

Residential Electricity Consumption Behavior Mining Based on System Cluster and Grey Relational Degree

Mengjia Xu, Yuhong Wang

School of Electrical Engineering and Information, Sichuan University, Chengdu, China

Email: marryxmj@163.com

How to cite this paper: Xu, M.J. and Wang, Y.H. (2017) Residential Electricity Consumption Behavior Mining Based on System Cluster and Grey Relational Degree. *Energy and Power Engineering*, 9, 390-400. <https://doi.org/10.4236/epe.2017.94B044>

Received: February 26, 2017

Accepted: March 30, 2017

Published: April 6, 2017

Abstract

In order to improve the utilization of the residential electricity consumption data which contains the information on the user's electricity consumption habits, a residential electricity consumption behaviors mining algorithm model is constructed. Firstly, according to the attribute, the collected data can be divided into the global data and the phase data, then the appropriate global variables are selected to mine the user's electricity consumption patterns in the near future on the system clustering algorithm. Based on the theory of grey relational analysis, combing phase data with the power modes to analyze the potential characteristics of residential electricity consumption behaviors deeply that verify the ability of latest power mode to predict household electricity consumption situation in the coming few days and the effect of dominant phase variables on the peak load shifting. Finally, from the actual data of a certain family, the proposed data mining algorithm is testified that it can effectively explore the electricity consumption behavior habits and characteristics of the family.

Keywords

Data Mining, Electricity Consumption Behavior, System Cluster, Grey Relational Degree

1. Introduction

In recent years, facing the growing problem of energy crisis and environment protection, as well as the increasingly tense relationship with the electric power supply and demand, smart electricity consumption which can improve the quality of electric power supply service and the electricity efficiency by means of intelligent measurement, high speed communication and high efficiency control,

has become the focus of global attention [1] [2]. With the construction of smart grid, advanced measurement devices are more and more installed in the user [3], as of 2013, China has installed a total of about 150 million smart meters [4], so that the users' power data is gathered to explosive growth now. These data imply the users' electricity consumption habits and real needs [5], which is conducive for the power sector to master the user's electricity consumption habits, and to draw up a detailed smart power strategy for the users to improve the passive response mode based on the time-of-use price before [6] [7], is advantageous to the power supply department to formulate economic power generation plan and improve the energy allocation efficiency. However, the data gathered is large and scattered, how to find the key variables in the messy data and extract the user's power consumption characteristics is a difficult point.

At present, the research on the analysis of the user's power using data has become a hot spot in the country that a variety of efficient data mining algorithms combined with the emerging computing model such as cloud computing platform has been studied and applied. References [8] [9] mine the massive power consumption data to get the user's common power mode by the improved k-means clustering algorithm. Reference [10] use cluster average diameter and minimum distance to evaluate the clustering effect of K-means, FCM and other clustering algorithms, so as to determine the optimal number of clusters and realize the analysis and statistics of power consumption data of large power customers ultimately. References [11] [12] based on cloud computing platform, utilize parallel Apriori algorithm and parallel k-means algorithm respectively which have greatly improved the efficiency of data mining to achieve the relevance mining of the users electricity consumption behaviors as well as the load classification. But most of the existing researches are mainly aimed at the analysis of user groups' electricity consumption behaviors, the individual user's is less. Furthermore the future power supply mode will be transformed from the one-way to the two-way power supply service mode dominated by the users, which is more intelligent and targeted, so it is necessary to excavate the power behavior characteristics of individual user.

In this paper, based on the data of household electricity consumption, the definition and division are achieved firstly. Then the system clustering algorithm which overcomes the problem of the selection of the clusters' number and the initial center in the traditional k-means algorithm is applied for the analysis and feature mining of the user's total power consumption in the near future to get the user's frequent electricity consumption mode. By the means of the gray correlation analysis, the household main power consumption behavior is gotten from each power mode, which is used to verify the user's electricity behavior predictability and peak household load shifting by the main power consumption behavior.

