Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 8
EMEHSS 2024

Research on User Profile and User Behavior of Integrating Big Data Platforms

Yaoxuan Wang
University of Toronto, Toronto, Ontario, Canada
wangyaoxuan0108@gmail.com

Abstract: This paper discusses the construction and analysis method of user behavioral portrait by the data provided by the electric power platform in the big data environment. Firstly, it introduces the construction and analysis of user profiles based on big data platforms, which covers the construction of user basic attribute profiles, user behavioral characteristics profiles, user product characteristics profiles and user interaction characteristics profiles from different dimensions. Secondly, for the electric power sector, the article discusses the analysis of big data provided by electric power platforms to better understand user behavior and trends in energy consumption. The article proposes a method for constructing a behavioral portrait of power users based on big data analysis, including the construction and management of a user label library and the process of constructing a behavioral portrait of power users based on the improved K-mean algorithm. Finally, the effectiveness and accuracy of the method of this paper are verified by experimental analysis. Overall, this paper provides some guidance and reference for the analysis of user behavior in the field of electric power by exploring the method of user behavior portrait construction with the data provided by the electric power platform in the big data environment.

Keywords: big data platform; user profiling; user behavior; power sector

1. Introductory

In the Internet big data situation, e-commerce platform, social platform, power platform formed by the big data for enterprises to analyze user needs to provide a new resource, this new resource has changed the traditional user demand analysis mode, the formation of a big data platform based on the user profile and user behavior analysis. User portrait based on big data is a user portrait set, from different dimensions can be constructed user basic attribute portrait, user behavioral characteristics portrait, user product characteristics portrait, user interaction characteristics portrait, etc. For example, literature [1] combines the algorithms of text mining and case-based reasoning technology to analyze the Apple App Store review data to achieve the function of identifying the opportunities for product innovation and benchmarking products. On the basis of user portrait construction and analysis, through big data user behavior knowledge extraction, user emotional tendency extraction is especially critical for user behavior analysis. Literature [2] collects eWOM review data and through computer-based sentiment analysis methods, emotionally classifies the review words and tracks the dynamic characteristics of word frequency to mine the evolution of customer demand.

In the power sector, analysis of big data on power platforms allows for a better understanding of user behavior and trends in energy consumption. The analysis of users' energy usage data allows electric utilities to understand users' peak electricity consumption at different time periods, so as to optimize energy supply and improve energy utilization efficiency. Literature [3] proposes a distributed clustering algorithm for the perception of massive users' electricity consumption characteristics, which is validated on the Irish measured electricity consumption dataset. While foreign scholars focus on the use of smart meters, literature [4] demonstrates the process of classifying residential electricity customers based on smart meter data, and analyzing and clustering residential users' energy behavior through smart meter data. It can be seen that research scholars at home and abroad have carried out very much research work in the field of electricity. This topic will be based on these results, and conduct the construction of the behavioral portrait of electricity users based on big data analysis.

Through the collection of real data related to the user's electricity consumption behavior, the user portrait is built containing the user's basic information, electricity consumption behavior, constructed based on the results of the division of different behavioral labels of residential electricity users and the overall regulation clusters, and presented in a visual way.

2. Research on the construction of behavioral portrait of residential electricity users based on big data analysis

2.1. Construction and management of user label library

In this paper, we believe that user profiles contain three elements, namely user attributes, user characteristics and user labels. The construction process of user label attribute system is as follows.

(1) Label Creation:Collect and analyze power business requirements, and then extract reasonable labels

(2) Label design: Combined with the actual situation of the power industry, the label classification rules and attribute definition to form the initial label were designed to identify the class, named class, continuous class, curve class and other data types in order to the law is not obvious, composite data acquisition.

Tagging with data mining.

(3) Labeling rules: coverage, accuracy and other metrics are used to evaluate the reasonableness of labeling rule definitions and attribute names.

