

Hybrid Features based K-means Clustering Algorithm for use in Electricity Customer Load Pattern Analysis

Pengyuan Liu¹, Chenye Yang¹, Jiang Wu¹, Xingbo Fu¹, Ruanming Huang², Yichao Huang², Fei Fei²

1. Ministry of Education Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China
E-mail: lp1001@stu.xjtu.edu.cn, yangchenye2015@stu.xjtu.edu.cn, xbfu@sei.xjtu.edu.cn
2. State Grid Shanghai Electric Economic and Technology Research Institute, Shanghai 200120, China
E-mail: harmjobhrms@163.com, huangyc@sgcc.com.cn, feifei@sgcc.com.cn

Abstract: In the power system, the load pattern of users is important for electricity price formulation, abnormal electricity consumption detection and load forecasting. Accurately identifying the load patterns effectively improve the overall load forecasting performance of the power grid. However, due to many factors affecting the load, identifying load patterns is not a simple job. In this paper, the features of load data are extracted and different load patterns of users are identified. In the pattern classification, K-means clustering is used. In feature extraction, PCA and prior knowledge were respectively used for dimensionality reduction to extract new features in the past. Combining the advantages of PCA and prior knowledge, a new method, namely, weighted combination of the obtained features from two former methods, is adopted to extract new features and identify the user's load pattern. In the similarity measure, the variance of each feature is taken as its weight, and the clustering result is superior to that of PCA and prior knowledge. In the numerical simulation, it is found that the clustering results using the new method in this paper are better than those using only PCA to reduce dimensionality or using prior knowledge to reduce dimensionality. In the experiment, the users in a certain area were divided into four categories according to the clustering result of load data: peak electricity type, partial peak electricity type I, partial peak electricity type II and abnormal electricity type.

Key Words: K-means, PCA, Priori Knowledge, User Classification, Load Pattern

1 Introduction

Nowadays smart meters make it possible to calculate and analyze load data of power users in the era of big data. In the power system, cluster analysis of load curves plays a very important role in demand-side management, optimal regulation on resource and so on. The specific applications are "load forecasting", "abnormal electricity detection" and "time-of-use price development", etc.[1] Although there are more than 300 million smart meters in China at present, the load analysis of their datasets has not yet been explored because of the limitations in bandwidth in communications and the huge amount of data[2,3].

If indirect methods, such as preprocessing data first and then clustering or doing pattern classification, are used, the time and space complexity of algorithms can be reduced. For example, using dimension reduction to do the feature extraction and then clustering the extracted features lead to the optimization of time and space resource of clustering algorithms.

There are mainly four methods for feature extraction of the data collected by smart meters[4]:

(1) Load data is a kind of time series. Electricity consumption has similarity over a period of time. Therefore, features can be constructed on the time scale to reduce the consumption of calculating resources. For example, under normal circumstances, there will be no drastic changes in power consumption from 7:00 am to 12:00 noon. Hence, it is possible to construct a feature in some way to represent the load information over this period.

(2) For the load curve, some features that can represent the daily load characteristics have been summed up, such as load rate[5], peak valley load rate and other indicators. Liu Si proposed a daily load curve clustering method based on the reduction of the characteristics. The method has low time complexity and good robustness[6].

(3) Some new features can be obtained by mathematical transformation, like, time series can be transformed using wavelet transform or Fourier transform[7] to get the corresponding harmonic coefficients.

(4) Use the principal component analysis (PCA), curve component analysis or typical analysis of variables to do dimensionality reduction, and then choose the data after dimensionality reduction as the features.

(5) Economic factors, residence, household appliances and other macro-factors will affect the electricity consumption of users[8]. We can calibrate the macroscopically facts influencing the power consumption by means of human intervention, and then use it as one of the features for subsequent clustering.

The rest of this paper is organized as follows. In Section II, the centralized K-means algorithm and some clustering validity assessments are introduced. In Section III, the combination of features based on PCA dimensionality reduction and prior knowledge model is given. In Section IV, results from simulation cases are provided. Finally, some relevant conclusions are given in Section V.

