Short-term bus load forecasting based on intelligent similar day selection and deviation self-correction

Qiuna Cai¹, Sijie Liu¹, Hui Zhou²*, Yang Wang², Qiaoyu Zhang¹, Binjie Yan¹ and Binghong Su¹

¹ Electric Power Dispatching Control Center, Guangdong Power Grid Corporation, Guangzhou, China
² Beijing Tsintergy Technology Corporation, Beijing, China
zhouhui8525@163.com

Abstract. Short-term bus load forecasting is of great significance for ensuring the stable operation of the power grid and the orderly conduct of electricity market transactions. In view of the problems such as complex bus load conditions and low prediction accuracy of short-term bus load forecasting, we firstly propose an intelligent strategy of the optimal similar day selection by considering the influence factors of short-term bus load in this paper, taking the bus load of similar day as the result of first forecasting step. The deviation caused by first forecasting step is analysed, on this basis, real-time feature vector is constructed by combining real-time meteorological status. Then, this paper presents a deviation self-correction model based on an efficient and accurate machine learning algorithm—XGBoost. The case study of a bus in Guangdong province of China shows that the model and algorithm proposed in this paper have high accuracy and stability in short-term bus load forecasting.

1. Introduction
The power system bus load refers to the sum of the terminal loads supplied by the main transformer of the substations, short-term bus load forecasting (STBLF) exerts a great importance on grid operation analysis, optimization scheduling and stability control [1]. Accurate STBLF results can effectively improve the level of refinement and intelligence of power grid dispatching.

In recent years, researchers have done more in-depth research on bus load prediction, and have made some achievements. In [2], authors study the bus load and meteorological factors related features, then built the bus load forecasting model which is based on numerical weather prediction and load classification. In [3], the concept of virtual bus load is introduced, and presenting a novel virtual bus load forecasting method, which is based on clustering method. In [4], the clustering technology and feed-forward artificial neural network are combined to construct a hybrid intelligent computing model for STBLF. The above literatures have used the system load forecasting method to some extent, however, there is a big difference between the bus load and the system load, for instance, the bus load has the characteristics of small scale, poor stability and more abnormal data. Therefore, the above methods are not fully applicable to STBLF, and the prediction accuracy is difficult to achieve the level of system load prediction.

Another outstanding characteristic of bus load is that it has large amount and wide coverage, loads vary from one bus to another, which leads to one STBLF model is difficult to apply to different bus lines. On basis of this situation, the intersection of load level set and curve shape set are token as the
result of similar day selection in literature [5], and the predictions of STBLF are calculated by an integrated forecasting model. In [6], the historical bus loads are analyzed by means of pattern clustering and pattern recognition, after that, the similar days are selected and the predicted values of historical loads are obtained based on the load derivation method. In [7], similar days of the day to be forecasted are selected based on the daily characteristics, and a method for selecting the combination forecasting model and a STBLF model with variable weights are proposed. The above researches show that the STBLF by searching for historical similar days can not only improve the accuracy of forecasting, but also have strong applicability. Unfortunately, existing methods also have some deficiencies, the main problems are as follows: 1) those methods fail to fully consider the relevant factors affecting STBLF, which increases the difficulty of extracting the most reasonable similar day from historical days; 2) after the similar days are selected, the bus load deviation of the similar day and the forecasting day has not been handled properly, even without further deviation correction, the bus loads of similar day are directly used as the prediction load of the day to be predicted.

To address the above questions, this paper comprehensively considers the date distance, day type, meteorological conditions and other relevant factors, the feature vector of relevant factors of historical days and the day to be predicted is firstly constructed. On this basis, an intelligent similar day recognition strategy is proposed. Then combining with real-time meteorological status, the deviation feature vector of the day to be predicted and similar day at each time point is constructed, and the internal relation between the real-time deviation feature vector and forecasting bus load deviation is trained with efficient machine learning algorithm—XGBoost. Finally, the self-correcting model of STBLF deviation based on similar day selection is proposed, combining the STBLF result of similar day and load deviation compensation to output the final short-term bus load of the day to be predicted.