2. Analysis Process of Household Electricity Consumption Behavior

For a single family, considering the influencing factors of user's electricity con-

sumption behavior such as real time, climate change and human activity cycle, the family's recent 10 - 15 days of electricity data is selected as the base of analysis. The collected information can be divided into two types: 1) Global data: including energy consumption, power, voltage and current, etc. 2) Phase data: the electricity data of the intelligent electrical appliances or the traditional electrical appliances with intelligent sockets, the electricity data of the home appliances with the similar attribute, and so on. Using the one day's data as a unit sample, the characteristics of the recent household electricity consumption behavior is excavated by the means of the system clustering and gray relational degree. **Figure 1** is a block diagram of household electricity behavior mining process, which is divided into three modules: variable screening and sorting, household power pattern mining, potential power characteristics mining.

1) Variable screening and sorting completes the division and definition of the gathered data. The variables that have great contribution to the mining purpose are chosen as the input parameters of the data mining algorithm, what reduces the calculation process and increases the validity of the data mining result.

2) Household power mode mining selects the appropriate global variables to cluster the household electricity consumption using the system clustering algorithm and digs the family's common electricity consumption modes to extract its characteristic curve.

3) Potential power characteristics mining. On the basis of the above-mentioned electricity consumption modes, combined with the phase parameters, household electricity consumption behavior characteristics are analyzed by the gray relational degree, which provides the basis for smart power strategy.

3. Household Power Modes Mining

3.1. Basic Theory of System Clustering

As an unsupervised learning process, clustering has a good effect on mining the

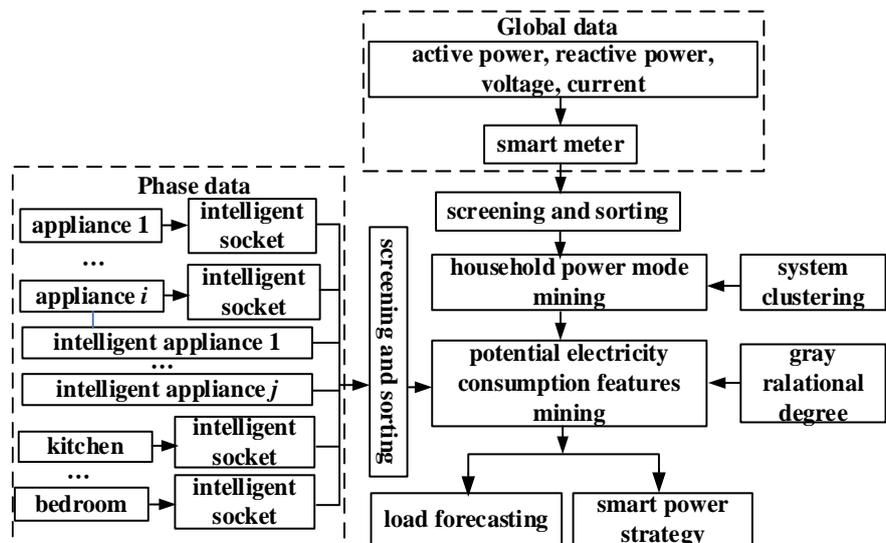


Figure 1. Mining process framework of residential electricity consumption behaviors.

inner relations of unmarked data sets [13]. Compared with the k-means algorithm, the system clustering don't need to set the number of clusters in advance, what is very suitable for processing such complex and unknown household electricity data and is one of the classical algorithms. This algorithm is divided into the cohesion method and the splitting method. Take the cohesion method as an example, each object is regarded as a cluster firstly. Then according to a similarity measure, two closest clusters are merged at every turn until all the clusters are merged into one and the affinity spectrum diagram is formed (specific principles are shown in **Figure 2**).The split rule is just the opposite.

3.2. Household Electricity Modes Mining Steps

Supposing the number of samples of household electricity consumption is s ($10 \leq s \leq 15$), global variables and phase variables are attribute parameters, and the k th sample can be described as: $f_k = [f_{kA}, f_{kB}] = \{A_{k1}, A_{k2}, \dots, A_{km}, B_{k1}, B_{k2}, \dots, B_{kn}\}$ ($k = 1, 2, \dots, s$), where A_{ki} ($i = 1, 2, \dots, m$) is a global variable, B_{kj} ($j = 1, 2, \dots, n$) is a phase variable, and both of which are time series parameters. Assuming that there are qacquisition points in the sample period, $A_{ki} = \{a_{ki}^{(1)}, a_{ki}^{(2)}, \dots, a_{ki}^{(q)}\}$, $B_{kj} = \{b_{kj}^{(1)}, b_{kj}^{(2)}, \dots, b_{kj}^{(q)}\}$. Because of the existence of different dimensions between different variables, standardized processing is carried out by Z-score method:

$$z = \frac{x - \mu}{\tau} \tag{1}$$

where: μ is the sample mean of x , and τ is the sample standard deviation. For the convenience of writing, standardized variable symbols remain unchanged.

