Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Research On The Daily Electricity Forecast Model Based On LightGBM

To cite this article: Shanshan Wu *et al* 2023 *J. Phys.: Conf. Ser.* **2477** 012036

Research On The Daily Electricity Forecast Model Based On LightGBM

Shanshan Wu[1], Haifeng Zheng[2], Chi Chen[3], Kun Zhang[4]

[1]State Grid Energy Research Institute Co.Ltd.

[2]State Grid Energy Research Institute Co.Ltd.

[3]Institute of Computing Technology, Chinese Academy of Sciences.

[4]Economic Research Institute, State Grid NingXia Electric Power Co., Ltd.

[1]wushanshan@sgeri.sgcc.com.cn,[2]zhenghaifeng@sgeri.sgcc.com.cn,[3]chenchi20s@ict.ac.cn,[4]zhangkun@jyy.nx.sgcc.com.cn

Abstract. Short-term electricity demand forecast is of great significance for guiding the Power Grid Company production operation, but the existing forecast mainly concentrated on load forecast, and most of them only use meteorological and date factors as input features, ignoring the economic factors as a basis to support the growth of electricity, also doesn't take into account the influence of recent significant events. To improve the accuracy and rationality of short-term electricity demand forecast, based on the Light Gradient Boosting Machine (LightGBM) model, in addition to the common meteorological and date factors, all kinds of economic and significant events are introduced into the model to forecast the daily electricity in this paper. The results show that economic and significant events provide useful information, making our model more accurate than previous short-term electricity demand forecasts. Moreover, due to the adoption of machine learning algorithms, our model is more stable and accurate than traditional measurement models with the same factors.

1. Introduction

The electricity demand forecast is based on the evolution of itself, combined with the influence of economic, social, meteorological, and other external factors to forecast the future electricity demand in a specific period. According to the length of the forecast cycle, it can be divided into long-term, medium-term, short-term, and ultra-short-term forecasts [1]; according to the forecast indicators, it can be divided into electricity forecast and load forecast.

Existing short-term electricity demand forecast studies mainly focus on load forecast, especially electric energy production forecast [2,3,4], and most of them only consider meteorological factors. For example, meteorological and date-type factors were considered to establish the QRNN model [5]. In [6], temperature and holiday factors were used to establish the combination model of LSTM and lightGBM for short-term electricity load forecast. However, the research on the daily electricity forecast is still limited [7]. With the deepening of electricity market reform, electricity suppliers and power grid enterprises will gradually participate in the spot market transactions, and the importance of daily electricity forecast is becoming more prominent. At the same time, under the background of frequent extreme climate and continuous intensification of the impact of significant events, daily electricity forecast will also become a necessary means for power grid enterprises to ensure safe supply[8,9] and can guide power dispatch strategy formulation[10,11].

In this paper, we adopted the Light Gradient Boosting Machine model. The time, meteorological, economic, and significant events factors are used as the input characteristics, and we used the Pearson correlation coefficient for screening and processing and constructed the LightGBM model to forecast the daily electricity in the next seven days. To compare the accuracy of results, we selected the representative Holt-Winters model[12,13] and the univariate LightGBM model in the same group of daily electricity data for the forecast. The forecast results show that the multivariate LightGBM model has higher accuracy and better effect than the above two models.

2. Features selection

2.1 Multiple input features

The change in short-term electricity indexes is affected by many factors. For example, literature [14] and [15]considers economic factors, government policies, economics, technology, and so on. Literature [16] analyzes the electricity in the two sessions. Literature [17,18,19] analyzes the impact of the epidemic and the Russia-Ukraine conflict on electricity.

Based on the existing literature and the latest situation at home and abroad, we comprehensively consider three influencing factors: meteorological, economic, and significant events. In terms of meteorological factors, we additionally used the apparent temperature[20] as the influencing factor. Compared with the temperature measured by meteorological stations, the apparent temperature(AT) can better reflect the electricity consumption behavior of people due to temperature perception. The lower the apparent temperature, the more effective the release of heating electricity will be. Its calculation Formulas (1) and (2) are as follows :

$$AT = 1.07T + 0.2\,e - 0.65V - 2.7 \qquad (1)$$

$$e = \frac{RH}{100} \times 6.105 \times \exp \frac{17.27T}{237.7+T} \qquad (2)$$

In the formula, T is the air temperature (℃), e is the water vapor pressure (hPa), V is the wind speed (m/sec), and RH is the relative humidity (%).

