ISSN: 2414-1895 DOI: 10.6919/ICJE.202007_6(7).0027

Data Correction of the Meter of Power IOT Based on PNN+Bagging and ELM

Kai Chen

College of Logistics Engineering, Shanghai Maritime University, Shanghai 200120, China.

1103342656@qq.com

Abstract

Nowadays, with the popularity of Internet of things and the arrival of big data era, IOT(Internet of Things) data has become the key to affect the accuracy of big data analysis. For the data of power IOT, how to correct the deviation caused by the meter in the process of power IOT (especially the deviation produced by the meter on the low-voltage side of the transformer is not close to the difference between the meter data on the high-voltage side and the standard loss) has become the problem to be solved in this paper. In this paper, a new method based on the combination of PNN+bagging and ELM is proposed to estimate the deviation degree of the data to be corrected, and then a new feature will be constructed, and the corrected data is obtained by ELM regression fitting. The simulation results show that the new method is effective.It is verified that the effect of this method is better than that of the traditional method.

Keywords

IOT; Meter; PNN; ELM.

1. Introduction

Nowadays, with the popularity of the Internet of things and the arrival of the era of big data, IOT data has become the key to affect the accuracy of big data analysis. For the data of power IOT, how to correct the deviation of electricity meter in the process of power IOT is the key problem to be solved. It is of great significance to correct the problem of power data in the process of power IOT meter to solve the problem of power IOT data quality.

When part of the real-time measurement data is eliminated as suspicious bad data, it must be corrected, such as load power data, to ensure the observability of the system, and how to improve the power data correction error has become the main difficulty.

At present, it is commonly used in artificial correction, statistical method correction, intelligent algorithm correction and so on. In reference [1], the author proposes a method to correct the bad data of electric power through longitudinal translation, which is simple but easy to be influenced by the average value of the selected historical data. In reference [2], the author proposes a data correction method based on automatic adjustment strategy, but it is easy to be affected by its set threshold. In reference [3], the author proposes a t-test method based on mathematical statistics to test and correct the bad data, but this method is easily affected by the set confidence interval range. In reference [4], the author puts forward BP neural network to modify the existing power data, but the traditional BP neural network is prone to lack of generalization ability and fall into the local minimum value problem. In reference [5], the author proposes to use the support vector regression method to build the model of identifying and correcting the bad data of electric power, but this method has a great impact on the large sample size, which consumes a lot of computing resources and is affected by the selection of free parameters.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202007_6(7).0027

For the third kind of solution, that is, through intelligent algorithm to correct the bad power data. Because of the shortcomings of single machine learning algorithm and the simpleness of traditional data processing, there is a big deviation between the corrected data and the real data. Therefore, in this paper, combining with the data deviation problem of power meter, we propose to solve the problem of excessive deviation in traditional correction method based on PNN + bagging algorithm (classification) and elm (regression). The method proposed in this paper is to determine the possible deviation degree of the high and low meter data in the test data set through the historical data set by PNN + bagging algorithm, and to build a new test set again by adding the test set, and to get the corrected height by elm regression fitting Count. The advantage of this paper is to solve this kind of problem. In this paper, we deeply mine the information of data set to reduce the deviation between the data to be modified and the real data as much as possible, which is the lack of consideration in the existing work.

The chapter arrangement of this paper is as follows: the first section mainly introduces the background and significance of this paper; the second section mainly describes the problems encountered and explains the motivation of the research; the third section mainly introduces the algorithm design of the data correction of the power meter. The fourth section mainly through the simulation analysis, and compared with the traditional methods, to verify the effectiveness of the improved method. The fifth section is the summary of the whole paper.

2. Problem Description

2.1 Problem Analysis

The power meter data used is based on the power consumption data generated by a transformer, including voltage, current and power data. In the power IOT system, all the energy consumption data are monitored by the power meter. Generally speaking, there are many abnormal situations in the power meter, especially in the monitoring meter at the low voltage side of the transformer, which leads to the difference between the data measured by the high-voltage meter and that measured by the high-voltage meter when considering the loss of the transformer. Generally speaking, for the active power, with the change of load rate, the standard loss of transformer nameplate is generally between 0-3kw. However, compared with the difference of the data monitored by the high and low meters, it is far beyond this range at some time nodes, and there is obvious deviation of the power meter, as shown in the figure below. Therefore, we need to correct such meter data (mainly for low meter).

