METHOD FOR ANALYZING AND SCREENING ELECTRICITY CONSUMPTION BEHAVIOR OF DEMAND SIDE USERS

TECHNICAL FIELD

The invention relates to the technical field of power system operation and control, in particular to a method for analyzing and screening electricity consumption behavior of demand side users.

BACKGROUND

The implementation of existing demand response depends on the accurate analysis of users' electricity consumption behavior, but the current technology still has shortcomings in data preprocessing, pattern recognition and user screening. Specifically, the original load data is vulnerable to noise and abnormal interference, while the traditional cleaning method lacks the correction mechanism based on the essential characteristics of the data, which leads to the distortion of the basic data. In the construction of user portraits, commonly used clustering algorithms mostly rely on random initialization and subjective feature selection, which easily fall into local optimum and cannot objectively reflect the differences in behavior patterns. In addition, the existing user screening strategies pay more attention to a single index such as response capacity, and lack a multi-objective optimization mechanism that comprehensively weighs price sensitivity and response cost, so it is difficult to achieve the best balance between economy and response effect under budget constraints.

SUMMARY

The invention discloses a method for analyzing and screening electricity consumption behavior of demand side users. This method uses competitive learning network to establish standard characteristic curve to correct abnormal data, and divides user behavior patterns based on multidimensional feature weighting and local density clustering algorithm. Further, a multi-objective optimization model is constructed according to the comprehensive price sensitivity of users, and the optimal screening list and suggested response load are determined by solving. The invention can effectively improve data quality and portrait accuracy, and realize accurate user screening with consideration of response cost and effect.

The invention provides a method for analyzing and screening the electricity consumption behavior of demand side users, which has the following beneficial effects.

1. According to the invention, the standard characteristic curve is established by using the competitive learning network, and the abnormal data is corrected by combining the peak-valley precedence ratio, so that the noise interference is effectively eliminated and the real power consumption law is restored, the problem of analysis deviation caused by low data quality is solved, and the accuracy of basic data is improved.

2. According to the invention, the entropy weight method is adopted for objective weighting, and the clustering algorithm based on local density is combined, so that the defects of subjectivity of human experience and the fact that the traditional algorithm is easy to fall into local optimum are avoided, and the discrimination degree of user electricity consumption behavior patterns is effectively improved.

3. The invention constructs a multi-objective optimization screening model, gives consideration to the comprehensive price sensitivity of users and the project budget limit, and screens the user group with the best response effect on the premise of strictly controlling the response cost, thus improving the economic benefit of demand side response.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 is a flowchart of the method of the present invention.

DETAILED DESCRIPTION

Referring to FIG. 1, the invention provides a method for analyzing and screening electricity consumption behavior of demand side users, which comprises the following steps:

In S100, the data correction module performs data cleaning and correction.

The data correction module obtains the original power load data of users in the target area. In order to ensure the accuracy of subsequent analysis, firstly, time alignment and missing value filling operations are performed for multi-source heterogeneous data. The specific implementation process is as follows:

S101, data segmentation and feature detection based on information entropy. The data correction module carries out discrete sampling and segmentation on the power consumption characteristic data in the target area. Using the principle of information entropy to measure the uncertainty of data. By calculating the change of information entropy before and after clustering, the feature points with the largest carrying information are selected. The calculation formula of information entropy H is as follows:

$H=-\sum_{i=1}^{n} P_i \log_2(P_i)$;

where: P_i represents the frequency probability of the i-th feature point appearing in the data set; n is the number of partitioned data subsets. According to the calculation results, the system preferentially selects the point with larger H value as the initial analysis object.

S102, the feature cluster is normalized by using the minimum distance principle. The data correction module uses the principle of minimum distance to normalize the characteristic data clusters (that is, clustering and merging). The distance $d(S_1,S_2)$ between two clustering sets S_1 and S2 S_2 is defined as the nearest point-to-point distance according to the minimum distance method:

$$d(S_1,S_2)=\min_{p\in S_1,p_m'\in S_2}\|p-p_m'\|^2;$$

where: p and p_m' are the core points or feature points in the cluster sets S_1 and S_2, respectively.

In order to eliminate the magnitude difference between different users (such as industrial and residential users), the range transformation method is used to map the data to a unified interval.

On the basis of normalization, the information gain I_{gain} generated by feature point merging is calculated:

$$I_{gain}\left(S_{p_m'}\right)=I\left(S_p+S_{p_m'}\right)-I(S_p).$$

Through iterative calculation, the feature points clustered for the first time in segmentation are processed until the number of clusters reaches the preset number of clusters or the information gain increment is lower than the convergence threshold. If the minimum distance between a data segment and the centers of all clusters exceeds the abnormal threshold (for example, 3σ) set based on the standard deviation of the data set, and the information gain after merging is abnormal or lower than the average level of the same batch, it is determined that the data segment is abnormal data.

