IOP Conf. Series: Earth and Environmental Science **121** (2018) 052095 doi:10.1088/1755-1315/121/5/052095

An analytics of electricity consumption characteristics based on principal component analysis

Junshu Feng

State Grid Energy Research Institute, Future Science and Technology Park, Beijing 102209, China.

fengjunshu@sgeri.sgcc.com.cn

Abstract. More detailed analysis of the electricity consumption characteristics can make demand side management (DSM) much more targeted. In this paper, an analytics of electricity consumption characteristics based on principal component analysis (PCA) is given, which the PCA method can be used in to extract the main typical characteristics of electricity consumers. Then, electricity consumption characteristics matrix is designed, which can make a comparison of different typical electricity consumption characteristics between different types of consumers, such as industrial consumers, commercial consumers and residents. In our case study, the electricity consumption has been mainly divided into four characteristics: extreme peak using, peak using, peak-shifting using and others. Moreover, it has been found that industrial consumers shift their peak load often, meanwhile commercial and residential consumers have more peak-time consumption. The conclusions can provide decision support of DSM for the government and power providers.

1. Introduction

With the improvement of informatization and intelligentialize of power system, a large number of advanced intelligent equipment have be widely used, which builds a good foundation of the DSM's transition from administrative orders to automatically demand response (DR) to some economic signals. Moreover, in order to attract much more users to participate in DR, figure out their types and how they act respectively should be done to target the best potential users.

Power consumer's characteristics can be used to describe their daily load curves, as presented in [1] and [2]. Different consumers have various load curves, however, there exists similarities between load curves of users from same type. For example, the peak load of residential consumers always concentrate in the morning and evening, and peak load of industrial consumers generally happen in the daytime.

Generally, one regional power load has seasonal characteristics, and daily load curves in one season are similar, such as daily load curves in summer often show as "double peak and double valley". Therefore, DSM needs to know load curves more specifically, just like figuring out the difference between each power consumption types, and accordingly implementing demand control with different means and levels. There are many researches about clustering load curves as shown in [3-7], where fuzzy method, neural network, mufti-objective method are used. However, most of the methods need big samples to extract load characteristics. So, PCA is considered in this article, which has advantages of controllable principle, easy operation, and so on.

In this paper, a comparison of different typical electricity consumption characteristics between different power consumption types (industrial consumers, commercial consumers and residents) is made, and decision support can be provided for the government and power providers to realize the overall coordination and classification control of DSM.

2. Model design

The method of electricity consumption characteristics analysis shown as figure 1 includes two steps, the first is to extract typical characteristics by classification of regional load curves, and the second is to build electricity consumption characteristics matrix.

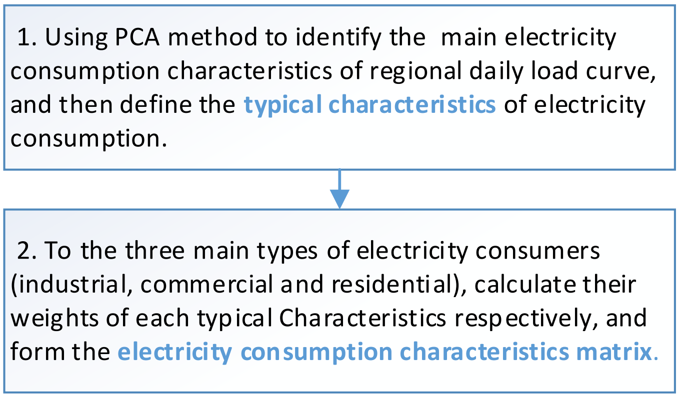

Figure 1. Model analysis steps.

3. Typical electricity consumption characteristics extraction based on PCA method

The role of PCA is to locate dominant ones from numerous variables, which can not only grasp the most important information expressed by original variables, but also reduce the number of variables which can simplify calculation and analysis. The basic principles of PCA method are as follows.

Assuming that data to be analyzed has p variables, as x_1, x_2 ... x_p respectively, which can constitute the p-dimensional random vector

$$X = (x_1, x_2, \ldots, x_p) \tag{1}$$

In formula (1), each observation can correspond to one value of X. Σ is the covariance matrix of X and there must exist an orthogonal matrix U, which

$$U'\Sigma U=\Lambda \tag{2}$$

Where Λ is a diagonal matrix, as diag(λ_1, λ_2, ..., λ_p), $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. Then, to do linear transformation on X, which

$$Y=U'X \tag{3}$$

So we can get a new random variable $Y=(Y_1,Y_2,\ldots,Y_p)$, which variance matrix is $U'\Sigma U=\Lambda$. $Y_1,Y_2,\ldots,Y_p$ of Y are uncorrelated, and the variance of Y_i is λ_i. In this case, we define Y_i as the i_{th} principal component (i = 1,2,...,p) of the random vector X.