Choosing the global variables as the analysis parameters, s samples are clustered together to calculate the average distance between clusters by formula (2) and (3):

$$dist_{avg}(C_i, C_j) = 1/n_i n_j \sum_{f_i \in C_i, f_j \in C_j} d(f_i, f_j) \tag{2}$$

$$d(f_i, f_j) = \sqrt{\sum_{r=1}^m \sum_{l=1}^q (a_{ir}^{(l)} - a_{jr}^{(l)})^2} \tag{3}$$

where: $d(f_i, f_j)$ is the Euclidean distance between samples f_i and f_j and $i \neq j$, ($i, j = 1, 2, \dots, s$), n_i and n_j are the number of objects of cluster C_i and C_j respectively.

Merging the two clusters whose distance is the smallest, then updating the data,

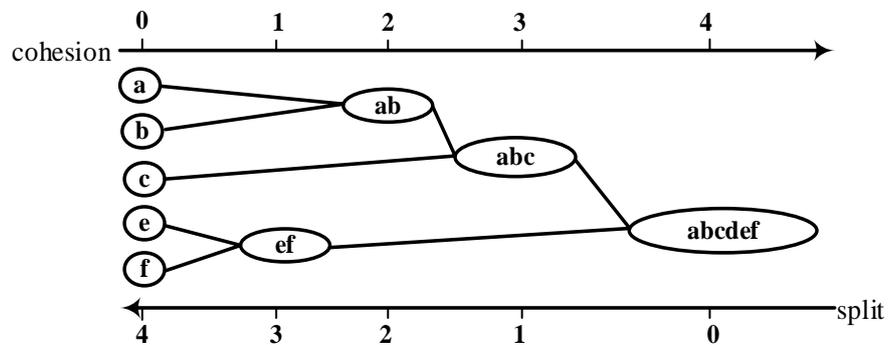


Figure 2. Hierarchical clustering process.

the calculation above is repeated until all the clusters are merged into one. According to the complicated relationship, the situation of the combination of two samples, and the combination of a sample and a cluster is classified as a class. Taking into account that the mean is susceptible to the discrete points and extreme values, the household energy modes M_1, M_2, \dots, M_p are obtained by utilizing the median eigenvalues to represent the general trend of a class of data sets:

$$M_p = \{A_{p1}^*, A_{p2}^*, \dots, A_{pm}^*, B_{p1}^*, B_{p2}^*, \dots, B_{pn}^*\} \tag{4}$$

where: A_{pi}^*, B_{pi}^* ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) are the medians of the corresponding variables in the p th class.

4. Potential Electricity Consumption Features Mining

4.1. Basic Theory of Gray Relational Degree

Gray relational analysis is used to quantitatively analyze and compare the dynamic development process of the system, and to excavate the main factors influencing its change, which is mainly composed of three elements as reference sequence, comparison sequence, and gray relational degree. Assuming that the reference sequence at the t th time is $x_0(t)$, $X_0 = (x_0(1), x_0(2), \dots, x_0(n))$. Comparison sequences are generally more than one, recorded as X_1, X_2, \dots, X_k , where $X_k = (X_k(1), X_k(2), \dots, X_k(n))$. The essence of gray relational analysis is to compare the similarity between the curves of X_1, X_2, \dots, X_k and X_0 with time respectively. The gray relational degree represents the relative order of the similarity of each comparison reference sequence to the reference sequence, and the more similar, the similarity degree is greater.

