Intelligent Identification of Potential Customers for Electricity Substitution

Chen Xiaofeng, Zhao Yadi[1], Wei Zhifeng, Gao Bingqiang

State Grid Information & Telecommunication Accenture Information Technology Co., Ltd, Beijing, 100032, China

Abstract. This paper proposes an intelligent method for the identification of potential customers for electricity substitution. This is developed on the basis of an original model, where a related indicator system of potential customers is constructed through exploratory analysis, improving the results. At the same time, ANOVA is used to screen the indicators and the XGBoost algorithm is employed to output the index importance score and identify likely electricity substitution customers. This method can accurately identify such customers, accelerate the fundamental transformation of energy development, and adapt to the new strategy of Energy Internet development.

Keywords. Electricity Substitution, Indicator System, XGBoost Algorithm, Accurate Identification

1.Introduction

Energy is the basic guarantee of economic and social development, and energy issues are related to human development, national economy, and people's livelihood [1]. With the rapid development of the global Energy Internet, the State Grid Corporation has proposed the strategic goal of developing an international leading Energy Internet enterprise with Chinese characteristics, as well as promoting the construction of an Energy Big Data Center [2-3]. Guided by both global trends and current strategic objectives, provincial and municipal power companies' in-depth research on commercial value and key technologies under the Energy Internet accelerate the construction of clean, low-carbon, efficient, and safe energy systems[4-6].

Electric energy substitution refers to the use of electric energy instead of coal and fuel energy consumption modes in terminal energy consumption, such as electric heating, agricultural electric irrigation, electric vehicles, etc. [7]. The implementation of electric energy substitution is of great significance for promoting the clean development of energy, and is an important measure to increase the proportion of electricity-coal and reduce air pollution [8]. With the rapid development of information technology and the national development of a new infrastructure strategy, power customers need higher standards and requirements in power products and power supply services [9-10]. Enterprises have begun to explore new methods to improve the efficiency and specialization of power substitution work in enterprises and provide

---

[1] Corresponding Author: Zhao Yadi, State Grid Information & Telecommunication Accenture Information Technology Co., Ltd, Email: 15510266633@163.com

support for the effective development of energy big data services [11-12].Shan, based on the STIRPAT (Stochastic Impacts by Regression on Population, Affluence, and Technology) ridge-regression model, used the size of the resident population, energy prices, and other factors to fit terminal power substitution, and quantitively evaluate the role of each factor [13].Using the K-means algorithm[2],Tu clustered users who have undergone power substitution, and constructed a power substitution user identification model by a cooperative filtering algorithm [14].Luo proposed to establish a network management system for an electric energy substitution project, which is based on Browser/Server(BS) and combines Hypertext Preprocessor(PHP) and MySQL to realize the whole process, with a multi-angle and diversified management of an power substitution project[15].

This paper introduces an intelligent customer identification method for potential power substitution. We use statistical and data analysis methods to construct customer identification indicators that affect power substitution potential and build models using the XGBoost algorithm. From comparisons with machine learning algorithms, we conclude that the model has strong applicability and expansibility, and can be used in marketing and other related business fields, providing theoretical and methodological guidance for the construction of an Energy Big Data Center.

2.Theoretical Methods

2.1 Variance Analysis

The basic principle of ANOVA is that there are two basic sources of difference between the mean of different processing groups: one is the deviation caused by experimental conditions, A, the other is the deviation caused by random error, B. The mean square error of A and B, MSb and MSw, are obtained by dividing by their respective degrees of freedom. If it is a deviation due to random errors within the group,we can establish that each group of samples is from the same population, therefore: $MSb/MSw \approx 1.$ If it is the error in the processing of the experiment, $MSb \gg MSw$. The MSb/MSw ratio constitutes the F statistic, and the F value is compared with its critical value to determine whether the difference between the two groups is due to the treatment of the experiment.

In this paper, we can use ANOVA to analyze the difference of the index before and after the substitution, and determine whether the index change comes from the development of electric energy substitution.Finally we can sort the index by F and P values[3] to determine the importance of the index.

2.2 XGBoost Algorithm

XGBoost is a kind of boosting algorithm; the main idea of boosting is to integrate many weak classifiers together to form a strong classifier. The strategy of XGBoost is

---

[2]K-means algorithm: A clustering method in which similar data members are grouped together.