In terms of economic factors, we consider a series of factors such as GDP, "troika", resident income/expenditure level, and population as the influencing factor set, which comprehensively reflects the basic driving force supporting electricity growth. In terms of significant events, frequent black swan events such as Sino-US economic and trade friction, the Ukraine crisis, and the COVID-19 epidemic have caused a great impact on short-term electricity demand[21,22], and the fluctuation of daily electricity is significantly greater than the annual average level. In this paper, the increasing number of COVID-19 cases and the energy commodity price index were respectively used as proxy variables for the COVID-19 epidemic and the Ukraine crisis, and the hosting of major political events such as "two sessions" and "Winter Olympics" was also considered[16,23]. The specific influencing factors are shown in Table 1.

Table 1. The set of factors affecting the short-term electricity

Factor type	Specific indexes	Index alias	Frequency
date factors	year	A1	/
	month	A2	/
	date	A3	/
	working day or holiday（T/F）	A4	daily
meteorological factors	maximum temperature	B1	daily
	average temperature	B2	daily
	minimum temperature	B3	daily
	wind speed	B4	daily
	humidity	B5	daily
	maximum apparent temperature	B6	daily

Factor type	Specific indexes	Index alias	Frequency
	average apparent temperature	B7	daily
	minimum apparent temperature	B8	daily
economic factors	resident population	C1	yearly
	residential investment value	C2	monthly
	residential investment year-on-year growth	C3	monthly
	real estate investment value	C4	monthly
	real estate investment year-on-year growth	C5	monthly
	industrial added value year-on-year growth	C6	monthly
	cumulative year-on-year growth of industrial-added value	C7	monthly
	cumulative year-on-year growth of completed investment in fixed assets	C8	monthly
	Per capita consumption expenditure quarterly value	C9	quarterly
	Per capita consumption expenditure year-on-year growth in the quarter	C10	quarterly
	Resident per capita disposable income quarterly value	C11	quarterly
	Per capita disposable income year-on-year growth in the quarter	C12	quarterly
	cumulative index of gross regional product	C13	quarterly
	gross regional product for the quarter	C14	quarterly
	year-on-year growth in gross regional product	C15	quarterly
	export amount	C16	monthly
	year-on-year growth in the export amount	C17	monthly
	total retail sales of consumer goods monthly value	C18	monthly
	total retail sales of consumer goods year-on-year growth	C19	monthly
significant events	new confirmed COVID-19 cases number	D1	daily
	monthly average of commodity price indices （Russia-Ukraine conflict）	D2	weekly
	during the two sessions or Olympic Games(T/F)	D3	/

2.2 Feature screening and processing

To avoid failure caused by multicollinearity, we used the Pearson correlation coefficient to calculate the correlation between pairwise input features[24,25], and redundant features were eliminated combined with expert experience. The formula is shown in Formula(3) :

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{s_X}\right)\left(\frac{Y_i-\overline{Y}}{s_Y}\right) \qquad (3)$$

We draw the correlation graph between input features as shown in FIG. 1(a) below. And then we delete one of the two feature variables with a correlation coefficient >0.8. After deletion, FIG. 1(b) shows that almost only the main diagonal has the deepest color.