2. Similar day selection strategy

2.1. Relevant factors feature vector construction
The effective selection of similar days is directly related to the prediction accuracy of STBLF, therefore, reasonable construction of relevant factors feature vectors and accurate selection of similar days are the preconditions for STBLF.

The main correlative factors which influence the bus load change include: day type, daily weather conditions, daily maximum temperature, daily mean temperature, daily minimum temperature, precipitation, humidity and wind speed, etc. The similar day selection strategy is based on the daily feature, and the relevant factors feature vector constructed as follows:
1) Date distance: calculate the date distance between the historical days and the day to be predicted.
2) Daily maximum temperature: Max Temp, (t=1, 2, 3, ..., 96)
3) Daily minimum temperature: Min Temp
4) Daily mean temperature: Mean (Σ Temp)
5) Daily mean humidity: Mean (Σ Hum)
6) Daily cumulative precipitation: Σ Pre
7) Daily mean wind speed: Mean (Σ Wind)
8) Day type: Short-term bus load has the characteristics of weekly periodicity, and the working day bus loads from Monday to Friday are obviously different from the weekend loads. The short-term load of a bus in Guangdong is selected as the research object in this paper, figure 1 shows the heat map of two weeks’ bus load from 2017.07.03 to 2017.07.16. As we can see from the figure, the bus loads are not only different between the working days and the weekend, but also between Saturday and Sunday are significantly different.
Figure 1. The heat map of one bus load in Guangdong of 2017.07.03 ~ 2017.07.16

Therefore, establishing the day type similarity value table as follows:

Table 1. The value table of day type similarity

| Historical day | Working day | Saturday | Sunday |
|----------------|-------------|----------|--------|
| Forecasting day| same day: 1, | 0.4      | 0.1    |
|                | different day: 0.7 | 1        | 0.2    |
| Working day    | 0.4         | 1        | 0.2    |
| Saturday       | 0.1         | 0.2      | 1      |

2.2. Similar day selection method

Based on the above $m$ daily relevant factors, the process of selecting similar days from historical days is as follows:

1) Taking the 30 days before forecasting day and days of the same season last year, a total of $N$ historical days as candidate days, constructing the relevant factors feature vector of each historical day $X_{ik}$ ($i = 1, 2, 3, ..., N; k = 1, 2, 3, ..., m$), the feature vector of forecasting day is $X_0$.

2) Calculating the similarity $r_{i0}$ of forecasting day and each historical day:

$$r_{i0} = \frac{\sum_{k=1}^{m} (X_{ik} - \overline{X_i})(X_{0k} - \overline{X_0})}{\sqrt{\sum_{k=1}^{m} (X_{ik} - \overline{X_i})^2} \sqrt{\sum_{k=1}^{m} (X_{0k} - \overline{X_0})^2}}$$

(1)

in which, $\overline{X_i} = \frac{1}{m} \sum_{k=1}^{m} X_{ik}$, $\overline{X_0} = \frac{1}{m} \sum_{k=1}^{m} X_{0k}$.

3) In this paper, we select the historical day with the largest similarity $r_{i0}$ as the similar day of the forecasting day.

3. Forecast deviation self-correction model based on XGBoost
3.1. **XGBoost algorithm**

The XGBoost algorithm is an efficient intelligent algorithm developed from the Gradient Boosting Decision Tree (GBRT), which is widely used in various fields of machine learning. Compared with the traditional GBRT algorithm, XGBoost can do parallel computing of multi-core CPU, so its computing speed is greatly improved [8]. In addition, when optimizing the objective error function, XGBoost performs second-order derivative expansion, and the accuracy of the model is greatly improved compared with the traditional first-order expansion.

The basic principle and calculation process of the XGBoost algorithm are as follows:

1) Initializing the parameters of the model and the weight of the sample, the weight coefficients of the training sample set are set to be the same.

2) The training samples are iteratively classified, and the error of the $m$th iteration is calculated according to the following formula:

$$
err_m = \frac{\sum w_i I(y_i \neq G_m x)}{\sum w_i}
$$

(2)

in which, $w_i$ represents the weight of the sample $i$; $G_m$ represents the classifier $m$.

