Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Unsupervised classification model of residential power users based on typical load patterns

To cite this article: Yingjie Tian *et al* 2022 *J. Phys.: Conf. Ser.* **2310** 012079

Unsupervised classification model of residential power users based on typical load patterns

Yingjie Tian[1a *], Zixue Zhai[2b], Fan Li[1c], Yi Wu[1d], Yingying Zhao[1e], Yun Su[1f]

[1] State Grid Shanghai Municipal Electric Power Company, Shanghai 200122, China

[2] School of Data Science, Fudan University, Shanghai 200433, China

[a*]13901712348@163.com, [b]121742607@qq.com, [c]tianda907@163.com, [d]wu.yi.christian@gmail.com, [e]872858177@qq.com, [f]oppenvi@163.com

Abstract—It is critical for the smart grid to mining the power users' consumption behaviors. The accumulation of power data provides the possibility of conducting experiments for research focus on behavior analysis. This paper proposes a residential power users' classification method based on typical load patterns called UCM-LP, to classify users' power consumption behavior. The method firstly performs fuzzy grouping on the typical power consumption characteristics of multiple users through two-stage clustering, then divides cluster label to obtain the final power consumption categories of the user. Finally, this paper verified the algorithm using the real household power consumption data, obtaining the specific power consumption category patterns and discussed the differences among power consumption categories.

1. Introduction

Power user classification is to classify various load patterns of power users [1]. Considering the complexity and variability of users' power consumption behaviours, users within the same power consumption category may have different power consumption characteristics, while users of different categories may share some characteristics in common. In the previous studies, machine learning, deep learning and other artificial intelligence algorithms are used to classify load patterns, such as ANN [2], self-organizing map (SOM) [3] and fuzzy model [4]. The one-stage classification framework can intuitively identify typical load patterns, however typical load patterns dealing with this kind of method cannot fully represent user characteristics, since the daily load patterns of users would fluctuate.

As a result, some scholars have also paid attention to the multi-stage classification framework. Panapakidis et al. [5] adopted a two-stage approach to achieve the characterization of consumer categories. The algorithm firstly obtains the load pattern of a single consumer by clustering the daily load curves of users and then performs the second clustering to form consumer categories. Noting the limitation of using a representative load pattern for each consumer, Mets et al. [6] proposed another fast wavelet transform and G-means algorithm for two-stage load pattern clustering, to retain as much information as possible. In order to classify the new consumers, Buitrago et al. [7] constructed a hybrid system consisting of a parameter estimation model, a clustering model and a neural network model. According to the above literature review, this paper proposes a multi-stage power user classification method based on typical load patterns. This method selects multiple typical power consumption patterns to reduce information loss, and uses two step clustering to fuzzy group them.

2. User Classification Model based on Typical Load Patterns

2.1. Structure of UCM-LP
In this paper, we propose a power user classification model based on typical load patterns, which we called UCM-LP. As shown in Fig.1, UCM-LP consists of three main stages. The first stage is the extraction of typical user load characteristics. This stage mainly clusters the load curves. The fusion clustering algorithm based on discrete wavelet transform (CC-DWT) [8] will be applied on the consumer level, taking all the daily load curves of each users as an input, and typical load patterns as the output. The second stage carries out a further clustering. In this stage, the fuzzy grouping means that the load patterns of all users are combined and clustered to obtain fuzzy user groups. The third stage will cluster the labels. In this stage, the label clustering is mainly performed on the fuzzy user grouping results obtained in the previous stage to clarify the user category and extract the typical power consumption characteristics.

