

Adaptive Momentum Neural Learning for Anomaly Detection of Power Distribution and Consumption Load Data

Yan Liu

North China Electric Power University, Beijing, 102206

alanncepu@foxmail.com

Keywords: Anomaly detection; Adaptive momentum; Neural network; Power distribution and consumption

Abstract. In order to overcome the high misstatement rate and false positive rate brought by the traditional anomaly detection technologies on power distribution and consumption load data, in this paper, a novel momentum adaptive neural learning-based anomaly detection method for power distribution and consumption load data is proposed by adding momentum adaptive mechanism into the error back propagation (BP) model. In the proposed algorithm, according to the gradient changes of experience generalization error in the anomaly detection process, the momentum factor is adjusted in the BP algorithm to avoid the training results of neural network learning converges to a local minimum value. The contrastive simulation experiments on real-time database of power distribution and consumption load data demonstrate that, our proposed method holds better performances on accuracy, misstatement rate and false positive rate, compared with other traditional algorithms.

Introduction

With the continuous development and deepening of power system information, each smart grid automation platform has produced a large number of diverse power load data. However, during the actual production run of network automation platform, the individual may occur with consumption data measured large errors, to follow-up analysis and prediction of the distribution network to bring great interference, safe and stable operation of the power grid. Therefore, accurate detection of abnormal points in the distribution of electrical load data has become the key to the reliable operation of the power system and the further improvement of the grid intelligence level.

In the 1980s, Hawkins proposed the concept of anomalous points and believed that the root cause of the anomaly phenomenon was due to the variability of the distribution of intrinsic indicators [1]. The literature [2] uses the fuzzy C-Means clustering method (FCM) to determine the existence of abnormal points. This method is effective for the identification of a single abnormal point, but it is easy to produce false alarms for the existence of large consecutive abnormal points. In literature [3], a fuzzy clustering method based on genetic algorithm is proposed, but the time performance is not good, and the adjustment of abnormal data is rough. In literature [4], the neural network algorithm is used to mine the data outliers. However, the BP algorithm has low learning efficiency and is easy to converge to the local minimum, which leads to higher false positive rate in the abnormal point detection process.

In order to overcome the phenomenon of false error rate and false positive rate in the traditional anomaly detection dealing with the electric load data, this paper adds the momentum item adaptive mechanism in the Error Back Propagation (BP) model and proposes an anomaly detection based on adaptive momentum neural learning. The method realizes the adaptive adjustment of the momentum term in the BP algorithm by calculating the gradient of the empirical generalization error function of the neural network model during the abnormal point detection process, and avoids the neural network training learning converges to the shallow local minimum value. Experiments show that the method is effective and has important reference value for the development of data analysis in distribution network.

Improvement of Neural Network Model for Adaptive Momentum Term

Overview of Error Back Propagation Network Theory. The Error Back Propagation network is a multiple layer feedforward neural network (MLFNN), which can establish a mapping relationship by learning a large number of samples of the input-output mode without the mathematical Eq.1 describing this mapping relationship. The BP neural network adopts the learning rule of the steepest descent method. It continuously adjusts the internal and external weights and thresholds of the network through the back propagation of the empirical generalization error, so that the neural network model minimizes the sum of the squared errors of the approximation of the data samples [5].

Generally speaking, BP neural network generally selects a three-layer neural network structure, namely Single Hidden Layer Neural Network (SLNN). The specific structure of the network includes: input layer, hidden layer, and output layer.

Suppose that the sample input sequence for a training sample for any N neural network models is $\{X_1, X_2, \dots, X_N\} \in R^n$, There is a corresponding sample output sequence $\{Y_1, Y_2, \dots, Y_N\} \in R^m$ corresponding to it, where the sample output satisfies $Y_s = (Y_s^1, Y_s^2, \dots, Y_s^m) \in R^m$. If the actual sample output sequence generated after the N samples are input to the neural network model is $\{O_1, O_2, \dots, O_N\} \in R^m$, the empirical generalization error of the s th sample is the sum of the squares of the error of each output unit, that is

$$E_s = \sum_{l=1}^m (O_s^l - Y_s^l)^2 / 2 \quad (1)$$

Among them, m is the number of output units, i.e., the sample output dimension. For all training samples, the total error of experience generalization is

$$E_T = \frac{1}{2} \sum_{s=1}^N \sum_{l=1}^m (T_s^l - Y_s^l)^2 \quad (2)$$

Let W be a connection weight in the network, according to the gradient descent method, the weight correction amount of the connection weight is

$$\Delta W = -\frac{\partial E_T}{\partial W} \quad (3)$$

Improved BP neural learning with adaptive momentum term. Aiming at the low efficiency of traditional BP algorithm and the tendency to converge to local minimum, this paper proposes an improved BP neural network algorithm (Momentum Adaptive BP Neural Learning Algorithm, MABP). Unlike the quasi-Newton method, the improved algorithm avoids the calculation of the Hessian matrix. When the function form evaluated is represented by a sum of squares, then the gradient can be expressed as