2 Clustering Technology

2.1 Centralized K-means Algorithm

Clustering is one of the most commonly used techniques in the field of data mining. Its main idea is to use similarity measurement to aggregate similar samples together. There

*This work is supported in part by the National Key R&D Program of China (2016YFB0901900), National Natural Science Foundation (NNSF) of China(61773308, U1766205), and the State Grid Science and Technology Program of China.

are many clustering algorithms, such as partition-based clustering algorithms (K-means), hierarchical clustering algorithms (BIRCH), density-based clustering algorithms (DBSCAN) and grid-based clustering algorithms (STING). K-means algorithm is concise, fast and its clustering result is satisfactory. Hence, K-means algorithm has been widely used[9].

K-means is a typical partition-based clustering method, the goal of which is to divide dataset $D=[d_1, d_2, \dots, d_n]$ into K clusters $[D^1, D^2, \dots, D^K]$. Let $C=\{c_1, c_2, \dots, c_K\}$ denote the set of centroids. The goal of K-means can be expressed as follows:

$$\min J = \sum_{i=1}^N \sum_{k=1}^K \mu_k^i \|d_i - c_k\|^2 \quad (1)$$

In the formula:

J — Sum of the square of the error of all the data.

N — The size of dataset D .

K — The number of clusters.

d_i — The i -th data point.

c_k — The k -th centroid.

$$\mu_k^i = \begin{cases} 1, & \text{if } d_i \in \text{Cluster } k; \\ 0, & \text{if } d_i \notin \text{Cluster } k; \end{cases} \quad (2)$$

The hyper-parameter k in the clustering depends on the distribution of data points, so it's difficult to find a suitable k value. Different k values have different effects on the clustering result. In this paper, we use the elbow method to determine the best number of clusters, namely, the best k value. Elbow method represents the degree of distortion. As the number of clusters increases, the average degree of distortion decreases. At this point, the number of samples contained in each cluster will decrease, which will lead to a closer distance between each sample and its related cluster centroid. However, as k increases, the effect of reducing the average distortion will continue to decrease. So the best value of k is the elbow where the effect of reducing average distortion drops greatly.

The K-means algorithm can be summarized in two steps: the first one is an assignment step and the second one is a refinement step. In the assignment step, what we need is to calculate the Euclidean distance between each data point d_i and the initial cluster center. After that, data points are grouped into its nearest cluster which can be represented by the nearest centroid. In the refinement step, the K centroids are updated using formula(3):

$$c_k(t+1) = \frac{\sum_{i=1}^N \mu_k^i(t+1) * d_i}{\sum_{i=1}^N \mu_k^i(t+1)} \quad (3)$$

After the centroids converges or the number of iterations reaches the specified number of times, the algorithm ends.

2.2 Clustering Validity Assessment

In the clustering evaluation, there are not only evaluations based on external data, but also evaluations based on clustering itself. Silhouette Coefficient is an validity

indicator which reflects the tightness of data within the cluster and the separation between the clusters. When we don't know the number of clusters, SI coefficient can be used to determine the optimal number of clusters and evaluate the quality of clustering. Let \bar{d}_i represent the average distance between data i and other samples in the same cluster. \bar{d}_i is used to characterize the degree of tightness within the cluster. The smaller the value, the closer the data points in the cluster. Let \underline{d}_i denote the minimum average distance between data i and other samples belonging to different clusters. The SI coefficient is defined as formula(4):

$$I_{SI}^i = \frac{\underline{d}_i - \bar{d}_i}{\max(\underline{d}_i, \bar{d}_i)} \quad (4)$$

The SI coefficient of simple i varies within $[-1, 1]$. If \bar{d}_i is smaller and \underline{d}_i is larger, the SI will be closer to 1, which means that the clustering result of sample i is better. The Silhouette Coefficient I_{SI}^i of all samples can be used to evaluate the clustering quality[10].

The other commonly used index which can evaluate the quality of clustering is Davies-Bouldin index. It is defined to be the maximum of the average distance within any two clusters divided by the distance between the two centroids of the two clusters, then adding them together and averaging them[10]:

$$DB = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \left(\frac{\bar{C}_i + \bar{C}_j}{\|c_i - c_j\|_2} \right) \quad (5)$$

In the formula:

\bar{C}_i ---- Average distance within cluster i

C_i ---- Centroid of cluster i

K ---- Number of clusters

However, due to the flaws in measuring the similarity using Euclidean distance, the clustering evaluation results are not ideal for the circular distribution of the data.