S103, extracting the standard characteristic curve by using Kohonen neural network. Aiming at the abnormal data identified in S102, the data correction module uses Kohonen competitive learning network to extract benchmarks, inputs historical normal load data into the network, adjusts the weights of winning neurons and their neighbors through the competition mechanism to fit the data distribution, and directly extracts the weight vector of winning neurons as the standard characteristic curve X_t after the training is completed.

S104, the data correction module constructs a shape correction model based on the standard characteristic curve X_t output by Kohonen network and the original load curve X_d to be corrected. Select the pre-order sampling points at the peak time p_1 and the valley time p_2 of the daily load curve (i.e. p_1-1 and p_2-1) as the anchor points for actual calculation, and calculate the revised load curve $X_r(i)$ as follows:

$$X_r(i)=\frac{X_t(i)}{2}\left[\frac{X_d(p_1-1)}{X_t(p_1-1)}+\frac{X_d(p_2-1)}{X_t(p_2-1)}\right];$$

where: i represents the currently calculated sampling time index; $X_d(p_1-1)$ and $X_t(p_1-1)$ respectively represent the load values of the original curve and the standard curve at the moment before the peak value.

In step S200, the cluster analysis module receives the modified $X_r(i)$ and performs user grouping based on the modified data morphological features. The specific implementation process is as follows:

S201, the system traverses each user's $X_r(i)$ and extracts key physical quantities: load rate (x_1), level coefficient (x_2), valley coefficient (x_3) and peak coefficient (x_4). In order to eliminate the influence of dimension, the following formula is used to standardize the range of each characteristic index, and the formula for calculating the normalized characteristic value x_{ij}' is as follows:

$$x_{ij}'=\frac{x_{ij}-\min(x_j)}{\max(x_j)-\min(x_j)+\epsilon};$$

where: x_{ij} represents the j-th characteristic index value of the i-th user; $\min(x_j)$ and $\max(x_j)$ represent the minimum and maximum values of the j-th feature in all m user samples, respectively. ϵ is the minimum value of floating-point number (for example, 10^{-6}) set to prevent the denominator from being zero, all characteristic indexes are dimensionless and mapped to the closed interval of $[0,1]$.

S202, after data normalization, the clustering analysis module introduces entropy weight method to objectively weight each characteristic index. The specific steps for generating the weighting and weighting matrix are as follows:

Step 1, calculating the proportion P_{ij} of the characteristic value of the i-th user to the sum of the characteristics under the j-th characteristic;

$P_{ij}=\frac{x_{ij}'}{\sum_{i=1}^{m} x_{ij}'}$;

Step 2, calculating the information entropy E_j of the j-th feature according to the specific gravity matrix;

$E_j=-k\sum_{i=1}^{m} P_{ij}\ln(P_{ij})$;

where k is the adjustment coefficient, defined as $k=\frac{1}{\ln(m)}$, and its physical function is to ensure that the range of information entropy E_j is within $[0,1]$. when $P_{ij}=0$, $\lim_{P_{ij}\to 0} P_{ij}\ln(P_{ij})=0$, or modifies P_{ij} to $P_{ij}+\delta$ (δ is a minimal positive number) in the calculation program.

Step 3, calculating the difference coefficient $D_j=1-E_j$ of each characteristic index according to the information entropy, and then normalizing to get the final weight W_j:

$W_j=\frac{(1-E_j)}{k-\sum E_j}$;

the formula shows that the smaller the information entropy E_j, the larger the difference coefficient D_j, the more useful information provided by this index, and the greater the weight W_j.

Step 4, weighting x_{ij}' by using the calculated weight vector to generate a weighting matrix Y of comprehensive electricity consumption information of users, and the element y_{ij} is calculated as follows:

$$y_{ij}=x_{ij}'\times W_j;$$

Through the above steps, the generated weighting matrix Y realizes the secondary reconstruction of the original data: it not only eliminates the dimensional influence, but also highlights the characteristic indexes with high discrimination according to the distribution characteristics of the data itself (for example, if the valley coefficients of most users are similar, but the peak coefficients are greatly different, the algorithm will automatically increase the weight of the peak coefficients), thus providing a high-dimensional measurement space that can truly reflect the heterogeneity of user behavior for the subsequent improvement of the K-means algorithm.