4.2. Potential Electricity Consumption Features Mining Steps

Combined with phase data, the gray relational grade is applied to analyze the latent influencing factors of user's power consumption modes, and the mode M_p is taken as an example:

1) The absolute difference between the phase and global variables is calculated by the formula (5):

$$\begin{cases} \beta_{pk} = \max_{i=1}^n \left\{ \max_{j=1}^q (b_{pi}^{*(j)} - a_{pk}^{*(j)}) \right\} \\ \alpha_{pk} = \min_{i=1}^n \left\{ \min_{j=1}^q (b_{pi}^{*(j)} - a_{pk}^{*(j)}) \right\} \end{cases} \tag{5}$$

where: $k = 1, 2, \dots, m$.

2) The gray relational coefficient of corresponding elements between the phase variables and the global variable A_{pk}^* is gotten by:

$$\xi_{ij}^{(pk)} = \frac{\alpha_{pk} + 0.5\beta_{pk}}{|b_{pi}^{*(j)} - a_{pk}^{*(j)}| + 0.5\beta_{pk}} \tag{6}$$

where: $\xi_{ij}^{(pk)}$ indicates the correlation between the i th variable B_{pi}^* and the global variable A_{pk}^* of mode M_p at the j th time point.

3) The relational degree of each phase variable to the global variable A_{pk}^* is

computed by:

$$\begin{cases} \lambda^{(pk)} = (\lambda_1^{(pk)}, \lambda_2^{(pk)}, \dots, \lambda_n^{(pk)}) \\ \lambda_i^{(pk)} = \frac{1}{q} \sum_{j=1}^q \xi_{ij}^{(pk)} \quad (i = 1, 2, \dots, n) \end{cases} \quad (7)$$

4) Comprehensive evaluation is applied to analyze and compare the relational degree of each variable in each mode finally.

5. Experiment and Results Analysis

5.1. Data Preparation

This data is from the University of California, Irvine (UCI) database [17], mainly consists of four parts: 1) the total household electricity consumption, including global active power, global reactive power, average voltage, average current. 2) Kitchen power consumption, including a dishwasher, a microwave oven and oven. 3) Laundry power consumption, including a washing machine, a dryer, a refrigerator and a lamp. 4) Living area of electricity, including the water heater and air conditioner. The database contains the electricity consumption of the family from 2006 to 2010, measured once every minute. Because the analysis of household electricity consumption focuses on the information mining process, so although this paper uses foreign data that does not affect the final conclusion, and as China's smart grid construction is maturing, household power split measurement is an inevitable trend. 10 days (2010/8/2 to 2010/8/11) of electricity data were selected randomly as an analysis of samples to build data cube.

5.2. Household Power Modes Mining

Considering the purpose of mining household electricity modes, and that household energy consumption is mainly based on active load, therefore, this paper chooses the daily load data as the analysis parameters, and uses the system clustering method in Section 2.2 to excavate the electricity consumption model to get the agglomeration schedule, as shown in **Table 1**.

Table 1. Agglomeration schedule.

Step	Aggregation cluster		correlation coefficient	first clustering step	
	Cluster 1	Cluster 2		Cluster 1	Cluster 2
1	9	10	131.640	0	0
2	3	4	180.789	0	0
3	5	6	189.296	0	0
4	1	2	207.926	0	0
5	8	9	220.040	0	1
6	5	7	269.404	3	0
7	1	3	289.126	4	2
8	5	8	329.282	6	5
9	1	5	343.351	7	8

The main purpose of clustering is to make the electricity consumption in the same mode as similar as possible, the electricity consumption of different modes as different as possible. Therefore, considering only the situation that the first agglomeration step is 0, agglomeration process of steps 7, 8, 9 are excluded. And the pedigree chart is shown in **Figure 3**.

From the figure above, we can see that, from 2010/8/2 to 8/11, the household electricity consumption can be roughly divided into four modes, and each mode is a clustering of days, which reflects the similarity of household electricity consumption in a short time. The median eigenvalue represents the corresponding modes, as shown in **Figure 4**. It can be seen that there are two peak powers in each household electricity consumption modes, where the peak period is gradually advanced from mode 4 to mode 1.