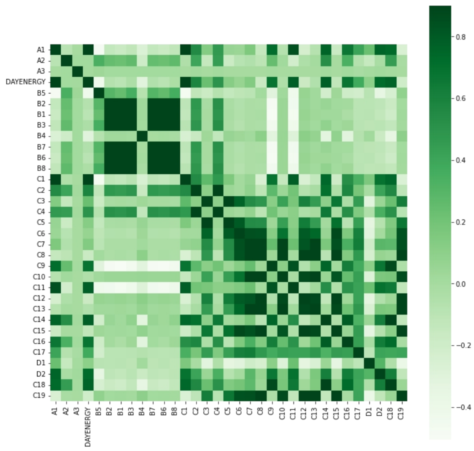

Figure 1(a). Original input feature correlation

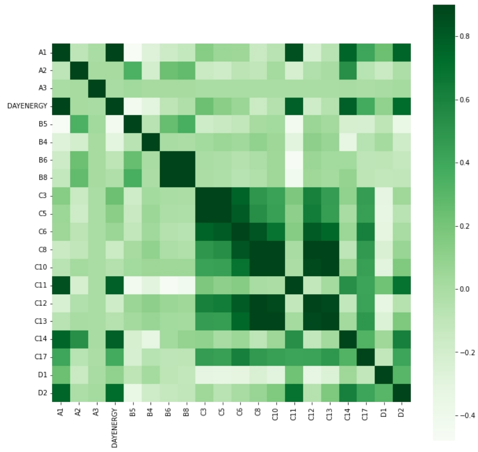

Figure 1(b). Feature correlation after deletion

3. LightGBM daily electricity forecast

3.1 The introduction of the LightGBM model

LightGBM[27] is an open-source framework proposed by the Microsoft team to implement the GBDT algorithm[26] in 2017. Firstly, LightGBM adopts Gradient-based One-Side Sampling (GOSS) to reduce the space and Exclusive Feature Bundling (EFB) to reduce dimension. In addition, it also uses a histogram algorithm and leaf-wise algorithm with depth limitation[28]. Due to a large number of factors and relatively high frequency in the daily electricity forecast, we selected the LightGBM model.

3.2 Construction of LightGBM model

The schematic diagram of the forecast process is shown in Figure 2:

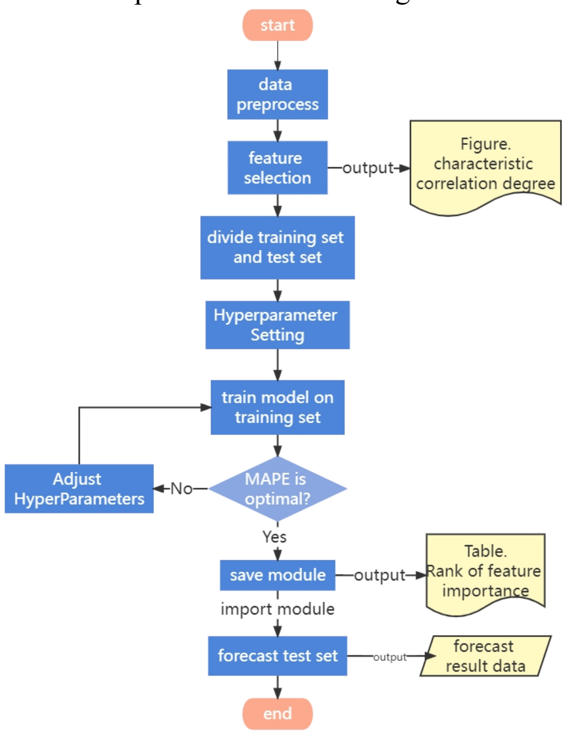

Figure 2. Flowchart of LightGBM model construction and forecast

4. Experimental evaluation

4.1 Experimental platform and data selection

The experimental environment is AMD Ryzen7-4800U, the software platform is implemented by python, and the open-source machine learning framework LightGBM is used.

The experimental data are from the State Grid and typical regions such as Jiangsu, Jilin, Ningxia, Henan, Beijing, and Sichuan, from January 1, 2015 to April 23, 2022, a total of 516 thousand pieces of data.

4.2 Experimental evaluation index

We choose MAPE as the main evaluation index and RMSE as a secondary reference. MAPE is the mean absolute percentage error, the calculation formula is shown in Formula (4) :

$$MAPE = \left(\frac{100}{\text{n}}\right)\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{4}$$

RMSE refers to the root mean square error. Using it as an auxiliary reference means, the calculation formula is shown in Formula (5):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{obs,i} - X_{\text{mod}el,i}\right)^2}{n}} \tag{5}$$

4.3 Hyperparameter Setup

The setting of hyperparameters has a significant impact on the final forecast effect. We use the grid search method[29] to obtain the optimal parameter combination of the model.

The learning rate of the LightGBM model is 0.1, the number of leaf nodes is 23, the maximum depth of the tree is 6, the minimum data amount of leaf nodes is 5, the number of iterations is 2500, and the regularization parameters of L1 and L2 are set as 0.5 and 0.3, respectively.

4.4 Comparison of forecast results in different models

The Holt-Winters model, single-factor LightGBM model, and multi-factor LightGBM model were used to forecast the data of seven typical regions respectively. Each model is divided into the training set and the test set with a ratio of 8:2. The forecast accuracy of the final test set is shown in Table 2. It can be seen that the multi-factor LightGBM model has lower MASE and RMSE than other models.

