Research on Key Technologies of Tag System Based on Power Data

Yan Liu[1, a], Ying Liu[1], Hanji Ju[1], Zhenyu Yue[2], Hongying Miao[3]

[1] Electric Power Research Institute, State Grid Jibei Electric Power Co., Ltd, Beijing 100045 China;

[2] Tangshan Power Supply Company, State Grid Jibei Electric Power Co., Ltd, Tangshan 063000 China;

[3] ChengDe Power Supply Company, State Grid Jibei Electric Power Co., Ltd, Chengde 067000 China.

[a] liuyan1989pl@126.com

Keywords: User label system; Large power data.

Abstract. Based on the characteristics of power users, this paper puts forward the user portrait technology based on big data and discusses the construction of power user portrait. this paper first introduces the concept of user's portrait and large data, principle and function of a picture of the user, and then introduces the steps and key technology of building user portrait, expounds the user label system. user portraits can help companies to refine marketing and improve user experience.

1. Introduction

In recent years, large data technology has developed very rapidly. More and more enterprises have made precise and differentiated marketing strategies by mining massive data and using large data technology to carry out fine analysis of business. Compared with extensive and unified traditional mode, fine and differentiated mode is more conducive to enhancing the core competitiveness of enterprises. As a traditional industry, the power enterprises should use large data technology to strengthen the continuous attention to the users, improve the service level of power supply and meet the customers' experience of electricity use, thus enhancing the marketing ability of the power products and improving the quality of the people's living electricity. Because of the accumulation of time and the increase of electricity users, electric power enterprises have abundant data resources. How to dig out valuable information as much as possible in the mass data, and finally extract the data assets which can improve the service level and marketing ability through analysis and integration and become one of the key points in the reform of the power industry.

At present, user tagging system is an effective technical means for enterprises to realize user personalization analysis and assist in fine operation decision making. User tagging technology based on big data has made great breakthroughs in recent years, and is widely used in insurance, e-commerce and other industries. Building the portrait of power users has also become one of the important research contents of electric power enterprises. The effective extraction of user features, a deeper and more extensive understanding of the user needs, the establishment and improvement of the user label system and the construction of a 360-degree omnidirectional user portrait will help to promote the construction of a new marketing model, optimize the service mode and upgrade the transformation and upgrading of the power grid enterprises.

2. Key Technology

2.1 User Tags and Big Data

The user label is based on the related data of power users. Through data mining, data analysis and other technical means, the accurate information about user attributes, behavior, habits and other dimensions is extracted as a user's label, and all tags are combined into a user label system that can reflect the overall characteristics of the user. The core of the user tag system is to label the user, which

is based on the depth analysis of the user information and generalizes the user's signature by a highly refined natural language.

Big data refers to the collection of data that cannot be effectively transmitted, managed and processed by conventional software. As for large data, traditional technology has not been met, and special technology is needed to effectively deal with large amounts of data in the time of tolerance, which we call large data technology. Large data technology has corresponding technology in all data processing links, such as the use of distributed database and distributed processing in data storage and processing, data mining and machine learning in data analysis.

In the process of constructing user label system and user portrait, it is necessary to use large data technology to efficiently process data, accurately analyze user attributes, behavior, habits and so on, so as to dig out valuable information, accurately and properly perceive user characteristics, and use the user portrait more stereoscopic and more accurate.

2.2 The Principle of Building User Labels

The construction of user tags should be carried out in combination with business scenarios and business requirements, and the dimensions of the data should also follow the business rules to avoid excessive unused data interference analysis. There are two principles to build user labels.

First, building user tags should start with business scenarios, and the ultimate goal is to solve practical problems. For example, in order to find the target customers, or to carry out the marketing activities, or to improve the experience of the user, different business scenarios have different requirements for the user's portrait.

The two is to accurately grasp the dimension of the data. User tags should be as plump and solid as possible, not the more dimensions of data, the better, but to meet business requirements and support business operations. In general, information that is strongly related to business should be used to ignore negligible information on business. The strong association information refers to the information directly related to the business scene, such as the power consumption of the power users is directly related to the information of the marketing business scene, and the level of the information is high. The weak correlation information is the information that has little influence on the business scene, and the existence or not of this kind of information will not affect the final picture. It will only interfere in the data analysis and should ignore the information in the process of processing. On the other hand, according to the business model and mathematical model, it is necessary to integrate the information of the minimum granularity of the user, and transform some quantitative information into qualitative information, and qualitative information is beneficial to the model to do information screening.