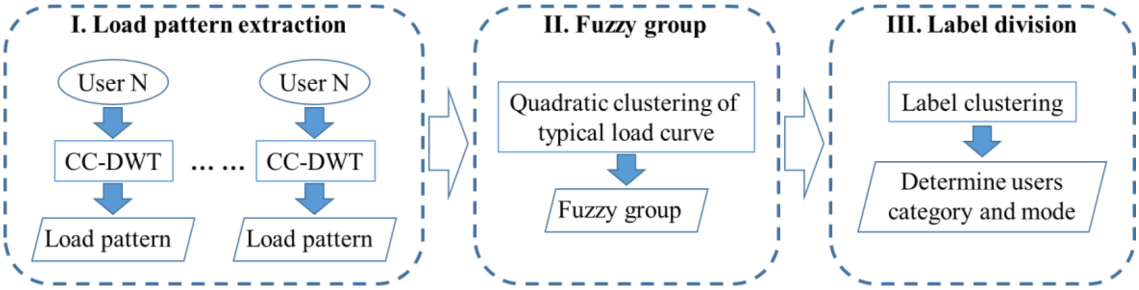

Fig.1 UCM-LP process

2.2. Fuzzy grouping
The power consumption of a single user usually has multiple load patterns. Referring to related works, when extracting the load pattern, the grid search method is used to find the optimal parameter combination. To reduce the complexity, the traditional method usually selects a specific load pattern for each user as the typical pattern, and then executes subsequent clustering. The user will be considered as a member of a certain group if its representative pattern is in the corresponding group. However, this will lead to a huge loss of information. Therefore, in the UCM-LP model, we keep all the typical power consumption features extracted in the first stage as the input of the second stage.

Since the data of all typical loads of N users may be too large for clustering, and if directly implementing clustering, the number of clusters may be too small. Accordingly, refer to relevant literature [9][10], we propose a load pattern clustering process shown in Fig.2. The typical power consumption characteristics of all users are randomly divided into n groups, then perform clustering within each group respectively. The clustering algorithm used here is CC-DWT too. Next, the clustering results of the n groups are mixed and clustered again to obtain m groups of clusters. Finally, the fuzzy grouping of all users is obtained by mapping the m clusters to the users' input data.

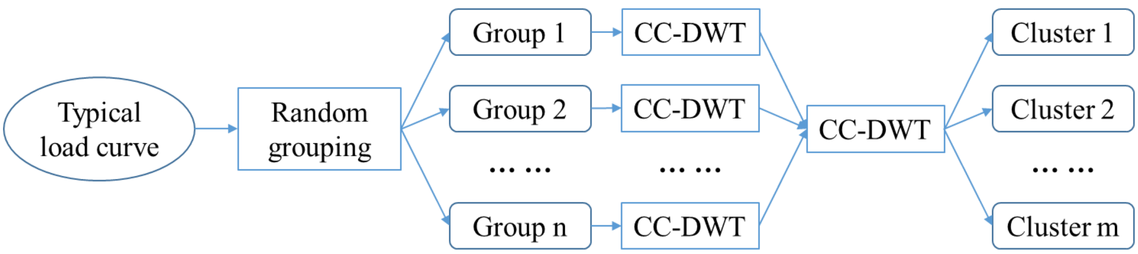

Fig.2 Fuzzy grouping process

2.3. Unsupervised label clustering

After the previous two steps, we propose a user category classification algorithm based on the obtained user fuzzy grouping results. The algorithm not only obtains a clear user classification, but also obtains the typical power consumption behavior characteristics of each user categories. The Algorithm 1 adds multiple labels to users who belong to multiple fuzzy groups, then K-modes clustering algorithm is adopted to obtain a clear division of user power consumption categories. Then, according to the cluster centers of K-modes and the user clustering center curve from fuzzy grouping, the center load curve corresponding to the cluster group is obtained, which is regarded as the typical power consumption behavior characteristic of the cluster group.

Algorithm 1: Label clustering

Input: fuzzy user category set C, corresponding cluster center C_{center}, user list $list_{id}$
Output: clear user category set L, corresponding category feature center $label_{center}$
for or id in $list_{id}$:
  for or C_k in C:
if id in C_k ***then*** $label_{id}(k) = 1$ ***else*** $label_{id}(k) = 0$
end for
 $Label = [label_{id} \quad for \;\; id \;\; \text{in} \;\; list_{id}]$
for or $\epsilon = 0.001, \epsilon \leq 1, \epsilon += 0.05$
  for $2 < k \leq 15,\; k++$ ***do***
    $L_k = KModes(Label,\; k)$;
    $SWC_k = SWC(L_k\;,\; metric =' hamming')$;
end for
$k_L = \; arg \max_k \{SWC_k\}$;
$L = KModes(Label,\; k_L)$;
Select the closest to each category L_{center}as $label$the feature label of the category $label_{center}$;
According to $label_{center}$the center curve in the corresponding selection , as the characteristic performance of the category;C_{center}
r etur n L, $label_{center}$.

