The model research of electric power information trading value-added service under the background of big data

He Yang^{1, a}, Shang Jincheng^{1,b} and Dai Yong^{2,c} ¹State Grid Henan Electric Power Company, Henan Province, China ²Beijing Kedong Power Control System Co. Ltd., Beijing, China ^ajinchengshang@qq.com, ^bhugesea78@sina.com, ^c kddaiyong@163.com

Keywords: The power market; information value; information mining; value-added services **Abstract.** With the advent of the big data era, as a resource to promote the development of economic and social, information has been more and more get people's attention, now has become a very important commodity. In electric power industry, different categories of information service can bring different benefit to the user expectations. By accurate prediction and effective mining, information can be divided into grades based on the value which could bring. And we can do further research in the pricing mechanism of electric power information, to form a complete of electric power information trading value-added service mode.

Introduction

Since the 1990s, Electricity market reform wave began to rise in the world, and most countries have carried out the reform of the electricity market. The emergence of GridWise provides a comprehensive and timely operation information of power grid. We are able to effectively use the information of big data and integrate all sorts of results analyses of information by a variety of automated function systems, to make the optimal response. Therefore, an accurate, fast, open and shared information system is the base of power grid, is also the biggest difference between the future power grid and traditional power grid. Information platform is a public platform and important means to support the construction of the strong and smart grid or even future global Energy Internet system. To realize the overall collection, transferring fluency and high-efficient processing of information, support the highly integration of electricity flow, information flow and business flow, it is necessary to build a collaborative and interoperability platform with information sharing and transparent, integration specification and powerful functions. For the guarantee of scientific analysis, prediction and decision, it is essential to ensure the reliable storage and management of big data information, fully tap the potential value of information, improve the intelligent analysis and decision support level.

Bases of Information Value Theory

Definition of Information Value

Information value is as an attribute of objective things. Information is the content to reflect the interconnection and interaction between things, and plays an important part in eliminating messiness, increasing order and improving the degree of self-organization for receiving system. Therefore, the value of information is actually the usefulness of information to the receiving system. For power consumers, the information value provided by energy Internet is characterized by the usefulness towards users and the effects that can happen in practice, which is also satisfaction that meets user needs in real life.

In view of the information value measurement, information commodity is often treated as a special commodity, the way information commodity forms and embodies also has particularity. Different from general goods and commodities, the embodiment of information value in the process of producing, exchange and application has apparent speciality. However because the special embodiment of information value and function, the quantitive measurement of information value has great difficulty^[1].

Grade Equivalent Value-added Service Mode

In the electricity market, To analysis the different categories of information service to users the benefits of expectations, in terms of the different value of different information, information value can be divided into several categories: original information of the electricity market, basic information of the electricity market, value-added service information of the electricity market and personalized value-added service information of the electricity market.



Fig. 1 The difference between information business mode and information hierarchical consumption (1)Original Information

The collection phase of original information of the electricity market is collecting information related to the development of the energy Internet. Original information is not sifted through and processed, only consists of a mass of data and information related to economy, geography, climate, electricity market, etc.

(2)Basic Information

At the stage of original information and primary processing, the main task is to sort all the original information, its work goal is to make the isolated data more meaningful, to meet the demand of the development of the energy Internet. Basic information mainly includes: meteorological data and resource data. This kind of information has a low opinion, which can be open free to consult.

(3)Value-added Service Information

Value-added service information mainly includes data about the electric system and the power industry decision-making information, the main block of power industry decision-making information includes: the development situation of the power industry annual analysis report, power companies bidding strategies research report, the influence of emissions trading of power companies report, etc. This kind of information service needs to be charged a fee. According to the demand, users choose to buy different types of information.

(4)Personalized Value-added Service Information

Personalized value-added service information covers customized package service and intelligent information push service. In terms of actual demands, power consumers are free to check the boxes to get the information type they hope to obtain within customized package service. Intelligent information push service set two main topics. There are two types of the information service subscription: the subscription according to the keywords and subscription by information type. Therefore, this kind of information service pricing is relatively high^[3].