The main steps to extract typical characteristics by use of PCA method of regional electricity load are as follows:

3.1. Data processing

The 96 or 24 variables to describe regional load are using 96 or 24 daily power load (collected per 15 minutes or 1 hour), so to the 365 days in one year, the 365 × 96 load matrix X, which represents regional electricity consumption, can be built.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{4}$$

In formula (4), p is the number of points collected per day and n is the number of days. And then standardize the original data x_{ij} to $x^{\cdot}_{ij}$

$$x^{\bullet}_{ij} = (x_{ij} - \overline{x_j}) / \sqrt{\mathrm{var}(x_j)} \quad (i = 1,2,\dots,n; j = 1,2,\dots,p) \tag{5}$$

Where

$$\overline{x_j} = (x_{1j} + x_{2j} + \cdots + x_{nj})/n \tag{6}$$

3.2. *Calculate the correlation coefficient matrix*

The correlation coefficient matrix R can be calculated by

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \tag{7}$$

Where

$$r_{ij} = \sum_{p=1}^{n} (x_{ki} - \overline{x_i})(x_{kj} - \overline{x_j}) / \sqrt{\sum_{p=1}^{n} (x_{ki} - \overline{x_i})^2 (x_{kj} - \overline{x_j})^2} \quad (i = 1,2,\dots,n; \mathrm{j} = 1,2,\dots,p) \tag{8}$$

3.3. *Calculate eigenvalues and eigenvectors*

Calculate the eigenvalues (λ_1, λ_2,..., λ_p) and eigenvectors $a_i = (a_{i1}, a_{i2}, \dots, a_{ip})$ $i=1,2,\dots,p$ for correlation coefficient matrix R.

3.4. *Choose main principal components*

For a region, the collected 96 daily power load points are 96 principal components. However, different information amount is contained in each principal component, therefore it is necessary to use contribution rate, which presents the importance of one principal component and can be calculated as the proportion of variance of one principal component to all variances, also as the ratio of one eigenvalue to sum of all eigenvalues.

$$\text{Contribution rate} = \lambda_i / (\lambda_1 + \lambda_2 + \dots + \lambda_n) \tag{9}$$

The bigger contribution rate is, the more information amount of original variable represented by this principal component has. Cumulative contribution rate is the sum of contribution rates of the first few principal components. Therefore, k principal components can be selected according to the amount of cumulative contribution rate of these principal components. That is, selection principle of the number of principal components k is based on cumulative contribution rate, and generally the value is 70% to 85%. Because the requirements of original data feature are much higher in this paper, we prefer 85%, and examples verification also give support to the high value. So, the following constraints should be met.

$$(\lambda_1 + \lambda_2 + \dots + \lambda_k) / (\lambda_1 + \lambda_2 + \dots + \lambda_n) \geq 85\% \tag{10}$$

To the k that can meet formula (10), it is reasonable to consider these k components can represent more than 85% regional components, and also because most of the other 15% components are interference, and what we want is the normal characteristics instead of some atypical ones.

3.5. *Define typical characteristics of electricity consumption*

Define the first to the k_{th} principal components as the typical characteristics of regional electricity consumption, and the load curves of k typical characteristics are same as eigenvectors $a_i = (a_{i1}, a_{i2}, ..., a_{ip})$ of k principal components, i=1,2,...,k. Then, according to the similarities and differences between typical characteristics and regional load, typical characteristics can be analyzed and described in detail.