The K-modes clustering algorithm is an extension of the traditional K-Means algorithm [11]. The traditional K-Means algorithm clusters numerical data (continuous data), while the K-modes clustering algorithm can process categorical attribute data (discrete data), such as gender, education level and so on. Different from the K-Means algorithm, the K-modes algorithm is based on a simple dissimilarity measure for clustering, where the simple dissimilarity (Eq.(1) and Eq.(2)) can be determined by the sum of the mismatch between the corresponding attribute categories of the two object. The smaller the dissimilarity, the more similar the two objects are.

$$d(X,Y) = \sum_{j=1}^{m} \delta(x_j, y_j) \tag{1}$$

$$\delta(x_j, y_j) = \begin{cases} 0, & (x_j = y_j) \\ 1, & (x_j \neq y_j) \end{cases} \tag{2}$$

Subsequently, we employ Silhouette Coefficient to evaluate the clustering of categorical attribute data. Among them, the distance of Silhouette Coefficient adopts Hamming distance to measure the distance between attribute data.

3. Experimental Results

We tested UCM-LP with the 15-minute power consumption data of residents in Shanghai Pudong. The data contains 536 households with a time span from April 1, 2016 to March 31, 2018.

3.1. Optimal layers of 1-D discrete wavelet transform

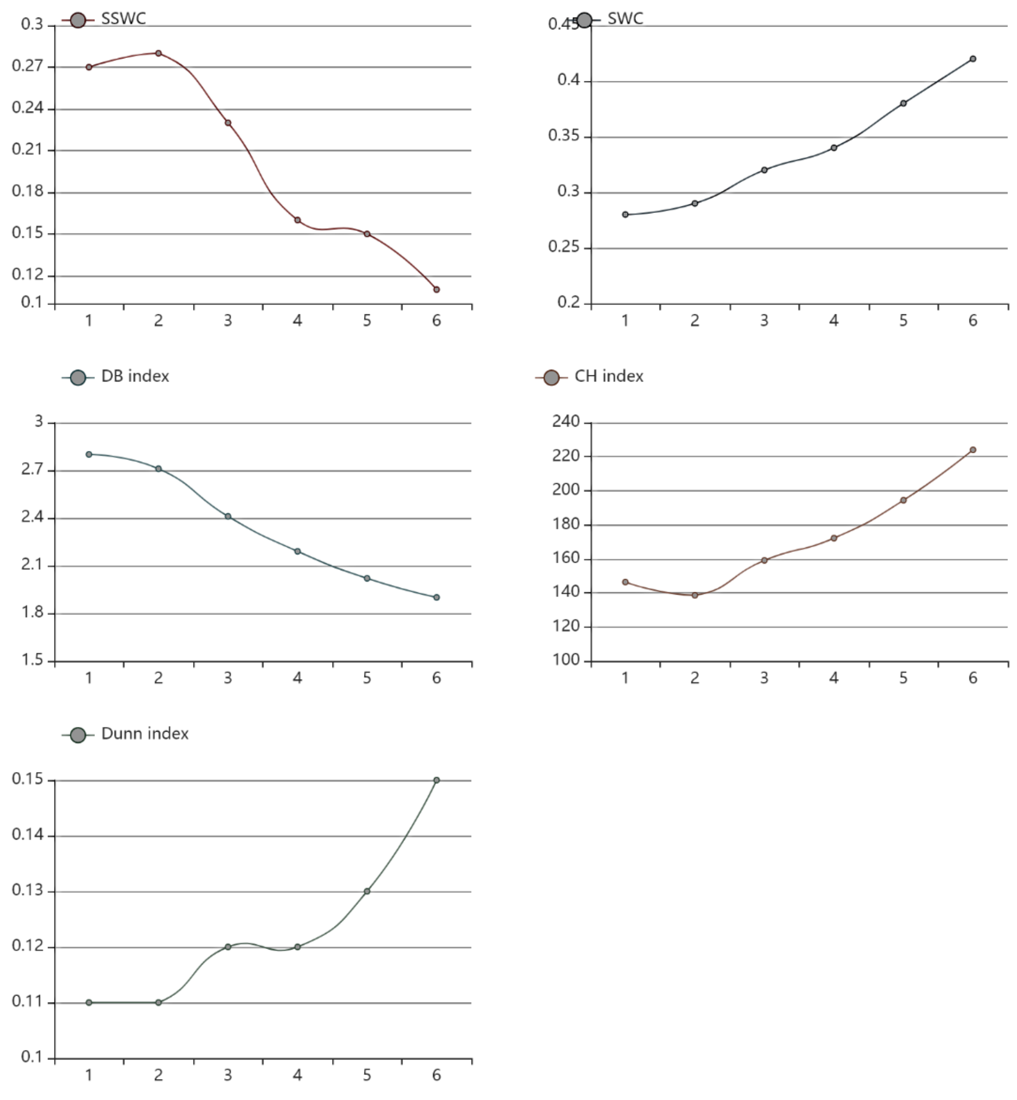

Fig.3 Optimal grid search results for parameters α

In the CC-DWT algorithm, the number of layers of 1- D discrete wavelet transform α is required in the dimension reduction stage. To determine the optimal α, we used a grid search method. Fig.3 shows the clustering performance index result corresponding to different α. Since our clustering data is unlabeled, Silhouette Coefficient (SWC), Davies-Bouldin Index (DB), Dunn Qualidity Index (Dunn), and the Calinski-Harabasz Index (CH) were selected for verifying the performance of the algorithm. These clustering evaluation indicators, except that a smaller DB index indicates better performance, all show the better clustering effect with larger values. Through Fig.3, we found that different α values have different performances in each indicator. We synthesized and determined that the optimal situstion is set α=6.

3.2. Load curve classification

Through experiments, the typical load patterns of 536 residential users can be divided into 6 categories by UCM-LP. Fig.4 shows the behaviour categories with the corresponding typical load patterns. We can see that the most common type of residential power consumption is mode 1, which is in line with the

 2310 (2022) 012079 doi:10.1088/1742-6596/2310/1/012079

living habits of residents who normally commute to get off work. Through the comparison between mode 1 and mode 2, it can be speculated that such a difference in power consumption mode may be due to seasonal weather. In summer or winter, residents will use high-power household appliances such as air-conditioners at night, so that the power consumption will increase relatively larger. If the resident 's power consumption category is mode 3, it is possible that the resident will travel for a long time for work reasons, or the house is lived by the elderly, because the power consumption is stable and very little. The residential power consumption mode 6 indicates that the resident may be more sensitive to temperature and weather.