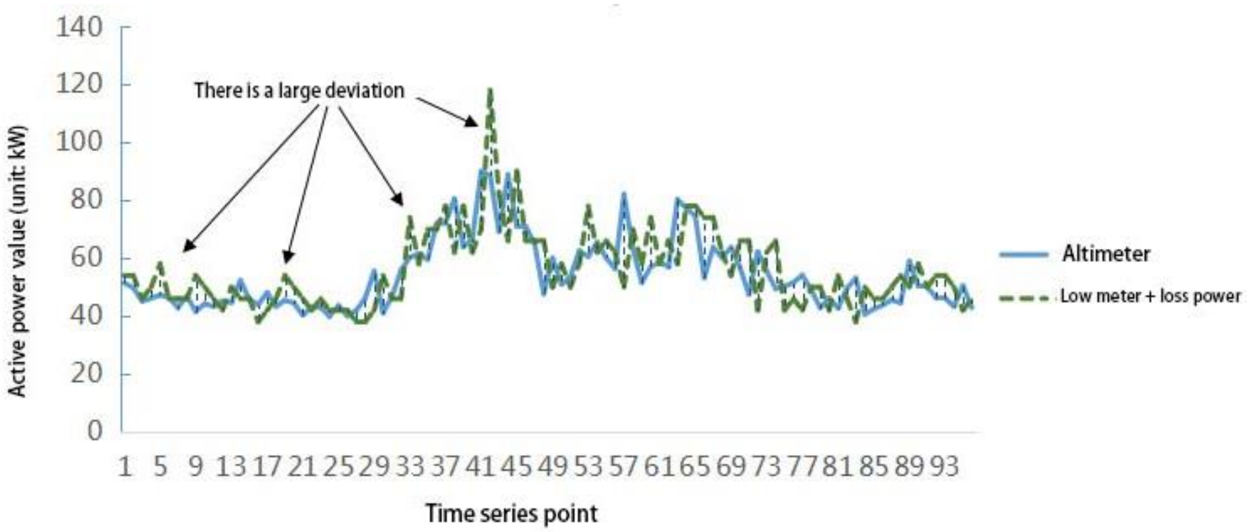

Figure 1. Deviation diagram of real high-voltage meter data and calculated high-voltage meter data

ISSN: 2414-1895 DOI: 10.6919/ICJE.202007_6(7).0027

2.2 Research Motivation

There are many problems in correcting the data of power meter, and many works have been done by predecessors. Generally, the power is the main part of power meter data correction (including active power and reactive power). In some references [4-5], the traditional method is usually used for power data correction, that is, the data to be corrected and the real data in the historical data are used as the training set, and then a single regression algorithm (such as BP neural network and support vector regression) is used) After training, the final result of the data to be corrected in the test set is obtained. Among them, the correction data obtained is the final data needed. The usual data correction processing flow is shown in the figure below.