4. Analysis of electricity consumption characteristics matrix

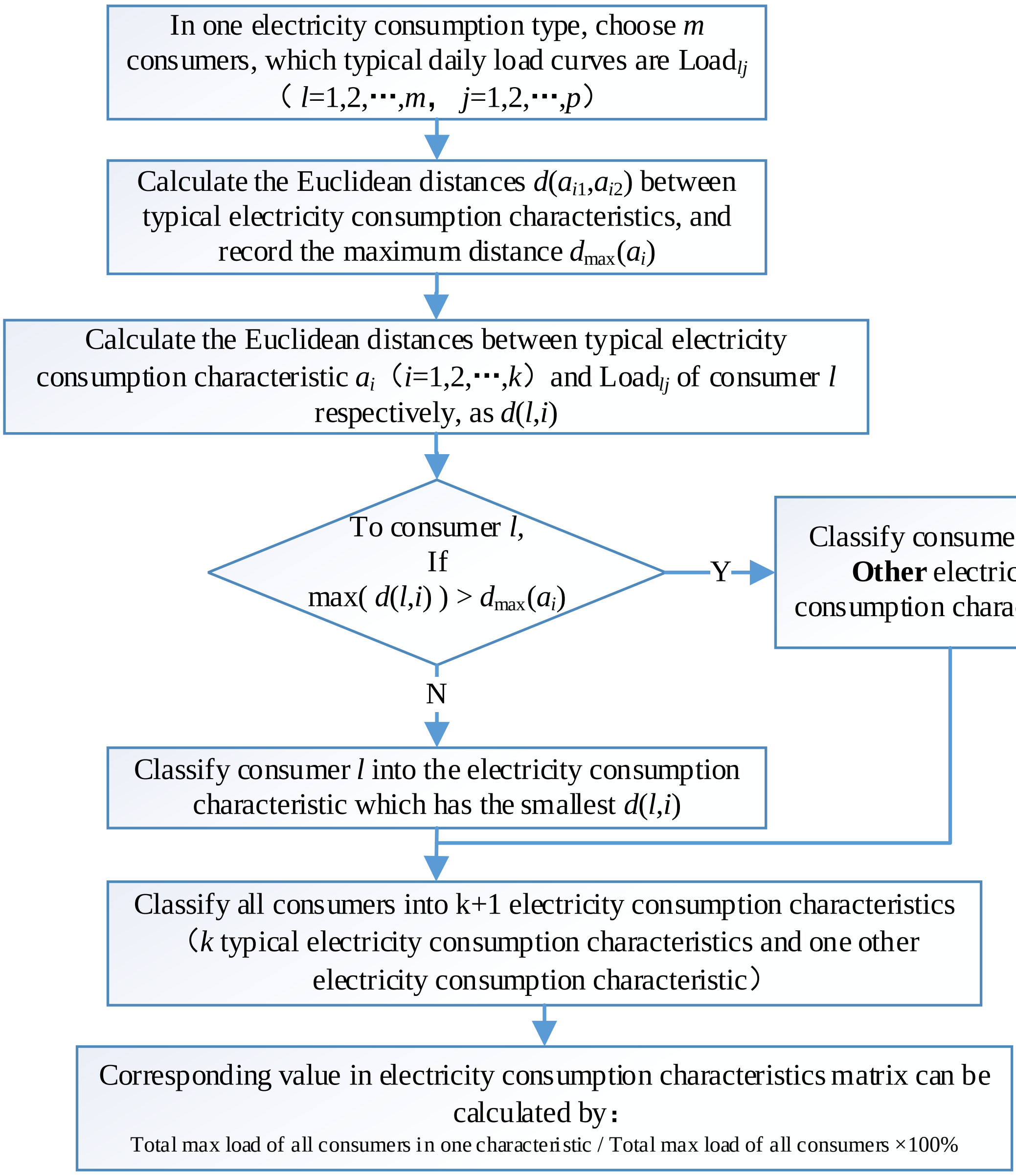

Figure 2. Framework of electricity consumption characteristics matrix.

After extracting regional typical characteristics, characteristics matrix including different types of consumers can be analysed. Rows of characteristics matrix represent different types of power users, including industrial users, commercial users and residential users. Columns represent different typical characteristics, including k typical characteristics and one characteristics. Element in the matrix represents the ratio of one typical characteristic in a certain users' type. Therefore, in the matrix, sum of each row is 100%, and sum of each column can represent and compare the proportions of each typical characteristic in regional electricity consumption.

The specific calculation framework of electricity consumption characteristics matrix is showed in figure 2.

5. Example

Based on the PCA analysis by use of daily load curves in 2014 of a south city in China, three typical characteristics are obtained. And though comparing the eigenvectors of these three characteristics to daily load curves of the city, define the typical characteristics as: extreme peak using, peak using, peak-shifting using and others, which contribution rate is 87.1% and shows in table 1 and figure 3.

