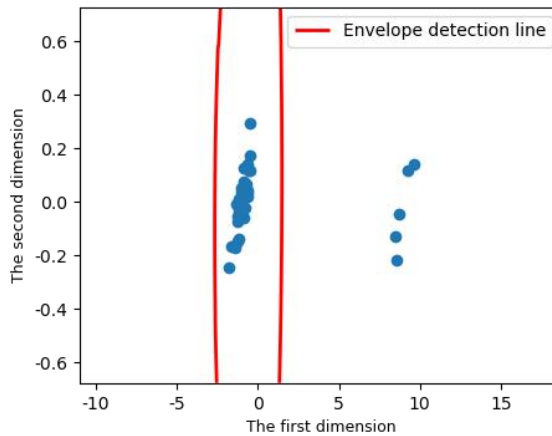


Figure 15. Low temperature data envelope detection

Combined with the high temperature and low temperature envelope detection maps, five abnormal points can be found. According to the product numbers, the abnormal product numbers are 45-50 respectively, and their detection data are quite different from other normal thruster products. According to the research of this experiment, the method can effectively detect abnormal products.

At the same time, in order to increase the accuracy, using the sampling method, 10 new data set, the 10 is the data set as the input of 10 local outlier factor model for training, get 10 different learning device, use the LOF - bagging algorithm and LOF algorithm for multiple tests, the final result shown in the table 2 below,

Table 2. Comparative Results

	Average accuracy (P)	Average recall (R)
LOF-bagging	95.8%	94.7%
LOF	92.5%	91.3%

The experiment shows that the algorithm proposed in this paper has higher prediction accuracy than the ordinary anomaly detection algorithm.

5. CONCLUSION

In order to solve the problem of abnormal detection of traditional power equipment, this paper proposes a method of predicting and checking power equipment test data based on LSTM-LOF model. According to the previous normal temperature data, the LSTM model predicts the high and low temperature data in the actual working environment. PCA is used for data dimensionality reduction extraction, and finally LOF algorithm is used for data detection. This algorithm solves the problem of predicting power equipment components by relying on experience knowledge, and can assist engineers to detect anomalies. Meanwhile, this algorithm has higher prediction accuracy than ordinary anomaly detection algorithms.

Subsequently, with the increase of various data of power equipment, more base learners can be used for model training. Meanwhile, the integration strategy can be adjusted to weighted average, and the weight coefficients of different base learners can be dynamically adjusted in the process of model training, so as to obtain better anomaly detection results.

REFERENCES

- [1] Wang, Y., Limmer, S., Olhofer, M., Emmerich, M. T. M., Back, T., Vehicle fleet maintenance scheduling optimization by multi-objective evolutionary algorithms. In 2019 IEEE congress on evolutionary computation, 442-449 (2019).
- [2] Ordóñez, C., Lasheras, F. S., Roca-Pardiñas, J., de Cos Juez, F. J., A hybrid ARIMASVM model for the study of the remaining useful life of aircraft engines. *Journal of Computational and Applied Mathematics*, 346, 184–191 (2019).
- [3] Wu, H. Y., Miao, W. W., Guo, B., et al. Research on State Prediction Algorithm of Power Communication Equipment Based on Improved Decision Tree[J]. *Computer and Digital Engineering*, 49(1):17-20,74 (2021).
- [4] Gao, Y. X., Sun, S. Z., Application of wavelet denoising combined with ARMA model in the prediction of failure rate of power equipment[J]. *Journal of Inner Mongolia University of Technology (Natural Science Edition)*, 38(2):122-128 (2019).
- [5] Liu, M. W., Application analysis of improved association rule method in power equipment fault prediction[J]. *Shandong Industrial Technology*, (3):175 (2018).
- [6] Wu, H. Y., Chen, P., Guo, B., et al., State prediction of power communication equipment based on attention mechanism and LSTM[J]. *Computers and modernization*, (10):12-16 (2020).
- [7] Wang, C. B., Chen, G., Zhou, R., et al. Fault prediction method of power equipment based on survival analysis model[J]. *Power Big Data*, 23(5):1-8 (2020).
- [8] Chen, H. J., Zhang, P., Jia, Y. F., et al., Research on operating temperature prediction method of power equipment based on BP neural network[J]. *Electronic world*, (10):40-41 (2018).
- [9] Yang, J. H., Liu, Y., Liu, J., et al., Parallel F-LSTM Model and Its Application in Power Communication Equipment Fault Prediction [J]. *Journal of Wuhan University (Science Edition)*, (3):263-268 (2019).
- [10] Zhang, Y. F., Meng, F. Y., Wang, Y. Q., et al., Research on temperature prediction method of power equipment based on improved LSTM[J]. *Journal of Electronic Measurement and Instrumentation*, 35(12):167-173 (2021).
- [11] Hochreiter, S., Schmidhuber, J., Flat minima.[J]. *Neural Computation* , 9(1) :1-42 (1997).