Advances in Computer Science Research, volume 88
International Conference on Computer, Network, Communication and Information Systems (CNCI 2019)

Customer Identification of Potential Energy Substitution Based on Big Data Method

Yunzhao Li[1, a]
[1]State Grid Xinjiang Electric Power Corporation, Wulumuqi, 830000
[a]406839653@qq.com

Abstract. The usage of clean energy such as electric energy will help promote energy conservation and emission reduction, optimize energy use structure and improve energy efficiency. The realization of the "electric energy substitution" strategy can realize the replacement of loose coal and direct fuel in the terminal energy consumption, and finally realize the fundamental transformation of the energy development mode. This paper analyzed the customer characteristics of electric energy substitution potential from multiple dimensions, built a multi-dimensional linear electric energy substitution evaluation prediction model based on logistic regression algorithm, and quantified the output customer comprehensive score based on the obtained model results to identify high potential customers. At the same time, from the perspectives of government, power grid and enterprise, we utilized big data mining technology to analyze characteristics of high-potential customers, tapped customer demand characteristics, and achieved precise services.

Keywords: Electrical energy substitution, Logistic regression algorithm, Score-card algorithm.

1. Introduction

At present, China's thermal coal proportion and electrification level are low, and a large amount of loose coal and fuel consumption is one of the main factors causing severe haze. The replacement of electric energy is in the end energy consumption link, using electric energy instead of the energy consumption mode of loose burning coal and fuel oil, such as electric heating, industrial electric boiler agricultural electric irrigation and drainage, electric vehicle, using shore power and electric storage can adjust the peak and so on. Electric energy has the advantages of being clean, safe and convenient. Implementing electric energy substitution is of great significance for promoting the energy consumption revolution, implementing the national energy strategy, and promoting the development of clean energy. It is an important measure to increase the proportion of coal, control the total consumption of coal and reduce air pollution[1-2].The eight ministries and commissions jointly issued the "Guiding Opinions on Promoting the Replacement of Electric Energy"(Development and Reform Energy[2016] No.1054)(hereinafter referred to as "Opinions"),the important significance, general requirements, key tasks and The guiding opinions were put forward in four aspects of safeguard measures, which provided a policy basis for comprehensively promoting electric energy substitution [3].Steadily promoting electric energy substitution is conducive to building a new power consumption market with higher levels and wider scope, expanding power consumption, improving China's electrification level, and improving people's quality of life. At the same time, it will drive the development of related equipment manufacturing industry and expand new economic growth point.

In recent years, in order to thoroughly implement the national demand for energy consumption revolution and prevention of air pollution, the State Grid Corporation has actively adapted to the new normal of economic development, increase marginal benefits, optimize energy structure and promote energy conservation and emission reduction. In accordance with the government-led, grid-driven, and social-participatory work guidelines, the company adheres to the principle of giving priority to benefits, piloting first and highlighting key points and vigorously implements the electric energy substitution strategy. According to the "State Grid Corporation's Opinions on Innovation and Development of Electric Energy Substitution" (State Grid Marketing[2018] No.89),"State Grid Marketing Department's Notice on Printing Power Generation Alternative Work Rules" (Marketing

Market [2017] No.17) and other documents In 2018, under the new normal of economic development, all power supply units should continue to carry out electric energy substitution, increase supply and expansion as the core work to enhance the company's efficiency and support the company's development, unswervingly push forward.

At present, when power supply enterprises promote electric energy substitution work, they adopt traditional house-to-house inspection methods, which have low efficiency and low actual work success rate; and equipment replacement, cost accounting, policy adjustment and other issues related to electric energy substitution [4],which puts forward higher requirements for staff [5-6]. Enterprises need to explore new methods, research new technologies, analyze the characteristics of electric energy replacement enterprises in depth, carry out special electric energy substitution consulting projects in combination with regional realities, use big data technology to accurately locate potential user groups, and improve the efficiency and professional level of electric energy substitution work[7],which provide support for the effective and in-depth development of the company's energy replacement.