[3]P values: An important evidence to determine whether the original hypothesis is correct; the smaller the value, the greater the probability of accepting the original hypothesis, which states that there is a difference between two variables.

to integrate many tree models together to form a strong classifier. The process is to constantly add trees and perform feature splitting to generate a tree. The tree model used in XGBoost is the CART (Classification And Regression Trees) model. We add a tree each time, representing learning a new function to fit the residuals of the last prediction. Each tree will fall into a corresponding leaf node and each leaf node corresponds to a score. Finally, we only need to add each tree corresponding to the score to get the predicted value of the sample.

The regression tree generation algorithm is as follows. Input training data set D, output regression tree $f(x)$. A typical CART regression tree produces an objective function:

$$\sum_{x_i \in R_m} (y_i - f(x_i))^2 \tag{1}$$

In the input space where the training data set is located, a binary decision tree is constructed by recursively dividing each region into two sub-regions and determining the output value for each sub-region:

Select the optimal tangent variable j and the tangent point s, calculate:

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \tag{2}$$

Traversing variables j, for fixed segmentation variables j, scan the segmentation points s. Select pairs of numbers (j,s) that minimize Formula(2).

Use the selected number pairs (j,s) to divide the region and determine the corresponding output value:

$$R_1(j,s) = \left\{ x \big| x^{(j)} \le s \right\}, \ R_2(j,s) = \left\{ x \big| x^{(j)} > s \right\} \tag{3}$$

$$\hat{c}_m = \frac{1}{N_m} \sum_{x_i \in R_m(j,s)} y_i \ , \ x \in R_m \ , \ m = 1,2 \tag{4}$$

Continue to perform steps 1 and 2 for both subregions R_1 and R_2 until the termination condition is satisfied.

Divide the input space into M regions $R_1, R_2, ..., R_M$ and generate a decision tree:

$$f(x) = \sum_{m=1}^{M} \hat{c}_m I(x \in R_m) \tag{5}$$

2.3 Error Theory

Using some form of error as an index to evaluate the performance of the model, the main indicators are as follows:

R^2 : The denominator is the degree of discretization[4] of the original data, and the numerator is the error between the predicted data and the original data.

$$R^2 = 1 - \frac{\sum (Y\_actual - Y\_predict)^2}{\sum (Y\_actual - Y\_mean)^2}$$

(6)

Mean Square Error(MSE): The average of the squared errors.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2$$

(7)

Mean Absolute Error(MAE): The average of the absolute error values.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|$$

(8)

Mean Absolution Percentage Error(MAPE): Average of relative errors.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| y_i - \hat{y}_i \right|}{y_i}$$

(9)

3. Descriptive Analysis

3.1 Data Preparation

We extracted basic data from power marketing information system, power information collection system and 95598 customer service system etc. in recent three years, as well as data from online platforms and government departments. The data covers three aspects: customer data, electricity use characteristic data, and other factor data.

(1).Customer data: user number, user name, date of establishment, power supply units, contract capacity, industry category, user classification, whether developing electricity substitution, etc.

(2).Electricity use characteristic data: electricity consumption, electricity load, electricity fee, service voltage,etc.

(3).Other factor data: credit value, extent of industry prosperity, business climate index, geographic position,etc.

3.2 Descriptive Statistic Analysis

There are great differences in electricity consumption in different industries, and the .potential of electricity substitution is also different. For example, the metal manufacturing industry can transform cupola into electric kilns with large capacity and strong replicability, low-voltage residential areas can   adopt electric heating which has

---

[4]The degree of discretization: The degree of difference between the individual values and the average, measuring the fluctuation of the overall data.

a small transformation capacity and a huge group scale. Therefore, industry categories will play an important role in identifying potential customers.

