Estimation Method of Line Loss Rate in Low Voltage Area Based on Mean Shift Clustering and BP Neural Network

Huang Tan¹, Yuan Li¹*, Liang Yu², Jing Liu¹, Linna Ni³, Xinping Diao¹

¹China Electric Power Research Institute, Peking, 100085, China;
²Shenzhen State Grid Technology Communication Corporation, Shenzhen, 518110, China;
³Marketing Service Center of State Grid Zhejiang Electric Power Corporation, Zhejiang, 311121, China.

*Corresponding author’s e-mail: liyuan3@epri.sgcc.com.cn

ABSTRACT: The main problems faced in the line loss management of the distribution network are the incomplete meter configuration, the difficulty of collecting operating data, and the excessive number of components and nodes. These problems lead to a very complicated calculation of line loss rate. This paper proposes an improved BP neural network estimation method for passive low voltage area line loss rate driven by low voltage area characteristic data, and realizes it through programming. First, the characteristic data required for calculating the line loss rate of the low-voltage passive station area is selected and classified according to the station area capacity after standardization. The BP neural network model improved by Mean Shift clustering method is used to calculate the line loss rate for the station area data of the same capacity, which provides scientific basis and data support for the calculation of the station area line loss rate and the management of high-loss stations in the future. A passive low-voltage station with a capacity of 315KVA in Zhejiang Province was used as the modeling object to perform simulation calculations to verify the accuracy of the proposed method.

1. Introduction
During the normal operation of the power system, different levels of line loss will occur in the four links of power generation, transmission, distribution and power consumption. The ratio of the line loss to the total power supply is the line loss rate. As an important indicator in the operation of the power grid, the accurate calculation of line loss rate has always been the focus of research for power companies [1]. Low-voltage station area usually refers to the power supply area of a transformer with low power measurement. The accurate calculation of the line loss rate of low-voltage passive station area is very important for the analysis of power grid operation, especially for high-loss station area and the identification of other abnormal line loss rates. Based on the above reasons, it is necessary to find an accurate method for calculating the line loss rate of passive low-voltage stations. At present, the estimation methods of station area line loss mainly include station area loss rate method, equivalent resistance method, Newton-Raphson method, regression analysis method [2][3]. Because the above-mentioned traditional line loss rate estimation method is too dependent on related information including grid structure, line length, line impedance, and actual power consumption [4], it is difficult to consider more complicated line loss rate factors, such as the proportion of residential electricity [5]. Therefore, according to the actual situation, it is necessary to screen and analyze the current passive low-voltage
station area statistical data based on the BP neural network.

We selected the characteristic data of the station area line loss in Section II. In Section III, we standardized the station feature data, and divided the station area data with the same capacity into several categories using Mean Shift clustering method. In Section IV, an improved BP neural network model is established to calculate the line loss rate, and conclusions are given in Section V.

2. Characteristic data selection
Theoretical line loss is the line loss calculated theoretically based on the parameters of the power supply equipment and the current operation mode of the power network, power flow distribution and load conditions. The traditional theoretical line loss calculation in station area includes root mean square current method, average current method, maximum current method, loss factor method, equivalent resistance method, etc. In this section, the main characteristic factors of the line loss are extracted from the theoretical calculation of the line loss in the platform area.

2.1. Percentage of residents in low voltage area
There is a big difference in the characteristic and habits of electricity consumption between residential users and non-resident users, and the number and ratio of the two will have an impact on the line loss of the station. Resident user load and non-resident user load have different energization characteristics, and the start, stop and power change of electrical equipment in each period have strong randomness. In addition, the increase in the total number of users in the station area will increase the line loss by affecting the power consumption scale and load factor of the station area. The proportion of residents in low voltage area can be calculated with the following formula:

\[
P = \frac{m}{n}
\]

In the formula, \( m \) is the number of residents in low voltage area, \( n \) is the number of total users in low voltage area.

2.2. Load factor
The load factor refers to the ratio between the actual power consumption of the electrical equipment and the rated power, and is an important operating parameter reflecting the load level. For the same station area, when the power supply is higher, the load is larger, and the current is larger, the power loss on the line is also larger. At the same time, in the long-term line loss management process, abnormal causes such as light load and heavy load are the more reliable field application experience summarized by the front-line of line loss managers. The load rate is the percentage of the daily (monthly) power supply under the station area to the full load capacity of the transformer, which can reflect the load situation in the station area. The load rate is selected as the key factor for the station area topology analysis and line loss calculation. The calculation formula as follows:

\[
\beta = \frac{W}{24S_e}
\]

Where \( W \) is the total daily electricity consumption, \( S_e \) is the rated capacity of low voltage area.

