Anomaly Detection Algorithm Based On Electric Equipment

Yu Sun¹, Gangfeng Yan¹, Xiasheng Shi¹
College of Electrical Engineering, Zhejiang University, Hangzhou, Zhejiang, 310027, China
*Corresponding author’s e-mail: 21760554@zju.edu.cn

Abstract. Traditional methods for detecting data anomalies of power equipment fail to fully mine the data characteristics. There are shortcomings such as complex calculation, poor flexibility and low accuracy. To solve the problem of abnormal fault detection of electric equipment, the abnormal detection algorithm based on statistical analysis and machine learning is used in the paper. The reconstruction method based on Long Short Term Memory (LSTM) time sequence is proposed for time series data abnormal detection. Experiments show that the new method is effective in detecting and correcting abnormal data, which reduces the detection time and improves the accuracy of state estimation results. There are huge differences within the positive samples in anomaly detection, so several methods are studied in the paper. One-class SVM algorithm is often used in novelty detection and isolation forest algorithm is used in outlier detection. Machine learning methods which is based on statistical analysis will play an increasingly important role in the fields of equipment monitoring and predictive maintenance, safety of electric equipment.

1. Introduction
With the development of the informatization and intellectualization of the power grid, the data volume of power industrial control equipment in the power grid becomes larger and larger, which brings great difficulties to data processing and analysis. In the processing of the application of smart grid, big data, data storage, efficient processing and real-time visualization of multi-source heterogeneous data integration and data aspect are facing serious challenges. It’s necessary to do further research on these aspects, and develop large data in ensuring safe and stable operation of electric equipment in the power grid. The influence of these abnormal data on the state estimation of modern power system cannot be ignored[1]. The traditional data anomaly detection method of power industrial control equipment fails to fully mine the data characteristics. It has the disadvantages of complex calculation, poor flexibility and low accuracy, which cannot meet the requirements of prediction speed and accuracy. The abnormal monitoring method for traditional data causes great difficulties, therefore it is necessary to break through the original traditional detection algorithm. It is of great significance to study the anomaly detection algorithm based on the time sequence data of power industrial control equipment by using machine learning method[2].

The rest of this paper is organized as follows. In Section 2, the anomaly detection problem analysis is introduced. It has unbalanced data and there are huge differences within the positive samples, so it a difficult to solve the problem. In Section 3, several anomaly detection algorithms are proposed. And the
proposed algorithm is verified by experiments is Section 4. Finally, the work about anomaly detection in electric equipment is concluded and provide future topics in Section 5.

2. Anomaly Detection

Compared with the classification problem of general supervised learning, the learning problem of anomaly detection lacks marked supervised data. Abnormal and normal courses are unbalanced (the balanced data is generally at least 1:5), and the data is autocorrelated. A data point depends on the early data points, which destroys the time series data. Generally the anomaly detection of training data, most of the exceptions are the latter to 1:10^6 level imbalance of events. And most of them come from electric equipment sensors.

All the data are autocorrelation, which makes it extremely difficult to use the traditional supervised learning classification method to solve the problem of anomaly detection[3].

In the problem of anomaly detection with a large amount of training data, the problem of class imbalance can be solved through the set constructed by multiple resampling data. You first create a new data set by taking all of the abnormal data points and adding a subset of the normal data points, for example, as four times the number of the abnormal data points. Then you use SVM or random forest to build classifiers for each data set and use ensemble learning to combine these classifiers. According to relevant data, this method works well and produces good results.

Simple classifiers will not work when data points are autocorrelated with each other. In this case, time series classification technique or recursive neural network processing is used to build the predictor and the residual is used to find outliers. The uninterpreted values of the model are the outliers. The next value can be predicted by building a regression, time series model or recursive neural network model. If the percentage of error and actual value is lower than a certain threshold[4], the class will be abnormal. Finally, abnormal detection of time series can be realized by learning abnormal alarm threshold.

For time sequence data, LSTM is a recursive model. It has been proved to has significant advantages on long-term learning and memory sequence information. LSTM is currently widely used in sequence learning tasks[5]. At the last of paper, one electrical device anomaly detection algorithm based on LSTM time series prediction is proposed.