Table 1. Industry Share

	Industry	Percentage of electricity consumption	Percentage of enterprise number
1	Manufacturing industry	42.52%	35.26%
2	Textile printing and dyeing industry	15.33%	12.78%
3	Business and residential services industry	10.67%	16.03%
4	Agriculture, forestry, and animal husbandry	9.63%	15.67%
5	Public utilities and management organizations	5.08%	6.58%
6	Transport and postal warehousing	4.86%	5.27%
7	Other industries	11.91%	8.46%

According to experience, enterprises with small electricity consumption and high percentage of enterprise number have potential for substitution. Table 1 shows the industry share according to the annual electricity consumption and the proportion of different industries. We can see that the manufacturing industry, textile printing and dyeing industry, and business and residential services industry in this area are more developed, accounting for about 70% of the total, and can be used as the target industries. At the same time, the agriculture, forestry and animal husbandry industry has a small peoportion of electricity consumption and a large proportion of the enterprise number, so this can also be used as a target in this research. By carding the standardized data, we preliminarily determine that the enterprises with the characteristics shown in Table 2 are potential customers, and assume that the geographical position, load factor, proportion of valley electricity and other attributes will play an important role in model training.

Table 2. Examples of Important Potential Customer Features

	Indicator Name	Numerical Information		Indicator Name	Numerical Information
Feature 1	Geographical Position	<=20	Feature 2	Geographical Position	>20 and <=30
	Load Factor	3.5%		Load Factor	3.7%
	Proportion of Valley Electricity	>20% and <=35%		Proportion of Valley Electricity	<=20%
	Standard Deviation of E.C.	28-35		Contract Capacity	350-450
				Contamination Index	15-20

The Figure1 shows the 24-hour load curves of enterprises using coal-fired boilers before and after transformation.We can see that the load of the unmodified enterprise from 22:00 to 8:00 the next day is higher than that of the reformed enterprises, and between 8:00 to 22:00 during the day is lower than that of the reformed enterprises.The results show that the load curve of unmodified enterprises is stable all day. However, the load scope of the modified enterprises changes greatly during the entire day, so we can obtain the conclusion that the load index can be used as a key index for the model.
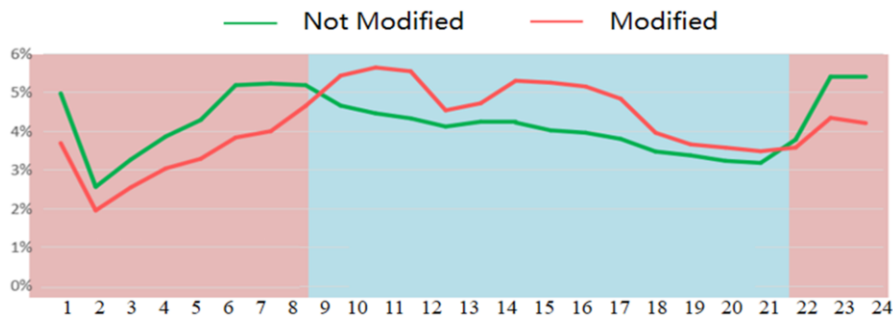
Figure 1. 24-hour Load Curves of Enterprises before and after transformation.

3.3 Indicator System Establishment

According to descriptive statistical analysis, the index system is preliminarily established from basic customer properties, electric characteristics of customers, and other factors as shown in Table 3.

Table 3. Indicator System

Indicator dimensions	Indicator name	Data sources	Indicator description
Basic Customer Properties	Customer Number	Power Marketing Information System	System automatic generation
	Customer Name	Power Marketing Information System	System input
	Household Age	Power Marketing Information System	The opening date until this year
	Power Supply Company	Power Information Acquisition System	System automatic generation
	Contract Capacity	Power Marketing Information System	The user capacity as agreed in the contract.
	Industry Category	Power Marketing Information System	Manufacturing industry/Textile printing and dyeing industry/Business and residential services industry/...
	Customer Classification	Power Marketing Information System	Class I/II/III customers.
	Supply Voltage	Power Marketing Information System	High pressure/Medium pressure/Low pressure.
	Electricity Substitution Implementation	Enterprise Survey	Implemented/Not implemented
Customer Electricity Characteristics	Electricity Type	Power Information Acquisition System	Large Industrial/General Commercial/Domestic/Agricultural
	Electricity Consumption	Power Information Acquisition System	Electricity consumption by day, month, quarter, and year.
	Standard Deviation of Electricity Consumption (E.C.)	Power Information Acquisition System	Standard deviation of electricity consumption by day, month, quarter, and year.
	Peak Electrical Values	Power Information Acquisition System	Average peak value by month, quarter, and year.
	Proportion of Peak Value	Power Information Acquisition System	Frequency of peak values by day, month, quarter, and year.
	Month of Peak Electricity Consumption (E.C.)	Power Information Acquisition System	The month with the highest percentage of electricity consumption.
	Valley Electrical Values	Power Information Acquisition System	Average valley value by month, quarter, and year.
	Proportion of Valley Values	Power Information Acquisition System	Frequency of valley values by day, month, quarter, and year.