The average MAPE of the multi-factor LightGBM model constructed in this paper is 1.78%, which is lower than the lowest MAPE value of daily electricity forecast obtained by the combined forecast model in literature [6], which is 2.03%, and also lower than the daily load curve forecast based on QRNN model in literature [5], with an average deviation of 5.81%.

Table 2. Comparison of forecast accuracy of different models

	MAPE/%			RMSE/kw		
region	Mul-LightGBM	Sing-LightGBM	Holt-Winters	Mul-LightGBM	Sing-LightGBM	Holt-Winters
State grid	1.20%	3.34%	1.56%	280933	463299	307504
Beijing	2.57%	3.09%	2.83%	5009	12447	12642
Jiangsu	2.42%	3.73%	2.31%	57691	82197	56151
Jilin	1.75%	2.35%	1.83%	4760	6561	4972
Henan	1.73%	3.65%	5.76%	25656	46313	57101
Sichuan	1.39%	3.39%	2.32%	19323	28551	21867
Ningxia	1.39%	2.18%	1.27%	5401	7493	5857

The above three models are used to forecast the data of several typical regions in a consecutive week (April 24, 2022 to April 30, 2022). The forecast results of the State Grid are shown in Figure 3. It can be seen that the forecast effect of the multi-factor LightGBM model is closer to the real value, while the single-factor LightGBM model shows a large deviation, and the results of the Holt-Winters method is smoother on the whole, which can not effectively deal with the implied nonlinear factors such as periodicity and randomness.

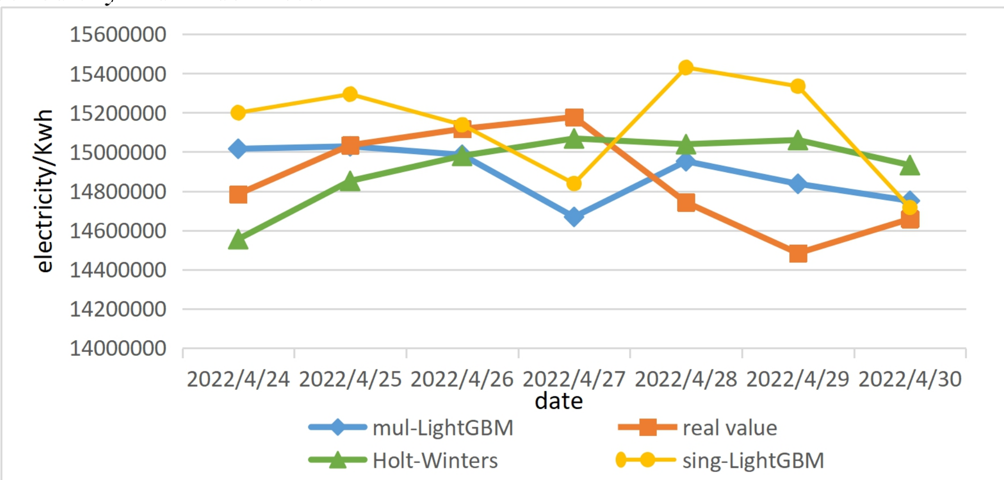

Figure 3. Comparison of the forecast of different models

In the case of multiple factors, we can obtain the weight of each influencing factor as shown in Figure 4. This figure also shows that in addition to meteorological and date factors, significant event and economic factors also play a vital role in the daily electricity forecast.

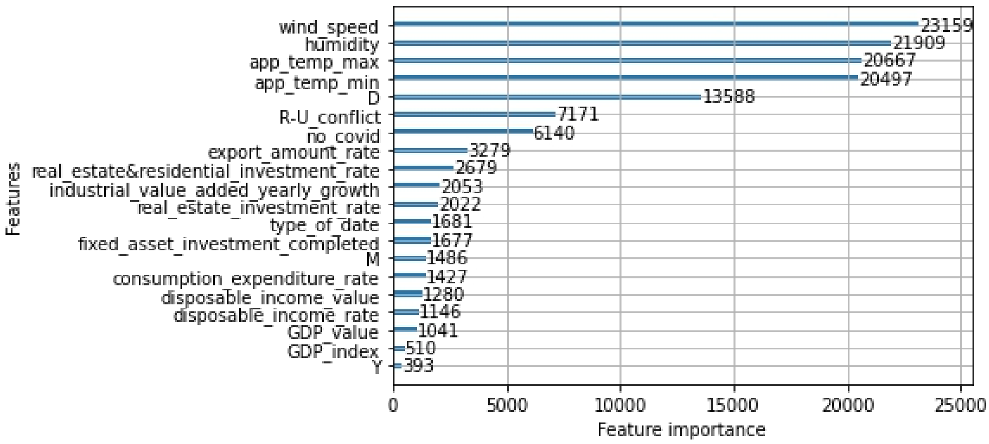

Figure 4. Ranking of the importance of influencing factors in the multi-factor LightGBM model