2. Current Status of Electric Energy Substitution Research

State Grid Xinjiang Electric Power Co., Ltd. formulated the "Guiding Opinions on Electric Energy Substitution in 2018" (New Electric Camp [2018] No.47) in the 2018 marketing key work. The "Opinions" clearly put forward the key tasks of electric energy substitution in 2018, and also gives a plan for the decomposition of electric energy substitutes in 2018.At the same time, the State Grid Xinjiang Power Marketing Key Work "Guidelines for Marketing Informationization Work in 2018" proposed that exploring the application of big data and promoting "Internet +" marketing services are two aspects of the work in 2018.Utilize big data analysis technology to continuously improve customer service level, service efficiency and service accuracy, and provide big data support for management decision-making. At the same time, relying on the company's big data platform and marketing massive basic data, through label combing, customer portrait, self-service analysis integration with data, customer segmentation, electricity cost risk prevention, default electricity analysis and thief analysis are the entry points, a customer tag library system covering application modules such as billing analysis, load analysis, safety analysis and electric energy replacement is built. China should fully apply Big Data Technology[8-9]to improve the level and effectiveness of refined service for electric energy replacement work, and accelerate the completion of the “Working Program for Accelerating the Electrification of Xinjiang”(New Deal Office [2016] No.161) During the 13th Five-Year Plan period, the overall tasks and objectives of electric energy replacement work will improve the level of electrification and promote steady economic growth.

Research on the potential of electric energy replacement can provide strong support and guidance for guiding electric energy replacement work. Sun Yi et al.[10] used the decoupling theory model to determine various model parameters under multiple scenarios, and analyzed the future medium- and long-term terminal energy substitution potential under various scenarios. Based on the electric energy replacement quantity emission time series data and the STIRPAT model, Shan Yuguo et al.[11]used ridge regression to obtain a multivariate linear model of terminal energy replacement and resident population, per capita GDP, energy price, etc.to predict the potential of electric energy replacement. Sun Yi et al.[12]defined the main influencing factors of electric energy substitution development based on multi-dimensional data, realized the fitting analysis of influencing factors and cumulative electric energy substitution by support vector machine, and optimized the particle swarm optimization for support vector machine parameters. Tu Ying et al.[13]clustered the users who have completed the electric energy substitution modification through the K-means algorithm, and used the collaborative filtering algorithm to construct the electric energy replacement potential user identification model to achieve accurate positioning of the energy substitution potential enterprises.

This paper introduces a method of customer identification of electric energy substitution potential. Firstly, some users who have completed the replacement of electric energy are trained by the decision

tree algorithm [14],and some rules for identifying potential customers are obtained initially, which provides a theoretical basis for constructing the indicators of the model. At the same time, based on the logistic regression algorithm, the surrogate potential customer prediction model is constructed, and the score of the customer is calculated by the scorecard function to predict the customer's potential level.

This method can provide an efficient method for Xinjiang electric energy replacement work, and provide a theoretical basis for the company to accurately identify potential customers and develop differentiated measures.

3. Main Methods

3.1 Optimal Variable Grouping Algorithm

Variable grouping is the process of merging certain categories of categorical variables to reduce their cardinality, or segmenting a numeric variable into a categorical variable. The method is based on the splitting of the decision tree model to find the optimal grouping scheme, and the predictive power index is maximized by combining the categories of variables. That is, the optimal binary segmentation point is first found by the principle of maximizing a certain predictive power index, and then the previous step is repeated in each subcategory, and the segmentation is stopped when the maximum number of packets is reached.

3.2 WOE (Weight of Evidence) Evidence Weight Conversion Algorithm

After grouping the variables included in the model, it is necessary to adopt a reasonable coding method for the variables, and convert the categorical variables into numerical variables to reduce the number of dummy variables in the model construction; and to facilitate the subsequent conversion of the logistic regression model to the standard score. Card format to facilitate interpretation and application of model results. For the i-th group of a classification independent variable, the calculation formula of WOE is as follows (1):

$$WOE(x_i) = In\left(\frac{p_i}{q_i}\right) = In\left(\frac{n_{i1}/n_{*1}}{n_{i2}/n_{*2}}\right) \tag{1}$$

This value represents the ratio of the responding customer to the unresponsive customer in the current group and the difference in this ratio across all samples. The larger the WOE, the greater the difference. For continuous variables, the WOE values must be arranged in the natural order of the segments to maintain the sequential relationship contained in the continuous original variables.