Electricity Load	Power Information Acquisition System	Electricity load by day, month, quarter, and year.
Load Factor	Power Information Acquisition System	Ratio of electrical load to maximum load by day, month, quarter, and year.
Electricity Demand Ratio	Power Information Acquisition System	Ratio of actual electricity consumption to maximum requirement by day, month, quarter, and year.
Demand Contract Ratio	Power Information Acquisition System	Ratio of maximum requirement to contract capacity by day, month, quarter, and year.
Electricity Costs	95598 Customer Service System	Average electricity costs by month, quarter, and year.
Standard Deviation of Electricity Costs	95598 Customer Service System	Standard deviation of electricity costs by quarter and year.
Maximum Daily Load Time	Power Information Acquisition System	Maximum load time period in the morning/a.m./p.m./night
Electricity Consumption (E.C.) Growth Rate	Power Information Acquisition System	Electricity consumption growth rate by month, quarter, and year.
Other Factors — Credit Value	95598 Customer Service System	Excellent/Good/Qualified/Unqualified
Industry Prosperity Extent	Power Marketing Information System	Volume of electricity for the whole industry compared with the same period of the previous year.
Geographical Position	Network Map	Distance from a central block of a city in a straight line in kilometers.
Population Density	Regional Statistical Offices	Based on the nearby population of two square kilometres.
Contamination Index	Network Weather	Frequency of higher district pollution index in a month.
Business Climate Index	Power Marketing Information System	Comparison of the company electricity volume with that of the same industry.
Estimated Construction Duration	Enterprise Survey	Estimated construction duration for electrical energy substitution.
Enterprise Economic Benefits	Enterprise Survey	The estimated ratio of the gross product of the enterprise to the cost.

4. Model Application

4.1 Model Construction

Taking the manufacturing industry as an example, the index variables are screened by correlation analysis and ANOVA in the enterprise has undergone electric substitution, and the index importance is obtained by finding the changes before and after transformation, so as to achieve a secondary screening of the index. The obtained index importance ranking is shown in Figure 2. We can see that the indexes ranked higher have darker colors, indicating that they have a larger change for before and after energy substitution, which can be screened as the index to enter the model.
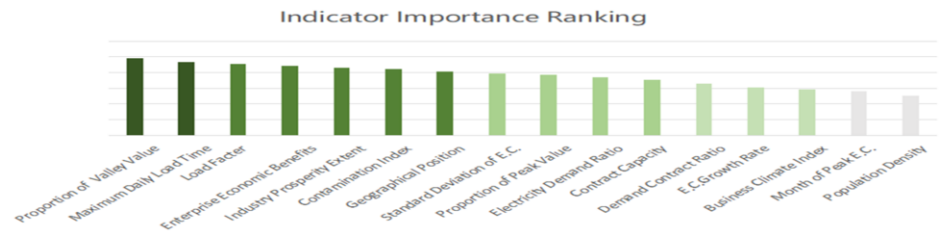

Figure 2. Indicator importance ranking.

In order to compare different data mining algorithms, we introduce five segmentation cross validation to test the generalization ability of the model. We divide the data into five parts, using four groups as training sessions to develop a model, and then using the remaining set as the test set, which requires five training sessions. We use linear SVMs, Random Forests, XGBoost, and derivative algorithm such as LSTM[16], DeepST[17] using the five segmentation cross-validation method; the validation results are shown in Table 4. We can see that the XGBoost algorithm has the least MSE and MAPE which fits the data well as well as faster than others, so we use the XGBoost algorithm to predict potential electricity substitution.