3) Calculate:

$$
a_m = \log \left( \frac{(1 - err_m)}{err_m} \right)
$$

(3)

4) The weight $w_i$ of each sample is updated according to the error $err_m$ of the $m$th iteration:

$$
w_i \times e^{a_m I(y_i \neq G_m x)}
$$

(4)

5) After each iteration, samples with large classification errors will be assigned a higher weight, while reducing the correct sample weight.

3.2. **Forecast deviation self-correction model**

The similar day selection strategy proposed in Section 2 of this paper can ensure that the selected similar day is the closest day among the historical days to the forecasting day. The case study below also shows that the bus loads of similar day are very close to the daily bus loads to be predicted. Therefore, in this paper, we take the short-term bus load of similar day as the result of first forecasting step.

Obviously, there will be a certain deviation between the first forecasting result and the actual value, factors that cause deviation at each time point include differences in daily characteristics vectors, differences in real-time meteorological conditions, effects of load growth, etc. When constructing the real-time feature deviations, in addition to the daily characteristics mentioned in Section 2.1, real-time meteorological conditions such as temperature and humidity at each time point should be considered [9].

The real-time feature vector format $Y_{m,t}$ obtained in this paper is shown in Table 2.

| Feature $m$ | Implication          | Feature $m$ | Implication          |
|------------|----------------------|-------------|----------------------|
| $Y_{1,t}$  | Day type             | $Y_{7,t}$   | Daily mean wind speed|
| $Y_{2,t}$  | Daily maximum temp.  | $Y_{8,t}$   | Real-time temp.      |
| $Y_{3,t}$  | Daily min. temp.     | $Y_{9,t}$   | Real-time humidity   |
| $Y_{4,t}$  | Daily mean temp.     | $Y_{10,t}$  | Real-time precipitation|
| $Y_{5,t}$  | Daily mean humidity  | $Y_{11,t}$  | Real-time wind speed |
| $Y_{6,t}$  | Daily cumulative precip. |

Figure 2 shows a forecast deviation self-correction model based on the XGBoost algorithm, this figure also illustrates the overall prediction process of STBLF proposed in this paper.
Start

Identify and repair historical abnormal data:
Initialize parameters: \( D_0 \), \( N_{\text{train}} \)

Select the similar day \( D_1 \) of forecasting day \( D_0 \) by similar day selection program:
GetSimilarDay(\( D_0 \))

Prediction result of first step: \( P_i^{(1)} \)

Calculate feature vector deviation \( \Delta Y_{i,t} \)
load deviation: \( \Delta P_i \)

Obtain the load deviation XGBoost calculating model:
\( \Delta P_i = f(\Delta Y_{i,t}) \)

Calculate the feature vector deviation of \( D_1 \) and \( D_0 \):
\( \Delta Y_{i,t} = y_{i,t}^{D_1} - y_{i,t}^{D_0} \)

Calculate load deviation compensation by XGBoost model:
\( \Delta P_i = f(\Delta Y_{i,t}) \)

Output the final prediction result:
\( \hat{P}_{i,t} = P_i^{(1)} + \Delta P_i \)

End

Figure 2. Calculation process of STBLF

Specific steps are as follows:

1) The data quality of bus load is poor, compared to the system load, therefore, it is necessary to identify and repair historical abnormal data before performing STBLF [1], including bus load data and meteorological data.

2) Initializing program parameters, including: forecasting day \( D_0 \), the number of samples for deviation training \( N_{\text{train}} \).

3) According to the content of the Section 2, the similar day \( D_1 \) of the forecasting day \( D_0 \) is selected, and the bus load of \( D_1 \) is taken as the result of first forecasting step \( P_i^{(1)} \).

4) Taking the \( N_{\text{train}} \) days before forecasting day as the deviation training sample set, selecting similar day respectively, then the daily real-time feature vector deviation \( \Delta Y_{i,t} \), and load deviation \( \Delta P_i \) can be calculated.

5) Training the relationship of \( \Delta Y_{i,t} \) and \( \Delta P_i \) by XGBoost, getting the XGBoost calculating model:
\[
\Delta P_{i,j} = f(\Delta Y'_{\alpha,j})
\]  
(5)

6) Calculating the real-time feature vector deviation \( \Delta Y'_{\alpha,j} \) of forecasting day \( D_0 \) and similar day \( D_1 \), getting load deviation compensation \( \Delta P_{0,j} \), by the formula (5).