3 Feature Combination based on PCA and Prior Knowledge

Electricity consumption data is measured by the electricity meter which records the load data. Today, the usual measurement acquisition interval is 30min or 1h. In some areas it can even reach 15 minutes, which means the number of features of the data are 96 per day. If we have data on a monthly time scale, the number of features are 2880! There are many methods to reduce dimensions, such as the classic Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA). However, sometimes initial data after the PCA dimensionality reduction don't have a good illustration of the physical meaning of the original data. After the dimension reduction using prior knowledge, the characteristics obtained have a more definite physical meaning. Therefore, in this paper, we use a feature extraction method that combines the data after PCA dimensionality reduction with the data obtained by prior knowledge.

3.1 PCA

Principal Component Analysis (PCA) is a method of data reduction to project high-dimensional data into low-dimensional space. Namely, it maps a k -dimensional feature to n ($n < k$) dimensions and the new n -dimensional feature is a brand new orthogonal feature, called the principal component. Specific practices are as follows:

- Feature centralization, that is, subtracting the average value of a feature from each value of the feature. For example, the average daily electricity consumption of a user minus the user's average monthly electricity consumption.
- Calculating the covariance matrix, that is, after feature centralization, we can get a matrix, and then calculate its covariance matrix.
- Calculating the eigenvalues and eigenvectors of the covariance matrix.
- Getting a new data set, that is, choosing the eigenvectors corresponding to the large eigenvalues, and those eigenvectors form a new data set.

Through the above four steps, we complete the dimensionality reduction using PCA.

3.2 Prior Knowledge

Load characteristics of power system refers to the nature and characteristics of electricity load. It can be reflected by the relevant characteristics of the indicators. Currently, there are no unified standards for load characteristics of power system in the world. China often uses 15 indicators such as maximum or peak load, minimum load, average load, mean square load, peak load duration and load factor, loss duration and loss factor[11] to analyze the characteristics of various types of electricity load. Commonly used indicators are presented in Table 1[12]:

Table 1: The Main Indicators

No.	Indicators	Computation expression
1	Peak Load	$d_{\max} = \max\{d(t); t \in (0, T_f)\}$
2	Average Load	$AL = \frac{1}{T_f} \int_0^{T_f} d(t) dt = \frac{\sum_{j=1}^{N_T} d_j \cdot \Delta t_j}{\sum_{j=1}^{N_T} \Delta t_j}$
3	Mean square load	$MSL = \sqrt{\frac{\int_0^{T_f} d^2(t) dt}{T_f}} = \sqrt{\frac{\sum_{j=1}^{N_T} d_j^2 \cdot \Delta t_j}{\sum_{j=1}^{N_T} \Delta t_j}}$
4	Load factor	$LF = \frac{1}{T_f \cdot d_{\max}^2} \int_0^{T_f} d(t) dt = \frac{\sum_{j=1}^{N_T} d_j \cdot \Delta t_j}{d_{\max} \cdot \sum_{j=1}^{N_T} \Delta t_j}$
5	Coefficient of Variation	$CV = \left(\frac{1}{N_T - 1} \sum_{j=1}^{N_T} (d_j - \bar{d})^2 / \bar{d} \right) * 100\%$
6	Skew	$SK = \frac{\sum_{j=1}^{N_T} (d_j - \bar{d})^3}{(N_T - 1) \sigma_x^3}$

7	Kurt	$KU = \frac{\sum_{j=1}^{N_T} (d_j - \bar{d})^4}{(N_T - 1) \sigma_x^4}$
---	------	--

3.3 The Weighted Combination of PCA Dimensionality Reduction and Prior Knowledge

Considering the method of dimensionality reduction in 3.1 and 3.2, although the result of PCA is better than that of prior knowledge, it has no obvious physical meaning. At the same time, using prior knowledge to reduce dimensionality has a greater loss of the information of the original data than using PCA. Therefore, this section uses a combination of the two dimensionality reduction methods, namely, using the weighted combinations of features obtained by the two methods. Multi-feature weighted combination refers to using a certain method to reflect the overall feature information of an object according to different features of the object. Since the ability of reflecting the overall feature information varies from feature to feature, the weight of every certain feature will directly affect the final result. The traditional K-means algorithm, in which the Euclidean distance is the measure of similarity, defaults that each feature has the same weight, ignoring the difference of the importance of different features. In this paper, we use the initial information provided by the features after dimensionality reduction to determine the weight coefficients.