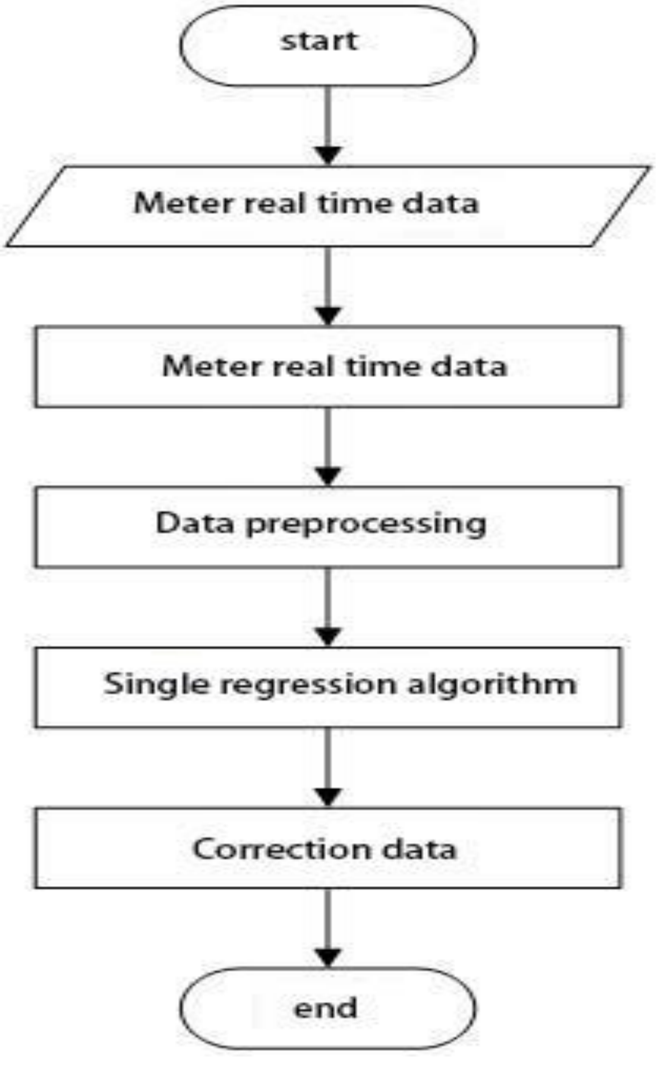

Figure 2. Flow chart of traditional data correction

Because of the simpleness of the traditional data feature processing and the shortcomings of the single machine learning algorithm, there is a big deviation between the modified data and the real data. In this paper, based on PNN + bagging algorithm (classification) and elm (regression) to solve the problem of large deviation in the traditional correction method. The method proposed in this paper is to determine the possible deviation degree of the high and low meter data in the test data set through the historical data set by PNN + bagging algorithm, and to build a new test set again by adding the test set, and to get the corrected height by elm regression fitting Count. The advantage of this paper is to solve this kind of problem. In this paper, we deeply mine the information of data set to reduce the deviation between the data to be modified and the real data as much as possible, which is the lack of consideration in the existing work. The flow chart proposed in this paper is as follows.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202007_6(7).0027