2.3. Three-phase unbalance
In the low-voltage station area power supply, the unbalance of the three-phase load will cause the load currents of the phases to be unequal, resulting in unbalanced currents between the phases. These unbalanced currents will cause losses on the phase line and also on the neutral line and increase the total line loss. We often use three-phase current unbalance to describe the unbalance of three-phase load, and its calculation formula is

\[
\varepsilon = \frac{I_{max} - \overline{I}}{\overline{I}} \times 100\%
\]
where \( I_{\text{max}} \) is maximum phase current, \( \bar{T} = \frac{1}{3} \times (I_A + I_B + I_C) \).

3. Data pre-processing

3.1. Characteristic data standardization

Data standardization is to remove the unit limit of the data and convert it into a dimensionless pure value, avoiding certain characteristics of the data from forming a leading role due to the difference in dimensions [7]. In the formula (4), \( n \) is the number of samples, the mean \( x_j \) and variance \( s_j \) of the feature data are respectively:

\[
\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij} \\

s_j = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2
\]

(4) \hspace{2cm} (5)

Calculate the standardized passive station area characteristic data \( z_{ij} \):

\[
z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}
\]

(6)

3.2. Feature data classification

Mean Shift clustering algorithm is an unsupervised and automatic machine learning algorithm that uses the (high-dimensional) spatial distribution of data points to generate reasonable clustering results. Its essence is to find local optimal solutions through probability density climbing. Compared with other clustering methods, this algorithm only requires the user to specify the "kernel function bandwidth" as a parameter, and requires less computer memory [8]. Mean Shift clustering algorithm is an iterative step, that is, first calculate the offset mean of the current centroid, move the centroid to the end position of the mean deviation vector, and then use this as the new centroid, and continue to move until the final condition is met.

![Fig.1 Mean Shift calculation process](image-url)
The Mean shift algorithm is used to cluster the characteristic data of passive stations with the same station capacity. The specific calculation process is as follows:

a) Randomly select a sample from all the unmarked passive station area feature data samples as the center point o of the cluster, and set the cluster radius h, threshold $\varepsilon_1$ and threshold $s$;

b) Temporarily classify the points whose distance from the cluster center point is less than h, mark it as H, and add 1 to the number of visits to all points in H;

c) Calculate the vector sum of all points from the center point o to H and record it as the vector d;

d) When $|d| > \varepsilon_1$, move the center point o along the vector d to obtain a new center point $o'$, then repeat the steps b)-d);

e) When $|d| \leq \varepsilon_1$, transfer all points in H to cluster $c_i$. If the distance between the center point of cluster $c_i$ and the center point of any known cluster c is less than the threshold $s$, merge the two clusters. Otherwise the cluster will be treated as a new cluster alone;

f) Repeat a)-e) until all feature data samples have been visited. At this point, count the number of visits of each sample, and assign the sample to the class with the most visits.

4. Improved BP neural network

The model structure of the BP network is shown in Figure 2, which is composed of three parts: input layer, hidden layer and output layer. The learning process of the BP algorithm consists of two parts: the forward propagation of the signal and the backward propagation of the error. Forward propagation means that the input samples are input from the input layer and passed to the output layer through various hidden layers. If the output result of the output layer does not reach the expected value, then go to the back propagation of error. Error backpropagation is to pass the output error back through the hidden layer, and adjust the weight and threshold of each neuron. The process of continuously adjusting weights and thresholds is the learning and training process of the network until the error reaches the expected range or reaches the set number of learning times $[^9][^10]$.

\[
\begin{align*}
    h_j(q) &= f \left( \sum_{i=1}^{m} w_{ji} x_i(q) - b_j^i \right) \\
    d_k(q) &= g \left( \sum_{j=1}^{l} w_{kj} h_j(q) - b_k^k \right)
\end{align*}
\]
where \( j = 1, 2, \ldots, l \), \( k = 1, 2, \ldots, n \), \( l \) is the number of hidden layer neurons, \( q \) is the sample number, \( b \) is the neuron threshold, the transfer function of the hidden layer of the network is:

\[
f = \frac{2}{1+e^{-2r}} - 1
\]  

(9)

then the error between the output and the expected output of the BP neural network is:

\[
E = \frac{1}{2} \sum_{k=1}^{n} (y_k - d_k)^2
\]  

(10)

Calculate the partial derivatives \( \partial_y (q) \) and \( \partial_h (q) \) of the error function to the weights of the output layer and the hidden layer, and modify the network weights by the following formula:

\[
w^{N+1}_{jy} = w^{N}_{jy} + \eta \delta_y (q) h_j (q)
\]  

(11)

\[
w^{N+1}_{ji} = w^{N}_{ji} + \eta \delta_h (q) x_i (q)
\]  

(12)

Where \( \eta \) is the learning step, \( \delta_y (q) \) and \( \delta_h (q) \) are the sensitivity of output layer and hidden layer neurons respectively. Repeat the above correction process until the relative error meets the accuracy requirements.