Information Value Forecast and Mining

As the global development of the energy Internet, a great deal of data information involved energy supply, energy environment, energy distribution and other aspects will emerge. But "Huge amounts of data, lack of knowledge" is the problem existing in the most enterprises in today's era of big data. If we fail to do reasonably accurate use of information value, This information will lose its value and the value of the derivative^[4]. From this, it is the necessary preparations to reasonably predict and accurately mine information value for its application and appreciation.

The Theory of Information Value Forecast and Mining

The goal of prediction tasks is to predict specific attributes values according to the values of other attributes. Forecast and estimate is one of the widely used in data mining. The function of data mining in solving practical problems is not completely alone, each of the different problem is suitable for different methods. In most situations, we have to use a variety of data mining methods to mine hidden information. In establishing a good prediction model process, we should pay attention to the following points:

(1)Selection of Data: Prediction model of ultimate goal is to use the existing value to predict the future value. So the selection of appropriate data model is necessary and critical.

(2)Model Testing and Evaluation: In building a forecasting model, only part of the data is to establish the initial model, the rest of the data is used for test and evaluation. The data can be divided into training set, test set and evaluation set.

(3)The Applicability of Prediction Model: A prediction model is not universal, when making prediction we must adopt suitable prediction model, otherwise it is difficult to achieve expected effect in cases where prediction model cannot be applied, although this prediction model is very good^[5].

The Information Value Prediction Model Based on Naive Bayesian Classification

In response to the application of information value prediction and mining, we will build the information value prediction model based on Naive Bayesian Classification. First of all, we should do attribute correlation analysis of information source, the difficulty of obtaining, obtaining way, reliability and so on, to find information attributes correlated with higher information value and information cluster. After that, we calculate the probability that certain information belongs to the information value ranks, normalized process obtained multiple probability, and assign the information to the level of information value information group with the highest probability according to the calculation results.

Attribute Correlation Analysis

In the energy Internet information database, such as statistics, geography, market and behavioral attributes used to describe the information characteristics are usually many kinds of ways. In order to improve the efficiency of computation, it is necessary to determine which attributes should be included in the data analysis.

Assume that the training data set C contains n data samples, class label attribute-"Level of

information value" covers k different values. Assume that n_i is the i class customer's sample (i=1, 2, ..., k), then for a data recording the expectations of the information needed to determine its class label

$$I(n_{1}, n_{2}, \dots, n_{i}, \dots, n_{k}) = -\sum_{i=1}^{k} \frac{n_{i}}{n} \ln \frac{n_{i}}{n}$$
(3-1)

Assume an attribute A in set C includes m different values $\{a_1, a_2, \dots, a_m\}$, C can be divided into m subsets according to the value of A $\{C_1, C_2, \dots, C_j, \dots, C_m\}$, the number of samples for each subset is n_j^A ($j = 1, 2, \dots, m$), thus we divided C entropy based on A.

$$E(A) = \sum_{j=1}^{m} \frac{n_j^{A}}{n} I(n_{1j}, n_{2j}, \dots, n_{kj})$$
(3-2)

In this one, n_{ij} is the i class data sample in subsets j.

Therefore, the encoding of information code on attribute A:

 $Gain(A) = I(n_1, n_2, \dots, n_k) - E(A)$ (3-3)

Using the above method to calculate each attribute of information gain in C, and set the attribute correlation threshold. Information gain value less than the attribute correlation threshold can be seen as irrelevant or weak correlation attributes for classifying (i.e., to determine the information value), which will not appear in the process of subsequent calculations that take advantage of naive bayesian classification to predict customer value.

Predicting The Value Level of Information Based on Naive Bayesian Classification

Bayesian classification is a statistical classification method, which is based on bayes' theorem. Assume for class label attribute, the influence of various attributes are independent, we can simplify the calculation process of bayesian classification, the specific process is as follows:

Step 1: Using a n d feature vectors $X = (x_1, x_2, \dots, x_n)$ to show the characteristic of each message, describe n attributes of customer (meet the measurement of attribute correlation analysis threshold).