(4) Label Update: Update label rule definitions and attribute names based on label evaluation results. Delete obsolete tags and add new tags.

(5) Labeling attribute system: Based on the existing basic marketing data and business requirements, the labeling attribute system for power users has been established in an all-round way.

From the technical level, the construction of 95598 user profile based on big data is divided into three steps: the first step is to mine the user data information base, such as internal data, external data, etc. The second step is to process the user profile data, cleaning all kinds of collected data to make it structured and standardized. The second step is user profile data processing, cleaning all kinds of collected data to make it structured and standardized. The third step is user portrait construction, including accurate identification of user variables, quantification of user static data and dynamic behavioral assessment, and determination of user label library.

(1) According to the labeling system of 95598 users in different levels and dimensions, refine the effective labels, combine different labels to form contextualized user characteristics through business requirements, and select multiple dimensions to describe user labels based on internal and external data of 95598 users, including social attributes of the user, power attributes, bill payment preferences, credit of electricity consumption, etc., so as to build the basic logic of user profile analysis. The result is an all-round user label library.

(2) Establish user labeling dimensions, effectively use big data mining theory to enrich 95598 user information and behavioral characteristics to facilitate the collection of various types of information on users by electric power enterprises.

2.2. User label management

The user tag library reflects a short and concise summary of power users with high precision. However, whether it is the change of other external factors or the change of power users' own attributes, the user label information needs to be updated periodically, and the old user data cannot reflect the future user characteristics. Therefore, the management of user labels is also an essential part of the research in this paper. It is mainly reflected in the periodic update and management of

power user data information. In this paper, the management of user data in the form of cataloging the label classification, the label is reasonably graded classification, easy to manage and maintain the label, simple and clear label management for the construction of the user image is more scientific and orderly.

Table 1 Label Classification

Level 1 labels	secondary label	Tertiary labeling	Four levels of labeling	Label Acquisition	Label Type
User Attributes	social property	name and surname	XXX	System labeling	named class
	Power Properties	Key Clients	Yes/No	Rule definition	Direct category category
	Value attributes	high consumption	Yes/No	Rule definition	Direct category category
	Internet Properties	Consumption frequency	High/Medium/Low	Rule definition	Cause description class
	Contribution preferences	APP Payment	Yes/No	Rule definition	Cause description class
Behavioral characteristics	Electricity credit	risk of non-payment	High/Medium/Low	Definition of forecasting	Cause description class
	Business characteristics	Outage sensitivity	High/Medium/Low	Definition of forecasting	Cause description class
	emotional identity	objectionable telephone call	High/Medium/Low	Rule definition	Direct categories

Table 1 gives the portrait labeling system in electric power enterprises. The clear hierarchical classification makes the management of user labels more comprehensive and avoids the loss and duplication of labels. When categorizing the dynamic labels of users, the dynamic labels consist of two parameters: attributes and weights. Therefore, when managing dynamic labels, it is necessary to assign certain weights to them, and the importance of different attributes is different, which also needs to be determined based on experience to determine the size of the weights. In addition, in the management process, it is necessary to dynamically manage the user labels according to the actual business needs, such as adding some important variables and expanding the dimensions of the labels. It can reflect the real problem and also improve the accuracy of user profiles.

2.3. Electricity based on improved K-mean algorithm to build user behavioral profile

Use the improved K-mean clustering algorithm to analyze the clustering of power users. Different types of user clusters are created to form a comprehensive portrait of power users, which is convenient for business personnel to accurately identify user information.

Consumption habits show the tendency of users' willingness to participate in interactions. Considering the existing tariff policy, we measure the consumption habits of power users in terms of

the composition of electricity costs and the assessment of smart power consumption awareness. Under the existing ladder tariff strategy, electricity cost and electricity volume no longer show a linear correlation, and the cost of electricity, regulatory potential and intention are all positively related. Denote the share of laddered electricity cost as.

$$L_{C2} = \frac{C_{j3}}{C_{j1} + C_{j2}}$$

The total cost of electricity is.

$$L_{C3} = \frac{C_i}{\sum_{i=1}^{n} C_i / n}$$

Where, C_{j1} , C_{j2} , C_{j3} ;in turn, represents the first, second and third step of the electricity user's tariffs, C_i is the total number of tariffs of the ith user, and the time span is one year.