3. Algorithm for different positive samples

3.1 One Class SVM

As a classification problem, there is only one type of sample, or there are two types of samples. But one of the types of sample size is far less than the other types of sample size (if using two classifiers, uneven training set of positive and negative samples, may cause large number classifier is too biased samples category, make the train out of the model has a bias). At this situation it is considered to use one-class SVM classification.

One-class SVM has the ability to capture the shape of the data set. It has a better effect on strongly non-Gaussian data, such as two distinct data sets. Strictly speaking, a classified SVM is not an outlier detection algorithm, but a singularity detection algorithm: its training set cannot contain abnormal samples. Otherwise, it may affect the selection of boundary during training. However, for a sample data set in a high-dimensional space, one-class SVM can be a powerful tool if it fails to make assumptions about distribution characteristics. One-class SVM is often used in novelty detection.

3.2 Isolation Forest

Isolated forest is an effective outlier monitoring algorithm. When the algorithm is used for detection, a feature will be randomly selected, and then a slice will be randomly selected at the maximum and minimum values of the selected feature. In this algorithm, the training of the whole training set is like a tree. The number of partitions is equal to the path distance D from the root node to the leaf node. The average value of D of all random trees (many trees are randomly selected to form a forest in order to enhance the robustness) is the final result of the detection function. The path D is smaller because it’s
further away from the main sample point distribution center. You can find outliers by looking for shortest path leaf nodes.

3.3 LSTM framework

LSTM network is a recursive model which is capable of learning and memorizing long-term sequence information. It’s composed of the storage module shown in figure 1.

The core of LSTM consists of three memory cells: input, forget and output, which can encode the input information at each moment. The behavior of each memory cell is controlled by gate, which controls whether the information is saved or not. It is 1 if it is saved, otherwise it is 0. In detail, forget gate f controls whether the cell information in the current state is saved, input gate I controls whether the input information is read, and output gate o controls whether the new cell information is output. LSTM training is robust and can be avoided gradient dispersion by means of multi-gate cooperation. LSTM network is used to represent and reconstruct the production data described by the time series in this project, and then abnormal detection is carried out[6].

As is shown in figure 1, LSTM cells refer to the same unit. The LSTM unit includes the forget gate $f_i$, input gate $i_i$, output $o_i$, unit status $c_i$, input $x_i$, output $h_i$. Time t indicates the preceding moment and t-1 indicates the previous moment. Forgetting gate determines how much cell state is retained to the current moment. The input gate determines how much input is retained to the cell state at the current time by the forgetting gate and the input gate unit state $c_i$. Output gate determines how much the unit state is output to the current moment.

The basic flow of the algorithm is to firstly represent the input time series with a given length of n, $X = \{x_1, x_2, \ldots, x_{n-1}, x_n\}$, and then use the underlying LSTM-1 to carry out vector representation of the
learning data. Then, the end state variable of LSTM-1 is taken as the initial input of LSTM-2, LSTM-3 for time series reconstruction. Finally, the error variance of the calculated reconstructed sequence and the original sequence is substituted into the Gaussian distribution estimated by the maximum likelihood estimation method (MLE), and the probability of abnormal points is finally obtained. It is judged to be abnormal when the probability is greater than the threshold value obtained by the training[7].

3.3.1 LSTM sequence reconstruction
In the above algorithm, it is necessary to train the LSTM-1, LSTM-2, and LSTM-3 networks including the parameters of the sequence reconstruction layer. The LSTM-1 network learns the vector representation of the input at each moment $t_i$. LSTM-2 and LSTM-3 network implicit the sequence of last moment (or input sequence) and the hidden layer output as the reconstruction of the sequence. And the output is sequenced as input to the next layer of network. As shown in Figure 2, given the input sequence $X = \{x_1, x_2, ..., x_{n-1}, x_n\}$, $h_{mi}$ is the hidden layer output of LSTM-1 at time $t_i$, $n$ is dimension of output vector in the network. At the moment $t_i$, sequence $x_i$ and the last moment output $h_{mi-1}$ is input into the LSTM cell. So the vector representation of sequence $x_i$, $h_{mi}[8]$ can be obtained.