3.3 Logistic Regression Algorithm

Logistic regression is a study of dichotomous variable Y and a series of influencing factors Xn, the multivariate analysis method of the relationship is further developed on the basis of the linear model. Its general form is: $p = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$, $P = \frac{1}{1+e^{-p}}$. In the formula, P is a probability occurrence of the variable Y, between 0-1.Logistic regression model has fast calculation speed, obvious results and good fitting effect. It is widely used in big data, machine learning, economics and other fields.

3.4 Clustering Algorithm

Cluster analysis is the classification of individuals or objects so that the similarities between objects in the same class are stronger than those of other classes. Its purpose is to maximize the homogeneity of objects within classes and to maximize the heterogeneity of classes and classes. K-means clustering is used for sample clustering, such as the division of different industries; system clustering is more suitable for variable clustering, and more for pre-processing of variables before modeling. The input is the index variable value and the corresponding threshold value, and the output is outputting the corresponding clustering result and the clustering map according to the index variable.

（1）System clustering

First calculate the distance between two indicators d_{ij}, referred as D=(d_{ij}) , structure n classes, each class contains only one sample; secondly ,the two closest classes are a new class and the updated

distances are calculated; finally, if the number of classes is equal to 1,the clustering graph is drawn and decided; if not equal to 1, return to the second step to continue.

（2）K-means clustering

First randomly selected k observations as the centroid (seed) of the initial cluster; secondly calculate the distance of each observation from the *k* centroids (such as the Euclidean distance), assign each observation to the nearest centroid class, when all the observations are classified, the position of the centroid of each category (total of k) is recalculated; finally, the second step is repeated until the centroids of the k categories are unchanged.

4. Model Construction and Application

4.1 Customer Definition

The specific energy replacement potential customers are defined as follows:

(1) Senior potential customers (customers with a combined score above 75 are defined as having the potential to replace customers);

(2) Intermediate potential customers (customers with a combined score between 51 and 75, defined as having the potential to not replace the customer);

(3) Non-potential customers (customers with a combined score<50, defined as customers without alternative potential). For author/s of only one affiliation (Heading 3): To change the default, adjust the template as follows.

4.2 Construction Ideas

Through the data extraction of customer application system,95598 business support system,SG186 marketing system, electricity load data energy service management platform and other business application systems, the customer characteristics analysis of electric energy substitution potential is analyzed from multiple dimensions, and multi-dimensional linear electric energy substitution evaluation and prediction is established in order to identify high potential customers. At the same time, based on big data mining technology, it analyzes the demand characteristics of high-potential large customers, and explores customer demand characteristics to achieve accurate services.

(1) Analysis of customer characteristics of electric energy substitution potential.

From the perspectives of government, power grid and enterprise, we conduct customer evaluation of electric energy substitution potential and construct an alternative evaluation prediction model. Clustering analysis based on indicators such as customer attributes in the sample, focusing on industries with more electricity consumption and enterprises. Exploratory analysis of customer behaviors and attributes of key industries and enterprises, in-depth exploration of key indicators related to alternative potential, and construction of model indicator systems.

(2) Establish a prediction model for electric energy substitution assessment

On the basis of constructing the indicator system, the characteristics of the enterprises to be predicted and the benchmarking enterprises are compared and analyzed through the model. Enterprises with similar characteristics as the benchmarking enterprises are found as enterprises with alternative potential. Using the methods of correlation analysis and analysis of variance, the index variables are screened to determine the final model. Based on the logistic regression model output variable weights, the scorecard algorithm performs weighted summation, obtains the electric energy to replace the customer comprehensive score, determines the level threshold according to the score distribution and outputs the customer grade result.

(3) Develop customized service strategies

According to the characteristics of high-level electric energy replacement potential demand, we will formulate personalized and customized alternatives to electric energy, improve customer satisfaction, improve the market competitiveness of power companies and innovate the value growth of enterprises.