2.3 The Role of Building User Tags

With the improvement of users' requirements, enterprises are eager to have a deeper understanding of users, resulting in user portrait. From the generation of users to the whole life cycle of loss, enterprises want to control it. To master the life cycle of the user, the enterprise can analyze it in depth according to the characteristics of each stage of the user, so as to take appropriate marketing strategy. The role of user portrait in enterprises is mainly reflected in the following aspects:

(1) promote precision marketing

According to the characteristics of the product, the enterprise accurately finds the target users, and formulates different marketing strategies for different target users. Fine marketing can increase profits for enterprises.

(2) improving the user experience

Fully tap the user's information, according to the needs of users to create more user-friendly products, to bring users a better experience and services.

2.4 Building a User Portrait Step

There are three steps to build the user portrait: data processing and analysis, establishing user tagging system, and developing user portrait.

2.4.1 Data Processing and Analysis

The purpose of user tagging is to maximize the potential information of users. The main source of these information is the source data owned by power grid enterprises. The power source data mainly come from the operation data and file data collected by the power system.

The quality of source data is not perfect. Data cleaning is needed to support subsequent work. Otherwise, the result of mining will be unsatisfactory. Data cleaning mainly includes handling outliers, completing null values, and removing duplicate values. Through effective means, the unusable, redundant and inconsistent data in source data can be transformed into standardized data, that is, data that meets the requirements of data analysis. On the basis of fully understanding the data, we ignore the weak correlation data and select strong correlation data. Ignore the weakly related data directly, and do not enter the next step, so as to avoid further difficulty in subsequent analysis.

2.4.2 Establishing User Tagging System

The power user tag is the basic component unit of the user's portrait. Through the analysis of the data collection system, the marketing business application system, the 95598-customer service system file data, the electricity data, the payment information, the measurement equipment data and so on, and the Internet trading platform and other external data, the highly refined user features are obtained. A user tagging system based on business needs can create user profiles. The core work of building user profiles is to match users with accurate labels. There are 2 main steps to establish a user tagging system: label design and classification, label application and feedback.

(1) Label design and classification

Label design is the definition of labels. The design of labels should be based on data and business to establish a comprehensive and applicable label library. Business personnel gather business-based scenario descriptions, function points, influencing factors, user characteristics, etc., to form a comprehensive user label requirement. Business experts instantiate abstract information after several times of decomposing requirements and evaluation requirements and refine the concise and comprehensive language to describe the user's characteristics and form a user label. Through the evaluation of the application effect of labels, we can optimize the description, type and quantity of labels. In accordance with the design process of tags, the user tags are maintained and managed throughout the lifecycle, and a standardized and efficient tag library is gradually formed.

In order to clearly and accurately display the various information of the user, it is necessary to sort out the user's data, summarize the user's label, and classify the label. Because of the strong professional user data, the classification of the label system is different from the Internet industry. It is necessary for business experts to classify the label system in a logical, hierarchical and standard way through a long period of professional training. At the same time, we need to maintain a certain degree of freedom for label classification, so as to ensure the relative stability and sustainable growth of the label system. Therefore, the power user tagging system adopts a hybrid classification method combining hierarchical enumeration classification with faceted component classification.

Some of the users' tags can be directly obtained from the data, such as the user's number of years, the quantity of electricity, capacity and so on. Some of them need to be obtained by means of large data algorithm analysis and data mining, such as the user's electricity stealing degree. The user tag system based on large power data mainly describes the user's label from 5 dimensions, including user attribute, power consumption behavior, performance, transaction behavior, equipment state and so on, and thus forms the tag library of the power users. User attributes mainly refer to the basic characteristics of users, including basic attributes such as age, gender, power consumption, power consumption and so on. Electricity consumption is mainly caused by the dynamic characteristics of users in electricity consumption, which mainly includes electricity usage, abnormal electricity consumption and so on. Equipment state mainly refers to the operation of metering equipment, whether there is abnormal state, including equipment file information, equipment failure and so on. Transaction behavior refers to the characteristics of users' payment, mainly including payment method, notification preference, frequency of payment, consumption level and so on. The ability to

perform refers to the credit risk of customers, mainly including the implementation of electricity charges, the risk level and so on. The contents are shown in Figure 1.

| User Attributes | Basic Attributes | Age | Sex | Address | Job | ...... |
| | Power Attributes | Voltage level | Power use category | Industry category | Load capacity | |

| Electrical behavior | Use of electricity | Grade of electricity | Power consumption level | Load property | ...... |
| | Electrical anomaly | Defaulting electricity | Suspicion of stealing electricity | Private expansion | |

| Equipment state | Equipment file | Date of production | Manufacturer | Production batch | Rated current | ...... |
| | equipment failure | Meter flying away | Voltage limit | Requirement overcapacity | Reverse electric quantity | |

| Transaction behavior | Payment method | Notification preference | Payment channel | Rate of payment | Consumption level | ...... |