After analyzing a large number of sample data, it is found that the family has a basic load with amplitude of about 0.4 kW. The load fluctuation is small, which is the minimum basic energy loss of the household, while the user electricity

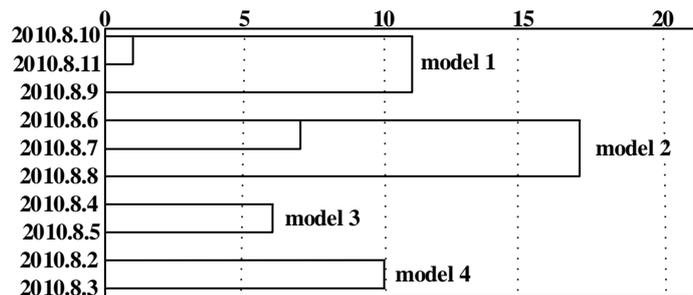


Figure 3. Pedigree chart.

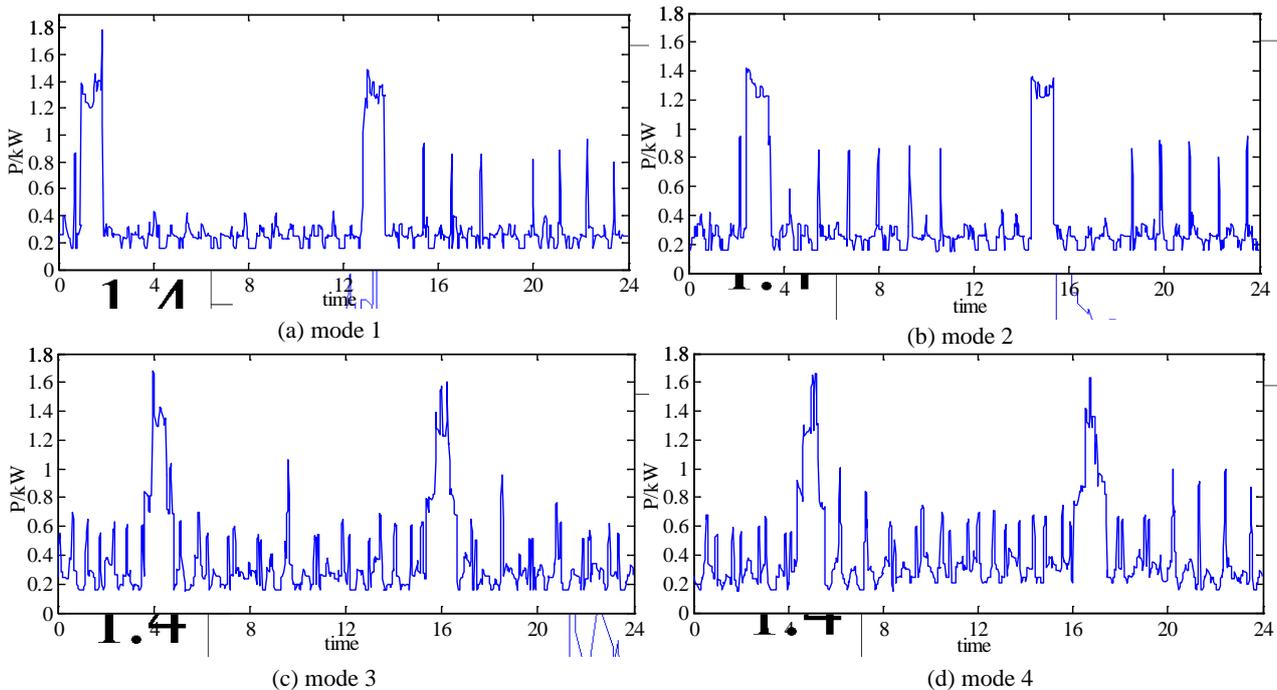


Figure 4. Graph of modes' feature.

consumption load which is mainly related to the use of electrical appliances is fluctuated obviously. In order to distinguish the user's electricity consumption load and household basic load, the load fluctuation threshold is set to be 0.5. Then the basic characteristics of the power consumption modes are obtained as shown in **Table 2**.

5.3. Household Electricity Feature Mining

1) Experiment 1

In order to verify the short-term similarity of the household electricity consumption behavior, the electricity data of 8/12 are adopted to compare the correlation between each power consumption modes and the power consumption in 8/12. Besides, based on the electricity data from 8/12 to 8/16, this paper compares the relationship between the past 5 days' power consumption and the most recent power consumption mode. Detailed data are shown in **Table 3** and **Table 4**.