4.3 Data Preparation

It mainly extracts relevant data from a provincial and municipal power marketing system, 95598 system, power collection back-end system and online platform. According to the previous business understanding, visiting business experts and business research results, theoretically looking for high substitution potential customer related variables, and getting a preliminary indicator set, including:

(1) Customer basic attributes: user number, user name, date of establishment, account name, power supply unit, contract capacity, industry category, user classification, etc.;

(2) Power consumption characteristics: line margin, power load, peak power month, power supply voltage, capacity increase rate, capacity reduction rate, etc.

(3) External competition index: business climate, geographical location, pollution index, economics, etc.

4.4 Exploratory Analysis

（1）Analysis of the overall characteristics of the customer

Based on the decision tree algorithm, a classification model for alternative potential users is constructed to grasp the comprehensive characteristics of customers with high substitution potential. Examples of classification results are shown in Table 1.

Customers with Feature Rule 1 or Rule 2 are high replacement potential customers. Therefore, it can be predicted that the geographical location, industry category, average daily load and other attributes of the enterprise will play an important role in model identification.

Table 1 Example of high substitution potential customer rule

	Variable name	**Feature rule**
Rule 1	Geographic location	>10 and <=20
	Industry category	Business service
	Average daily load	3.5%
	Daily peak electricity ratio	37%
Rule 2	Geographic location	>20 and <=30
	Industry category	Manufacturing
	Average daily load	3.8%
	Line margin	18%
	Daily peak electricity ratio	26%
	Historical 6-month electricity bill average	>60000 and <=70000
	Pollution index	5

（2）Analysis of the overall customer industry area

There are differences in the alternative equipment used in the industry, the proportion of various energy sources, and the total energy consumption. The examples of alternative energy transformation in different industries are given. It is seen that the difference in the industry determines the direction of different alternative energy research, and its total market capacity determines the market size of electric energy replacement.

Table 2 Examples of characteristics of transformation under various industries

Industry	**Transformation characteristics**
Metal foundry	Cupola-electric kiln (intermediate frequency furnace), large capacity for reconstruction, strong reproducibility
Manufacturing (processing, transportation)	Coal-fired boiler-electric boiler
Large public service agency	For example,the government department uses heat pump utilization as the main research object.
Low pressure resident	Electric heating,small capacity for reconstruction,but large group size
...	...

For the purpose of industry-wide analysis, under each industry, select representative companies as key areas. The division of the industry sector is based on the characteristics of the industrial economy combined with the region, and the characteristics of the electric energy replacement technology combined with the field. Judging from the electricity consumption and the proportion of enterprises in different industries, the industry has a wide coverage, and the manufacturing, industrial, textile printing and dyeing industries, commerce, agriculture, forestry and animal husbandry are relatively developed, accounting for about 75% of the overall proportion.

Table 3 The proportion of various industries

Industry	**Percentage of electricity used**	**Percentage of business**
Manufacturing (transportation, optics, processing)	22.1%	12.1%
Chemical manufacturing	18.3%	17.0%
Metal foundry	13%	12.0%
Textile printing and dyeing industry	11.6%	2.4%
Business and residential services	7.8%	4%
Transportation industry	4.5%	5%
Construction industry	3.9%	5.6%
...	...	...

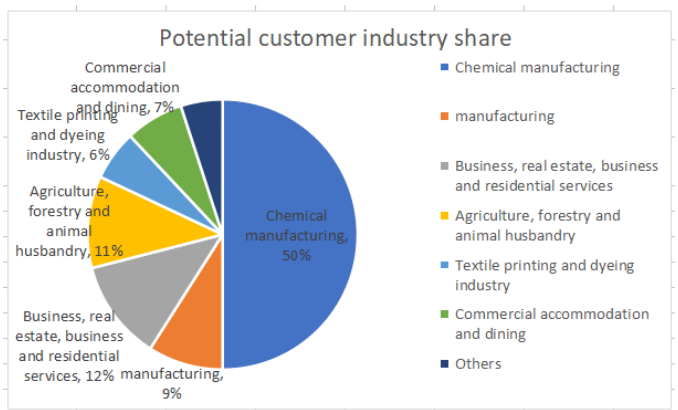

Fig. 1 Proportion of potential customer industry

Based on the above analysis, the electricity consumption in different industries is very different, and the potential of electric energy substitution is also different. Therefore, it is expected that the industry category attribute will play a significant role in the identification process of potential customers.