Based on the behavioral data of the power network under the fusion of multi-source data to implement the smart electricity awareness assessment", including the number of times of cell phone App login, the number of times of online business hall page browsing and the number of times of dialing 95598 to inquire about electricity bills, as shown in the following formula.

$$L_{ES} = N_{App} + N_{net} + N_{598}$$

The user profile global regulation feature is to analyze user behavior by season and time period guided by the needs of grid interaction objects. The methodology covers the following 4 processes.

(1) Data standardization

The integrated regulation characteristics of the same time period, including three categories of 14-dimensional labels, in order of behavioral labels 1-dimensional L_1 , electricity consumption characteristics labels 8-dimensional L $L_{2\sim9}$ consumption habits labels 5-dimensional $L_{10} \sim L_{14}$,then the N user data samples expression is:

$$\begin{bmatrix} L_1(1), L_1(2), \cdots, L_1(k), \cdots, L_1(14) \\ L_2(1), L_2(2), \cdots, L_2(k), \cdots, L_2(14) \\ \vdots \\ L_n(1), L_n(2), \cdots, L_n(k), \cdots, L_n(14) \end{bmatrix}$$

Describe the data of the same type for all users as a vector of columns (L(1),L(2), L(k),... ,L(14)).

The 14-dimensional data contain different physical meanings, the range of quantitative values varies a lot, and there are a large number of anomalous data within the data, so before the comprehensive cluster analysis, data cleaning should be completed, and for the column vector data, data standardization is completed by using the Z-score.

$$L(k) = \left(L(k) - \overline{L}(k)\right) / \sigma\left(L(k)\right)$$

where $\overline{L}(k)$ represents the mean of L(k) and σ(L(k) is the variance of L(k).

(2) Analyze the value of the composite indicator for each category

The user behavioral feature label is a logical quantity that is substituted into the four types of control objects, and the value of this segment at home is 1, and vice versa is -1. For example, the value of the control type for weekday workers is -1, -1, 1, 1. The behavioral labels are the original choices of the

user groups, and the customers with the value of this segment of -1 are directly excluded. The electricity consumption characteristic labels cover the user characteristic labels L2 to L9, and there are 8 in total.

dimensional data and solves for the integrated labeled value LE of electricity usage characteristics:

$$L_E = \sum_{k=2}^{9} \lambda_k L(k)$$

(3) Improved K-mean clustering calculation

The best class of the clustering algorithm is identified using the contour coefficient CH metric. After the completion of process 2, the target dataset is transformed into two-dimensional data [LE(i),LC(i)] with i representing the ith user. The operational idea of using the improved K-mean clustering algorithm is to cluster ml using the K-mean clustering idea under the Euclidean distance with respect to the electricity usage characteristic feature label LE(i), and in this way solve for the centroid of each class:

$$L_E(k) = \min \sum_{k=1}^{m1} \sum_{i=1}^{n} \left(L_E(i) - L_E(k) \right)^2$$

Similarly, with respect to the consumption feature label LC(i), it is turned into m2 classes using K-mean clustering means under Euclidean distance, and the centroid LC(j) of each class is solved in turn.

Fusion is implemented according to LE and LC centroids to obtain (LE (k),LC (j)), which constitutes m1 × m2 planar centroids corresponding to m1 × m2 clusters in turn. With all planar centroids specified, the clusters of each user are evaluated using the Manhattan distance sum equation . $L_E = \sum_{k=2}^{9} \lambda_k L(k)$

(4) Exploration of user clusters

LE and LC were categorized as high, medium, and low, and the clusters were increased in vertical order using the criterion of objectivity over subjectivity, as shown in Figure 1.

