Distributed Learning Algorithm for Distributed PV Large-Scale Access to Power Grid Based on Machine Learning

Zhen Lei[1], Yong-biao Yang[2(✉)], and Xiao-hui Xu[3]

[1] State Grid Jiangsu Electric Power Company, Nanjing, China
[2] Southeast University, Nanjing, China
danghongen2017@163.com
[3] China Electric Power Research Institute, Nanjing, China

Abstract. Due to the long prediction time and the large range of data filtering, the traditional algorithm has low system operation efficiency. For this reason, distributed learning based on machine learning is widely used to predict the power grid output. First, establish a grid output prediction model to limit the system's line loss and transformer losses. Secondly, based on the distributed photovoltaic power generation output prediction model, the vector moment method and the information method are used to narrow the search space. Based on the data concentration and fitness function values, the calculation formula of voltage output prediction of distribution network nodes with distributed photovoltaics is derived to realize the power grid output prediction algorithm. Finally, it is proved by experiments that distributed PV large-scale access to power grid output prediction algorithm can effectively improve system operation efficiency.

Keywords: Operation efficiency · Photovoltaic capacity · Radial structure · Power system

1 Introduction

With the rapid development of the power industry, the access of distributed photovoltaics and charging piles for electric vehicles in distribution networks increases the complexity of the power usage mode. The core idea of grid output forecasting is to study the changing law of historical load data, as well as the impact of meteorological factors, economic factors and other related factors on the load. Describe the relationship between load and influencing factors by establishing a suitable mathematical model, and then make reasonable guesses about the load in the future period. According to the different time periods, the grid output prediction can be divided into four short-term, short-term, medium-term and long-term forecasts. The short-term grid output forecast period is generally one to several days, and the grid dispatcher specifies the regional power generation plan based on the results of the short-term forecast. If the forecast result is too high, the system operation efficiency will be too low, resulting in waste of resources. If the forecast result is low, there will be insufficient power supply to meet the social production needs and the people's power demand. Therefore, the importance

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 301, pp. 439–447, 2019.
https://doi.org/10.1007/978-3-030-36402-1_47

of short-term load forecasting is self-evident. Therefore, finding the right method to reduce the prediction error is crucial. Literature [1] proposed an improved big bang algorithm for the grid-connection capacity of distributed pv, and solved the above model. Consider the impact on the power distribution system after the distributed photovoltaic (pv) grid, from the point of view of power distribution network planning, establishes a distributed photovoltaic (pv) grid acceptance ability as objective function, to run load voltage level, feeder rate, voltage total harmonic distortion rate and the short-circuit current level for the optimization model and application of the constraint of 33 nodes distribution network testing, comparing with other algorithm results this method has a certain error. In order to reduce the average prediction error of short-term load by 1%, a distributed PV large-scale power grid output prediction algorithm based on machine learning is proposed to ensure that the prediction result falls within the established range.

2 Distributed Photovoltaic Large-Scale Access Grid Output Prediction Algorithm

2.1 Power Grid Output Prediction Model

After distributed PV access to the distribution network, the impact on system network loss depends mainly on the location of access and the capacity of access. First, the impact of distributed PV access capacity on network loss. When the small-capacity photovoltaic is connected to the distribution network, the local compensation of the load absorption power can be realized, and the power flowing from the power supply node to the load node on the feeder line is reduced, thereby reducing the line loss and the transformer loss of the system. When the photovoltaic capacity is large, the unidirectional flow in the distribution network will have two-way flow and generate a new power distribution, which will increase the local network loss, but the total network loss of the system may be reduced. Second, the impact of distributed PV access locations on network loss. When the photovoltaic access is close to the head end of the line, a larger photovoltaic output can have a significant effect on the reduction of the network loss; As the photovoltaic position moves toward the end of the line, even if the access photovoltaic capacity is small, the network loss can be significantly reduced; Near the end of the line, the increase in PV capacity may increase the network loss, but still less than the network loss when there is no PV access [2].