（3）Comparative analysis of indicator variables before and after transformation

The left side of the figure below shows the change of the distribution of the electricity in the valley before and after the transformation, they are different so the indicator is selected; the right side is the change of the unit contract capacity before and after the transformation, there is almost no difference, so we don't choose into the indicator. （The red part indicates that it has not been modified, and the green part indicates that it has been modified.）

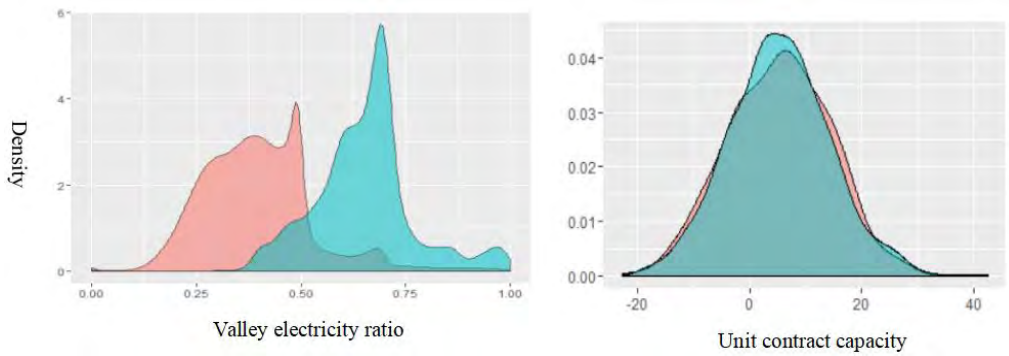

Fig. 2 Comparison of indicators before and after transformation

（4）Conclusion

Through the feature analysis of high-substitution potential customers, the main features are as follows:

High replacement potential customers have some overlapping attributes, such as daily load, enterprise geographical location, industry category, pollution index, etc., which provide a theoretical basis for the subsequent construction of model indicators.

High-value major customers are mainly concentrated in the manufacturing, industrial, textile printing and dyeing industries, commerce, agriculture, forestry and animal husbandry industries, and the manufacturing industry is in a leading position in terms of electricity consumption and the proportion of enterprises.

We can use the method of comparison before and after the transformation to make preliminary prediction and analysis on other variable indicators, such as the geographical location of the enterprise, and the density of residents, the line margin, etc. will have different degrees of impact on whether there is alternative potential.

In summary, the model index system can be constructed from the three dimensions of enterprise basic attributes, electricity consumption information and external competitive factors.