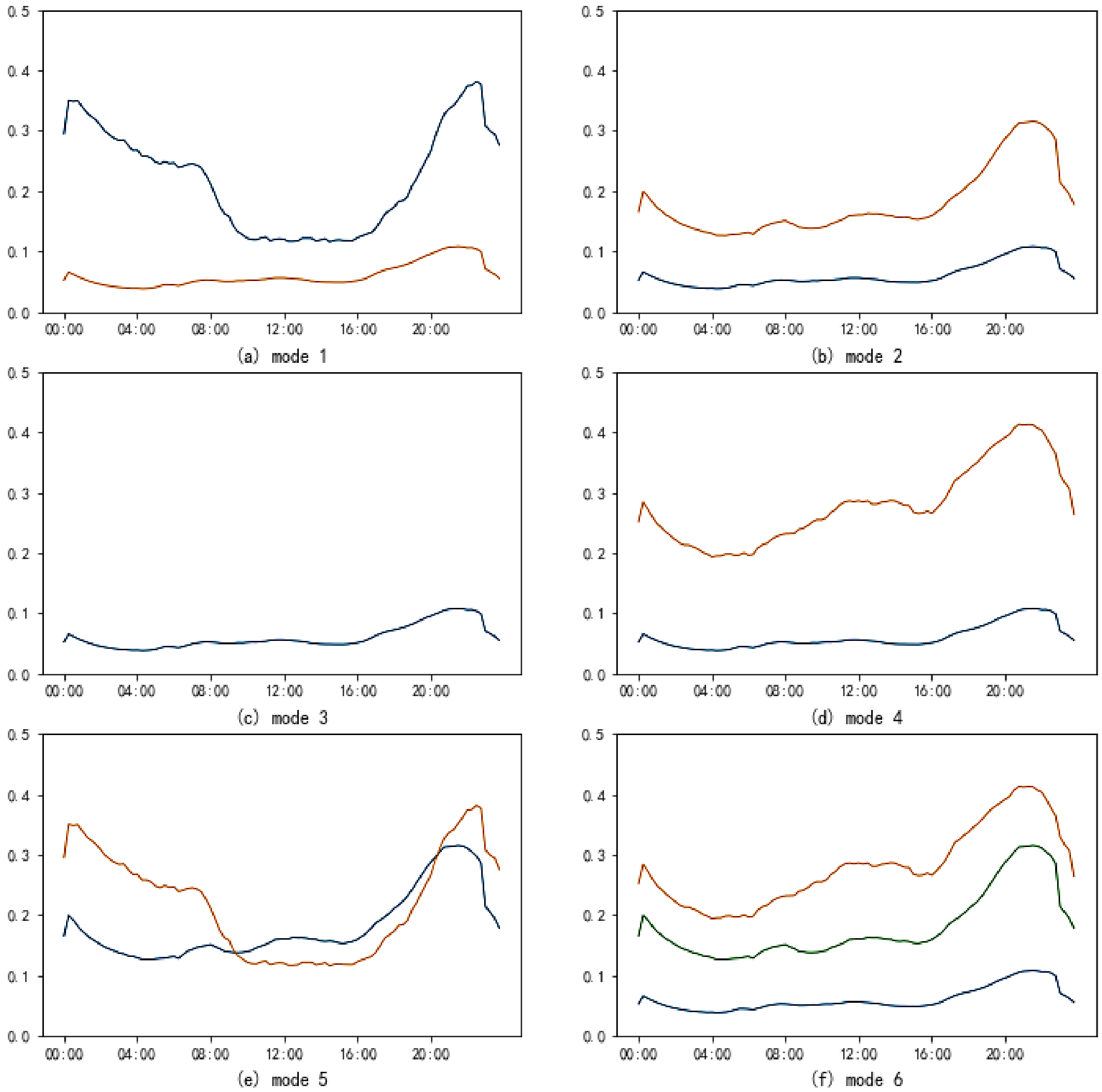

Fig.4 Residential power categories

4. Conclusion

In this paper, a new classification model based on typical load characteristics is proposed for the classification of power consumption categories of power users. On the basis of extracting the typical power consumption characteristics of a single user, UCM-LP randomly divides the typical power consumption characteristics of all users into *n* groups, and uses two-stage clustering to obtain the fuzzy grouping of users, then perform label clustering to get user classification. This method overcomes the fuzzy grouping problem that may be caused by direct clustering of power consumption characteristics,

Journal of Physics: Conference Series **2310** (2022) 012079 doi:10.1088/1742-6596/2310/1/012079

which can avoid information loss as much as possible and retain more user categories. We conducted an empirical test on the real residential power consumption dataset in Shanghai, and identified 6 types of residential power consumption modes. Such user classification is conducive to power supply providers to allocate power supply, plan the distribution network rationally, meet the user's increasing load demand and provide users with the better demand programming.

Acknowledgments
This work was financially supported by the Science and Technology Project of State Grid Corporation of China (No. 52094020006M) and Shanghai Electric Power Artificial Intelligence Engineering Center (No. 19DZ225280).

References
[1] Zhou, K., Yang, S., Shen C. (2013) A review of electric load classification in smart grid environment. Renewable & Sustainable Energy Reviews, 24: 103-110.
