Data Prediction and Real-time State Estimation of Distribution Network

Wen-si HUANG¹, Xin LU¹, Hui LI⁰, Zhao-yan LIU², Xiao-wei QI³, Liang AI³, Ren-jun WANG⁴,* and Hong-jun GAO⁴

¹State Grid Xintong Yili Technology Co. Ltd, Fujian, 350003, China
²State Grid Beijing Electric Power Company, Beijing, 100031, China
³State Grid Beijing Electric Power Company Haidian Power Supply Company, Beijing, 100142, China
⁴College of Electric Engineering, Sichuan University, Chengdu, 610065, China

*Corresponding author

Keywords: Data prediction, Fuzzy C-means, BP neural network, State estimation method.

Abstract. With the improvement of real-time measurement technology in distribution network, the data collected by power dispatching department is explosively growing. As a result, the real-time detection and analysis of the operation status of distribution network is facing new challenges. In this paper, a prediction model based on the combination of fuzzy c-means (FCM) algorithm and BP neural network is proposed to describe the range of future distribution network operation data. The purpose of this work is to provide a reference range for real-time operation state analysis. Then, considering the problem of massive measurement data in distribution network, a robust state estimation method based on exponential function weighted least squares (EFWLS) is proposed, which has better accuracy and efficiency than traditional weighted least squares (WLS) method. Finally, the accuracy of the prediction model is verified by comparing the predicted data with the real data, and the advantages of the proposed state estimation method are verified by comparing the results of different methods.

Introduction

In distribution network, the measurement data usually contain redundant data, error data, invalid data and missing data, which reduces the accuracy and reliability of operation state detection results of distribution network [1]. Besides, the lack of a reasonable description of the range of future normal operation data makes it impossible to support the real-time operation status judgment of distribution network. Therefore, how to describe the reasonable range of normal operation data, and then how to quickly and accurately detect the real-time operation state based on the range, is of great significance.

It is obvious that the various operation data are affected by lots of factors such as weather change, seasonal characteristics and user's living habits, which make the operation data fluctuate within a certain range. BP neural network is a commonly used method in prediction algorithm [1]-[3]. This method has strong ability of non-linear mapping, self-learning and fault tolerance. When applied to small data set prediction, it has the advantages of high prediction accuracy and fast training speed [2]. However, the traditional BP neural network obviously cannot adapt to the work of this paper with massive data sources. Therefore, a prediction model based on BP neural network and Fuzzy C-means (FCM) algorithm is proposed to describe the fluctuation range of future normal data. After processing the data by FCM, the accuracy of the prediction model is guaranteed, which provides a basis for real-time operation state detection of distribution network.

The function of state estimation is to process the measured data. In many literatures, the research of state estimation methods mainly focuses on the algorithm level [1]. Various improved algorithms based on weighted least squares (WLS) theory and robust estimation are favored by many experts [8]-[10]. Fang el al. [8] use WLS method to solve the state estimation model of active distribution system with photovoltaic power supply, distribution network and battery energy storage load, and uses regularization residual method to deal with bad data. But the traditional WLS algorithm is not...
reliable for the data deviating from normal distribution in the actual system. In view of this, some scholars put forward robust estimation, which means that when outlier error is unavoidable, appropriate estimation method is chosen to minimize the impact of outlier error on parameter estimation and obtain the best estimation under normal mode. Chen et al. [9] construct the robust state estimation model based on WLS method, and uses the maximum regularized residual method to identify bad data. A robust estimation method for voltage-power sensitivity of intelligent distribution network based on WLS is proposed in [10], and false data are processed by combining cap weighting and minimum absolute value estimation. Based on above analysis, this paper uses exponentially function weighted least squares (EFWLS) robust estimation method on state estimation, and then combines FCM algorithm with EFWLS method to test and correct false data in measurement data.

Data Prediction of Distribution Network Operation

In this part, a large number of discrete scene data sets are clustered by clustering method. According to the normal operation data of distribution network, normal data sets based on historical data of distribution network are selected from the clustered typical sample data sets. Finally, a prediction model for future state data quality of distribution network is constructed, and the normal fluctuation range of future state operation data is predicted by the prediction model.