7) Combining the result of first step \( P^{(1)}_{0,j} \) and deviation compensation \( \Delta P_{0,j} \), to obtain the short-time bus load prediction result of \( D_0 \):

\[
\hat{P}_{0,j} = P^{(1)}_{0,j} + \Delta P_{0,j} 
\]  
(6)

4. Case study

Taking a bus in Guangdong mentioned in Section 2.1 as the research object, selecting the bus load and meteorological data of the summer of 2016 and June to 15th August of 2017 to construct the historical daily relevant factors feature vectors. The forecast target is the 96-point bus load from 16th August to 22nd August of 2017. The evaluation indexes of the prediction result are mean absolute percentage error (MAPE) of the true load value \( P_t \) and the predict value \( \hat{P}_t \), and daily bus load prediction accuracy \( Z \). The formulas are as follows:

\[
\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{P_t - \hat{P}_t}{P_t} \right| \times 100\% 
\]  
(7)

\[
Z = \left(1 - \frac{1}{n} \sqrt{\sum_{t=1}^{n} e_t^2} \right) \times 100\%
\]  
(8)

In (8), \( e_t \) is reference error, \( e_t = |P_t - \hat{P}_t|/P_{\text{base}} \times 100\% \), \( P_{\text{base}} \) is bus load reference value, for bus of 220kV voltage class, \( P_{\text{base}} \) usually takes 305MW.

According to the method proposed in Section 2, the similar day of 16th August of 2017 is 26th July of 2017. The comparison of the daily feature vectors of such two days is shown in Table 3, the similarity \( r_{0,1} \) reaches 0.9986. It can be seen from Table 3 that the day type and meteorological conditions of the two days are very close, which can verify the effectiveness of the similar day selection strategy proposed in this article.

| Date       | Day type | Maximum temperature/°C | Minimum temperature/°C | Mean temperature/°C | Mean humidity/% | Cumulative precipitation/mm | Mean wind speed/m/s |
|------------|----------|-------------------------|------------------------|---------------------|----------------|-----------------------------|-------------------|
| 2017.08.16 | Wed      | 30.475                  | 23.706                 | 26.204              | 80.94          | 5.54                        | 2.50              |
| 2017.07.26 | Wed      | 30.145                  | 23.688                 | 26.314              | 82.18          | 5.17                        | 2.51              |

The result of the first forecasting step, that is, the MAPE value of the similar day’s bus load on 26th July and the actual value of forecasting day 26th August is 4.70%, and the bus load prediction accuracy \( Z \) is 98.14%, it can be seen that the bus loads of the two days are similar when the daily feature vectors are similar.

On this basis, the load deviation self-correction method proposed in Section 3 is used to obtain the deviation compensation value, after that, the MAPE and \( Z \) reach to 3.31% and 98.82%, which increase 1.39% and 0.58% respectively. The final prediction accuracy is in the leading position in the existing literatures.

Figure 3 is a comparison of the true value, result of first forecasting step and the final prediction value after deviation correction.
In order to verify the validity of the STBLF method proposed in this paper, the bus load for one week from 16th August to 22nd August is continuously predicted, and the current common forecasting method Support Vector Machine (SVM) and the XGBoost algorithm mentioned in this paper are used directly to predict bus load. The results of the three methods are listed in Table 4.

Table 4. Comparison of prediction accuracy of three methods (%)