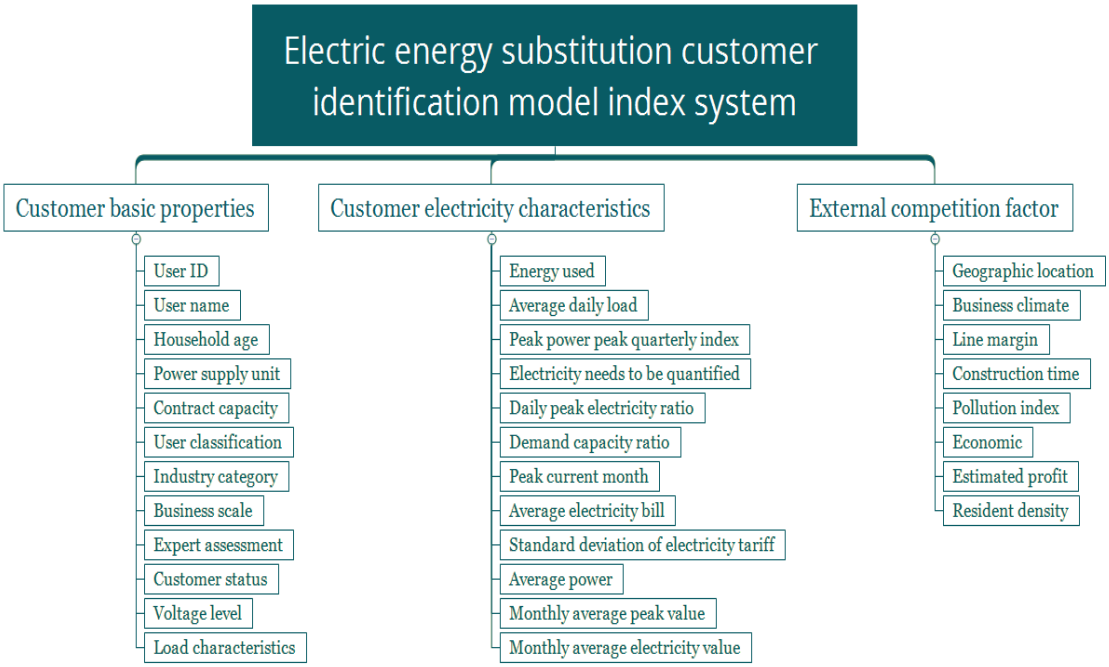

Fig. 3 Model construction indicator system

4.5 Model Construction

.1 Indicator Screening: Data Preprocessing

(1) Variable predictive ability screening:

IV values can be used to measure the predictive power of independent variables. Similar indicators include information gain, Gini coefficient, and so on. The definition of Info Value is（2）:

$$IV = \sum_{i=1}^{r} (p_i - q_i) In\left(\frac{p_i}{q_i}\right) \tag{2}$$

r represent the kind of the independent variable X, p_i and q_i are respectively represent the proportion of responses and the unresponses.

(2) Classification variable grouping (declining)

One is the redundant merging, which combines the less frequent categories into one new category; the second is the optimal grouping, that is, based on the clustering or decision tree algorithm, the predictive power index is maximized by merging the categories of variables.

(3) Numerical variable segmentation

One is equidistant grouping, that is, grouping variables according to the width interval and converting the numerical variables into categorical variables; the second is the optimal segmentation, that is, ordering the variables in order, first grouping them by the smallest particles, then according to the principle of "optimal grouping", find the optimal segmentation.

(4) Analysis of variance

After constructing the indicator system, a quantitative method is used to select the combination of indicators that effectively build the energy substitution model. Through the analysis of variance, the influence of the effect on the outcome variable is calculated, and finally 20 indicators with practical influence on the electric energy substitution are selected to enter the model. Blue indicates that the indicator has a large correlation with the outcome variable, and the variance analysis has significant differences. （Arranged from large to small according to the degree of influence.）

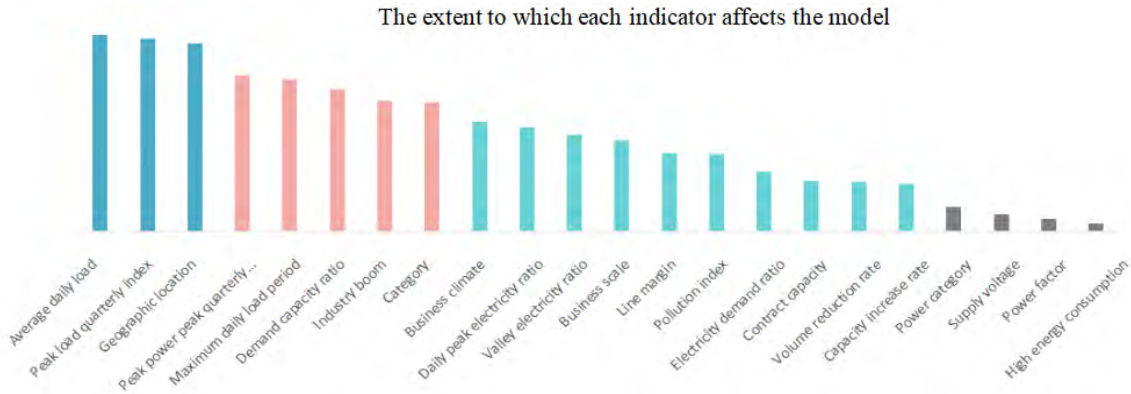

Fig. 4 indicator ranking

.2 Model construction

Based on R software, logistic regression algorithm is used to realize the construction of alternative potential customer identification model. According to the results of the variable screening, there are 20 indicators that finally enter the model. At the same time, the results are quantified and the scorecard score is given. It was verified that the model passed the goodness of fit test and the coefficient of each variable was significant at the 0.05 level. The index coefficient with strong influence is shown below, and the intercept term is 3.521.