Most of the distribution systems in urban and rural areas in China are mainly radial structures, and the power supply mode is simple. For medium voltage distribution networks in urban power distribution systems, radial operation is also performed under normal mode [3]. Therefore, this paper selects the radial distribution network as the research object, and the circuit diagram of the photovoltaic distribution network is as follows (Fig. 2):

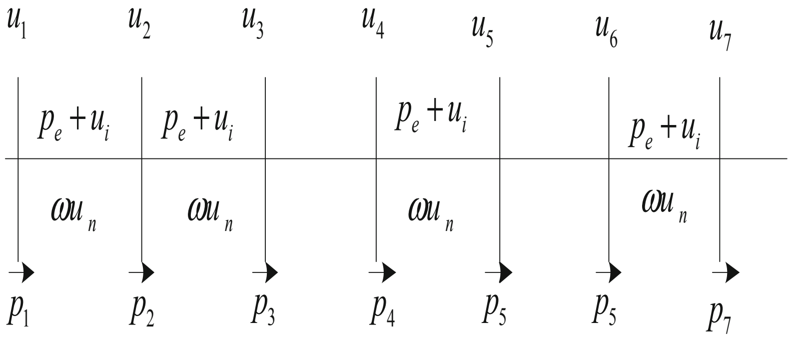

Fig. 1. Circuit diagram of photovoltaic distribution network

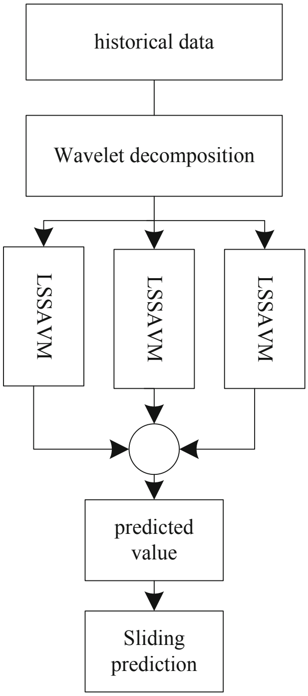

Fig. 2. Grid output prediction model

Since the wind power output has randomness, the fluctuation is large within a certain time range. If the machine learning algorithm is directly used, the prediction effect is not good. The low-frequency component obtained by using the wavelet to decompose the Fig. 1 is a circuit diagram of a photovoltaic distribution network. As can be seen from Fig. 1, the node voltage and line loss of the distribution network mainly depend on the load power, distributed PV output, distributed PV access position, line unit impedance and line length. Assuming that the power factor of each load in the line is constant, the load active power and reactive power depend on the apparent power of the load. When the capacity or power factor of distributed PV changes, it will affect the active output and reactive output of distributed PV [4].

Due to the intermittent nature of distributed photovoltaic power generation and the difference in load characteristics in the power grid, it is impossible for the power grid to accept photovoltaic power sources without restriction. The existing research on PV capacity of distribution network mainly obtains the maximum access capacity of PV in distribution network through various safety and stability operation constraints of grid after grid connection, or enhances the access capacity of PV through certain measures. However, the research on photovoltaic power transmission capacity of distribution network with PV penetration rate is scarce, and the existing research perspective is relatively simple. The impact of photovoltaic power on the distribution network and the support of photovoltaic energy to the distribution network cannot be fully reflected. Especially, the photovoltaic capacity planning for the distribution network is rarely involved on the basis of permeability. For a known regional distribution network, when planning the photovoltaic capacity in advance, it is necessary to conduct a macro assessment of the distribution of distributed photovoltaics from the perspective of power and energy balance according to the current situation in the region or the overall load curve in the future. And considering the specific grid structure and safe and stable operation restrictions, how to determine the PV installation location, installation method and installation capacity, and what control scheme is adopted to improve the PV maximum access capacity of the distribution network is a key consideration in the grid connection design phase. On the basis of reasonable penetration rate, the remaining distributed PV acceptance of the distribution network is announced in time. In order to improve the power of distributed photovoltaics, a power grid output prediction model is established, and the power grid output prediction model is as follows;