When LSTM-1 and LSTM-2 are jointly trained, the sequence is reconstructed in reverse order, that is, the end output of LSTM-1 is used as the initialization input of LSTM-2, so that most of the information of the sequence can be used for reconstruction. The sequence is reconstructed in reverse order, $\{x_n, x_{n-1}, ..., x_2, x_1\}$.

The $h_{si}$ in the LSTM-2 network represents the output of LSTM-2 at time $t_i$, and $h_{si}$ in the LSTM-3 network represents the output of LSTM-3. The represents of the reconstruction sequence of the moment $t_i$ is $X_{ri}$. The computing of $X_{ri}$ is as follows:

$$X_{ri} = W h_{ri} + b$$

$W, b$ is the parameters to be trained.

Because of differences in training and test objectives, LSTM-2, LSTM-3 uses different inputs in both processes. During the training, at time $t_i$, the LSTM-3 cells receive the output $X_i$ and $h_{si}$ as inputs at the previous time[9]. During the test, at time $t_i$, LSTM-3 receives the last time outputs $h_{si}$ and $X_{ri}$ as inputs. The model training goal is to minimize the loss function as follows:

$$L = \sum_{X \in R} \sum_{i=1}^{\infty} e_i^2, e_i^2 = \|x_i - x_{ri}\|^2$$

$e_i$ is the reconstruction error, $R$ is all the sequence set in training set.

3.3.2 LSTM anomaly detection
The anomaly detection is divided into two steps. The first step assumes that the time series anomaly score is Gaussian with the reconstruction error. The MLE algorithm is used to estimate the parameters of the Gaussian distribution and the probability of the time series anomaly is calculated. The second step is to determine the threshold of the training anomaly. And last the sequence abnormal state can be determined.

Generally speaking, when the abnormal score $S_i$ is bigger than a certain threshold $S_a$, it is considered that the point is abnormal, otherwise it is normal. Since the setting of threshold $S_a$ has a great influence on the anomaly detection, the $F_β$ operator is used to estimate it. The $F_β$ operator is a common indicator for measuring classification results in machine learning. In order to make the estimation more accurate, it is necessary to have enough normal and abnormal samples to estimate[10].

The objective function of this parameter estimate is set to:
\[ F_p = (1 + \beta^2) PR / (\beta^2 P + R) \]
\[ P = T_p / (T_p + F_p) \]
\[ R = T_p / (T_p + F_N) \]

In the formula, P is the accuracy rate. It is the number of samples predicted to be positive and predictively correct \( T_p \) which is proportional to the total number of samples \( T_p + F_p \) predicted to be positive. The definition of "abnormal" is positive and "normal" is negative. R is The recall rate indicates the proportion of the number of samples that are predicted to be positive and predictively correct, and the total number of samples that are truly positive. In general parameter \( \beta < 1 \), the proportion of abnormal samples in the total sample is small[11].

4. Experimental Result

4.1 Experiment data
In this paper, Numenta Anomaly Benchmark (NAB) data is used for algorithm verification. In the simulation experiment, the LSTM-1 hidden layer dimension is set to 8, LSTM-2 hidden layer dimension to 32, LSTM-3 hidden layer dimension to 15. The batch size is 100, 50, 210. The network parameters is trained by using the Adam optimizer[12].

4.2 One Class SVM
One class SVM is used for collective anomalies (unordered). It’s good for novelty detection (no anomalies in the train set). The algorithm performs well for multimodal data.

In the experiment, it should take useful features and standardize them, then one class SVM model is trained and the data is added to the main model. Visualization of anomaly throughout time is as follows:

![Figure 3. one-class SVM anomaly detection](image)

The red dot in the figure is the identified anomaly data. As is shown, one-class SVM performs well for multimodal data.

4.3 Isolation Forest anomaly detection
Firstly, it takes useful features and standardizes them, then it needs to train Isolation Forest anomaly detection model and add the data to the main model. Visualization of anomaly throughout time is as follows.
Figure 4. Isolated forest anomaly detection

The red dot shown in the figure is the identified anomaly data. Isolated forest algorithm works well with different data repartition and efficient with high dimension data.

4.4 LSTM

The process of using LSTM algorithm for test scheme is as follows:

- Read the original data, construct the elastic data set, extract the features, and finally export the features to Dataframe. In the original data set, in order to train the deep learning model, each second of statistical data is extracted as characteristic data.