5. Conclusion

To improve the accuracy of short-term electricity demand forecast, this paper proposes a multi-factor LightGBM forecast model integrating date factors, meteorological factors, economic factors, and significant event factors, which makes up for the inaccurate forecast results caused by only considering the date and meteorological factors in previous work. Moreover, for temperature characteristics, the apparent temperature is used in this paper instead of meteorological temperature, which is more instructive for electricity consumption. The results show that, compared with the single-factor model and the traditional measurement model, the multi-factor LightGBM model has higher forecast accuracy and a more stable forecast effect.

In the future, it can be considered to expand the data set to other regions of the country for forecast, and make classification forecast of electricity consumption, such as residential electricity consumption, agricultural electricity consumption, etc.

Acknowledgments
This work was financially supported by the State Grid Science & Technology Project No. 5100-202257028A-1-1-ZN (Research on energy and power supply and demand interactive simulation technology for the new power system).

References

[1] Dong Shuang. Overview of Short-term Load Forecast Methods for Power System [J]. Heilongjiang Science and Technology Information,2009(29):81.
[2] Hu Yao, Shu Zhengyu, Li Huangqiang, Yao Qin, Li Shichun, Xu Buzhe. Based on grid runoff of radial flow small hydropower power load forecast [J/OL]. Water resources and hydropower technology (both in English and Chinese): 1-11 [2022-11-01]. http://kns.cnki.net/kcms/detail/10.1746.TV.20220921.1930.016.ht ml
[3] Jiang Xuchu, Xu Yucheng, Song Chao. A new method for short-term wind power load forecast [J]. Journal of Beijing Normal University (Natural Science Edition),202,58(01):39-46.
[4] Lin Lin, Huang Yichen, Zhang Yang, Liu Cheng, Yang Junqing. Correlation analysis based on support vector machine (SVM) of wave energy power generation power load forecast [J]. Journal of Nanchang university (science edition), 2019 lancet (5): 504-510. The DOI: 10.13764 / j.carol. cnki NCDL. 2019.05.015.
[5] Yan Wei, Li Dan, Zhu Jizhong, Ren Zhouyang, Zhao Xia, Yu Juan. Probabilistic forecast and stochastic scenario Simulation of Month-ahead Daily Load Curve [J]. Automation of Electric Power Systems,2017,41(17):155-162.
[6] Weinan Chen, Zhijian Hu, Jingpeng Yue, Yixing Du, Qi Qi. Short-term Load Forecast Based on Long Short-term Memory Network and LightGBM Combination Model [J]. Automation of Electric Power Systems,2021,45(04):91-97.
[7] Fang Zhiqiang, Wang Xiaohui, Xia Tong. Electricity forecast model based on long and short-term memory network study [J]. Electric power engineering, 2018 ((03): 78-83. The DOI: 10.19464 / j.carol carroll nki cn32-1541 / tm. 2018.03.014.
[8] CAI Xia, Xing Jun. A review of Load Forecast Methods for Power Systems [J]. Information Research,2010,36(06):5-7+25.
[9] Peng Xiangang, Hu Songfeng, Lu Dayong. A survey of Short-term Load Forecast Methods Based on RBF Neural Network [J]. Protection and Control of Electric Power Systems,2011,39(17):144-148.
[10] Kang Chung-Qing, Xia Qing, ZHANG Boming. Overview and Development Direction of Load Forecast in power Systems [J]. Automation of Electric Power Systems,2004(17):1-11.
[11] Wu Qiao, Lu Shufeng, Yang Shihai, Chen Jian, Li Zhixin. Demand side response under the condition of electricity management research [J]. Jiangsu electrical engineering, 2016, 35 (5): 28-31. DOI: 10.19464 / j.carol carroll nki cn32-1541 / tm. 2016.05.007.
[12] Yang Guohua, Zheng Haofeng, Zhang Honghao, Jia Rui. Short-term Load Forecast Based on Holt-Winters Exponential Smoothing and temporal convolution Network [J]. Automation of Electric Power Systems, 222,46(06):73-82.
[13] Yang Shouhui, Chen Chuanbin, Wang Xuejing, Li Qingwei, Wu Yuanlin, Chen Jing. A Short-term Load Forecast Method Based on Wavelet Transform and Seasonal Holt-Winters Model [J]. Power Demand Side Management,2021,23(05):70-75.
[14] Kong Sheng, Wang Yu, Zhang Chengwei. An empirical study on the impact of the economic crisis on Mid - and long-term electricity forecast [C]//. Proceedings of the 5th (2010) China Annual Management Conference -- Business Intelligence Session.[Publisher Unknown],2010:3-7.
[15] Zhang K N, Gao H B. Power market forecast and analysis [J]. Northeast Electric Power Technology,2001(11):36-40.