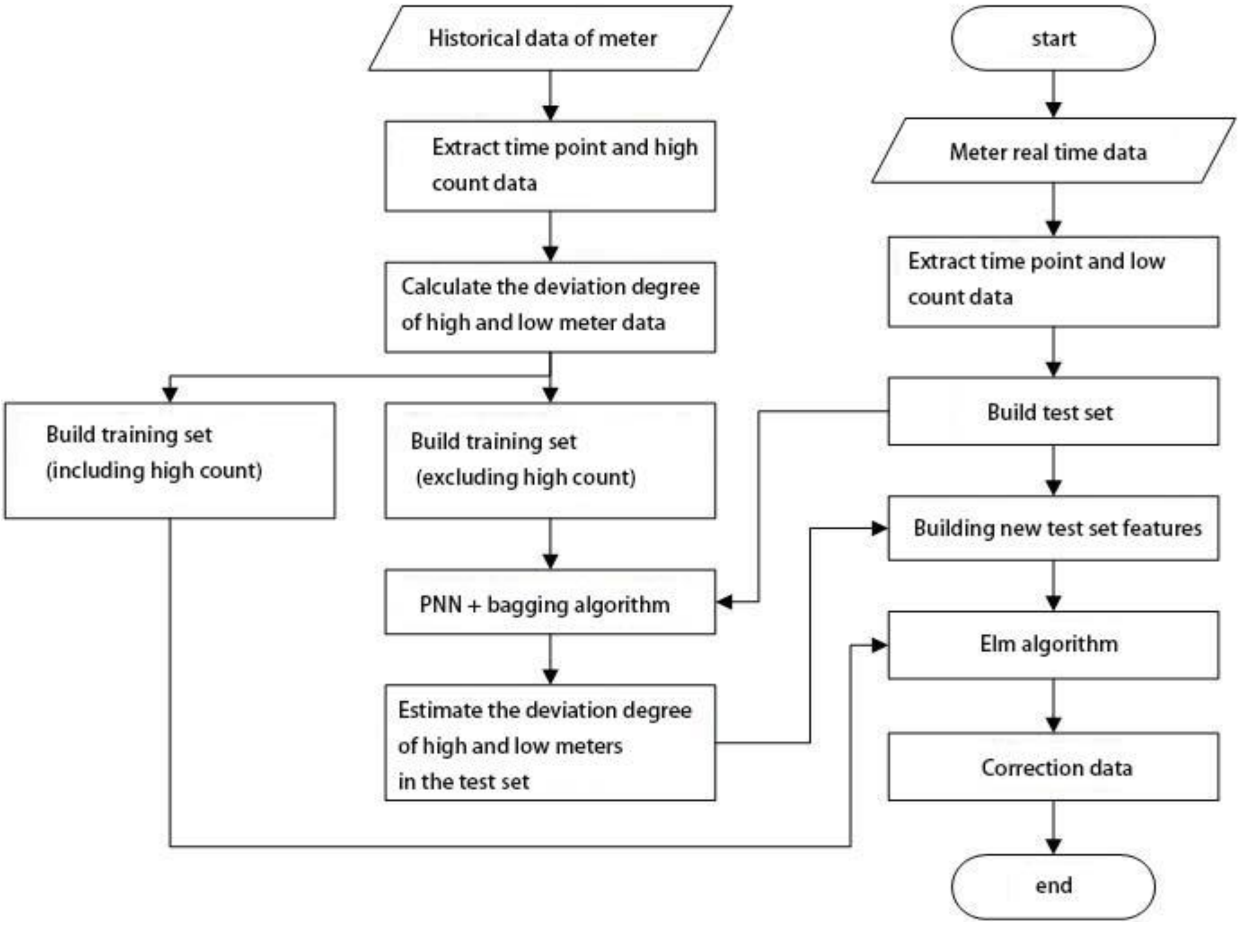

Figure 3. Improved data correction flow chart

3. Meter data correction of power IOT based on PNN + Bagging and ELM algorithm

In view of the data deviation of the low meter of the power meter, here, the power data monitored by the meter is mainly analyzed. In this paper, according to the different difference between high and low meters, it can be roughly divided into three situations. The first is that the power data of the high-voltage side meter is smaller than that of the low-voltage side meter; the second is that the power data of the high-voltage side meter is 0-3kw larger than that of the low-voltage side meter, which belongs to the loss range shown on the normal nameplate; the third is that the power data of the high-voltage side meter is much higher than that of the low-voltage side meter, which is beyond the normal loss range. For these three cases, labels 1, 2 and 3 are respectively assigned here. Then, through the classification algorithm, we calculate the difference between the high count and the low count and set the deviation degree label at different time points in the training set. Then, according to the low count data at different time points in the test set, the deviation degree category corresponding to the low count values at different time points in the test set is predicted.

3.1 Estimation method of deviation degree based on PNN + bagging

3.1.1 PNN and bagging algorithm

PNN (probabilistic neural network) is a neural network based on Bayesian optimization rules. In essence, PNN can get classification results by probability density estimation [6]. Generally speaking, the hidden layer of PNN uses Gaussian function as the probability density output, and then obtains the category with higher probability value as the category of output layer through the comparison of competition layer. The traditional probabilistic neural network is mainly divided into four layers. Its structure is shown in the figure below, from left to right are input layer, mode layer, summation layer (competition layer) and output layer.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202007_6(7).0027

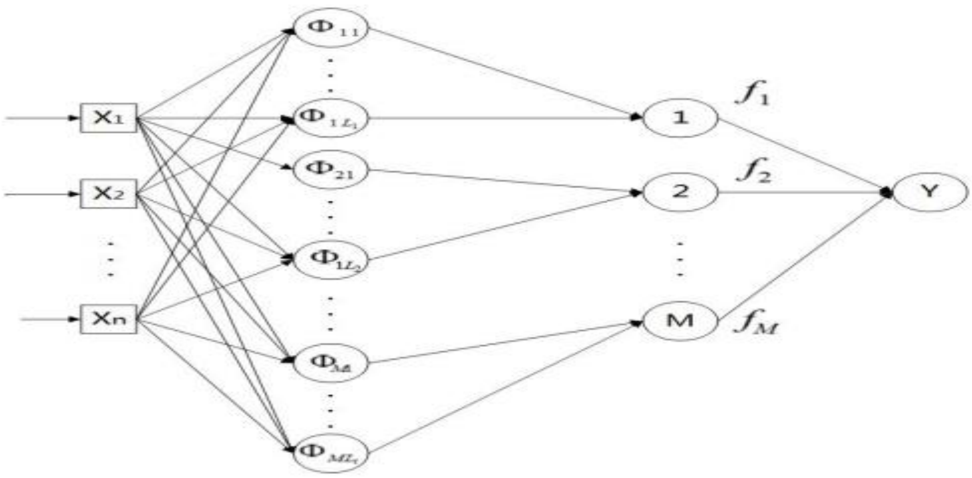

Figure 4. PNN structure

In the figure, the input layer of the first layer is the input vector, which is mainly responsible for importing the input vector into the mode layer; while the mode layer of the second layer is mainly composed of Gaussian radial basis functions, and the number of neurons is equal to the number of samples in the training set, each neuron has a center, and the probability value is calculated with each test set sample, and the output of the probability value is:

$$\Phi_{ij}(x) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma^{d}} e^{-\frac{\|x-x_{ij}\|}{2\sigma}}, i = 1, 2, ..., M; j = 1, 2, ...L_i \tag{1}$$

Among them, D is the sample dimension; it is the number of class I samples in the training set; m is the number of classes; it is the smoothing factor; it is the distance from the input sample to the sample center, generally the Euclidean distance.

In the summation layer (competition layer), the number of neurons taken is the same as the number of categories, and the mode layer

The output of hidden neurons belonging to the same class is weighted average, and the probability density function of class I obtained by kernel function estimation method is as follows:

$$f_j = \frac{\sum_{j=1}^{L_i} \Phi_{ij}(x)}{L_i} \tag{3}$$

In the output layer, the class output with the maximum probability is finally obtained, as shown in the following formula:

$$y = \arg\max(f_j) \tag{4}$$

From the point of view of its working principle and structure, it saves the back propagation time of BP neural network, has the ability of fast learning, and is suitable for dealing with the situation of fast identification of samples. Although PNN has some advantages, it also has the disadvantage that it does not consider the overlapping and interleaving of models in different categories, resulting in the lack of generalization ability [7].