Table 4 Partial indicator coefficients

Model index	Coefficient	Model index	Coefficient
Constant term	3.521	Demand capacity ratio	-0.317
Peak power peak quarterly index	0.684	Line margin	0.257
Geographic location	0.1825	Economic	0.406
Average daily load	-0.521	Peak load in the early morning	-0.879
Peak load quarterly index	-0.402	Peak load in the afternoon	-0.563
Industry boom	0.1608		

.3 Model evaluation

(1) Model results

Based on the result of the potential substitution recognition model, the standard scorecard for the electric energy substitution level prediction is constructed according to the score card function, and the enterprise comprehensive score is given according to the score card standard. A user's rating can be expressed as:

$$score = A + B\beta_0 + (B\beta_1\omega_{11})\delta_{11} + (B\beta_1\omega_{12})\delta_{12} + ... + (B\beta_r\omega_{p1})\delta_{r1} + (B\beta_r\omega_{p2})\delta_{r2} + ... \quad (3)$$

among them $B = PDO / In(2), A = BaseScore - B * In(odds)$, β is the model parameter. ω is the WOE conversion value, δ is a binary variable (1 or 0) which indicates whether the variable takes a certain value. r is the number of model variables. The user score is controlled between 0-100 points. The algorithm based on the standard score card converts the logistic regression model result into a score card form, and the user's final score is the sum of the corresponding scores of the variables.

Through the model calculation of 352 household data of 7 counties in the city, the industry involves construction, mining, manufacturing, chemical industry, etc.; the technical field involves heat pump, building electrical kiln, industrial electric boiler, auxiliary electric power etc., the following results are obtained:

Table 5 Customer Score Results

	Expert assessment	Average daily load	Industry boom	Geographic location	Line margin %	pollution index	Economic	Valley electricity ratio
Business 1	87.84	3.5%	75.00	50.00	22.00	360.00	41.00	0.29
Business 2	64.47	4.1%	44.00	73.00	14.00	130.00	20.00	0.42
Business 3	63.73	4.9%	50.00	36.00	9.00	185.00	25.00	0.45
Enterprise 4	61.18	4.0%	45.00	19.00	15.00	320.00	23.00	0.44
Enterprise 5	76.97	3.8%	61.00	53.00	14.00	358.00	25.00	0.39
Business 6	76.10	3.7%	45.00	65.00	14.00	348.00	32.00	0.43
Enterprise 7	79.09	3.8%	70.00	21.00	12.00	402.00	41.00	0.26
Enterprise 8	65.90	4.1%	70.00	22.00	16.00	115.00	25.00	0.48
Business 9	68.99	4.1%	45.00	46.00	22.00	145.00	25.00	0.42
Enterprise 10	77.93	3.8%	42.00	73.00	15.00	330.00	25.00	0.41
		.....	.....	.....	.....			
Enterprise 352	75.28	3.7%	53.00	82.00	9.00	266.00	30.00	0.43

(2)Model evaluation

It has been verified that the hit rate and coverage of the alternative potential prediction model are 35.2% and 73.8%, respectively. That is: model hit rate = predicted correct number/total forecast

number = 35.2%, model coverage = predicted correct number/objective existence correct number = 73.8%.

5. Conclusion

(1) Specialized service

For grid companies, the first is to use the top score customers as senior potential users. Through on-site visits and one-to-one communication platforms, actively promote relevant energy substitution policies and package transaction policies, calculate the profit margin of electric energy substitution, and improve customers. Confidence in transformation; second, the introduction of preferential tariff policies, subsidies to upgrade equipment purchases; third, increase the economic benefits of electricity generation, while considering the impact of the corresponding power supply line margins, truly ensure electricity and quantity. For example, different science information can be pushed to customers in different industries to meet the needs of customers in different industries; free load, real-time electricity price, power quality and other relevant information can be pushed to customers in real time; green channel service is opened to speed up customer service processing.