[2] Sarava, F., Bernades, W., Asada, E. (2015) A framework for classification of non-linear loads in smart grids using Artificial Neural Networks and Multi-Agent Systems. Neurocomputing, 170 : 328-338.
[3] Hernandez, L., Baladron, C., Aguiar, J., et al. (2012) Classification and Clustering of Power Demand Patterns in Industrial Parks. Energies, 5(12): 5215-5228.
[4] Vigas, J., Vieira, S., Melicio, R., et al. (2016) Classification of new power customers based on surveys and smart metering data. Energy, 107: 804-817.
[5] Panapakidis, I., Alexiadis, M., Papagiannis, G. (2012) Power customer characterization based on different representative load curves. 2012 9th International Conference on the European Energy Market. IEEE, 1-8.
[6] Mets, K., Depuydt, F., Develder, C. (2016) Two-Stage Load Pattern Clustering Using Fast Wavelet Transformation. IEEE Transactions on Smart Grid, 7(5): 2250-2259.
[7] Buitrago, J., Abdulaal, A., Asfour, S. (2015) Electric Load Pattern Classification Using Parameter Estimation, Clustering and Artificial Neural Networks. International Journal of Power and Energy Systems, 35(4): 167-174.
[8] Li, F., Tian, Y., Wu, Y., et al. (2021) A method of mining power consumption behaviour based on CC-DWT. IOP Conference Series: Earth and Environmental Science, 791.
[9] Tsekouras, G., Hatziargyriou, N., Dialynas, E. (2007) Two-stage pattern recognition of load curves for classification of power customers. IEEE Transactions on Power Systems, 22(3): 1120-1128.
[10] Panapakidis, I., Alexiadis, M., Papagiannis, G. (2013) Three-stage clustering procedure for deriving the typical load curves of the power consumers. 2013 IEEE Grenoble Conference. IEEE: 1-6.
[11] Huang, Z. (1998) Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery, 2(3): 283-304.