[16] Explore the road of high-quality development of the power industry in the new era -- Observation of power energy in the National Two Sessions in 2021 [J]. Contemporary Electric Power Culture,2021(03):10-11.
[17] Shan Baoguo, Ji Xingpei, Xu Chuanlong, Liu Zhilin. Analysis of the recent global energy supply and demand situation and China's energy and electric power supply guarantee strategy [J]. China Electric Power, 222,55(10):1-13.
[18] Liu Zehong, Yan Zhipeng, Hou Yu. Conflict of Ukraine and its influence on the development of the world's energy and enlightenment [J]. Journal of global energy Internet, 2022, 5 (4): 309-317. The DOI: 10.19705 / j.carol carroll nki issn2096-5125.2022.04.001.
[19] Analysis of the impact of the Russia-Ukraine conflict on the global energy industry [J]. Economic Guide,2022(Z1):68-73.]
[20] Xu Shuo, Luan Le, Xu Zhong, Liu Tian, Guo Qianwen. Based on the correlation of apparent temperature time-sharing LSTM load forecast algorithm [J]. Electric technology, 2021 (18): 47-51. DOI: 10.19768 / j.carol carroll nki DGJS. 2021.18.017.
[21] Lu Delong, Guo Juyi, Wu Yang. Power system load forecast method based on multi-source data driven under the influence of COVID-19 [J]. Power supply, 2022, 33 (01) 6:74-80. The DOI: 10.19421 / j.carol carroll nki. 1006-6357.2022.01.11.
[22] Zhao Zhongbu, Yin Haizhao, Wu Houke. Impact of COVID-19 on the overseas business of power enterprises and countermeasures [J]. International Engineering and Labor Service,2021(01):57-59.
[23] Li Chunhui. Brief analysis on the power guarantee of media operation of Genting Ski Park of Beijing Winter Olympic Games [J]. Modern Television Technology,2022(06):55-58.
[24] Chen Gongping, Wang Hong. Personalized recommendation algorithm based on improved Pearson correlation coefficient [J]. Journal of Shandong Agricultural University (Natural Science Edition),2016,47(06):940-944.
[25] Xu Weichao. Review of correlation coefficient research [J]. Journal of Guangdong University of Technology,2012,29(03):12-17.
[26] Zheng Kaiwen, Yang Chao. Decision tree (GBDT) short-term load forecast based on the iterative study [J]. Journal of Guizhou electric power technology, 2017, 20 (02): 82-84 + 90. DOI: 10.19317 / j.carol carroll nki. 1008-083 - x. 2017.02.019.
[27] Wang Huayong, Yang Chao, Tang Hua. GBDT short-term load forecast based on LightGBM improvement study [J]. Automation instrument, 2018, 33 (09) 6:76-78 + 82. DOI: 10.16086 / j.carol carroll nki issn1000-0380.2018020003.
[28] Zhou Binbin, Jiang Yan, Zhao Zhenyu, Duan Ruiqin, Liu Liming. Research on LightGBM Ultra-short-term Load Forecast Based on Particle Swarm Optimization Algorithm [J]. Energy and Energy Conservation,2021(02):2-6.DOI:10.16643/j.cnki.14-1360/td.2021.02.001.
[29] Liu Jiaxing. Support vector regression based on grid search hyperparameter optimization [J]. Science and Technology Innovation,2022(13):71-74. (in Chinese)