For sample $d_i = (d_1^{(i)}, d_2^{(i)}, \dots, d_{N_T}^{(i)})$ and sample set $D = \{d_1, d_2, \dots, d_n\}$, the weight coefficients corresponding to N_T features are recorded as $W = \{w_1, w_2, \dots, w_{N_T}\}$. For any two features w_i, w_j , if the feature dispersion of w_i is larger than w_j , then it is considered that the feature w_i occupies a more important position in the clustering. Adjusting the feature space by giving w_i more weight can reflect the similarities within the cluster more precisely and get better clustering results. In the dimensionality reduction of data, the optimal dimensionality reduction reduces the n -dimensional feature to m -dimensions, and the variance of the samples in each dimension is large. This is the mapping process from high-dimensional data to low-dimensional data. Let the projection length of the sample set on the unit vector $u = (u_1, u_2, \dots, u_{N_T})^T$ (representing the direction vector of a straight line) be:

$$\langle d_i, u \rangle = d_i^T u = u^T d_i, \quad d_i \in D \quad (6)$$

Assume that the average value of each dimension after data preprocessing is 0 and remains 0 after projection, thus, the variance is:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i^T u)^2 = \frac{1}{n} \sum_{i=1}^n u^T x_i x_i^T u = u^T \left(\frac{1}{n} \sum_{i=1}^n x_i x_i^T \right) u \quad (7)$$

In the above equation, the last item contains the covariance matrix of the sample. The covariance matrix is expressed as A, then the above equation can be rewritten as:

$$\sigma^2 = u^T A u \quad (8)$$

Because $u^T u = 1$, we can get:

$$A u = \sigma^2 u \quad (9)$$

Therefore, the variance is the eigenvalue of the covariance matrix and u is the corresponding eigenvector. The best projection direction is the corresponding eigenvector when the variance is maximum. Here, the standard deviation is used to calculate the degree of dispersion of each feature of the sample set to determine the value of the feature weight w_r of the N_T -dimensional data[13]:

$$w_r = \sqrt{\frac{1}{n} \sum_{r=1}^{N_T} (d_i^{(r)} - \overline{d^{(r)}})^2} \quad (10)$$

In the formula(10), $\overline{d^{(r)}}$ is the average of the r -th dimension of the features of sample d_i . Accordingly, the Euclidean distance is updated as:

$$d(d_i, d_j) = \sqrt{\sum_{r=1}^{N_T} w_r |d_i^{(r)} - d_j^{(r)}|^2} \quad 1 \leq i \neq j \leq N_T \quad (11)$$

After the weights are determined, the weighted combination of features can be finally determined as:

$$f = \sum_{r=1}^{N_T} w_r d_i^{(r)} \quad (12)$$

The final optimization goal of the centralized K-means algorithm with multi-feature weighted combination is:

$$J = \sum_{i=1}^n \sum_{k=1}^K \sum_{r=1}^{N_T} \mu_k^i \|w_r d_i - c_k\|^2 \quad (13)$$

4 Numerical Results

4.1 Data Cleaning

The data in the real world is dirty. There are many problems with the original data, such as incompleteness, repetitiveness, noise, non-uniformity. The incomplete values should be complemented by inference, which means that the average of the attributes, the average of the same type of sample, or the most likely predicted value can be used. Another method is to ignore the data directly. The method used in this paper is to use the average of the attributes. A missing attribute data of a user in one day should be replaced by the mean of the attribute data for the first 5 days and the last 5 days of the day. The formula used is as follow:

$$d_k = \frac{(d_{k-m} + d_{k-m-1} + d_{k-2} \dots + d_{k-1}) + (d_{k+1} + d_{k+2} + \dots + d_{k+m-1} + d_{k+m})}{2m} \quad (14)$$

$$\begin{cases} k-m \geq 1 \\ k+m \leq \text{days of month} \end{cases}$$

For the days in the beginning or at the end of one month, the number of days calculated can be reduced, that is, till the first or last day of the month. We directly delete the data of users with more than one-third of blank data.