4) Input the test set into the model, compare the calculated line loss rate with the actual line loss rate, calculate the relative error, and verify the calculation accuracy of the model.

5. Experimental simulation analysis

In order to verify the feasibility and accuracy of the algorithm, the data of passive low-voltage area in Zhejiang province were selected to calculate and analyze the line loss rate. The areas were classified according to their capacities, and the stable platform areas with capacities of 200KVA and 315KVA were selected for experimental verification in combination with the line loss data of day and month. Each area selected 1100 samples, each sample data contains residents proportion of users, load rate, three-phase imbalance degree, the maximum load and line loss. After the order is scrambled, randomly selected from 1000 samples as training set for line loss calculation model building, the remaining 100 samples tested model as a test set, and the actual line loss, the standard BP neural network model without Mean Shift clustering and the improved BP neural network model after Mean Shift clustering were compared, as shown in Figure 3 and 4.

(a) Comparison of algorithm results
(b) Error analysis

Fig.3 Calculation results of line loss in 200KVA area
6. Conclusion
This paper proposes a estimation method for the line loss rate of the station area and applies it to the calculation in a certain area, and the following conclusions are obtained:
1) According to the electrical characteristic parameters of the samples, a mean shift clustering algorithm is proposed to classify the original data, which effectively solves the problem of poor BP neural network training accuracy caused by the dispersion of the line loss rate in low voltage area.
2) Taking a low-voltage passive station area in Zhejiang as the modeling object, based on the feature data set clustered by the mean shift clustering algorithm, the BP neural network station area line loss calculation model is established. The calculation results show that compared with the standard BP neural network, the relative error of the improved BP neural network is small, only 16.67%, which meets the research needs.

Acknowledgments
This paper was supported by the Research Projects of State Grid “Research on key technologies of complex low-voltage station area topology recognition and line loss calculation application based on cross-platform multi-source data fusion” (5600-201919168A-0-0-00).

References
[1] Yang T., Zhao L.Y., Wang C.S. (2019) Overview of application of artificial intelligence in power system and integrated energy system. Automation of Electric Power Systems, 43(01): 2-14 (in Chinese).
[2] Ya L I, Liping L, Baiqing L I, et al. Calculation of Line Loss Rate in Transformer District Based on Improved K-Means Clustering Algorithm and BP Neural Network[J]. Proceedings of the Csee, 2016, 36(17)(in Chinese).
[3] Bingbing Q U, Xiangtao Z, Xue C. Line loss calculation of distribution network based on power flow iteration[J]. Engineering Journal of Wuhan University, 2015(in Chinese).
[4] Shouxiang W, Pengfei D, Yingjie T. A Novel Method of Statistical Line Loss Estimation for Distribution Feeders Based on Feeder Cluster and Modified XGBoost[J]. Energies, 2017, 10(12):2067.
[5] ZHONG Xiaoqiang, CHEN Jie, JIANG Minmin, et al. A Line Loss Analysis Method Based on Deep Learning Technique for Transformer District[J]. POWER SYSTEM TECHNOLOGY, 2020, 44(2): 769-774.
[6] Wu W, Zhou Y, Li P, et al. Survey on negative line loss rate of transformer region: Rectification measures and challenges[J]. AIP Advances, 2020, 10(4):045124.
[7] Liu T, Wang S, Zhang Z, et al. Newton-Raphson method for theoretical line loss calculation of low-voltage distribution transformer district by using the load electrical energy[J]. Power System Protection and Control, 2015.
[8] Meerimatha G, Rao B L. Novel reconfiguration approach to reduce line losses of the photovoltaic
array under various shading conditions[J]. Energy, 2020, 196(Apr.1):117120.1-117120.18.

[9] ZHANG Yitao, WANG Zezhong, LIU Liping, et al. A 10 kV Distribution Network Line Loss Prediction Method Based on Grey Correlation Analysis and Improved Artificial Neural Network[J]. POWER SYSTEM TECHNOLOGY, 2019, 43(4): 1404-1410.(in Chinese).

[10] Ling-Xiao Z, Xiang-Dong T, Pu-Zhong Y, et al. Research on Power Line Theoretical Line Loss Calculation and Loss Management Based on Power Big Data Platform[J]. Digital Technology & Application, 2019.