(2) Policy services

The target for electric energy replacement in the 13th Five-Year Plan is 450 billion. One of the main concerns of the government is the air pollution caused by energy combustion. With the accurate identification of potential customers, it is possible to issue relevant policies to selected enterprises, and actively urge enterprises to renovate equipment. For densely populated areas, we must also consider seasonal and wind direction issues, and take corresponding measures in a targeted manner.

In summary, a data analysis and modeling analysis method introduced in this paper can guide Xinjiang companies to improve work efficiency more effectively in the future field of electric energy substitution, and timely forecast future business, so as to effectively control risks and take targeted measures. The application of the logistic regression model is perfected, with strong applicability and scalability. It can be widely used in the field of energy replacement market potential measurement, potential user portraits, marketing services and other related business areas, providing design for the subsequent alternative energy management platform. Theoretical method guidance. Through practical application, it demonstrates the ability of the model to integrate data associations, ensuring that electric energy replaces customer identification accurately and efficiently, and works in an orderly manner.

References

[1]. Yan Taishan, Cheng Haozhong, Zeng Pingliang, et al.Energy Internet Architecture and Key Technologies[J]. Power System Technology, 2016, 40(1):105-113.

[2]. Huang M, He Y, Cen H. Predictive analysis on electric-power supply and demand in China[J]. Renewable Energy, 2007, 32(7):1165-1174.

[3]. Wang Wei, Huang Ke. Alternative Strategy of Electric Energy: Opportunity, Challenge and Policy Choice[J].Journal of North China Electric Power University (Social Science Edition), 2014(4):1-5.

[4]. Zhou Youxue.Discussion on the Promotion Measures of Electric Energy Substitution Technology[J]. Agricultural Power Management, 2016(2): 42-43.

[5]. Xue Yusheng, Lai Yening. The Integration of Big Energy Thinking and Big Data Thinking (I) Big Data and Power Big Data[J]. Automation of Electric Power Systems, 2016, 40(1): 1-8.

[6]. Han Ying, Zhu Zhu. Smog weather reflects energy dilemma pollution and forced energy structure adjustment [EB/OL].[2017-06-20].http://www.gd.xinhunnet.com/newcenter/2013-01/18/.

[7]. Zhao Yuhui. Research on energy substitution analysis and optimization model under Beijing environmental emission reduction target [D]. North China Electric Power University, 2014.

[8]. Chen Linhan. Application of Big Data in Electric Energy Substitution[J]. Chinese and Foreign Entrepreneurs, 2016(36).

[9]. Yang Huafei, Li Donghua, Cheng Ming. Analysis and Research on Key Technologies and Construction Ideas of Power Big Data[J]. Electric Power Information and Communication Technology, 2015, 13(1):7-10.

[10]. Sun Yi, Zhou Shuang, Shan Yiguo, et al. Analysis of Electric Energy Substitution Potential under Multiple Scenarios[J]. Power System Technology, 2017, 41(1): 118-123.

[11]. Shan Yiguo, Zhao Jia, Jia Dexiang, et al. Analysis method of electric energy substitution potential based on STIRPAT-ridge regression[J]. For Power Supply, 2018(1): 68-73.

[12]. Sun Yi, Graphite, Shan Yiguo, et al. Analysis Method of Electric Energy Substitution Potential Based on Particle Swarm Optimization Support Vector Machine[J]. Power System Technology, 2017, 41(6): 1767-1771.

[13]. TU Ying, LIU Qiang, WANG Qingjuan, et al. Research on user mining model of electric energy substitution potential based on collaborative filtering algorithm[J]. Electric Power Information and Communication Technology, 2017(12):25-31.

[14]. HUANG Wensi, HAO Hanyong, LI Jinhu, et al. Risk Decision of Power Customers' Arrears Based on Decision Tree Algorithm[J]. Electric Power Information and Communication Technology, 2016(1):19-22.

[15]. FELDMAN R, SANGER J. Text mining handbook: advanced approaches in analyzing unstructured data[M]. Cambridge University Press, 2006.