Bus ultra-short-term load forecasting considering the impact of distributed photovoltaic power supply

Changyong Yu¹, Mingli Zhang¹, Qiankun Hu²,¹, Heyan Zhu¹, Li Jiang¹, Zhe Nan¹, Fang Li¹
¹Limited Economic Research Institute State Grid Liaoning Electric Power Company Shenyang, China
²School of Electrical Engineering Northeast Electric Power University Jilin, China
¹3844218772@163.com

Abstract. Bus load base is small and volatile, so it is difficult to predict ultra-short-term load. Distributed photovoltaic power supply connected to bus power supply area further increases the uncertainty of bus load and directly affects the prediction accuracy. Aiming at the ultra-short-term load forecasting problem of bus with photovoltaic power supply under complex weather conditions, a method of bus ultra-short-term load forecasting based on feature selection and XGBoost is proposed. Firstly, a high-dimensional feature set is constructed with various influencing factors, and the XGBoost predictor is trained with the original feature set in the typical meteorological day historical data such as rainy and sunny days, and the degree of response between the bus load and the multi-factor is determined during the training process. Then, the MAPE value is used as the evaluation index. Forward search is used to select features to determine the optimal feature subsets of different typical weather day predictors. Finally, according to the selected optimal feature subsets, the optimal XGBoost prediction model for different weather days are established to predict the bus output of different weather days. Experiments on bus data with distributed photovoltaic in a certain area show that the new prediction model has higher prediction accuracy.

1. Introduction
With the continuous development of renewable energy such as solar energy and wind energy, the use of distributed generator such as photovoltaics and wind power has increased greatly, effectively achieving low carbonization and cleanliness of the terminal power supply. However, wind, light and other DG are sensitive to meteorological factors and are susceptible to it. Their output has the characteristics of strong volatility and randomness, and distributed photovoltaic generator output is affected by factors such as cloud volume and temperature [1-2]. Bus ultra-short-term load forecasting, as an important part of power system load forecasting, is an important basis for ensuring the safe operation of power grids. It is necessary to have high-precision prediction conclusions to support operation and control [3-4]. The large number of distributed photovoltaic power sources makes the bus load appear more irregular, which increases the difficulty of bus load forecasting.

Some methods currently used in load forecasting include fuzzy theory [3], neural network [5], and Support Vector Machine (SVM) [1]. In the literature [5] and the literature [6], meteorological factors are
the most important influencing factors for bus load forecasting. By selecting similar days to select the most relevant factors, the effectiveness of input features is improved as much as possible. In the literature [7], by improving the neural network, solar radiation, ambient temperature, season and other factors are combined with thinking algorithms to effectively improve the prediction accuracy of photovoltaic output. In the literature [8], the neural network method is used to perform principal component analysis on the data, and the principal component is used as the model input, which effectively reduces the prediction error. The above documents specifically solve the basic problems of photovoltaic output and bus load prediction, but the access volume of distributed photovoltaic generator sources is gradually increasing, and the historical output value of a large number of small power distributed photovoltaic generator sources is difficult to control. Under the condition of unknown photovoltaic output, it is a difficult point to study the accurate prediction of the bus load under the influence of distributed photovoltaic power source by directly using the known bus load.

In the prediction of photovoltaics, Liu Nian [9] considered the influence of radiation intensity, temperature and rainfall factors on photovoltaic output. The input contains a lot of similar information, which is easy to cause information redundancy and lead to low prediction efficiency. In the literature [10], the tensor-flow deep learning framework is applied to construct a long-short-term memory artificial neural network prediction model, which lacks consideration of meteorological factors. In the above literature, in the modeling of photovoltaic output prediction or bus load prediction, the characteristics are either too much or too little. Therefore, how to reasonably handle the feature input in predictive modeling is one of the issues that need to be considered at present.

Aiming at the above problems, in order to improve the accuracy of bus load forecasting under the influence of distributed photovoltaics generator, a new method for bus ultra-short-term load forecasting considering the influence of multi-factors is proposed. Firstly, the correlation analysis is carried out on the bus load of different weather types, the meteorological and power consumption behavior of the distributed photovoltaic generator, and the training device is established by XGBoost to sort the importance of the original features of the training. Secondly, according to the feature importance degree, the forward feature selection is carried out to construct the optimal feature subset. Finally, the selected optimal feature subset is used to construct the prediction model for different weather types.