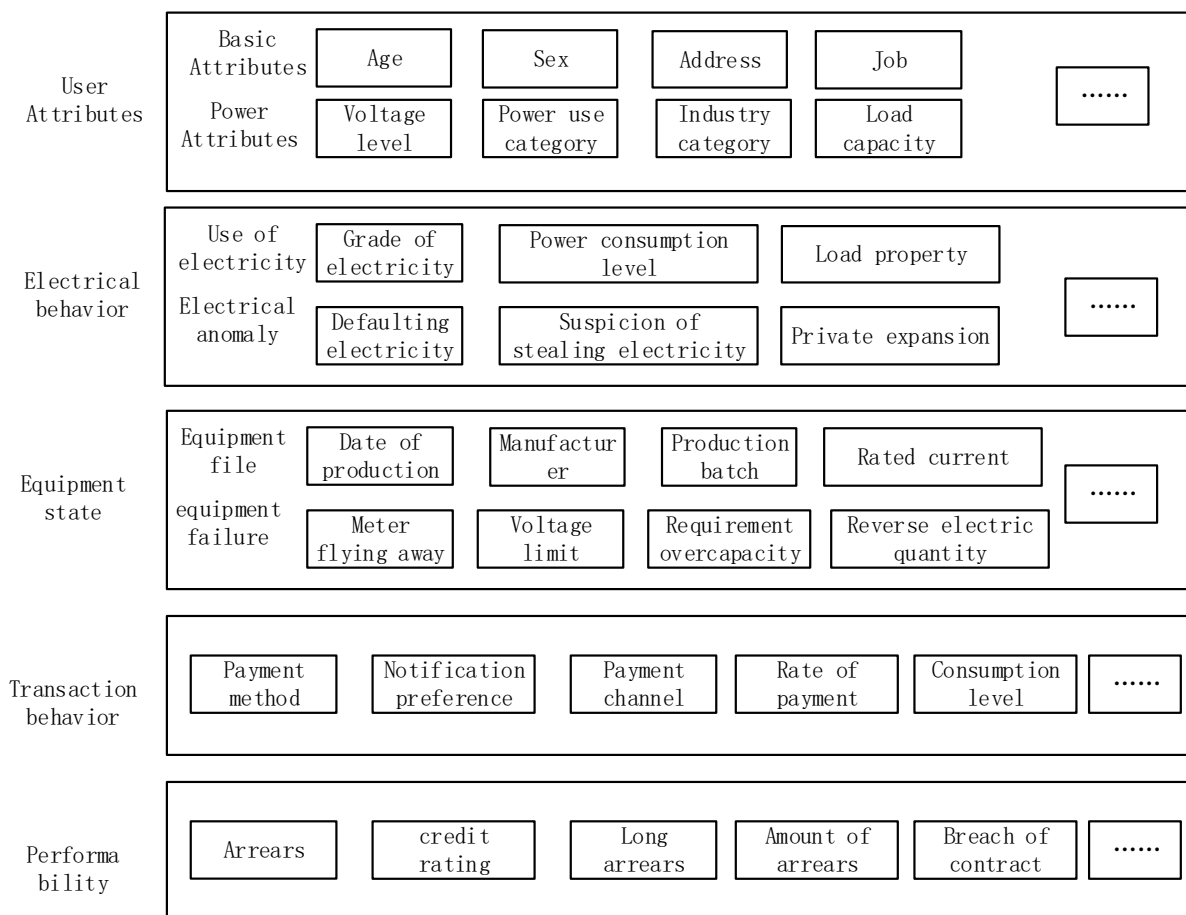| Performa bility | Arrears | credit rating | Long arrears | Amount of arrears | Breach of contract | ...... |

Fig.1 User label system

(2) Label application and feedback

Tag application is to provide user attributes and labels to the application layer through the application of the system tag library and carry out differentiated business. In the application, the actual problems are recorded, including data problems and business problems, in order to feedback the problems in the follow-up, correct the data, and optimize the application strategy of the label.

Label feedback is the evaluation of the labels in the application, the summary of the problems and feedback. The contents of the evaluation include the push of labels, the number of tags used, etc. The main way of evaluation is to set cycle automatic assessment. According to the evaluation effect, the weight of the label, the data model and the structure of the label system are adjusted to delete the label which is not applicable to ensure the efficiency of the label system.