For noise, we can use data binning method to smooth the data according to the average (boundary value, median value), or we can use the regression method to smooth the data by using the fitting function.

Duplicate data can be omitted. It should be noted that there are two different sets of data of the same user at the same time in the data set. There are only a limited number of different but similar values in the two sets of data. These values are replaced by the average.

4.2 Data Normalization

Scaling data proportionally to a specific interval eliminates the differences caused by different numerical attributes. There are many methods such as z-score normalization and maximum / minimum normalization. In this paper, the maximum / minimum normalization method is used to normalize an attribute based on the maximum and minimum value of the attribute:

$$v' = \frac{v - v_{\min}}{v_{\max} - v_{\min}} \quad (15)$$

Where:

v represents the value of one attribute on a certain day in one month.

v_{\max} and v_{\min} respectively represent the maximum and minimum values of the attribute in this month.

v' is the value of the attribute after normalization.

4.3 Feature Combination Based on PCA and Prior Knowledge

The dataset used in this paper is the 96 sampling points per day of 3305 users' load in one month in a certain area. Silhouette Coefficient is used to evaluate the clustering result. Clustering is implemented separately according to the feature extraction method elaborated in the third section of this article, that is, clustering is carried out separately using PCA, load characteristics indicators and the combination of those two above.

In order to choose a k , the number of clusters in K-means, we use the "Elbow Method", which is to draw the cost function in different situations with different k values. With the increase of k , the average distortion degree will decrease, and the number of samples in every cluster will reduce too. Thus, samples are much closer to their cluster centroids. However, there isn't a liner relationship between the decrease of the distortion degree and the increase of the k . With the increase of k , the change of the average distortion degree will continue to decrease. The "Elbow" is the k value corresponding to the position where the improvement of the distortion decreases the most. We use the "Elbow" as the number of clusters. Fig. 1 shows the process using the "Elbow Method" to determine the best k value.

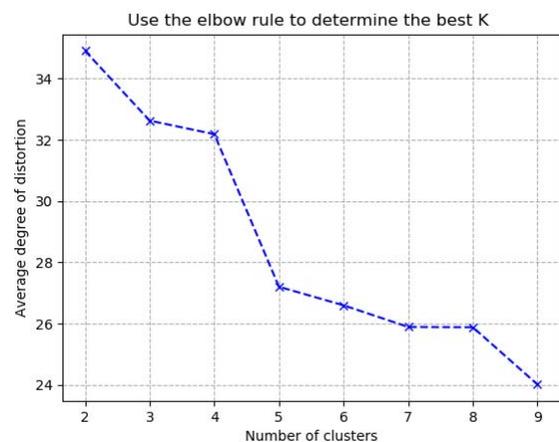


Fig. 1: Use the elbow rule to determine the best K

From Fig. 1 above, the primal best number of clusters can be determined as 4. Because the averaging distortion degree decreases fastest when k value increases from 4 to 5.

a) Clustering after dimensionality reduction using PCA

For PCA dimensionality reduction, the variance contribution rate is selected as 85%. The variance contribution rate of each component is shown in Fig. 2, and it is arranged in descending order.