2. Research methods

2.1. The basic principle of XGBoost

Let the model have a $T$ decision tree, sample feature set is $\mathcal{D} = \{(x_i, y_i)\} \in R^m$, $i = 1, 2, 3, \ldots, n$, $n$ represents number of samples, $m$ represents dimension of sample. The prediction model is:

$$\hat{y}_i = \sum_{t=1}^{T} f_t(x_i) = y_i + f_t(x_i)$$

$y_i$ Represents sample raw output value, $\hat{y}_i$ represents the output of the current model at the $i$ sample point, $x_i$ represents input characteristics of the model at the $i$ sample point, $f_t$ represents the correlation between the leaf weight $\omega$ corresponding to the $b$ tree and the tree mapping structure $q(x)$. The objective function can be defined as:

$$obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(f_t) + C$$

(2)
\[ \sum_{i=1}^{n} I(y_i, \hat{y}_i) \] Represents training error, \( \sum_{i=1}^{T} \Omega(f_i) \) represents the sum of the complexity of the tree, \( C \) is a constant. Define the complexity model of the tree:

\[ \Omega(f) = \gamma P + \frac{1}{2} \lambda \| \omega \|^2 \]  

(3)

\( P \) represents number of leaf nodes, \( \| \omega \| \) represents modulus value of leaf node vector, \( \gamma \) represents distinguish difficulty, \( \lambda \) represents regularization coefficient. The objective function can be expressed as:

\[ \text{obj}^{(k)} = \sum_{j=1}^{T} \left( \sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \sum_{i \in I_j} h_i + \lambda \| \omega_j \|^2 \right] + \gamma T \]  

(4)

Define the sample set of leaf node \( j \) as \( I_j = \{ i | q(x_i) = j \} \), and \( G_j = \sum_{i \in I_j} g_i \) and \( H_j = \sum_{i \in I_j} h_i \) respectively represent the sum of the first derivative of the \( j \) node and the sum of the second derivatives, the minimum estimated value of \( \omega \) is \( \omega_j^* = -\frac{G_j}{H_j + \lambda} \), the final form of the objective function:

\[ \text{obj}^{(k)} = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \]  

(5)

The optimal prediction model can be established by constraining the objective function to a minimum.

2.2. Grid search optimization algorithm

The typical XGBoost algorithm model has multiple parameters, and different parameter values can be set to achieve different control of the model. The max_depth determines whether the model will have over-fitting problems, and the appropriate learning rate (eta) value can make the model more robust.

Grid search is an exhaustive retrieval method in the interval. Compared with the particle swarm optimization algorithm [15], which is prone to premature and local optimal problems, it has the advantages of simple, effective and parallelization, which can effectively solve the uncertainty and randomness when selecting parameters.

3. Analysis of Factors Affecting Bus Load of Distributed Photovoltaic Power Supply

The output of a distributed photovoltaic generator is directly related to the type of weather. For the bus A with actual distributed photovoltaic generator, and the load curve with different weather types is shown in Figure 1.
It can be seen from Fig. 1, that between 6 and 18 o'clock of photovoltaic output, the bus load value fluctuates very little under sunny and rainy days; however, the load on cloudy and rainy days is significantly higher than sunny days, and the difference is around 20 MW. In another word, the photovoltaic output of sunny days is 20MW; for sunny and cloudy, the load fluctuation during the period of photovoltaic output is large, and the overall load value is slightly lower than that of rainy days. It’s true that the amount of cloud also has an important impact on the output of photovoltaic power. Overall, for different weather types, the output of distributed photovoltaic power supplies is significantly different, and the bus load has a similar trend.

4. Feature Selection of Bus Short-term Load Forecasting Model

4.1. Original feature set

According to the above analysis, each factor has different degrees of influence on photovoltaic output and bus load. In order to verify the effects of different features, the importance of each feature is analyzed. The original set of experimental choices is shown in Table 1.