From the point of view of its working principle and structure, it saves the back propagation time of BP neural network, has the ability of fast learning, and is suitable for dealing with the situation of fast identification of samples. However, the output results are easily affected by the expansion speed (smoothing factor) of the radial basis function in the hidden layer.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202007_6(7).0027

Bagging algorithm is one of the integrated learning methods [8], which is suitable for improving the overall effect through bagging method when the output effect of some basic learning devices is not good, so as to achieve the effect of reaching a strong learning device with poor basic learning effect.

3.1.2. Process of deviation degree estimation based on PNN + bagging algorithm

In this section, the main purpose is to mine the historical power meter data provided, and get the estimated high and low meter deviation degree of the data to be tested by PNN + bagging algorithm. The specific process is as follows.

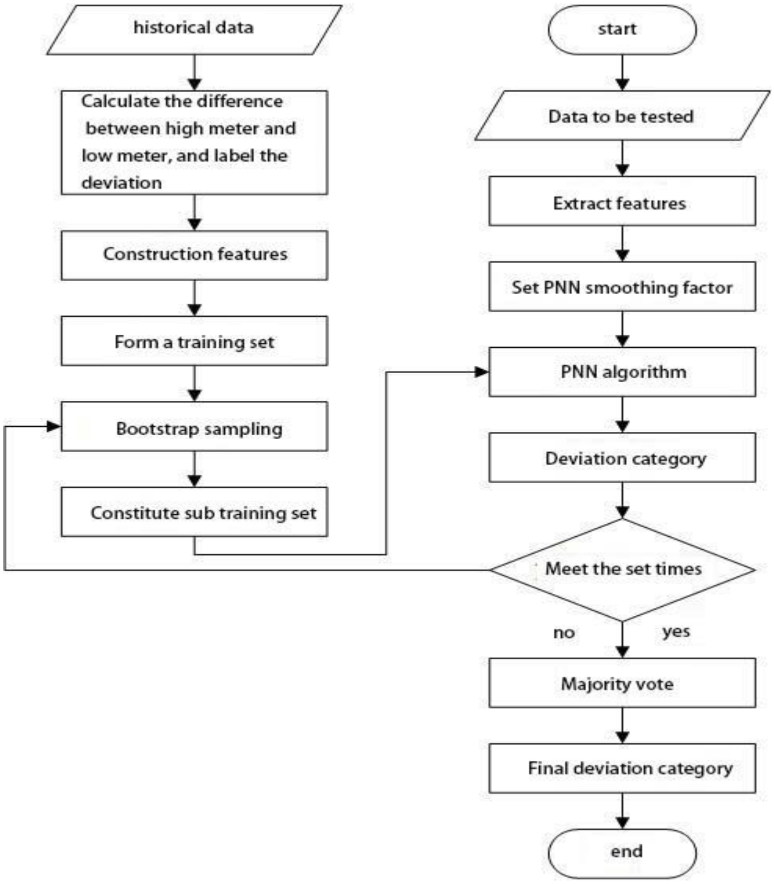

Figure 4 PNN + bagging processing flow chart

Based on PNN + bagging data correction process introduction:

1) The time point, the lower active power data and the corresponding higher value in the historical data of several days are extracted as part of the training set;

2) Calculate the deviation degree of high and low count data in the training set, and label the deviation degree, as shown below.

$$H(i)=\begin{cases}1 & g(i)-d(i)<0\\ 2 & 0\le g(i)-d(i)\le 3\\ 3 & g(i)-d(i)>3\end{cases} \tag{5}$$

Where, H (I) is the deviation degree of the high and low meter data at a certain time; G (I) is the real high meter data at a certain time; D (I) is a low meter data.

3) The new training set features include time point, low count value and deviation degree;

4) Then the PNN smoothing factor is set, and the time point and low count value in the test data set are regarded as the test characteristics;

5) Set the number of learning bases of PNN in the way of bagging integration;

6) Finally, PNN + bagging is used to test the data in the test set.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202007_6(7).0027

3.2 Data correction method based on PNN + bagging + ELM meter

3.2.1 Introduction to elm algorithm [9]

Elm (Limit learning Machine) is a kind of feed-forward neural network. Compared with BP neural network, it has the advantages of simple structure and short operation time. When it is used, it only needs to determine the number of hidden layer neurons and activation function, which can be used as either classification algorithm or regression fitting algorithm.

3.2.2 Data correction process based on PNN + bagging + elm meter

In the traditional power data correction, the historical data is usually directly used for regression correction of test data. In this paper, after PNN + bagging processing, the historical data set is fully mined to form a new data set, and then the modified high count value is obtained by elm regression algorithm.The process is as follows.