Table 4. Algorithm Evaluation

	Algorithm	R^2	MSE	MAE	MAPE	Calculation times (s)
1	Support Vector Machine (SVM)	0.79	6.314	4.369	0.173	0.560
2	XGBoost	0.95	6.232	4.325	0.098	0.512
3	Random Forests	0.77	6.398	4.483	0.109	0.893
4	LSTM Algorithm	0.83	6.218	4.619	0.204	0.501
5	DeepST	0.91	6.292	4.332	0.126	0.439

We select the top eight indexes for the model, 70% of the estimated enterprise data with electricity substitution potential is extracted as the training set, with 30% as the test set. The model is constructed based on XGBoost to identify potential customers, and at the same time output indicator weights. The potential customer identification output results are shown in Table 5, and the index weight output results are shown in Figure 3. We standardize the output results to the range of 0-100; customers with a score of more than 65 are regarded as potential alternative enterprises represented by the number "1" , and the rest by the number "2".

Table 5. Customer Identification Output

Enterprise Serial Number	Load Factor	Maximum Daily Load Time	Proportion of Valley Value	Business Prosperity Index	Geographical Position	Contamination Index	Enterprise Economic Benefits	Standard Deviation of E.C.	Output
1	3.5%	p.m.	29%	75	50	360	41	28	1
2	4.3%	a.m.	41%	46	79	150	26	56	2
3	3.7%	p.m.	42%	51	58	250	21	19	1
4	3.8%	p.m.	28%	80	34	280	37	22	1
5	4.0%	morning	45%	46	19	320	22	61	2
6	3.7%	night	43%	45	65	308	32	26	1
7	3.8%	p.m.	27%	70	28	388	41	31	1
8	4.1%	a.m.	48%	70	22	115	25	51	2
9	4.6%	a.m.	56%	40	26	158	23	53	2
10	4.3%	morning	51%	48	32	133	20	55	2
11	3.6%	p.m.	42%	60	36	105	18	58	2
12	4.0%	p.m.	51%	37	46	428	29	41	1
13	3.7%	p.m.	42%	65	22	288	36	38	1
14	4.3%	a.m.	53%	43	32	110	17	49	2

15	3.8%	p.m.	48%	45	33	335	29	30	1
16	4.1%	night	40%	51	43	156	26	24	1
17	3.3%	p.m.	29%	78	47	350	46	29	1
18	3.5%	p.m.	38%	36	51	320	28	23	1
...	...	...	...	...	...	...	...	...	...
1598	3.7%	p.m.	45%	42	41	332	25	30	1
1599	3.5%	a.m.	36%	62	37	309	38	25	1
1600	3.4%	a.m.	43%	74	39	391	29	25	1
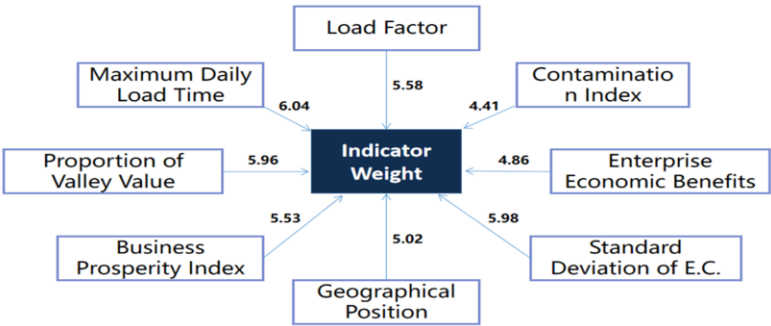
Figure 3. The index weight output results.

We can determine that 1039 of the 1600 companies are predicted as potential replacement customers. At the same time, the weight values of six indexes are greater than 5, which can be regarded as important explanatory variables. We can also derive the conclusion that when the load factor is small, the maximum daily load time is in the night, the business prosperity index is high and the proportion of valley value, the geographical position, and the standard deviation of electricity consumption values are low, we think that the enterprise has a higher likelihood of developing electricity substitution than others.

4.2 Model Evaluation

For the evaluation of the classification model, we use performance metrics. The cross matrix between the predicted results of the model and the actual values is shown in Table 6.