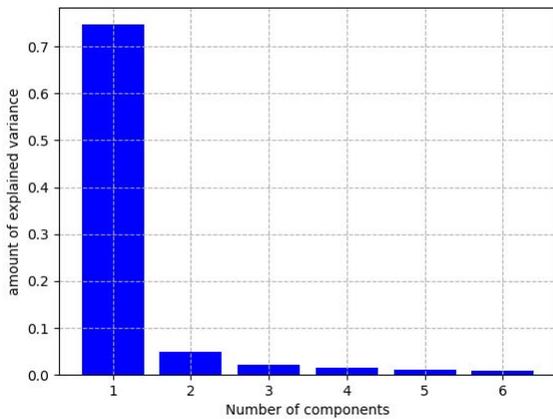


Fig. 2: The variance contribution rate of each component

It can be seen from Fig. 3 that after the PCA dimensionality reduction, the dimension of the original input data is reduced to 6 dimensions when the cumulative contribution rate of variance is 85%. The data are arranged according to the variance contribution rate, and the first two columns are selected for the visual display of the clustering results. The x-axis represents the first column of data and the y-axis represents the second column. The clustering results are as follow (select $K = 2, 3, 4, 5$):

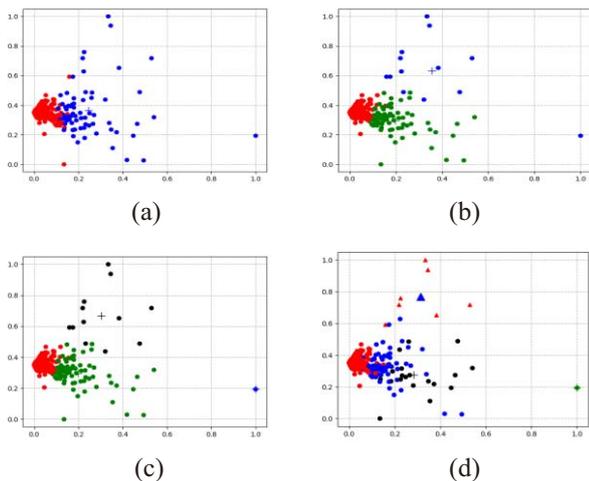


Fig. 3: Clustering results : (a) $K=2$ (b) $K=3$ (c) $K=4$ (d) $K=5$

Intuitively from the clustering results, when the K value is small, the result is good, except for a few points. When the K value exceeds 5, the clustering result becomes poor and the points in the dense area can not be separated. On the other hand, there is no definite physical meaning of the data features after dimensionality reduction by PCA, and the

concrete meaning of horizontal and vertical coordinates can not be explained when the clustering result is displayed. Therefore, the clustering result using only PCA is not very satisfactory. Calculate the Silhouette Coefficient of each point corresponding to K value from 2 to 9, to evaluate the clustering performance. The results are as follow figure:

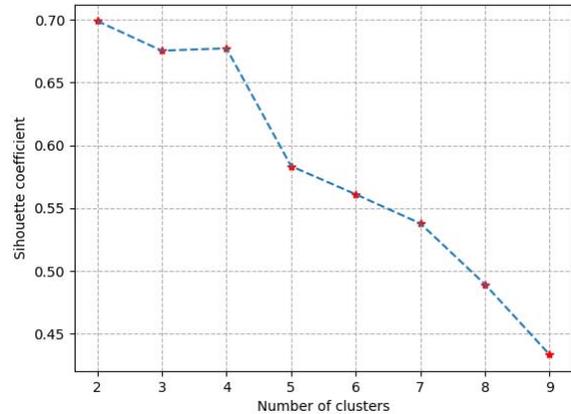


Fig. 4: SI Coefficient

b) Clustering after dimensionality reduction using Prior Knowledge

The load characteristic indicators and static characteristics of time series, a total of six features, are extracted from the given original training data in time domain: average daily electricity consumption, load rate, peak-trough difference, coefficient of variation (CV), skew and kurt. They are new features after the data is reduced in dimension. Clustering after multi-dimensional feature calculation, using SI Coefficient to evaluate the clustering results, as shown in Fig. 5:

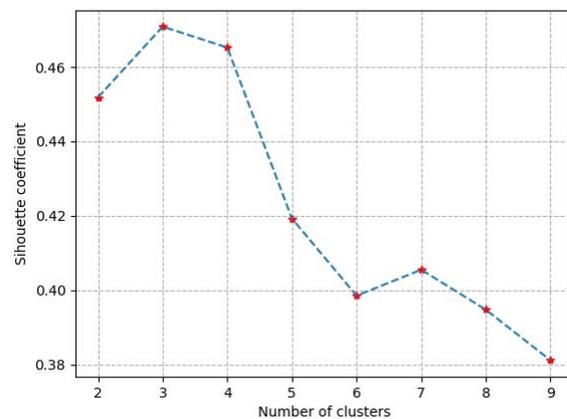


Fig. 5: SI Coefficient

c) Clustering using the weighted combination of PCA dimensionality reduction and Prior Knowledge