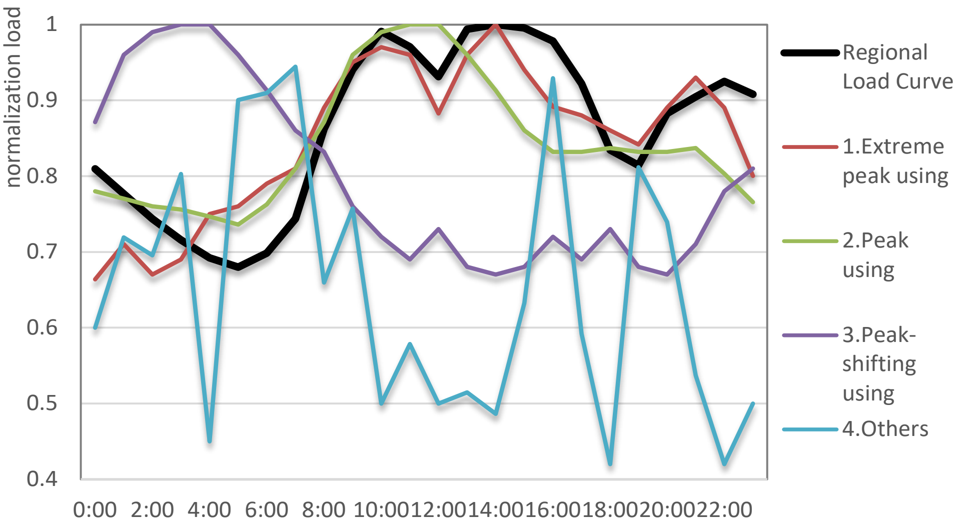

Figure 3. Classification of electricity consumption.

Table 1. Classification of electricity consumption characteristics.

	Classification	Description
1	Extreme peak using	The time interval of user's peak load and regional peak load is within 30 minutes.
2	Peak using	In addition to the extreme peak using, there exists user's one peak load and one regional peak load, which happen nearly, and their time interval is in 60 minutes and less.
3	Peak-shifting using	The number of user's peaks is not bigger than regional peaks, and all time intervals between every users' peak and every regional peak are more than 60 minutes.
4	Others	Except the three types above, this kind of users have no regular rules.

Then, 50 consumers (20 industrial users, 20 commercial users, 10 residents) are selected and their electricity consumption characteristics matrix is analysed. The results are showed in figure 4. Which we can see, the industrial consumers shift their peak load often, meanwhile commercial and residential consumers have more peak-time consumption.

IOP Conf. Series: Earth and Environmental Science **121** (2018) 052095 doi:10.1088/1755-1315/121/5/052095

	extreme peak using	peak using	peak-shifting using	others
industrial users	0.1	0.2	0.4	0.3
commercial users	0.2	0.5	0.1	0.2
residents	0.3	0.5	0.1	0.1

Figure 4. Example of electricity consumption characteristics matrix.

6. Conclusions

By use of PCA method, in our case study, a regional electricity consumption has been mainly divided into four characteristics: extreme peak using, peak using, peak-shifting using and others. Moreover, its electricity consumption characteristics matrix is calculated, and it's found that 50% of industrial consumers would shift their peak load backward regional, and more than 70% of commercial and residential consumers' load peak happen near regional load peak. Therefore, the DSM should focus on commercial and residential consumers much more on account of their big potential, and more tariff mechanisms or incentive mechanisms should be carried out to promote these power consumers to shift their peak-time load.

References

[1] KIRSCHEN D S 2003 Demand-side view of electricity market *IEEE Trans on Power Systems* **18(2)** 520-527
[2] Steven Stoft 2002 Power system economics: designing markets for electricity *IEEE Press & WILEY-INTERSCIENCE*
[3] LIU Youbo, et al 2009 Calculation of characteristic attributes of consumer aggregations based on mufti-objective clustering *Automation of Electric Power Systems* **33(19)** 46-51
[4] LI Xinran, et al 2010 A classifying and synthesizing method of power consumer industry based on the daily load prolile" *Automation oI Electric Power Systems* **34(10)** 56-61
[5] GU Yundong, et al 2015 Multi-model fuzzy synthesis forecasting of electric power *Transaction of China electrotechnical society* **30(25)** 110-115
[6] LI Peiqiang, et al 2005 The characteristics classification and synthesis of power load based on fuzzy clustering *Proceedings of the CSEE* **25(24)** 73-78
[7] ZHANU Hongbin, et al 2003 The characteristics clustering and synthesis of electric dynamic loads based on KOHONEN neural network *Proceedings of the CSEE* **23(5)** 1-5