- Further process these feature data in the elastic distributed data set, including wavelet denoising, normalize and sliding average processing of numerical values. For example, expand the feature data sequence based on 50 seconds, so that the deep learning model can predict the next data point through the mode of the first 50 seconds.

- Use the Keras API provided in Tensorflow to create the time series anomaly detection model, including the three LSTM layers as shown in the figure and a full connection layer, and train the model with data, such as using the first 50 points to train the next point.

- The next step is model evaluation: use test data or all data to detect anomalies, and abnormal data refers to data points far away from the prediction of LSTM model. It is assumed that the specified abnormal data is 10% of the overall data set. That is, the 10% data farthest from the predicted value of the model is abnormal data. The screening ratio is set as an adjustable parameter, which can be adjusted according to the actual situation[13].

Figure 5. abnormal detection and processing flow of timing data

In the experiment, keras API in tensorflow is used. Firstly, it should select and standardize data, then set important parameters and train/test size. Next it need to train LSTM network model and add the data to the main model. LSTM learn to recognize sequence in the data and then make prediction based on the previous sequence. An anomaly is considered when the next data points are distant from LSTM prediction. Aggregation, size of sequence and size of prediction for anomaly are important parameters to have relevant detection.

Here it should learn from 50 previous values, and then predict just the 1 next value. Next it need to
create the list of difference between prediction and test data, then plot the prediction and the reality for the test data[14].

Figure 6. LSTM prediction and the reality

Figure 7. LSTM anomaly detection

Figure 6 shows a comparison of raw data and LSTM model prediction data. The red dot shown in the figure is the identified anomaly data, the orange line is the predicted value of the LSTM model, and the blue line is the original value. As shown in figure 7, the trained model finally predicts the failure of the device, and some fluctuations in the early time series can be used as warning information for device failure.

5. Conclusion

By leveraging unsupervised deep learning and the end-to-end processing flow provided in this article, the anomaly can be detected effectively in data sets. In this paper, the corresponding algorithms are proposed for outlier detection, singular detection and time series anomaly detection. Experiments in the data set also demonstrate the effectiveness of the method. Next the anomaly detection algorithm need to be improved so that it’s more accurate and robust.

References

[1] WEI Z. (2003) The Application of ISODATA in the Bad Data Identification of the State Estimate of the Distribution Power System. Journal of Hohai University, 24(2):97-100.
[2] Hawkins D M. (1980) Identification of outliers. Chapman and Hall, London.
[3] Page E S. (1954) Continuous inspection schemes. Biometrika, 41(1/2): 100-115.
[4] Ma J, Perkins S. (2003) Time-series novelty detection using one-class support vector machines Neural Networks. In: Proceedings of the International Joint Conference on, pp. 1741-1745.
[5] Salehfar H, Zhao R.A (1995) Neutral Network Pre-estimation Filter for Bad Data Detection and Identification in Power System State Estimation. Electric Power System Research, 34:127-134.
[6] Ma X, Hovy E. (2016) End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1064-1074.
[7] Bock R D, Aitkin M. (1981) Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46(4): 443-459.
[8] Malhotra P, Vig L, Shroff G, et al. (2015) long short term memory networks for anomaly detection in time series. In: Proceedings of Presses universitaires de Louvain, pp. 89.
[9] Goutte C, Gaussier E. (2005) A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In European Conference on Information Retrieval. Springer, Berlin, Heidelberg, pp. 345-359.
[10] Le Q V, Ngiam J, Coates A, et al. (2011) On optimization methods for deep learning. In: Proceedings of the 28th International Conference on International Conference on Machine Learning. Omnipress, pp. 265-272.
[11] Cho K, van Merrienboer B, Bahdanau D, et al.(2014) On the properties of neural machine translation: Encode-Decoder approaches. In Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, pp.103-111.

[12] Cho K, Van M B, Gulcehre C, et al.(2014) Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724-1734.

[13] Cybenko G.(1989) Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems, 2(4): 303-314.

[14] Graves A.(2012) Long short-term memory. Supervised sequence labelling with recurrent neural networks. Springer, Berlin, Heidelberg.