Table 6. Forecasting Cross Matrix

		Actual Result	Actual Result
		0	1
Prediction Results	0	True Negative (TN)	False Negative (FN)
Prediction Results	1	False Positive (FP)	True Positive (TP)
Total Number		Negative (N)	Positive (P)

We define Accuracy = (TP+TN)/(N+P), Precision = TP/(FP+TP). The prediction results of the model are shown in Figure 4. We can conclude that the accuracy of the model is 83.00%, the precision is 91.72%, which indicate that the model is accurate and can be applied in practice.
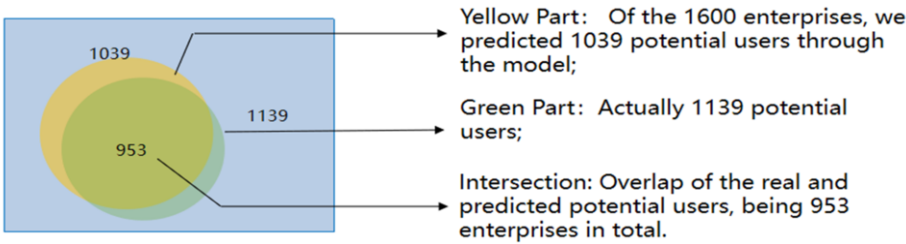
Figure 4. Prediction results.

We have also compared the method proposed in this paper with Collaborative Filtering Algorithm in reference[14], SVM, LSTM and DeepST algorithms; the results are shown in Table 7. From this we can see that in the case of equivalent precision, compared with other algorithms, XGBoost has a higher correct rate, which can be extended to more applications.

Table 7. Algorithm Evaluation

	Algorithm	Precision	Accuracy
1	XGBoost	91.72%	83.00%
2	Collaborative Filtering Algorithm	96.23%	68.73%
3	Support Vector Machine (SVM)	87.65%	58.93%
4	LSTM Algorithm	89.31%	64.39%
5	DeepST	85.09%	67.33%

4.3 Business Applications

Technology for identifying customers most likely to adopt electricity substitution can be used in all aspects of enterprise data applications, including intelligent operation of grid enterprises, modernization of government governance, sustainable development of social economy, and so on.

4.3.1 Intelligent Operation of Grid Enterprises

This technology can help the implementation of enterprise label libraries and energy big data center construction. Figure 5 shows the business management process of power substitution; we tap customer demand through data analysis and sharing application, actively propagate power substitution policy to customers with high potential, and improve the company's intelligent operation efficiency.
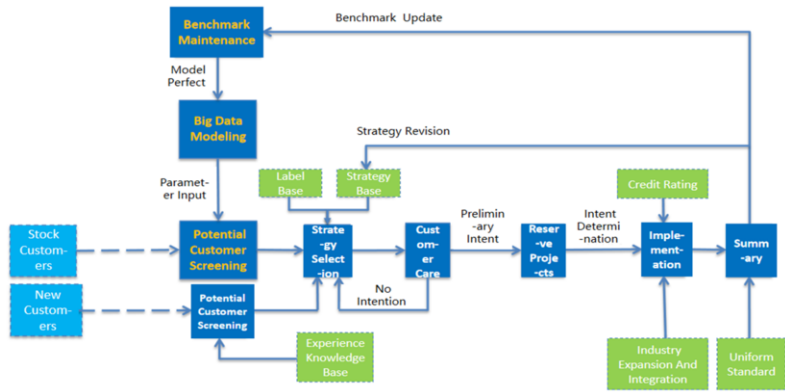
Figure 5. Business management process of power substitution

4.3.2 Modernization of Government Governance

Electric substitution customers identification technology can help the government to implement preferential policies and urge enterprises to transform equipment. We give users economic subsidies to mobilize the enthusiasm of users to use clean energy.

4.3.3 Sustainable Development of Social Economy

The development of electricity substitution identification technology is conducive to improving the level of national electrification and improving the quality of life. At the same time this can cultivate and develop new industries, new forms of business, and new models for the development of the power economy. References[18-19] exploit sophisticated models to capture hidden traffic and environment characteristics from substantial historical mobility data and then makes use of trained models to predict traffic conditions and environmental quality in the future. In the future, we can apply the above research results to this paper ,increase the traffic conditions and environmental quality indicators to make the results more accurate.

5.Conclusion

Based on the XGBoost algorithm, we establish a full set of indicators through descriptive analysis, which makes the calculations quantifiable and accurate; an accuracy of 83% can be achieved. At the same time, compared with other algorithms, it is concluded that the method has good application potential and can be applied to various business areas such as urban computing, follow-up we can enhance further discussion of related work[20-21].

Reference

[1] WEI Qinglai, SHI Guang, SONG Ruizhuo, et al. Adaptive Dynamic Programming-Based Optimal Control Scheme for Energy Storage Systems With Solar Renewable Energy[J]. IEEE Transactions on Industrial Electronics, 2017, PP(99):1-5478.