Step 2: For k given level of customer value (in keeping with the specified number of clusters when to cluster information), for each new customers or potential customers, taxonomy will predict the probability of the client belonging to each level of customer value, according to bayes' theorem, the client belonging to certain level of customer value.

$$P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X)} \quad \forall_i$$
(3-4)

Step 3: For a lot of training data sets with many attributes, the calculation cost may be expensive, in order to simplify the calculation, do class conditional independence assumption is allowed. For a given level of information value (class label), Assume that each attribute is conditionally independent to each other, namely there is no dependencies between attributes, so there is

$$P(X | C_i) = \prod_{k=1}^n p(x_j | C_i) \quad \forall i$$
(3-5)

The probability of $p(x_j | C_i)$ can get estimation value by the training sample. If A_j is discrete

attribute, $p(x_j | C_i) = \frac{s_{ij}}{s_i}$. s_{ij} is in the attribute A_j with the number x_j the number of training samples in C_i , s_i is the number of training sample of C_i . If A_j is continuous attribute, this attribute is often assumed to obey gaussian distribution, so

$$p(x_{j} | C_{i}) = g(x_{j}, m_{C_{i}}, s_{C_{i}}) = \frac{1}{\sqrt{2ps_{C_{i}}}} e^{\frac{(x_{j} - m_{C_{i}})^{2}}{2s_{C_{i}}^{2}}}$$
(3-6)

Step 4: Calculate the probability of certain information belonging to each information value level, normalized process obtained k probability, and assign the information to the level of information value information group with the highest probability according to the calculation results.

Research of Pricing Mechanism of Electric Power Information Trading

The choice of pricing strategy is pretty tightly linked to the type and level of information service. Value-added information products and services should price differentially based on user group characteristics and value-added degree. For professionally personalized value-added information service, it can be carried out in accordance with the complete market standard fee, to form the information cost recovery and pricing level of the pyramid structure.

Next, put the user to distinguish, distinguish the difference in profit of service provider between unified pricing and differential pricing. Under a hypothetical condition, implement discrimination pricing model, make a different pricing for different users. According to the price of certain types of information service, the value of information service (assuming the higher level information service value containing lower level information service value) and the will of consumers willing to pay, we reasonable prices.

In addition, there is a kind of bundling pricing model, bundling pricing as important sales way of ordinary goods and information services, is widely used by businesses. In the information service pricing model, there are bound and not bound way. Whether in bundles or without bound, information service providers will take at the same time or before and after the price decision. Order of decision will also have a major impact on the final price^[6].

Conclusion

The realization of open fair and just of electricity market and electric power resource optimal allocation effectively are inseparable from the science and proper analysis of the various of electric market main body behavior and all kinds of data of electricity market. Access to effective power market data is the premise and foundation to deploy follow-up analysis of electricity market behavior and electricity data. Researching based on the pattern of electric power information value transaction value-added services, in view of the information itself and value-added products and services, to classify pricing, is conducive to enhance the use efficiency of personalized value-added information service, dig information value to a great extent, for the development of the whole electric power industry and even the future energy Internet provide effective support.

References

[1] He Shaohua, Kang Bin. Evaluation on Value of Information and Value of Information Service [J]. Library And Information Service, 2005, 49(5):72-75.

[2] Liu Zhenya, in: Global Energy Internet [M]. China Electric Power Press, 2015.

[3] Ye Qing, Tuo Guangzhong, Zhang Yong. Transation Information Publication System of Power Market and Its Operating Mode [J]. North China Electric Power, 2009(12):1-4.

[4] Lu Jianchang, Fan Weiguo. Application of Data Mining Technology in Electric Power Enterprises in Era of Big Data [J]. Guangdong Electric Power, 2014(9):88-94.

[5] Li Cunbin, Dong Fugui, Chen Yongquan, Tan Zhongfu, Guo Xiaopeng. Platform Research on Electricity Market Information System Structure and Application Intergration [J]. Automation of Electric Power Systems, 2004, 28(20):63-67.

[6] Sun Ke, Xiao Qing. Effectiveness of Electricity Market Information Disclosure and Selection of Market Transaction Mode [J]. Automation of Electric Power Systems, 2008, 32(6):60-65.