Table 1. Original feature set and its meaning

| feature number | Characteristic quantity meaning |
|----------------|---------------------------------|
| f0, f1, f2     | temperature at k moment, temperature at k-1 moment, temperature at k-2 moment |
| f3-f675       | f3 represents load at k-1 moment, f4represents load at k-2 moment, ..., f675represents load at k-96×7 moment |
| f676           | date type at k moment |
| f677, f678, f679 | wind direction at k moment, wind direction at k-1 moment, wind direction at k-2 moment |
| f680, f681, f682 | Wind speed at k moment, wind speed at k-1 moment, wind speed at k-2 moment |
| f683, f684, f685 | humidity at k moment, humidity at k-1 moment, humidity at k-2 moment |
| f686, f687, f688 | air pressure at k moment, air pressure at k-1 moment, air pressure at k-2 moment |
| f689, f690    | latitude of the bus, longitude of the bus |
| f691          | electricity price at k moment |

* Data collection interval is 15 minutes

4.2. Feature importance analysis

Based on historical load data, the analysis shows the importance of features under three weather types: sunny, cloudy, and rainy. Table 2 gives the importance of the first 25-dimensional feature number.
It can be seen from Table 2 that the bus are affected by different weather types, and the same type of characteristics have different effects on the load in different weathers. For example, in the case of sunny weather, the temperature has the greatest influence, among which f0, f1, and f2 are all ranked in the first five dimensions; When cloudy, the wind speed has the greatest influence, where f680 and f681 are in the top three, while the temperature is in the first five; At the same time, humidity has the greatest impact. The results show that it is necessary to carry out feature selection.

Table 2. Most important 25-dimensional feature content

| weather type | top 25 features |
|--------------|-----------------|
| sunny        | f24 f1 f2 f0 f674 f6 f8 f12 f16 f20 f3 f28 f32 f33 f35 f40 f48 f52 f86 f87 f91 f90 f123 f236 f254 |
| cloudy       | f0 f680 f681 f3 f4 f5 f8 f12 f16 f20 f24 f18 f32 f34 f35 f41 f148 f1 f2 f88 f91 f93 f123 f236 f676 |
| rainy        | f0 f677 f674 f3 f4 f6 f9 f12 f16 f20 f22 f18 f37 f34 f35 f341 f148 f52 f676 f188 f11 f93 f123 f36 f683 |

4.3. Forward feature selection
The prediction accuracy of the XGBoost prediction model is used as the decision variable, and the forward feature selection is performed. Figure 2 shows the error curves of MAPE as the evaluation index in the feature selection process under three weather types. It can be seen from the results of Fig. 2 that the increase of the feature dimension makes the prediction error of each curve decrease significantly, but when the feature dimension increases to a certain number, the prediction error no longer decreases or even increases. The comparison shows that the number of optimal feature subsets of each weather type is different under the three weather types, about 280 dimensions in cloudy weather, and about 120 dimensions in the other two weather types.

![Figure 2. Feature selection process](image)

5. Case Analysis

5.1. Parameter optimization
The experimental optimization parameters are eta and max_depth. The parameter range is set: eta takes the range of (0.1, 0.5), and max_depth takes the range of (2, 10). For other parameters: min_child_weight has a value of 6.82, gamma has a value of 0.624, and alpha is set to 0.4 [11]. The results of the parameter optimization process are shown in Figure 3. The position at the triangle in the figure represents the optimal combination of parameters, and the result is that max_depth is 5 and eta is 0.28.
5.2. Analysis of prediction results
The prediction curve of bus A with distributed photovoltaic generator under different weather types is shown in Fig. 4. Sub-picture a) is cloudy weather. Combined with Table 3, it can be seen that during the period between 7 o'clock and 18 o'clock, the change in cloud amount causes the photovoltaic output to fluctuate greatly, causing the bus load to fluctuate greatly, and the demand for the model becomes higher. Although the prediction curve under the new method fluctuates greatly, the overall fit is very high. Subgraph b) and subgraph c) are rainy and sunny days respectively. Since there is no influence of cloud amount change, the output of photovoltaic generator has no large fluctuation, and the bus load does not fluctuate greatly. The MAPE values for three weather types were below 3%.

![Figure 4. Bus load forecasting results under different weather types](image-url)
Table 3 shows the new method and other methods corresponding to the prediction error of Figure 4. The relevant parameters of the SVM according to document 1 are determined, and the relevant parameters of the conventional XGBoost according to document 2 are determined. It can be seen from Table 3 that the new method effectively reduces the MAPE and RMSE values of bus load prediction under different meteorological types. The MAPE value under the new method is decreased by about 33% compared with the traditional method, and the RMSE value is significantly lower than the other two methods.