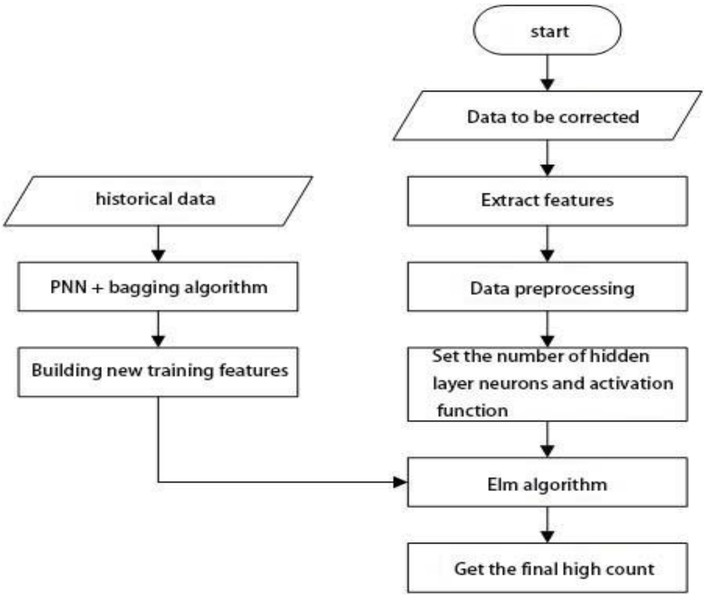

Figure 5. Data correction process based on PNN + Bagging + ELM

4. Simulation analysis

The simulation environment of this paper is matlab2015b, and the data comes from the altimeter data provided by Zhongheng Cloud Energy Company under the IOT platform. In this simulation test, No. 4.1-4.10 data is used as the historical data, 96 pieces of data and 960 pieces of training sample data are collected every day, and one day data between 4.11-4.30 is taken as the test sample data.

In this experiment, in the PNN + bagging classification algorithm used, the PNN smoothing factor is set as 0.06; the number of bagging integration is 10; in the elm regression algorithm used, the number of elm hidden layers is the same as the number of inputs, and the activation function is "sigmode".

This simulation analysis will be divided into two parts: one is to use PNN + bagging algorithm to judge the deviation degree of low-level data under different time points in the test set; the other is to use elm to test on the basis of PNN + bagging, and to verify and compare through traditional data correction methods. The quality of the evaluation and correction results can be measured by absolute error (E_{AE}), as shown in formula 2. For the data in a day, the average absolute deviation (E_{MAPE}) is used for comparison, and the formula is as follows.

$$E_{AE} = \frac{|y - y^{'}|}{y} \tag{6}$$

ISSN: 2414-1895
DOI: 10.6919/ICJE.202007_6(7).0027

$$E_{MAPE} = \frac{1}{M}\sum_{i=1}^{M}\frac{\left|y_i - y_i^{'}\right|}{y_i}\times 100\% \tag{7}$$

Where, M is the number of samples; y_i is the target of test samples; $y^{'}$ is the estimated value of test samples.

4.1 Simulation Analysis of Deviation Degree based on PNN + bagging algorithm

Here, PNN + bagging is used to train the training set (including time point, low count value and high-low count deviation degree), and the test set (including time and low count value) is compared with PNN + bagging algorithm respectively. The results are as follows. A single PNN algorithm as a classifier, as shown in Figure 5, has an effect of 68.8%. As shown in Figure 6, through the bagging integration of a single PNN algorithm, the accuracy reaches 74%, which is nearly 5% higher than that of a single classifier.

