Calculation method of line loss rate of photovoltaic station based on PCA-GRNN

Yuan Li¹, Jing Liu¹, Huang Tan¹, Yajie Li¹, Xinping Diao¹, Li Yao², Zhiguang Xu³

¹China Electric Power Research Institute, Peking, 100085, China;
²Marketing Service Center of State Grid Zhejiang Electric Power Corporation, Zhejiang, 311121, China;
³Power technology branch of Nari Technology Corporation, Nanjing, 210061, China.

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

ABSTRACT: Line loss rate of substation area is a comprehensive economic and technical index of power companies. With the continuous expansion of grid connected scale of distributed generation, accurate calculation of line loss rate of substation area with distributed generation is imminent. In this paper, considering the basic operation attributes and the grid connection attributes of distributed energy, a calculation method of line loss rate of substation area based on principal component analysis and generalized regression neural network is proposed. Firstly, the influence factors of line loss are extracted for the area containing photovoltaic distributed generation, and the feature dimension is reduced through principal component analysis; secondly, the reduced features are used as the feature input of generalized regression neural network for training, and the error between the training results and the real line loss is analyzed; finally, the ability of the proposed algorithm in the area identification of high line loss rate is tested. The results show that the generalized regression neural network trained in this paper has good generalization ability, the calculation error of online loss rate is small, and the identification of high loss area has high accuracy.

1. Introduction
Line loss rate of substation area is a comprehensive economic and technical index of power company [1]. As an important part of line loss management, it involves distribution network planning management, operation management, maintenance management, marketing power management, metering management, reading and verification management and other aspects, which comprehensively reflects the management level of power company for low-voltage substation area equipment and users [2]. The existing algorithms for calculating the line loss rate of the substation area with distributed generation have high requirements on the low-voltage substation area network, distributed energy parameters and operation data. Under the current technical bottleneck of obtaining and calculating the operation parameters of the low-voltage substation area with distributed generation and the accuracy of calculation, it is difficult to realize the calculation of line loss rate and the accurate identification of high-loss area through the existing feasible data [3].

At present, some scholars have carried out research on line loss calculation in substation area. In [4], according to the line loss data requirements of local distribution transformer, the author analyzes the causes of abnormal line loss in substation area, and discusses the calculation method of line loss caused by distribution transformer and intelligent measuring equipment, but it does not consider the influence of operation parameters.
of abnormal user behavior on line loss calculation. In [5], some scholars proposed a back propagation neural based on improved k-means clustering algorithm and Levenberg Marquardt algorithm, aiming at the complicated calculation process of line loss rate caused by incomplete meter configuration, difficult collection of operation data, and excessive number of components and nodes. Network model can quickly calculate the line loss rate of low-voltage substation area, but it does not consider the influence of power flow inversion caused by a large number of distributed generation access to substation area on the accuracy of line loss calculation model. In [6], a multi scenario model of wind, sunlight and electricity is established on the PSCAD simulation software. Based on a large amount of historical data, the multi case experimental simulation system is designed and developed to simulate the operation state of typical urban substation area, and the line loss of multi case area is calculated in detail. However, this multi case experimental simulation model needs a lot of historical data and station area. The data quality is too high for running equipment parameters. In [7], the author studied the photovoltaic grid connection position in the station area containing photovoltaic power generation. If the photovoltaic grid connection point is set at the beginning and the end of the line, the line loss and average voltage deviation are large. Therefore, this paper proposes a method to connect the photovoltaic grid more evenly to minimize the whole line loss, but the operation parameters such as three-phase imbalance and power factor are not considered in this method Influence on line loss of station area.

2. Feature selection of line loss rate in station area

When the characteristic parameters of substation area with distributed generation are not enough, the line loss rate can be calculated by using the operation characteristic parameters of substation area master table, including power variance, three-phase imbalance and load rate.

1) Power variance $z_1$: represents the power fluctuation of the total meter in the station area

$$z_1 = \frac{\sum_{i=1}^{n} (P_i - P_{ave})^2}{n}$$

Where $P_i$ is the $i$-th power measured by the regional master meter, $P_{ave}$ is the average power detected by the regional master meter every day, and $n$ is the daily power detection times of the regional master meter.

2) Three-phase unbalance degree $z_2$: indicates the three-phase load balance of operation in the station area

$$z_2 = \frac{I_{max\Phi} - I_{ave}}{I_{ave}} \times 100\%$$

Where $I_{max\Phi}$ is the maximum phase load current of substation area master meter, and $I_{ave}$ is the average phase load current of station area master meter.