Based on the weighting matrix Y of user comprehensive electricity consumption information generated in step S202, which can truly reflect the heterogeneity of user behavior, this embodiment uses an improved K-means clustering algorithm to perform user grouping, and the specific implementation process is as follows:

S203, calculating the average distance $MD(x_p)$ of each sample point x_p in the weighting matrix Y;

$$MD(x_p)=\frac{1}{n(n-1)}\sum_{q\neq p}\|x_p-x_q\|;$$

Then calculate the local density $DS(x_p)$ of the sample point x_p, and the calculation formula is as follows:

$$DS(x_p)=\sum_{q=1}^{n}H\left[MD-d(x_p-x_q)\right];$$

$$H(x)=\begin{cases}0, & x\leq 0\\ 1, & x>0\end{cases};$$

where: n is the total number of samples, and $\|x_p - x_q\|$ is the Euclidean distance between sample points x_p and x_q; $d(x_p - x_q)$ is the Euclidean distance between any two objects in the dataset. $DS(x_p)$ is the density of the object x_p, and the density set is $D=\{DS(x_1),DS(x_2),\cdots,DS(x_n)\}$; $H(x)$ is a step function.

According to the principle of maximum local density, the maximum density point and the second largest density point are selected as the clustering centers of the first and second categories in turn, and so on until the category K value is met.

S204, the system defines the clustering validity index $\rho(k)$ to determine the optimal k value, and the index is defined as the ratio of intra-class compactness to inter-class separation:

$$\rho(k)=\frac{1}{k}\sum_{p,q=1}^{k}\max_{q\neq p}\left\{\frac{MD(p)+MD(q)}{d_{p,q}}\right\};$$

where: $MD(p)$ represents the average distance within the class of the p-th cluster; $d_{p,q}$ represents the Euclidean distance (separation degree) between the center of the p class and the center of the q class; select the k corresponding to the minimum value of $\rho(k)$, that is, determine the optimal number of clusters by minimizing the clustering effectiveness index.

For step S300, based on the clustering results, the demand response user screening module first identifies the target cluster with high peak clipping and valley filling potential (for example, the cluster with high average peak coefficient x_4), and takes the users in this cluster as the candidate set. Subsequently, a screening mechanism based on price elasticity matrix is constructed for users in the candidate set to screen high-potential users. The specific implementation process is as follows:

S301, the demand response user screening module identifies the time point when the electricity price changes step by step in combination with the time-sharing electricity price strategy implemented by the power grid in the target area, and traverses the electricity price vector in the all-day sampling period to form the electricity price conversion time set T_{trans}.

S302, for each electricity price conversion time $t \in T_{trans}$, the system calculates the single-point price sensitivity ed_{ij} of user i at time t, and the calculation formula is as follows:

$ed_{ij} = \frac{(L_t - L_{t-1})/L_{t-1}}{(C_t - C_{t-1})/C_{t-1}}$;

where: i represents the unique identification of the user or user category; t represents the moment when the electricity price changes; $L_{i,t}$ and $L_{i,t-1}$ respectively represent the actual power load value of user i at time t and the previous time $t-1$; C_t and C_{t-1} respectively represent the electricity selling price of the power grid at time t and the previous time $t-1$;

S303, calculating the comprehensive price sensitivity ed_i of users:

$ed_i = \sum_{j=1}^{n} \omega_j \cdot ed_{ij}$;

where: ω_j is the weight coefficient of the j-th transition moment.

S304, multi-objective optimization screening.

Based on the calculated single-point price sensitivity, this embodiment further constructs a multi-objective optimization screening model to screen out the user combination with the highest comprehensive sensitivity and the lowest total quotation. The objective function is as follows:

$\min F = \omega_{ed} \frac{1/ed}{1/ed_{max}} + \omega_B \frac{B}{B_{max}}$;

constraints include:

$P_{min} \le \sum k_i P_i \le P_{max}$; $ed = \sum_{i=1}^{M} \frac{k_i P_i}{\sum_{i=1}^{M} k_i P_i} k_i ed_i$;

$B = \sum_{i=1}^{M} k_i B_i$; $P_{tmin} \le \sum_{i=1}^{M} k_i P_i \le P_{tmax}$; $k_i = 0,1$;

where: ed_{max} is the maximum comprehensive sensitivity; ed is the comprehensive sensitivity of power users; B is the response total quotation; B_{max} is the maximum total quotation; ω_{ed} and ω_B are the comprehensive price sensitivity weight and the total quotation weight respectively. $ed_i \in ed_\Pi$ is the weighted sensitivity of user i considering all transition points, $B_i \in B_\Pi$ is the user's load adjustment, and P_{tmin} and P_{tmax} are the minimum overall target load and the maximum overall target load respectively; k_i is the screening situation, and $k_i=0$ indicates that power is not selected to participate in the project, whereas $k_i=1$.

By solving the model, the list of end users is determined.