6. Conclusion
In order to overcome the problems of unknown photovoltaic output, many load influencing factors and difficult prediction in the prediction of distributed photovoltaic power bus, the new method has the following improvements:

- Through feature selection, the adverse effects of redundant features between different meteorological factors and complex power usage behaviors on the model can be effectively avoided;
- The factors affecting photovoltaic output are directly used as input characteristics, which effectively avoids the problem of insufficient power output data of small power distributed photovoltaic generator;
- The XGBoost parameters are optimized by the grid optimization algorithm to improve the prediction accuracy.

The experiment proves that the new method effectively reduces the error of bus short-term load forecasting with the influence of distributed photovoltaic generator, and has good application value.

Acknowledgements
First of all, I would like to thank the State Grid Liaoning Electric Power Company Limited Economic Research Institute for providing experimental data in the accurate forecasting of grid load based on meteorological data under the trend of diversified electricity use (SGTYHT/18-JS-206). In addition, I would also like to thank my brothers and sisters for their patient guidance on the key technologies of the article.

Table 3. Comparison of prediction errors in different methods

|          | SVM      | XGBoost  | Improved XGBoost |
|----------|----------|----------|------------------|
|          | MAPE/%   | RMSE/MW  | MAPE/%           | RMSE/MW  | MAPE/% | RMSE/MW  |
| sunny    | 6.5487   | 13.23    | 3.2482           | 10.38    | 2.8214 | 9.48     |
| cloudy   | 7.0235   | 14.06    | 3.4231           | 11.87    | 2.9672 | 10.98    |
| rainy    | 6.4875   | 14.02    | 3.6548           | 11.23    | 2.5241 | 10.02    |

References
[1] Zhang Yujin, Yang Lingfan, Ge Shuangzhi, Zhou Hangxia. Short-term photovoltaic power forecasting based on Kmeans algorithm and support vector machine [J]. Power System Protection and Control, 2018, 46 (21): 118 - 124.
[2] Zhang Qian, Ma Yuan, Li Guoli, Ma Jinhui, Ding Jinjin. Applications of Frequency Domain Decomposition and Deep Learning Algorithms in Short-term Load and Photovoltaic Power Forecasting[J]. Proceedings of the CSEE, 2019, 39 (08), 2221 - 2230+5.
[3] Li Pin, Qin Fanglu, Wu Yin, Huang Jia. Short-Term Daily Load Curve Forecasting Based on Fuzzy Information Granulation and Multi-Strategy Sensitivity [J]. Transactions of China Electrotechnical Society, 2017, 32 (09): 149 - 159.
[4] Panapakidis, Ioannis P. Clustering based day-ahead and hour-ahead bus load forecasting models [J]. International Journal of Electrical Power & Energy Systems, 2016, 80: 171 - 178.
[5] Niu Dongxiao, Wei Yanan. Short-term Power Load Combinatorial Forecast Adaptively Weighted by FHNN Similar-day Clustering [J]. Automation of Electric Power Systems, 2013, 37 (03): 54 - 57.
[6] Wang Xuewei, Cheng Ruofa, Yang Hongchao, Lü Caiyan. Support Vector Machine Short-term Load Forecasting Based on Similar Days [J]. Computer Measurement & Control. 2018, 26(01): 47 - 50, 83.

[7] Tan Jianbin, Duan Chunyan, Ban Qun. Research on short-term prediction method of photovoltaic power based on improved neural network [J]. Renewable Energy Resources, 2019, 37 (08): 119 2 - 1197.

[8] Hong Songbao, Wang Kanhong, Shi Kaibo, Kong Li, Cao Hui. Short-term output power forecast of Photovoltaic power generation system based on similar day and principal component analysis [J]. Renewable Energy Resources, 2018, 36 (01): 15 - 21.

[9] Liu Nian, Zhang Qingxin, Li Xiaofang. Distributed photovoltaic short-term power output forecasting based on extreme learning machine with kernel [J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30 (04): 152 - 159.

[10] Chen Liang, Wang Xia, Wang Gang. Application of LSTM Networks in Short-Term Power Load Forecasting Under the Deep Learning Framework [J]. Electric Power Information and Communication Technology, 2017, 15 (05): 8 - 11.

[11] Shi Jiaqi, Zhang Jianhua. Load Forecasting Based on Multi-model by Stacking Ensemble Learning [J]. Proceedings of the CSEE, 2019, 39 (14): 4032 - 4042.