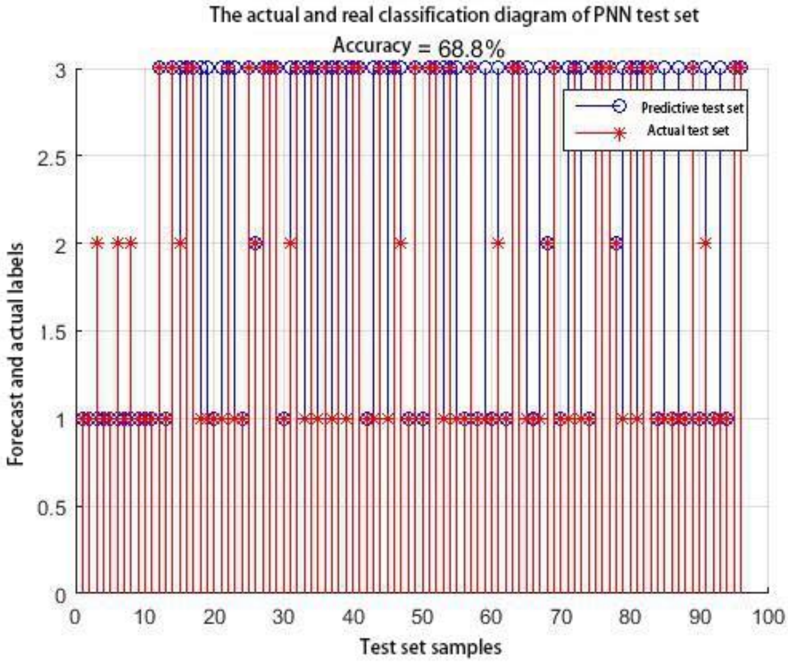

Figure 6. PNN accuracy

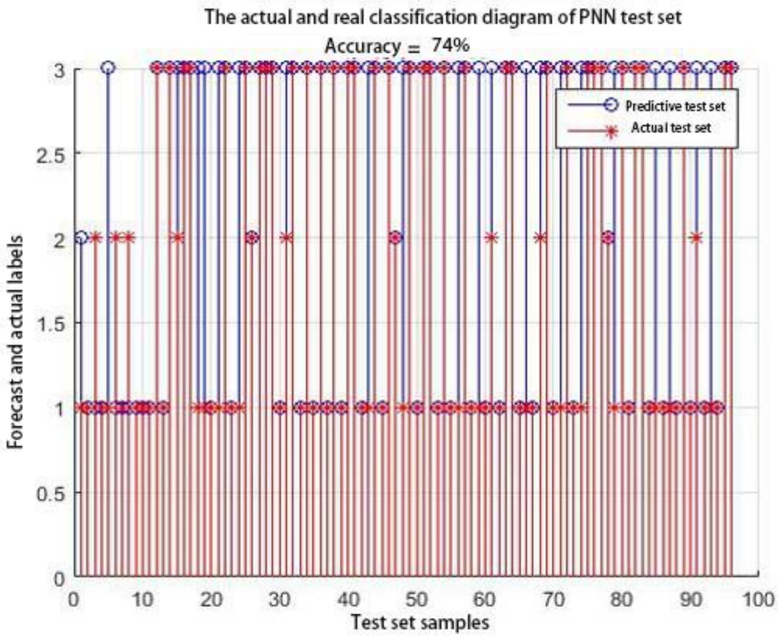

Figure 7. PNN + Bagging accuracy

ISSN: 2414-1895
DOI: 10.6919/ICJE.202007_6(7).0027

4.2 Simulation analysis of data correction based on PNN + Bagging + ELM meter

The deviation degree obtained by PNN + bagging algorithm is added to the original training set, and a new test set is formed. The results obtained by PNN + bagging combined with elm algorithm are as follows. Figure 5 shows the modified results of the traditional single algorithm, and the average deviation of the test results is 14.2% The average deviation of the improved test is 9.12% , the correction effect has been improved to some extent.