| Date       | SVM MAPE | SVM Z  | XGBoost MAPE | XGBoost Z | Method proposed in this paper MAPE | Method proposed in this paper Z |
|------------|----------|--------|--------------|-----------|-----------------------------------|-------------------------------|
| 2017.08.16 | 5.97     | 97.58  | 4.18         | 98.43     | 3.31                              | 98.72                         |
| 2017.08.17 | 10.44    | 95.65  | 4.42         | 98.30     | 3.60                              | 98.31                         |
| 2017.08.18 | 5.50     | 97.34  | 6.79         | 97.58     | 3.36                              | 98.62                         |
| 2017.08.19 | 2.93     | 98.47  | 1.44         | 99.37     | 3.64                              | 98.68                         |
| 2017.08.20 | 7.17     | 97.51  | 2.03         | 99.24     | 3.18                              | 98.87                         |
| 2017.08.21 | 16.33    | 93.73  | 3.46         | 98.49     | 3.06                              | 98.84                         |
| 2017.08.22 | 11.43    | 94.41  | 8.89         | 96.55     | 4.37                              | 98.27                         |
| Mean       | 8.54     | 96.38  | 4.46         | 98.28     | 3.50                              | 98.62                         |

Comparing the prediction results of the three methods in Table 4, it is easy to see that the mean value of MAPE and the daily bus load prediction accuracy Z of the multi-day continuously predict using the STBLF method proposed in this paper are 3.50% and 98.62% respectively, which are superior to the SVM and XGBoost predictions. Although the prediction accuracy of XGBoost is slightly higher than the method proposed in this paper on some day, however, the fluctuation of prediction result is large. This is especially worse with SVM predictions, with a minimum MAPE of 2.93% and a maximum MAPE of 16.33%, which can be clearly seen from the figure 4. It shows that the SVM and XGBoost prediction methods are difficult to adapt to the complex conditions in actual production, such as different day types and sudden change in weather.

Such facts illustrate that the STBLF method proposed in this paper not only has outstanding prediction accuracy, but also has stable prediction effect and high reliability.
5. Conclusion
In this paper, an STBLF method based on intelligent similar day selection strategy and prediction deviation self-correction is proposed. Firstly, by establishing the relevant factors feature vector of the forecasting day and historical days, the historical day with the largest similarity is selected as the similar day, and the bus load of the day is taken as the result of first prediction result. Then, the XGBoost intelligent algorithm is used to train the relationship between the real-time feature vector and the load deviation, realizing the load deviation self-correction, to get the final prediction bus load.

The example shows that the STBLF method and the algorithm proposed in this paper can effectively select similar days, greatly improve the prediction accuracy of STBLF, and the operation is simple and reliable. It has important practical guiding significance for STBLF of grid dispatchers.

References
[1] Kang, C.Q., Xia, Q., Liu, M. (2007) Power System Load forecasting. China Electric Power Press, Beijing.
[2] Li, B., Men, D.Y., Yan, Y.Q., et al. (2015) Bus Load Forecasting Based on Numerical Weather Prediction. Automation of Electric Power System, 39: 137-140.
[3] Tong, X., Kang, C.Q., Chen, Q.X., et al. (2014) Virtual Bus Technique and Its Application (II): Virtual Bus Load Forecasting. Proceeding of the CSEE, 34: 1132-1139.
[4] Panapakidis, I.P., Christoforidis, G.C., Papagiannis, G.K. (2015) Hybrid Computational Intelligence Model for Short-Term Bus Load Forecasting. In: 15th International Conference on Environment and Electrical Engineering. Rome. pp. 2029-2034.
[5] Sun, Q., Yao, J.G., Zhao, J. (2013) Short-term Bus Load Integrated Forecasting Based on Selecting Optimal Intersection Similar Days. Proceeding of the CSEE, 33: 126-134.
[6] Zhao, D.M., Wei, J., Zhang, X. (2013) Probabilistic Interval Prediction of Ultra-Short-Term Bus Load Based on Similar Days. Electrotechnical Application, 11: 36-40.
[7] Yin, X.L., Xiao, X.Y., Sun, X.L. (2015) Bus Load Forecasting Model Selection and Variable Weights Combination Forecasting Based on Forecasting Effectiveness and Markov Chain-Cloud Model. Electric Power Automation Equipment, 35: 114-119.
[8] Yang, X.D., Wang, J.M., Zhang, L.N. (2017) Application of XGBoost in ultra-short load forecasting. Electric Drive Automation, 39: 21-25.
[9] Wang, J.X., Zhong, H.W., Lai, X.W., et al. (2017) Exploring Key Weather Factors from Analytical Modelling toward Improved Solar Power Forecasting. IEEE Transactions on Smart Grid, 99: 1-11.