Wind power output has a high amplitude, is relatively stable, has a certain regularity, and can describe the general trend of the wind power output, and is dominant in each component. If a machine learning algorithm is used for the low frequency component, the prediction effect is good. Although the high-frequency components obtained by wavelet decomposition have large fluctuations, the amplitude is small. If a machine learning algorithm is used for the high-frequency components, even if a large prediction error occurs, the influence on the overall prediction effect is small. Therefore, a grid output prediction model is established. The model first uses wavelet decomposition to decompose historical wind power data, then puts the high and low frequency components into the prediction model to obtain the predicted values of the components, and finally adds all the component prediction values to obtain the final result [5].

When the distributed photovoltaic power generation system adopts the current-controlled inverter as the grid-connected interface, the output active power and the current amplitude are constant, and can be converted into multiple nodes for processing in the power flow calculation of the distribution network [6]. In order to output as much active power as possible, the distributed photovoltaic system can maintain a pure resistive power factor when it is integrated into the grid. In this control mode, distributed photovoltaics only output active power [7]. Its calculation formula is as follows:

$$\mathrm{h}^{e} = \sqrt{n^{2} \times (m-1)^{2}} \tag{1}$$

In the formula (1), h represents a node where the reactive power is 0, n represents the active power of the output, and m represents the active power. This formula is used to maintain a pure resistive power factor in a defined spatial range. From the simulation of the voltage influence of the distributed photovoltaic access distribution network, it can be known that the influence of the access location and capacity of the distributed photovoltaic on the voltage of the system node is very different. On the other hand, there are asymmetry in the distribution network, the transmission line is not cyclically transposed, and the three-phase load is asymmetrical [8]. If the access to the single-phase distributed PV is not restricted, the voltage and current three-phase unbalance of the distribution network may be aggravated. The model solves the problem of three-phase imbalance by reducing line loss and extending equipment life. The distributed photovoltaic power grid output prediction model is to access distributed photovoltaics as much as possible under the premise of satisfying the constraints of power balance [9], voltage deviation and three-phase unbalance, and reasonably configure the capacity of each node to access distributed photovoltaics [10]. So far, the construction of the distributed photovoltaic power output prediction model is completed.

2.2 Realize Power Grid Output Prediction Algorithm

After distributed PV access to the distribution network, the power supply network structure changes from a single power supply to multiple power supplies, and the magnitude and direction of the power flow may change, further affecting the voltage and loss of the distribution network. For the typical topology of the distributed photovoltaic access distribution network, the calculation formula for the voltage output prediction of the distribution network node with distributed photovoltaic is derived as follows;

$$|\langle f_e, t_u \rangle| = \max |\langle a_e, h_u \rangle| \tag{2}$$

In the formula (2), f represents a low frequency signal, t represents a signal to be decomposed, a represents a voltage consumption time, and h represents a position. When using the grid output prediction algorithm, each decomposition requires a large amount of inner product operations for each atom in the atom library and the signal to be decomposed. The vertical projection of the residual component of the signal on the selected atom has non-orthogonality, which makes the result of each decomposition not optimal, but a suboptimal solution. This will result in an increase in the number of iterations, an increase in the amount of calculations, and a lengthy calculation. In order to solve the above problems, based on the matching pursuit algorithm, this paper orthogonalizes the selected best matching atom and the selected best matching atom for each step decomposition. This allows for faster convergence on the basis of the same accuracy. In order to improve the convergence performance of the original algorithm, a speed expression for updating the system transmission data set is proposed. The formula is as follows:

$$a_i(y+1) = \Re(a_i \max(t)) + y_i \tag{3}$$

In Eq. (3), a represents the convergence speed, y represents the spatial data set, and t represents the search space. The power grid output prediction algorithm has a vector moment method and an information method. Compared with the information method, the vector moment method can directly correspond to the fitness function of the optimization problem solution, which can effectively narrow the search space and avoid double counting. Therefore, a vector moment based concentration selection method is adopted in the improved grid output prediction algorithm. Suppose m antibodies form a non-spatial system set u. The specific flow of the concentration selection operation of the vector moment is as follows:

The first step is to initialize various parameters in the vector moment method, and the parameter data set is $t_1, t_2, t_3, \cdots, t_n$. In the second step, the data transmission speed in the solution space is initialized, and the initial position is set, and the initial position data set is $v_1, v_2, v_3, \cdots, v_n$. The third step is to determine whether the number of iterations satisfies the division criterion. If the number of iterations is $y_i \in 1$, the data is divided into two types. If the number of iterations is $y_i \notin 1$, then it is necessary to continue the iteration without satisfying the condition. In the fourth step, the data transmission speed is calculated according to the fitness function. The fitness function formula is as follows:

$$u_i = \Re(u_i \max(t)) \tag{4}$$

In the formula (4), u represents the transmission speed, and t represents the data parameter. The fifth step is to evaluate the data results and update the speed and location of the data. The inertia weight and learning factor adopt a dynamic adjustment strategy. Record the optimal position of the data and the current optimal position of the data set. If the data set is divided, the worst position of the data and the current worst position of the population are recorded at the same time. In the sixth step, the concentration and fitness function values of each particle are calculated, and the individual extreme values of the particles and the extreme values of the population are updated according to the fitness function value. For the data set, the optimal extreme value of the data and the worst extreme value are updated at the same time. In the seventh step, the vector moment immune selection is performed according to the data transmission speed. The eighth step is to generate the next generation data set. In the ninth step, it is judged whether the convergence condition is satisfied, and if it is satisfied, the optimal solution is output, otherwise it returns to the third iteration to start the next iteration evolution.

In the iterative optimization process, the failed data set experienced is represented by the data itself or the data location with poor fitness in the population. For the speed and position update of the failure experience, it can be represented by formula (3), and the calculation formula is as follows:

$$j_i(y+1) = \Re(j_i) + k_i \tag{5}$$

In formula (4), j represents the data transmission position, y represents the transmission failure position, and k represents the worst position searched. Due to the large amount of data transmitted by the system, it is necessary to learn from successful experiences. For the failure experience, it only absorbs the effective information to guide the next iterative search and avoid returning to the failed position again. The final update position of the data set is still determined by Eq. (4).

Article method can reflect the intrinsic link between economy and safety when photovoltaics are connected to the distribution network. By calculating the current distributed PV permeability index group and photovoltaic utilization efficiency and utilization cost, it can effectively guide the orderly grid connection of distributed PV in the future, and provide an important reference for the pre-distribution planning of distributed PV in regional distribution network. For a specific distribution network, the geographical environment is basically determined by its photovoltaic output characteristics. Without considering other measures, its load characteristics have a greater impact on the PV capacity of the region. The correlation coefficient can reflect the correlation between the load characteristics and the photovoltaic output characteristics, and initially grasp the PV consumption capacity of the grid, and carry out reasonable photovoltaic grid layout to provide guidance for regional PV capacity planning. So far, the grid output prediction algorithm is implemented.

3 Experimental Results

In order to verify the effectiveness of the machine-based distributed PV large-scale access grid power prediction algorithm, the algorithm optimization, transmission speed and accuracy were tested. In the experiment, by comparing the response speeds of the two algorithms and analyzing the performance of the algorithm, the accuracy of the large-scale distributed PV output prediction algorithm based on machine learning is proved.

The machine learning-based distributed photovoltaic large-scale access grid output prediction algorithm is tested. The test results are as follows:

Figure 3 is a comparison of experimental results. From the above test data, it can be seen that it is more advantageous than literature [1] Method and literature [2] Method. From the results of the optimized scheduling in Fig. 3, the dispatching power at 100 min is 0, indicating that the grid-connected operation of the algorithm does not affect the voltage quality. Although the grid-connected operation of distributed photovoltaics can play a certain role in improving the network loss and voltage quality of the system, it is not superior from the economic perspective. This is mainly because the data transmission speed of the machine-based distributed photovoltaic large-scale access grid output prediction algorithm established in this paper is faster.