Using the method in section 3, that is, weighted combination of PCA dimensionality reduction and prior knowledge, to extract the features and do clustering. The clustering results were evaluated by the Silhouette Coefficient, as shown in the following figure:

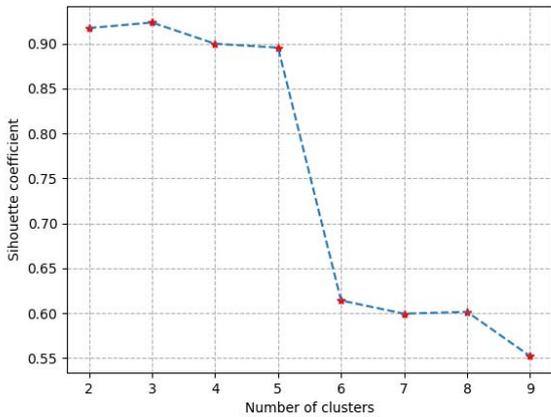


Fig. 6: SI Coefficient

It can be seen from the figure above that the clustering results using weighted combination of features is obviously better than the result of clustering using only PCA and only prior knowledge. Based on the clustering results, taking the cluster centroids as the typical load curves, we can draw four typical load curves:

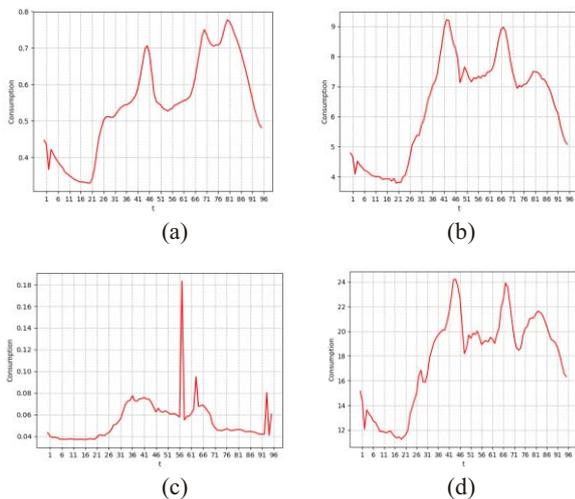


Fig. 7: Typical load curve

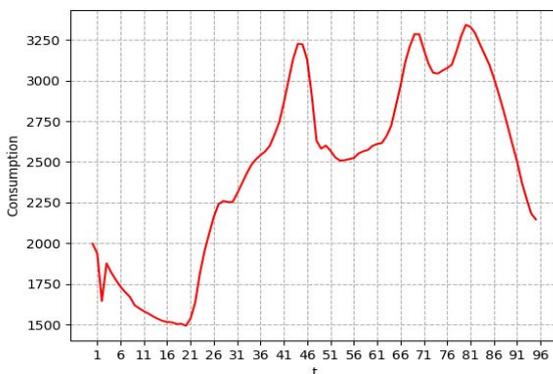


Fig. 8: Total load

In order to facilitate the description of the user's classification, the load centers and the total load of different clusters are compared. Then the types of users are divided into (according to the load curve): peak electricity type, partial peak electricity type I, partial peak electricity type II

and abnormal electricity type. The total load curve of users is shown in Fig. 8. The peak period of peak electricity type (shown in Fig. 7 (a)) basically coincides with the peak period of total load of users. There are three obvious peaks in peak electricity type, and the evening peak is slightly higher than the midday peak. It can be seen that the load consumption of users belonging to this type is low. Ordinary residential users and some small business users belong to this category. Partial peak electricity type I is shown in Fig. 7 (b). This type of users has only two peaks in load curve, and the midday peak is slightly higher than the evening peak. Some non-residential lighting users belong to this category. Partial peak electricity type II has three distinct peaks. These users have high load levels. The third peak appears at around 8 pm. Some large industrial users belong to this category. The last one has low load level and poor regularity of electricity consumption. This phenomenon may be caused by some abnormal data generated during the measurement process. Specific classification results are shown in Table 2.