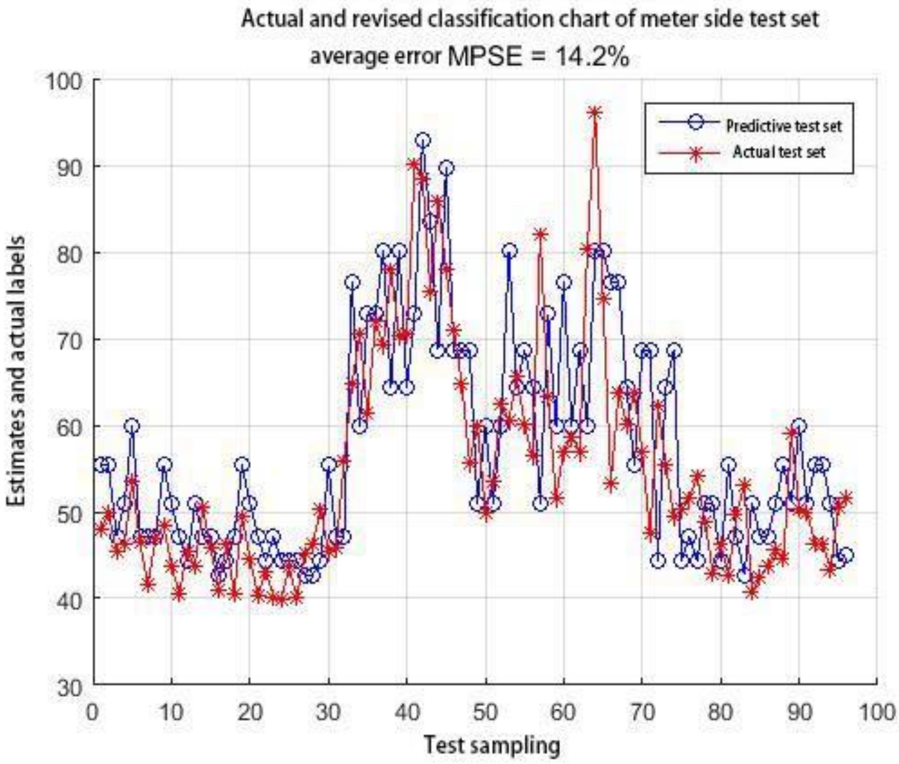

Figure 8. Results from traditional single algorithm

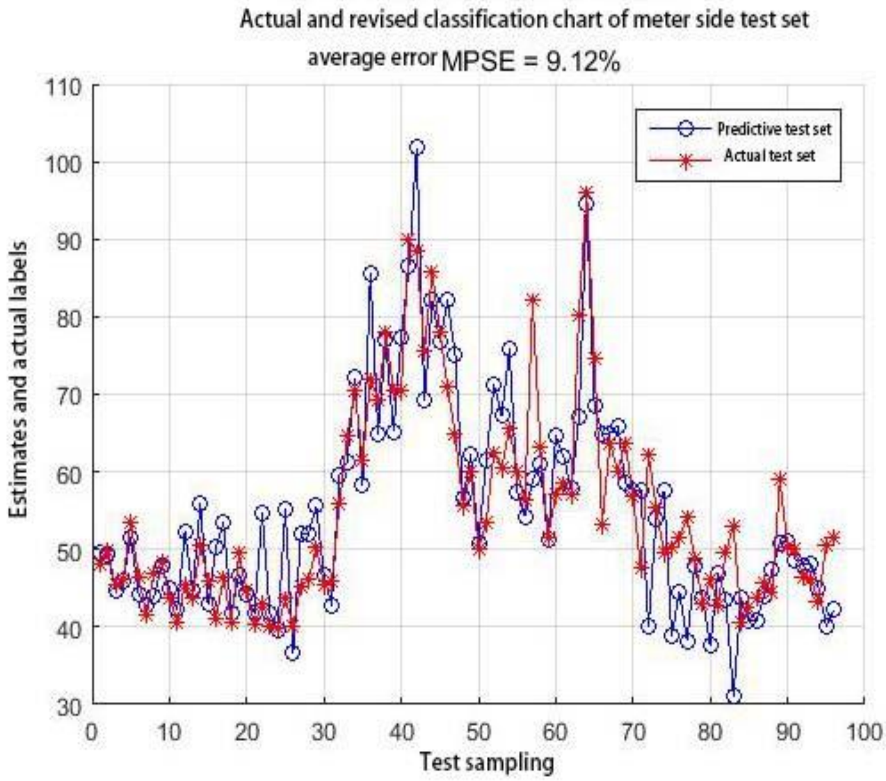

Figure 9. Results based on PNN + bagging and elm algorithm

ISSN: 2414-1895 DOI: 10.6919/ICJE.202007_6(7).0027

5. Conclusion

This paper analyzes the deviation of meters in the process of power physical connection (especially the deviation of meters at low voltage side of transformer is not close to the difference between high meter side and nameplate). In this paper, a new method based on the combination of PNN + bagging and elm is proposed to estimate the deviation degree of the revised data, and then a new feature is constructed, and the revised data is obtained by elm regression fitting. The simulation results show that the new method is effective It is verified that this method is better than the traditional method based on single algorithm.

Acknowledgements

Here, I especially want to thank Zhongheng Cloud Energy Company for the data support and all my friends who helped and inspired me in writing this article.

References

[1] Tian Zengyao, Zhang Mingli, Zhao Rui. Identification and correction of abnormal load data in short-term load forecasting (Jilin electric power, 2004), 000 (006): 21-23.

[2] Jiang Youhua, Huang Zhimin, Wan Yong. Research on state detection and data correction technology of power cable (Optical communication research, 2014), 40 (5).

[3] Zhang Jian. T-test identification and correction of abnormal power load data. DSM, 2005 (01): 16-18.

[4] Ye Xueyong, Wu Junji, Yang Wei, et al. Correction of bad data in power system based on neural network (Grid technology), 2007 (S2): 173-175.

[5] Bao Yongjin. Online detection, identification and correction of bad data in power system based on support vector machine (Sichuan electric power technology), 2013 (01): 65-69.

[6] Yang Lingxiao and Zhu Yali. High Voltage circuit breaker fault diagnosis based on probabilistic neural network J. Power System Protection and control, 2015 (10):62-67.

[7] Meng liu, zhou jinzhi. Human health monitoring method based on improved PNN model (Automated instrumentation), 2017,38 (11): 1-48.

[8] Eric, BauerRon, Kohavi. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants (Machine Learning), 1999.

[9] Miche Y, Sorjamaa A , Bas P , et al. OP-ELM: Optimally Pruned Extreme Learning Machine. (Neural Networks IEEE Transactions), 2010, 21(1):P.158-162.