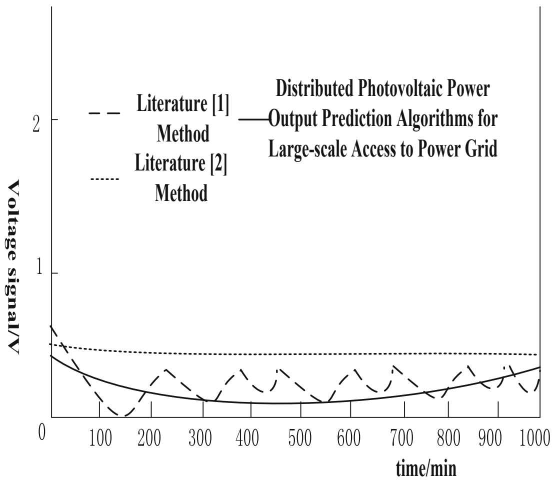

Fig. 3. Comparison of experimental results

4 Conclusion

Short-term grid forecasting is one of the indispensable tasks of the power dispatching department, and it is also a necessary process for the power dispatching department to realize modern management. Since the power generation plan and operation mode of the power grid are mainly determined by the load, high requirements are placed on the accuracy of load prediction. In this paper, based on machine learning, distributed PV large-scale access to power grid output prediction algorithm is established to establish a power grid output prediction model. Start with two aspects to improve the predictive effect of the model. The lack of objectivity for the model input selected by experience, and also affect the accuracy of prediction, the power grid output prediction algorithm is proposed, and the two algorithms are compared and analyzed through experimental simulation. The experimental results show that the input variables of the distributed PV large-scale access to the grid output prediction model will directly affect the prediction effect of the algorithm. In this paper, the grid output prediction algorithm can directly extract the features of numerical data. When selecting useful variables, the effects of redundant variables can be eliminated.

References

1. Liu, S., Luo, F., Wang, C., et al.: Calculation method of hosting capacity for distributed grid-connected photovoltaic based on improved big bang-big crunch algorithm. Proceedings of the CSU-EPSA **11**, 5–10 (2017)
2. Cheng, M., Dai, X., Wang, S., et al.: Study on voltage characteristics of distributed photovoltaic access to active distribution network. Northeast. Power Technol. **38**(5), 2118–2221 (2017)

3. Wang, H., Ge, L., Li, H., et al.: Summary of characteristics analysis and prediction methods of distributed photovoltaic power generation. Electr. Power Constr. **38**(7), 2321–2329 (2017)
4. Yang, F., Tian, C., Li, Z., et al.: Distributed photovoltaic spatial-temporal distribution prediction considering space capacity saturation. Power Grid Technol. **23**(12), 3917–3925 (2017)
5. Li, L., Wang, J., He, Y., et al.: Multi-point PV-DG day-ahead allocation plan based on K-means clustering particle swarm optimization. High Volt. Technol. **44**(4), 1263–1270 (2017)
6. Huang, W., Gao, Y., Zhang, Y., et al.: Limit capacity assessment of distributed photovoltaic access distribution network considering uncertainty. Power Syst. Prot. Control. **11**(14), 1453–1654 (2018)
7. Jin, Z., Xiang, T., Chen, H., et al.: Distributed photovoltaic access planning method considering power quality problems [J]. Power system protection and control **45**(9), 1231–1238 (2017)
8. Yang, C., Pan, Yu., Zeyang, W., et al.: Output optimization and capacity allocation method of distributed photovoltaic-energy storage system considering power loss. Renew. Energy **35**(2), 245–251 (2017)
9. He, X., Zhang, M., Zhu, Y., et al.: Application analysis of centralized and string inverter in photovoltaic power plant. J. Xi'an Polytech. Univ. **04**, 443–448 (2018)
10. Zhang, W., Qi, W., Xu, Q., et al.: Study on location and capacity of distributed photovoltaic access distribution network considering time series characteristics. J. Nanjing Norm. Univ. (Eng. Ed.) **17**(3), 1122–1128 (2017)