$$G = J^T e \quad (4)$$

Then the Hessian matrix can be approximated as

$$H = J^T J \quad (5)$$

Among them, e represents the vector of the network error; J represents the Jacobian matrix, which contains the first derivative of the network error with respect to the weight and the deviation, can be directly calculated by standard backpropagation methods. By using the approximate matrix of the Hessian matrix, the weight adjustment rate of the improved algorithm is

$$\Delta W = [J^T J + \mu I]^{-1} J^T e \quad (6)$$

When μ is relatively large, Eq.6 becomes a gradient descent method with a relatively small step size; when $\mu=0$, Eq.6 represents the Newton method using an approximate Hessian matrix. By adaptively adjusting the parameter μ , the network performance of the improved algorithm is excellent, and each step of the operation will make the evaluation function develop in a good direction [6-7]. Specifically, the improved BP algorithm learning process is as follows:

Step 1: Define the sum of the squares of the difference between the actual output value of the output unit and the expected output value as the error function r :

$$r = \frac{1}{2} \sum_j (V_j^m - y_j)^2 \tag{7}$$

Among them, V_j^m represents the actual output and y_j represents the desired output.

Step 2: The adjustment weight can be expressed as:

$$\Delta W_{ij} = -\varepsilon \cdot d_j^k \cdot V_j^{k-1} \tag{8}$$

Among them, ε represents the learning step and takes positive parameters. V_j^{k-1} represents the output of the upper layer of neurons. The value of d_j^k has two cases:

(1) If j is the m th output neuron, then

$$d_j^m = V_j^m (1 - V_j^m) (V_j^m - y_j) \tag{9}$$

(2) If j is the hidden layer of neurons, then

$$d_j^k = V_j^k (1 - V_j^k) \sum_l W_{jl} d_l^{k+1} \tag{10}$$

Step 3: Repeat the operation until all inputs are completed.

In order to suppress the problem of excessive calculation and slow convergence, the improved algorithm adopts the method of adding momentum items. The model is as follows:

$$\Delta W_{ij}(t+1) = -\varepsilon d_j^k V_j^{k-1} + \alpha \Delta W_{ij}(t) \tag{11}$$

Among them, α represents the momentum factor, which obviously can achieve the improvement of convergence [8-9].

MABP-based algorithm for abnormal point detection of electric load data

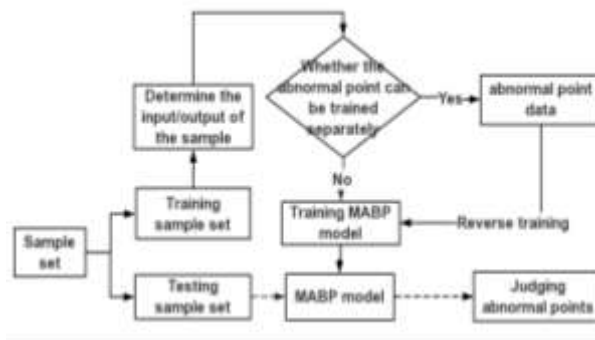


Figure 1. Schematic diagram of the abnormal point detection algorithm

As shown in Fig. 1, the sample set is first divided into a training set and a test set, and the input and output indicators of the sample are determined. Then, for the training samples, if the abnormal points have been marked and the individual points can be trained separately, the penalty points are used to inversely train the abnormal points of the electric load data; otherwise, the use of the

abnormal points is directly used. The electrical load data is trained using the MABP algorithm. Finally, the MABP neural network model obtained by the training is used to predict the sample in the test sample set point by point, and determine whether the absolute value of the difference from the true value has exceeded the threshold, and if it is exceeded, it is an abnormal point. The specific experimental operation flow is shown in Section 4.2.

Numerical simulation experiment and result analysis

Identification problem simulation on NCI dataset. All the tests in this paper are carried out under the Matlab 2012 platform. The hardware device parameters are: CPU Intel® Core™ i7 2.30GHz, RAM 8.00GB, Win 8 environment. In this paper, 350 experiments in the NCI1-09 data set are used to conduct experiments. 250 randomly selected training sample sets, and the remaining 100 are test sample sets. The traditional BP algorithm is used as the control group to verify the recognition accuracy of the MABP algorithm proposed above. Among them, the training target (Goal) in the BP algorithm and the MABP algorithm is set to 0.001, the training iteration step threshold (Epochs) is set to 10000, the input layer uses 23 neurons, and the output layer uses 2 neurons. The learning rate (Lr) of the revised weight is 0.45, and the number of hidden layer neurons is set to 12. The results of the comparative test are shown in Table 1 below.

Table 1 Experimental comparison of BP and MABP algorithms on NCI1 dataset

Algorithm	Positive recognition rate	Negative recognition rate	Iteration step
BP	87.35%	79.32%	8454
MABP	91.77%	92.56%	3524

The data shown in Table 1 above shows that the MABP algorithm overcomes the defect that the traditional BP algorithm has low learning efficiency and is easy to converge to the local minimum. In the comparison experiment on the NCI1-09 dataset, the MABP algorithm is superior to the traditional BP algorithm in both the positive recognition rate and the negative recognition rate, indicating that the MABP algorithm can avoid convergence to some local minimum points, thus improving the recognition accuracy. In terms of iteration steps, the iterative convergence step of MABP algorithm is significantly lower than the traditional BP algorithm, which shows that the learning efficiency of MABP algorithm is improved.