The fuzzy c-means (FCM) algorithm to classify multiple sample datasets. FCM algorithm obtains the membership degree of each sample point to all class centers by optimizing the objective function, so as to determine the category of sample points and realize the automatic classification of sample data. The principle of FCM algorithm is as follows:

Initialization: specify the number of clustering categories C, \(2 \leq C \leq n\) (n is the number of data), and set the iteration stop threshold \(\varepsilon\), then initialize the cluster center \(V_0\) and set the iteration counter \(b=0\);

Step 1: calculate or update partition matrix \(U\) according to the following formula:

\[
U_i = \left[ \sum_{j=1}^{N} \left( \frac{\left\| x_j - v_i \right\|}{\left\| x_j - v_k \right\|} \right)^{\frac{1}{m}} \right]^{-\frac{1}{m-1}}, 1 \leq i \leq C, 1 \leq j \leq N
\]  

(1)

Step 2: update the clustering center \(V_{(b+1)}\) according to the following formula:

\[
v_i = \frac{\sum_{j=1}^{N} x_j (u_{ij})^n}{\sum_{j=1}^{N} (u_{ij})^n}, 1 \leq i \leq C
\]  

(2)

Step 3: If \(\|V_b - V_{b+1}\| \leq \varepsilon\), the algorithm stops and outputs partition matrix and clustering center \(V\), otherwise \(b = b + 1\), turn to step 1.

Then, a prediction model based on BP neural network is established in this paper by taking the normal sample set clustered by FCM as input layer and the normal operation data of distribution network as output layer. The number of neurons in the input layer of the neural network is set to \(d\). When selecting the number of neurons in the hidden layer and in each layer, the experiment is gradually increased from two to ten by trial-and-error method. The complexity and accuracy of the model are weighed. In the output results, the optimal network structure is selected; for the output layer of the neural network, a neuron is set up. Firstly, the training set is used to train the BP neural network, and the prediction model is obtained. Then, the accuracy of the prediction model is tested by the test set.

State Estimation Model of Distribution Network

Real-time measurement data of distribution network include node voltage amplitude, branch power and current amplitude, and pseudo-measurement of node injection power. Based on these data, it can
be judged whether the distribution network is in the normal operation range. But there are many false measurements in these data. It is difficult to judge whether the distribution network is running normally because of the low precision of these pseudo-measurement data. Therefore, we use the state estimation technique based on the exponential function weighted least squares (EFWLS) principle to obtain the optimal state estimation of the system.

In this part, a robust state estimation model based on exponential weighted least squares (EFWLS) is established for a large number of pseudo-measurements in distribution networks. Considering network connection, branch parameters and measurement data, system state variables and measurement values satisfy the following relationships:

\[ \tilde{z} = h(x) + r \]  \hspace{1cm} (3)

where \( \tilde{z} \) denotes m-dimensional measurement vector; \( r \) denotes measurement residual vector.

The EFWLS method obtains the optimal solution of the state variable by calculating the minimum weighted square sum of the measurement error, as follows:

\[
\begin{align*}
\min J(x) &= \frac{1}{2} \left[ \tilde{z} - h(x) \right]^T W \left[ \tilde{z} - h(x) \right] \\
\text{s.t.} \quad c(x) &= 0
\end{align*}
\]  \hspace{1cm} (4)

where \( c(x) \) denotes zero injection equality constraint. \( W \) is an exponential weight function diagonal matrix, that is, \( W = \text{diag} \{ w_1(r_1), W_2(r_2), \ldots W_m(r_m) \} \), whose element expression is:

\[
w_i(r_i) = R_i^{-1} \exp \left( -\frac{r_i^2}{2\sigma^2} \right)
\]  \hspace{1cm} (5)

where \( R_i^{-1} \) is the standard residue of measurement \( i \), and its formula is:

\[ r_{Ni} = \frac{r_i}{\sqrt{\left( KR \right)_i}} \]  \hspace{1cm} (6)

According to the EFWLS state estimation model (5), Lagrange multiplier lambda is introduced to deal with equality constrained equation to form unconstrained optimization problem of augmented Lagrange function \( L(x, \lambda) \):

\[
\min L(x, \lambda) = J(x) + \lambda^T c(x)
\]  \hspace{1cm} (7)