3) Load rate $z_3$: represents the daily power supply of the station area

$$z_3 = \frac{W}{24S_N}$$

Where $W$ is the daily power supply of the substation area and $S_N$ is the rated capacity of the transformer in the substation area.

It is also necessary to extract the line characteristic factors of photovoltaic distributed generation operation, mainly including the daily photovoltaic power generation, photovoltaic grid connection location, and the number of photovoltaic users.

1) Daily photovoltaic power generation $z_4$: refers to the total daily power generation of photovoltaic distributed generation in the station area

$$z_4 = \sum_{i=1}^{m} \sum_{j=1}^{86400} \frac{P_{PVij}}{3.6 \times 10^6}$$

Where $P_{PVij}$ is the $j$-th second photovoltaic power of the $i$-th photovoltaic power supply, and $m$ is the number of photovoltaic power sources in the station area.
2) Percentage of PV users $z_5$ represents the proportion of PV users in the station area

$$z_5 = \frac{N_{PV}}{N}$$  \hspace{1cm} (5)

Where $N_{PV}$ is the number of photovoltaic users directly absorbed in the station area, and $N$ is the total number of users in the station area.

3) Photovoltaic grid connection position $z_6$ refers to the grid connection position, and represents the distance between photovoltaic grid connection point and bus.

Next, principal component analysis (PCA) is used to reduce the dimensionality of the features to make full use of the feature information and simplify the calculation complexity of the model.

Firstly, the correlation coefficient matrix is calculated using the station data:

$$R = \left( r_{ij} \right)_{m \times m} = \frac{Z'Z}{m-1}$$  \hspace{1cm} (6)

Where $m$ is the number of original features, $R$ is the correlation coefficient matrix with $m$ rows and $m$ columns, $r_{ij}$ is the element in the $i$-th row and $j$-th column of the matrix $R$, and $Z$ is the station area with $m$ rows and $n$ columns feature matrix.

According to the correlation coefficient matrix, the eigenvalues and eigenvectors are calculated using the following formula:

$$|R - \lambda E| = 0$$  \hspace{1cm} (7)

$$\left( R - \lambda_j E \right)u_j = 0$$  \hspace{1cm} (8)

Where $\lambda$ is the eigenvalue of the correlation coefficient matrix $R$, $E$ is the identity matrix, $\lambda_j$ is the $j$-th eigenvalue, $u_j$ is the eigenvector corresponding to the $j$-th eigenvalue, $j=1,2,\ldots,m$, used to compose the transformation matrix.

Secondly, the number of principal components is determined to make the information expressed by principal components account for more than 85% of the original information.

$$\sum_{k=1}^{h} \frac{\lambda_k}{m} > 0.85$$  \hspace{1cm} (9)

Where, $h$ is the main component.

Finally, using the transformation matrix composed of $h$ eigenvectors corresponding to the principal component, the original station area characteristic matrix is transformed into the main component station area characteristic matrix:

$$Y = UZ$$  \hspace{1cm} (10)

Where, $U = [u_1, u_2, \ldots, u_h]^T$, $Y$ is the transformed principal component characteristic matrix.

3. Calculation model of line loss rate in station area

The calculation model of line loss rate in the station area includes distributed generation. Due to the problem of power flow reverse transmission, the common equivalent resistance method and voltage loss method are no longer applicable. Under the condition that the line parameters and operation parameters of the substation area are sufficient, Newton power flow method can be used to calculate the line loss of single station area [8], but there are many parameters in a large number of stations with distributed generation. In the case of insufficient data acquisition, the advanced neural network algorithm can be used to train the line loss calculation model of the same type of substation area. For the station area that needs to calculate the line loss rate, only a small number of total meter measurement characteristics of the substation area and the grid connection characteristics of the distributed generation can be used to calculate the line loss rate. The number of training samples of neural network is reduced due to the less number of stations with distributed generation than that of passive ones, and the output of distributed
generation fluctuates greatly. At this time, the deviation of line loss rate calculated by traditional neural network algorithm is large. The general regression neural network (GRNN) based on radial basis function is often used for regression fitting, and has obvious advantages in processing small sample unstable data. Its internal structure is shown in Fig.1.