Table 2: Final classification results

User Type	Load morphological characteristics	Category contains user groups
Peak electricity type	Three obvious peaks. Last two peaks are near. Slightly higher evening peak than midday peak. Low load level.	Cluster 1
Partial peak electricity type I	Two peaks. Middle load level. Slightly higher midday peak.	Cluster 2
Partial peak electricity type II	Three peaks. Obviously low last peak than the first two peaks. High load level.	Cluster 4
Abnormal electricity type	Poor regularity of electricity consumption. Extremely low load level.	Cluster 3

5 Conclusion

In this paper, we considered the different degree of information retained from the original data by different features, when the dimensionality of data was reduced. As a result of that, the proportion of different features in clustering was different. Therefore, the traditional Euclidean distance was updated by taking the variance as the weight, and the model of weighted combination of features, based on PCA and prior knowledge dimensionality reduction, was obtained. The numerical simulation results showed that the proposed method was better than using PCA alone or using prior knowledge alone to reduce dimensionality. Finally, the users in a certain area were divided into four categories according to their load data: peak electricity type, partial peak electricity type I, partial peak electricity type II and abnormal electricity type. And we got their typical load

curves respectively, which provided the basis for the following load forecasting and electricity price making in this area.

References

- [1] *China Power big data development white paper*, Beijing: China Electric Power Press, 2013.
- [2] Xiao Bai, Zhou Chao, Mu Gang, Review and prospect of the spatial load forecasting methods, in *Proceedings of the CSEE*, 2013, 33(25): 78-92.
- [3] Zhang Tiefeng, Gu Mingdi, Overview of Electricity Customer Load Pattern Extraction Technology and Its Application, *Power System Technology*, vol.40, no.3, pp.704-811, 2016.
- [4] Reem Al-Otaibi, Nanlin Jin, Feature Construction and Calibration for Clustering Daily Load Curves from Smart-Meter Data, *Proceedings of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, vol.12, no.2, pp.645-654, 2016.
- [5] Liu Si, Li Linzhi, Wu Hao, Cluster Analysis of Daily Load Curves Using Load Pattern Indexes to Reduce Dimension, *Power System Technology*, vol.40, no.3, pp.797-803, 2016.
- [6] N. Jin, P. Flach, T. Wilcox, R. Sellman, J. Thumim, and A. Knobbe, Subgroup discovery in smart electricity meter data, *IEEE Trans. Ind. Informat.*, vol.10, no.2, pp. 1327–1336, 2014.
- [7] S. Verdu, M. Garcia, C. Senabre, A. Marin, and F. Franco, Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps, *IEEE Transactions on Power Systems*, vol.21, no.4, pp. 1672–1682, Nov. 2006.
- [8] Tang Jing, Li Ruixuan, Huang Yuhang, Research on the accurate forecasting of monthly electricity demand based on the multi-feature analysis, *Power System Protection and Control*, vol.45, no.16, pp.145-150, Aug. 2017.
- [9] Jain A K, Data clustering: 50 years beyond K-means, *Pattern recognition letters*, 31(8): 651-666, 2010.
- [10] Gianfranco Chicco, Overview and performance assessment of the clustering method for electrical load pattern grouping, *Energy*, 42: 68-80, 2012.
- [11] *National Development and Reform Commission*, State Grid, Guide to demand side management of electric power. Beijing: China Electric Power Press, 2007: 237-245
- [12] B.Neagu, G.Georgescu, M.gusa, Load Curves Characteristics of Consumers Supplied From Electricity Repartition and Distribution Public System, *Buletinul Institutului Politehnic din Iasi Tomul LVII(LXI) Fasc 1*, pp. 141-157, 2011.
- [13] Ji Shuyao, Lu Hongfa, Combined Weighted Multi-Feature K-means Clustering for Wireless Sensor Nodes, in *Proceedings of the JOURNAL OF SHANGHAI DIANJI UNIVERSITY*, vol.20, no.4, pp.226-23, 2017.