The measurement equation \( h(x) \) and the zero injection equality constraint equation \( c(x) \) are linearized, and the Newton method is used to solve the optimum conditions, so that the matrix form of the iteratively modified equation can be obtained:

\[
\begin{bmatrix}
\Delta x^{(k)} \\
\Delta \lambda^{(k)}
\end{bmatrix} = 
\begin{bmatrix}
H^T W H & C \\
C & 0
\end{bmatrix}
\begin{bmatrix}
H^T W \Delta z^{(k)} \\
-c(x^{(k)})
\end{bmatrix}
\]  \hspace{1cm} (8)

It is obvious that detection and recognition of false data is an important part of distribution network state estimation. Its purpose is to detect and eliminate a small amount of false data occasionally appearing in the sampled data according to the redundant information provided by the system. The basis of state estimation is obtained. This paper presents a new method which combines EFWLS robust estimation with FCM algorithm to correct false data. The main calculation steps of false date identification method are as follows and Figure 1 is the corresponding flow chart.
Step1: input all measurement data into the state estimator to obtain the best estimation of all measurements. And calculate the relative measurement estimation error of the system measurements

\[ \Delta z_i = \left| \frac{z_i - \hat{z}_i}{z_i} \right| \]  

(9)

where \( \Delta z_i \) represents relative measurement estimation error.

Step2: classify the relative error matrix according to the above clustering partition method.

Step3: find the best classification, take a measurement as the known accurate measurement value. The best classification that does not contain the measurement is false date sets, and output the measurement contained in false date sets.

Finally, when the data is corrected, the false date is corrected by the method of node power balance, and the state estimation is carried out again. After repeated fuzzy clustering analysis, all the data are classified into one group, that is, all the data are reliable values, that is, the correction is completed.

**Case Study**

The IEEE 33-bus system is used to simulate to verify the validity of the model, where the three-phase power reference value is 10MW and the voltage reference value is 12.66kV. In order to obtain the raw data, the method of applying 2% Gauss noise to the power flow calculation results is adopted in this paper. All the proposed algorithms are programmed in MATLAB 2016a. Besides, experimental platform is a 3GHz personal computer with 8 GB RAM. The convergence tolerance of the state estimation algorithm is \( 10^{-6} \).

In this study, the structure model of BP neural network in Fig. 1 is adopted. Seven normal data sets are selected from 20 data sets by using FCM variable clustering method as input of BP neural network, and the "trial-and-error" method is used to obtain two-layer hidden layer BP neural network for this
The number of neurons was 6 and 5, respectively. The number of neurons was 1, and the output predicted the survival time after operation. The maximum number of parametric training is set to 3000. Firstly, six groups of samples from the whole sample set are randomly selected as the training set to train the BP neural network and generate the prediction model. Then, using the remaining set of samples as the test set, the prediction model is tested and its accuracy is calculated.

Figure 2. A-phase voltage amplitude prediction diagram.

The prediction results are shown in Figure 2. From the figure, it can be seen that the upper and lower limits of three-phase voltage amplitude prediction error is small, which proves the effectiveness of the method.

In order to characterize the proximity between the estimated value and the true value of the system under different evaluation indexes, $\xi_1$, $\xi_2$ and $\xi_3$ can be used to represent the ratio of the absolute value of the difference between the true value and the estimated value in the system to the total number of the measured points within 1, 2 and 3 times of the standard deviation. The reasonableness of the proposed method and WLS method is evaluated and analyzed as shown in Table 1 by using the evaluation indexes of $\xi_1$, $\xi_2$ and $\xi_3$ indices in turn, the larger the indices, the more reasonable the estimation method.