[2] DONG Z, ZHAO J, WEN F, et al. From Smart Grid to Energy Internet: Basic Concept and Research Framework[J]. Dianli Xitong Zidonghua/automation of Electric Power Systems, 2014, 38(15):1-11.

[3] RAO Wei, JIANG Jing, ZHOU Aihua, et al. Research on Power Big Data Basic Architecture and Standard System for Global Energy Interconnection[J]. Electric Power Information and Communication Technology, 2016(04):21-29.

[4] TIAN Shiming, LUAN Wenpeng, ZHANG Dongxia, et al. Energy Internet Technology Form and Key Technologies[J]. Proceedings of The Chinese Society for Electrical Engineering, 2015,35(14):3482-3494.

[5] ZHOU Lin, FU Xueqian, LIU Shuo, et al. Day-Ahead Market Clearing Model of Integrated Energy System for Promoting Renewable Energy Consumption[J]. Electric Power, 2019(11):11-18.

[6] WU Jielong, YANG Jian, et al. Research on the Current Situation and Development Trend of the Technology of Energy Internet[J].Power System and Clean Energy, 2016,032(003):8-12.

[7] SUN Yi, ZHOU Shuang, SHAN Baoguo,et al. Analysis of Electric Power Alternative Potential Under Multi-Scenario[J]. Power System Technology, 2017(01):126-131.

[8] FU Chengran, CHEN Rongjun, HE Yongxiu. Analysis on Electricity Substitution Economics of Multi-Agent Electric Cooker[J]. northeast electric power technology, 2018(04):16-23.

[9] Li Boyang , Kisacikoglu M C , Liu Chen , et al. Big Data Analytics for Electric Vehicle Integration in Green Smart Cities[J]. IEEE Communications Magazine, 2017, 55(11):19-25.

[10] XUE Yusheng, LAI Yening, et al. Integration of Macro Energy Thinking and Big Data Thinking Part One Big Data and Power Big Data[J]. Automation of Electric Power Systems, 2016,40(1):1-8.

[11] WU Kehe , YAO Peng. Research on Energy Substitution Terminal Energy Application Based on CRITIC Algorithm[C]// 2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA). IEEE, 2017(10):34-38.

[12] XIA Huaijian, LIN Haiying, ZHANG Wen, et al. Research on Development Potential of Regional Electric Energy Substitution Based on Multi Model[J]. science and technology management research, 2018(04).

[13] SHAN Baoguo, ZHAO Jia, JIA Dexiang, et al. Electric Power Substitution Potential Analysis Method Based on STIRPAT-Ridge Regression[J]. Distribution & Utilization, 2018,035(001):68-73.

[14] TU Ying, LIU Qiang, WANG Qingjuan,et al. Research on Potential User Mining Model of Power Substitution Based on Collaborative Filtering Algorithm[J]. electric power information and communication technology, 2017(12):11-18.

[15] LUO Jun, DING Canhui, et al. Development of Electric Energy Alternative Information Management System Based on BS Technology[J]. Popular Utilization of Electricity, 2017(03):25-26.

[16] Essien A, Petrounias I, Sampaio P, et al. A deep-learning model for urban traffic flow prediction with traffic events mined from twitter[J]. World Wide Web, 2020(8):13-38.

[17] Wang Leye , Geng Xu , Ma Xiaojuan , et al. Cross-City Transfer Learning for Deep Spatio-Temporal Prediction[J]. 2018.

[18] LIU Lun, Elisabete A.Silva, WU Chunyang, WANG Hui. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment[J]. Computers, Environment and Urban Systems, 2017,65: 113-125.

[19] LIU Zhidan, LI Zhenjiang, WU Kaishun, LI Mo. Urban Traffic Prediction from Mobility Data Using Deep Learning[J]. IEEE Network,32(4):40-46.

[20] QI Zhongang, WANG Tianchun, SONG Guojie, et al. Deep Air Learning: Interpolation, Prediction, and Feature Analysis of Fine-Grained Air Quality[J]. IEEE Transactions on Knowledge & Data Engineering, 2018:1-1.

[21] Gomes E F , Batista F . Classifying Urban Sounds using Time Series Motifs[C]// Security-enriched Urban Computing & Smart Grids. 2015(97):52-57.