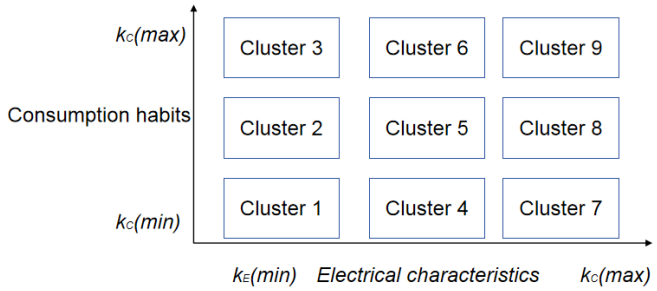

Figure 1 Distribution of residential electricity user behavior

3. Experimental analysis

In a provincial electric power company, 1,000 residents who have installed non-home terminals and are bound to the WeChat client of State Grid Power are selected as experimental subjects, and the total load and sub-identified load curve data of these residential electric power users are collected for any three weeks in the spring of 2021, and the total load and sub-identified load curve data for the summer of 2021 from mid-July to mid-August, as well as the electricity volume, electricity bill and network behavior statistics for one year of 2021 are collected. and network behavior statistics. Using the method in the paper, these data are calculated and analyzed to complete the construction of the residential electricity user profile.

Through the above data on residential power user behavioral portrait analysis, the establishment of each user's behavioral label library, to carry out user behavioral clusters of clustering analysis, on the power consumption characteristics of the indicators, taking into account the direct correlation between the indicators and the user's behavioral potential to set the weighting coefficients, weighting value of 0.2 belongs to the general characteristics of the power users are K2 and K9, weighting value of 0.5 ~ 4 users are K3 - K8. Since the 5-dimensional data is an indirect characterization of each different angle, the weights are taken to be the same.

The results of the cluster analysis of the users in the midday and evening hours can be visualized in a radar chart, see Figure 2.

According to the analysis found that, in the payment behavior, because the selection of the experimental object is based on the premise of binding the client, so the group is on the younger side, the consumption consciousness is stronger. According to the consumption habits of residential power users in high school and low 3 and so on distribution is relatively uniform, consumption awareness of residential power users have a certain advantage; because half of the experimental subjects are office workers, so the midday period of the overall weak regulatory capacity, with a certain advantage of the user is the cluster 1, 2, 3, the evening hours have a certain advantage of the user is the cluster 5, 6.

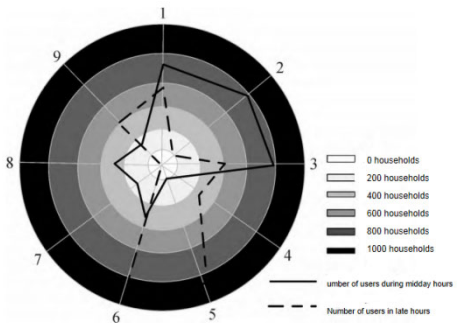

Fig. 2 Radar distribution of user clusters

Users have certain advantages; because half of the experimental subjects are office workers, the overall regulation ability is weak in the midday period, and the users with certain advantages are clusters 1, 2 and 3, and the users with certain advantages in the evening period are clusters 5 and 6.

The number of system accesses by residential electricity users at different times of the day from May 12-15, 2021 was counted,and the results are shown in Fig. 3.