| False data ratio | Method         | Iteration numbers | time | Evaluation index of state estimation |
|------------------|----------------|-------------------|------|-------------------------------------|
|                  |                |                   |      | $\xi_1$    | $\xi_2$    | $\xi_3$    |
| 0                | WLS algorithm  | 12                | 84   | 1.000      | 1.000      | 0.943      |
|                  | The proposed method | 8          | 105  | 1.000      | 1.000      | 0.943      |
| 2.84%            | WLS algorithm  | 12                | 93   | 0.783      | 0.730      | 0.547      |
|                  | The proposed method | 8          | 109  | 0.980      | 0.973      | 0.952      |
| 6.13%            | WLS algorithm  | 12                | 121  | 0.597      | 0.530      | 0.384      |
|                  | The proposed method | 16          | 153  | 0.984      | 0.953      | 0.927      |

As can be seen from the table, 1) false data will have a greater impact on the accuracy and computation time of state estimation; 2) Compared with WLS algorithm, the proposed method guarantees higher accuracy of state estimation with the number of iterations and computation time.

EFWLS robust estimation is used to detect false data based on FCM, so as to minimize the impact of false data on state estimation. Table 2 shows the identification results of false data in IEEE 33 distribution network.

Table 2. Identification results of false data in IEEE 33 distribution network.

As can be seen from Table 2, the proposed method guarantees higher accuracy of state estimation with the number of iterations and computation time. EFWLS robust estimation is used to detect false data based on FCM, so as to minimize the impact of false data on state estimation.
It can be seen that when false data is very serious, the method proposed in this paper can identify all false data, and there is no false detection, missed detection and other phenomena, so it has a strong ability to identify false data.

Summary

In this paper, a data range prediction model is established based on FCM algorithm with BP neural network for distribution network. According to the predicted interval results, the real-time running state can be judged. Then, a robust estimation algorithm based on EFWLS is used to estimate the distribution network state. Through the simulation analysis, the following conclusions are drawn: 1) The prediction error of this method is small, which proves the validity of this method. 2) In the case of more false data, although the EFWLS method still has errors, it is smaller than the traditional WLS method. 3) The method combining FCM with EFWLS can identify all false data and has strong ability to identify false data.

Acknowledgement

This work was supported by Science and Technology Project of State Grid Corporation of China (Research and Application on Power Grid Planning Data Automatic Collection and Intelligent Execution Management Technology Based on Novel Information and Communication Technology) under Grant SGIT0000YXJS1900324.

References

[1] Zaijun W U, Junjun X U, Xinghuo Y U, et al. Review on State Estimation Technique of Active Distribution Network[J]. Automation of Electric Power Systems, 2017, 41(13):182-191.
[2] Ping Z, Xueping P, Wenchao X. Short-term load forecasting based on wavelet decomposition, fuzzy gray correlation clustering and BP neural network [J]. Electric Power Automation Equipment, 2012.
[3] Xueneng S U, Tianqi L, Hongqian C, et al. A Multiple Distributed BP Neural Networks Approach for Short-term Load Forecasting Based on Hadoop Framework[J]. Proceedings of the CSEE, 2017, 37(17):4966-4973.
[4] Ye T. A forecasting model for current load of neural network group based upon momentum factor[J]. Power System Protection and Control, 2016.
[5] Chen W, Xiaoyang T, Ruihan W. Wide-area Backup Protection Algorithm Based on Least Squares Estimation with Limited PMU[J]. Automation of Electric Power Systems, 2015.
[6] Cheng Y, Chao L U. Application of Complex Field Weighted Least Square Algorithm in Phasor Measurement State Estimation [J]. Power System Technology, 2014, 38(3):700-706.
[7] Zhichao H, Xia X, Bin W. Power system state estimation combined with fuzzy comprehensive evaluation and decision-making[J]. Power System Protection & Control, 2015, 43(7):65-69.
[8] Zhi F, Shaojian S, Yuzhang L, et al. State Estimation for Active Distribution Systems Incorporating Photovoltaic Plant and Battery Energy Storage System. Automation of Electric Power Systems, 2019, 43(13):71-83.

[9] Feng L, Jin M, Yanbo C, et al. Robust State Estimator Based on Maximum Exponential Absolute Value[J]. IEEE Transactions on Smart Grid, 2017, 8(4):1537-1544.

[10] Peng L, Hongzhi S, Chengshan W, et al. Robust Estimation Method of Voltage to Power Sensitivity for Smart Distribution Networks Based on PMU Measurements. Power System Technology, 2018, 42(10):3258-3267.