Simulation of Outlier Detection Problem on Short-term Load Data Set. Then, the experiment randomly selected 1000 short-term load samples containing only a single abnormal point in the Hbase database of the Southern Power Grid with the big data platform as the sample data. A random selection of 600 of them is a training sample set, and the remaining 400 are test sample sets. Then, 400 short-term load samples containing only a large number of consecutive abnormal points (15-22 sampling points) in the Hbase database of China Southern Power Grid Power Data Platform were extracted as sample data. A random selection of 350 is a training sample set, and the remaining 50 are test sample sets. Assume that only the first 15 sampling points are related to the next sampling point. The traditional FCM algorithm and BP algorithm are used as the control group to verify the performance of the proposed MABP algorithm, as shown in Table 2. Among them, the training target (Goal) in the BP algorithm and the MABP algorithm is set to 0.05, the training iteration step threshold (Epochs) is set to 10000, the input layer uses 15 neurons, and the output layer uses 1 neuron to modify the weight. The learning rate (Lr) was 0.3 and the number of hidden layer neurons was set to 25. Further, for the normalized electric load data, the abnormal point threshold is set to 0.05.

Table 2 Comparison of the three algorithms on the Hbase database

Algorithm	Single abnormal point recognition rate	Large continuous abnormal point recognition rate	Number of false positives
FCM	95.50%	46.00%	0/27
BP	93.75%	62.00%	0/19
MABP	96.75%	76.00%	0/7

The data shown in Tab.2 above shows that the MABP algorithm is superior to the traditional FCM algorithm and BP algorithm in the recognition rate of single abnormal points and the recognition rate of large continuous abnormal points; while the FCM algorithm performs better in the detection of single abnormal points., poor performance in large continuous abnormal point tasks. In terms of the number of false positives, the FCM algorithm, the BP algorithm and the MABP algorithm proposed in this paper are all good for a single abnormal point detection task, and there is no false positive accident. In the large continuous abnormal point task, the FCM algorithm, BP algorithm and The MABP algorithm proposed in this paper has 27, 19, and 7 times of misjudgment accidents on normal load sampling points, but the MABP algorithm proposed in this paper can reduce the false positive rate by more than 60%, and the performance is ideal.

Summary

In order to overcome the phenomenon that the false alarm rate and the false alarm rate are high in the traditional abnormal point detection technology, the momentum backpropagation (BP) model is added to the error back Propagation (BP) model. An adaptive point data detection technique based on adaptive momentum term neural learning is used. In the algorithm, the momentum term in the BP algorithm is adaptively adjusted according to the gradient change of the empirical generalized error function in the abnormal point detection process, so as to avoid the neural network training learning converge to the shallow local minimum value. Experiments show that the feasibility and effectiveness of the method provide a new idea for the distribution network data analysis technology.

References

- [1] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey [J]. *ACM computing surveys (CSUR)*, 2009, 41(3): 15.
- [2] Li Peiqiang, Li Xinran, Tang Waiwen et al. Application of Fuzzy C-means Clustering in Power Load Modeling[J]. *Journal of Hunan University (Natural Science)*, 2006, 33(3): 41-45.
- [3] Shen Yuhui, Liu Huawen, Xu Xiaodan et al. Anomaly detection algorithm based on neighborhood dispersion[J]. *Journal of Frontiers of Computer Science and Technology*, 2015. DOI: 10.3778/j.issn.1673-9418.1509075.
- [4] Han S J, Cho S B. Evolutionary neural networks for anomaly detection based on the behavior of a program [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2005, 36(3): 559-570.
- [5] Shi Hongtao, Yang Jingling, Ding Maosheng et al. Short-term wind power prediction method based on wavelet-BP neural network[J]. *Automation of Electric Power Systems*, 2011, 35(16): 44-48.
- [6] Chen Haiping, Zhang Hang, Lu Wei et al. Fast Blind Source Separation Algorithm for Adaptive Momentum Term of Time-Varying Hybrid System[J]. *Computer Science*, 2013, 40(z2): 15-17, 45.

- [7] Ma Lixin, Wang Yuzhu, Wang Yuexiao et al. An Adaptive Harmonic Detection Algorithm Introducing Momentum Term[J]. *Control Engineering*,2013,20(6):1077-1080,1083.
- [8] Wang Hua, Cheng Haiqing. Adaptive Balancing BP Neural Network Blind Equalization Algorithm[J]. *Computer Engineering and Design*, 2010, 31(6): 1297-1300.
- [9] Chen Haiping, Zhang Hang, Lu Wei et al. Fast Blind Source Separation Algorithm for Adaptive Momentum Term of Time-Varying Hybrid System[J]. *Computer Science*, 2013, 40(z2): 15-17, 45.