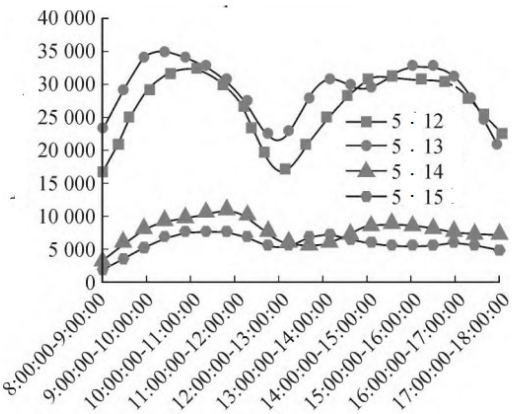

Figure 3: On-line situation of residential electricity users at different times of the day

The number of visits by residential power users is higher from 9:00 a.m. to 10:00 a.m., from 15:00 p.m. to 16:00 p.m., and lowest from 12:00 p.m. to 13:00 p.m. The number of visits to the system is at its peak from 8:00 a.m. to 9:00 a.m., and the number of visits decreases gradually from 16:00 p.m. every day.

Taken together, it can be seen that the user profile constructed using the method in the paper can effectively analyze the electricity consumption of residential electricity users in different time periods.

In order to further validate the advantages of the methods in the paper, simulation methods are used to analyze and validate the method with the residential power user's electricity behavior regulation as the test objective. Three scenarios are selected, of which Scenario 1 is the multi-dimensional fine-grained behavioral data-based residential user portrait method proposed in [5], Scenario 2 is the fusion of multi-source data researchers' portrait construction method proposed in [6], and Scenario 3 is the big data analysis-based residential power user behavioral portrait construction method proposed in the paper, and the three scenarios are utilized to randomly select 300 residents and construct user behavioral portraits, and to verify the application of the method in the paper by comparing the regulation potentials under the three scenarios. By comparing the regulation potential of residential power users under the three scenarios, the application performance of the method in the paper is verified, and the simulation rules are assumed to be Eq:

$$\Delta Q = \sum_{i=1}^{300} a_{Ei} a_{Gi} Q_i$$

Where: the objective regulation coefficient is aEi , defined as 0.4 for high-grade, 0.2 for mid-range, and 0.1 for low-grade; the subjective regulation coefficient for residential power users is aCi , defined as 0.7 for high-grade, 0.5 for mid-range, and 0.3 for low-grade; Qi is the load power. According to the above equation, it can be seen that the lower the overall regulation power, the lower the electricity consumption and the better the regulation potential.

In order to verify the validity of the methods in the paper, the researchers' portrait construction method fusing multi-source data proposed in Scheme 2; the residential user portrait method based on multi-dimensional fine-grained behavioral data proposed in Scheme 1 and the residential power user behavioral portrait construction method based on big data analysis proposed in Scheme 3 are compared in terms of the regulation effect of the residential power users, and the comparison results are shown in Fig. 4.

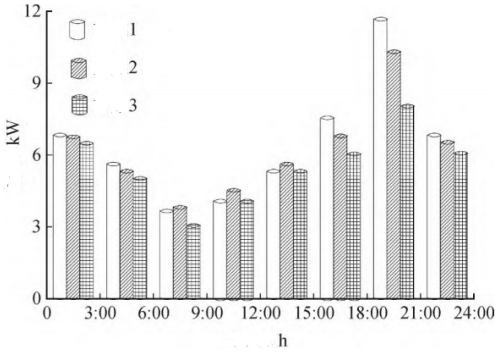

Fig. 4 Simulation analysis of user supply and demand interaction effect

According to the experimental results show that in the same time period using the method of this paper, the overall regulation power of Scheme 3 is lower than that of Scheme 1 and Scheme 2 in the method of this paper, the overall regulation power is lower, which indicates that the electricity consumption is less regulation potential, based on the method of this paper to build the behavioral image of residential electricity users, can accurately find out the potential for regulation of the residential electricity users will be accurately defined as the accuracy of the user image is consistent

with the behavior of the residential users of electricity, the ratio of the number of features of the image to the number of features of the total behavioral image of residential users. The accuracy of the user profile is defined as the ratio of the number of features of the profile to the number of features of the total residential electricity consumption behavior profile. The user profile accuracy formula is:

$$D_z = \frac{Z_t}{Z_z} \times 100\%$$

Where Zt denotes the number of features that match the behavioral profile of residential electricity users; Zz denotes the number of features in the total residential electricity user behavioral profile. The researchers' portrait construction method fused with multi-source data proposed in Option 2, the residential user portrait method based on multi-dimensional fine-grained behavioral data proposed in Option 1, and the residential electricity user behavioral portrait construction method based on big data analysis proposed in Option 3 are selected to conduct a comparative analysis of the behavioral portrait accuracy of 1,000 residents, and the comparison results are shown in Fig. 5:

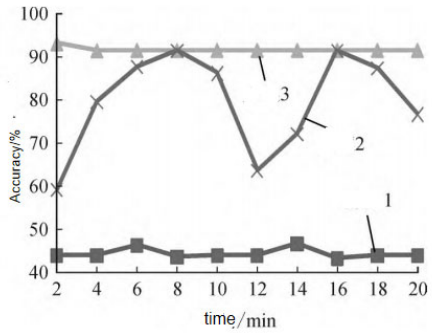

Fig. 5 Comparison of accuracy rate of behavioral portrait of residential electricity users

According to Fig. 5, the accuracy rate of the residential electricity user behavioral portrait of the method in the paper is high and smooth, while the accuracy rate of the researchers' portrait construction method fusing multi-resource data proposed in Scheme 2; the residential electricity user behavioral portrait of the residential user portrait method based on multi-dimensional and fine-grained behavioral data proposed in Scheme 1 is lower than that of the residential electricity user behavioral portrait of the residential electricity user portrait construction method based on big data analysis proposed in Scheme 3. The accuracy rate of the residential electricity user behavioral portrait proposed in Scheme 3 is lower than that of the residential electricity user behavioral portrait construction method based on big data analysis.

4. Concluding remarks

In this study, under the big data environment, based on the data provided by the electric power platform, the behavioral portrait of electric power users is constructed by constructing a user label library, performing user label management, and adopting the improved K-mean algorithm. The results of the study show that by labeling multiple dimensions such as social attributes, power attributes, value attributes, and Internet attributes of power users, an all-around user label library can be formed, which provides a more accurate user portrait for power enterprises. In the process of user behavioral portrait construction, the cluster analysis of power users is realized by improving the K-mean algorithm to form different types of user clusters, which provides accurate user information for business personnel. The experimental analysis shows that the constructed user portrait can effectively analyze the electricity consumption of residential power users in different time periods, which provides guidance for power companies to optimize energy supply and improve energy utilization efficiency. Comparative experimental results show that the residential power user behavioral portrait construction method based on big data analysis proposed in this study exhibits better performance in

terms of regulation potential and user portrait accuracy, which is superior compared to the other two schemes.

References

[1] LI Jinrui, ZHANG Jiabao, PENG Mei. Research and design of job-seeking user profiling system based on big data technology[J]. Industry and Technology Forum, 2019, 18(4):2. DOI:CNKI:SUN:CYYT.0.2019-04-040.

[2] Ding W, Wang T, Liu XH, et al. Research on cell phone user portrait and credit collection based on big data technology[J]. Post and Telecommunication Design Technology, 2016(3):6. DOI:10.16463/j.cnki.issn1007-3043.2016.03.014.

[3] Hua Z , Dawei W .Research on User Learning Behavior of the National Library Open Course Based on Big Data[J]. 2019.

[4] Chen Q .USER-BASED FRAUD BEHAVIOR ANALYSIS AND RESEARCH ON E-LEARNING PLATFORMS: AN EMPIRICAL STUDY[J].Journal of nonlinear and convex analysis,. 2019, 20(6).

[5] Yang Guirong. Intelligent user profiling method and device based on IoT big data platform:CN 201410531377[P][2023-12-11].

[6] Wei, Loan Mei. Research on user profiling technology integrating social media content and behavioral data [D]. Northwest Normal University [2023-12-11].