After the weighted iterative calculation of hidden layer and output layer, the line loss rate of the station area is output, and the relative error is selected as the GRNN evaluation index. The calculation formula is as follows:

$$E_i = \frac{\hat{y}_i^l - y_i^l}{y_i^l}$$  \hspace{1cm} (11)

Where, $E_i$ is relative error of the $i$-th station area, $y_i$ is the true value of line loss rate of the $i$-th station area, $\hat{y}_i^l$ is the predicted value of line loss rate of station area, $i=1,2,\ldots,r$, $r$ is the number of test station areas.

The flow chart of the whole calculation process is shown in Fig.2.

4. Experimental simulation analysis

In order to verify the effectiveness of the proposed method, this paper calculates the line loss rate of photovoltaic stations in South China, and selects 100 normal test samples and 100 high loss test samples.
as the result display.

The back propagation neural network (BPNN) algorithm, which is commonly used in neural network, is selected as the comparison algorithm, and the two algorithms run in the same software and hardware environment. The experiment was carried out on Windows 10 of 64 bit operating system with Intel processor R7-4800CPU@3.60Hz. The running memory is 8GB, the software platform is Python 3.8.3, and tensorflow machine learning framework is adopted. The comparison of line loss calculation results of the two algorithms with the real value is shown in Fig. 3, and the relative error comparison is shown in Fig. 4.

As shown in Fig.3, for the calculation of line loss rate in the substation area with distributed generation, the prediction results of GRNN algorithm used in this paper are in good agreement with the real value. Specifically, the average relative error of GRNN algorithm and BPNN algorithm are 18.47% and 27.36%, and the average relative error of GRNN algorithm is 32.49% lower than that of BPNN algorithm. According to the box diagram and the normal distribution fitted in Fig.4, it can be seen that the error distribution of GRNN is more concentrated and the fluctuation is smaller, and the accuracy of line loss calculation in the station area is maintained at a high level. In addition, the maximum relative error of GRNN algorithm is 131%, which is 15% less than that of BPNN algorithm.

Next, the reasonable range of line loss rate is determined according to the station samples with normal line loss rate, and then the data set containing high line loss is selected for test to test the identification ability of high loss station area of GRNN algorithm. The test results are shown in Fig.5.

It can be seen from Fig.5 that the proposed algorithm can effectively identify high loss area, and the line loss rate of real station area is in good agreement.

5. Conclusion
In this paper, a calculation method of line loss rate is proposed and applied to the data calculation of
some certain areas. According to the electrical characteristic parameters of the sample areas, the principal component analysis algorithm is used to reduce the dimension of the original data, which effectively solves the problem that the neural network training model does not converge due to more input features. Due to the small sample size and obvious fluctuation of characteristic data, the generalized regression neural network model used in this paper can calculate the line loss rate more accurately than the traditional back propagation neural network. In addition, in the aspect of high line loss rate area identification, the generalized regression neural network performs well, which can be used to identify the abnormal station area of distribution network line loss rate.

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] Wang S X, Zhou K, Su Y, (2017) Estimation method of reasonable line loss rate based on random forest algorithm, Power Automation Equipment, 37(11): 39-45(in Chinese).
[2] Li J, (2020) Analysis of causes and calculation methods of distribution network line loss, Power Equipment Management, 1(09): 71-72(in Chinese).
[3] Liu X H, Ma L, et al. (2014) Improved Equivalent Resistance Method for Low-Voltage Distribution Line Loss Calculation. Advanced Materials Research, 2014, 3383.
[4] Zhang B., Zhu Z T., et al. (2015) Discussion on the treatment method of abnormal line loss data in operation distribution through station area. Automation of Electric Power Systems, 34(02): 9-11(in Chinese).
[5] Li Y, Liu L P, Y, et al. (2016) Calculation method of line loss rate based on improved k-means clustering and BP neural network, Proceedings of The Chinese Society for Electrical Engineering, 36(17): 4543-4552(in Chinese).
[6] Li H T, Yu X J, et al. (2019) Simulation Modeling and Line Loss Research Based on Typical Urban Distract Area, 2019 IEEE 3rd International Electrical and Energy Conference (CIEEC). IEEE.
[7] Sirojiddin R, Chorshanbiev, et al, (2019) Structural Analysis of Power Losses in (6-10 / 0.4 kV) Urban Distribution Electric Networks of the City of Dushanbe, the Republic of Tajikistan, IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering.
[8] Khosravi M, Monsef H, Aliabadi M H, (2018) Loss allocation in distribution network including distributed energy resources (DERs), International Transactions on Electrical Energy Systems: 2548.