2.4.3 User Portrait

User tagging system is the core of user portrait, and user portrait is developed according to user's label system. The user's portrait is based on four dimensions: user attributes, electricity consumption, trading behavior and performance capability. According to the individual characteristics of users, personalized and differentiated service strategies are made for users. It is worth noting that the granularity of the label of the user's portrait should be reconciled. The granularity is too small to guide the marketing system business; the size is too small, on the one hand, it will increase the workload, increase the cost, and on the other hand, the service target is too single. Therefore, we must combine the actual demand and cost, grasp the granularity of user portrait, help to build a rich, three-dimensional, clear user portrait.

2.5 Technology Roadmap

The core of user portrait is to build user tagging system. The key of tag system is data mining information. The power system has a large number of user groups and massive data, and there is a complex relationship between the basic labels. Only through complex logic operation can the user picture be built. Based on so many data and complex operations, the traditional database can not be satisfied. The data storage, mining and computing can be efficiently realized by large data technology.

Apache Hadoop, Hive and Hbase are used to store and calculate data. The data is drawn from the relational database to the Hive as the original database by Sqoop, and the HiveQL is compiled according to the requirements of the business and the algorithm is used for the logical operation of the data. The different Hive middle tables are set up for the label data of different users. Using MR to integrate the basic label, the model of user portrait is stored in Hbase, mainly because the data storage structure is more discrete, and the storage space is small. On the other hand, the Hbase is based on the rowkey lookup, and the speed of the tag is faster than that of the other products. ElasticSerach (distributed search engine) is used to create two level indexes for tag data, so as to realize label query. The flow of data is shown in Figure 2.
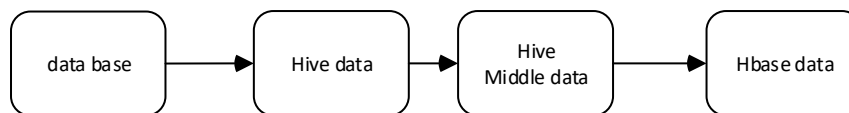

Fig.2 Data flow chart

In user tags, parts can be obtained through simple calculation or direct extraction, and some tags are obtained through data mining. The main data mining methods are regression, correlation, clustering and classification analysis. Through all kinds of data mining methods, we can fully excavate the information of the data and the characteristics of the users, so that the deep and effective information hidden in the data can be reflected in the user portrait.

3. Summary

User portrait can help business people to better analyze the characteristics of the user, according to the different characteristics of the user can be targeted to work, the power industry and the users both play a positive role. This paper describes the application background, main concepts, basic steps and key technologies of user portrait based on large power data. It is a new exploration and attempt for the application of large data in the power industry. In this era of rapid development of the big data, only with the times, continuous innovation, deep mining and business closely related information, can we fundamentally improve the service level and realize the precision marketing, and can adapt to the trend of the times and promote the development of the industry.

References

[1]. LIU Hai, LU Hui, RUAN Jin-hua, et al. Research on precision marketing segmentation model based on mining "persona" [J]. Journal of Silk, 2015, 52(12): 37-42.

[2]. YU Meng-jie. Data modeling of user portrayal in the product development[J]. Design Research, 2014, 4(6): 60-64.

[3]. JIANG Ying, WANG Zhi-qiang, DAI Bo. Research and analysis of residential electricity consumption behavior based on big data[J]. Electric Power Information and Communication Technology, 2015, 13(11): 7-11.

[4]. ZHU Chao-yang, WANG Ji-ye, DENG Chun-yu. Research and design of electric power big data platform[J]. Electric Power Information and Communication Technology, 2015, 13(6): 1-7.

[5]. YANG Hua-fei, LI Dong-hua, CHENG Ming. Analysis and research of key technologies and construction ideas of power big data[J]. Electric Power Information and Communication Technology, 2015, 13(6): 7-10.

[6]. ZHAO Yong-liang, FU Xin. Research on the monitoring and analysis of the sale of electricity business based on big data[J]. Electric Power Information and Communication Technology, 2016, 14(8): 52-57.

[7]. SONG Cai-hua, LAN Yuan-juan, FAN Ting, et al. Research on customer segmentation model of power enterprise[J]. Modern Electronics Technique, 2014, 37(2): 91-94.

[8]. LIANG Xiao-bo. Research on electric power marketing service risk based on customer perspective[J]. Modern Electronics Technique, 2016, 39(15): 95-98.