It can be seen from **Table 3** and **Table 4** that the gray relational degree between each mode with 8/12 household electricity consumption is basically reduced as the time interval larger, that is, the power characteristic of mode 1 is most similar to that of 8/12. In addition, the closer the date to mode 1, the greater the degree of correlation is, and the difference between the recent 3 days of

Table 2. Basic characteristics of power mode.

mode	peak period	characteristics of the power consumption
1	1:00-2:00 12:45-14:00	After the early peak, there is no obvious electricity consumption behavior and is mainly the basic load power consumption. After the latter peak, there is the same kind of electricity behavior or the same kind of electrical appliances is used.
2	2:15-3:30 14:30-15:30	There is a similar power consumption behavior or the use of same electrical appliances after the power peak. There is no obvious power consumption behavior after latter peak.
3	3:30-5:00 15:15-17:00	The electricity consumption is frequent and similar in the whole day.
4	4:15-5:30 16:00-17:30	The electricity consumption is frequent in the whole day. 2 hours after the early peak and 2.5 hours after the latter the peak exist active electricity consumption behaviors.

Table 3. Gray relational degree of household electricity in 8/12.

	Mode 1	Mode 2	Model 3	Model 4
Grey relational degree	0.8776	0.8274	0.8196	0.8223

Table 4. Gray correlation degree of model 1.

Data	8/12	8/13	8/14	8/15	8/16
Grey relational degree	0.8883	0.8850	0.8710	0.8165	0.7978

gray relational degree is not significant. The above shows that in the short term, the household electricity behavior has certain continuity and similarity. Through the large number of fitting experiments on the random number of days' load, it is found that the similar electricity behavior cycle is 2 - 3 days, so utilizing the recent household electricity consumption model to predict this family 2 - 3 days power consumption is feasible.

2) Experiment 2

In this experiment, the data of phase parameter in each mode are taken to analyze the correlation between the phase variables. As the kitchen power consumption is zero, we select the two phase variables as the laundry and living area to analyze, and ultimately get the potential impact of each power model, as shown in **Table 5**.

It can be seen that in the mode 1 to mode 4, the gray relational degree of living area is greater than the laundry's. Compared with the energy consumption of the laundry, the load curve of the living area is more similar to the characteristic curve in the corresponding mode, indicating that the electricity consumption habits of living area is more dominant than the laundry's in the whole household.

From **Table 2**, we can see that the power peak hours is about 1.5 h, so taking the mode 3 for example, the living area's peak time is delayed 1.5 h to get the curve shown in **Figure 5**, which shows that compared with the original load curve, living area electricity of mode 3 is mainly concentrated in the peak hours, after delaying the living area, the household peak load is obviously reduced and the load curve is relatively gentle. In summary, appropriate adjustment on the dominant phase variables is beneficial to improve the overall electricity load

Table 5. Gray relational degree of phase variables.

	Mode 1	Mode 2	Mode 3	Mode 4
Laundry	0.7884	0.7852	0.7798	0.8445
Living area	0.9293	0.9276	0.9262	0.9432
Greater influencing factors	living area	living area	living area	living area

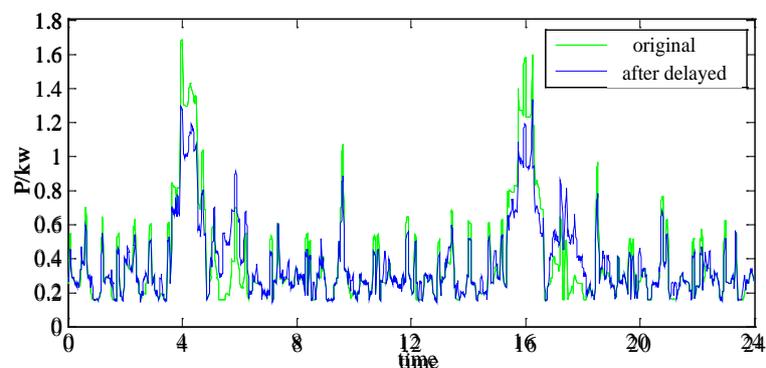


Figure 5. Comparison of the load curve after adjusting the power time of living area.

fluctuation, is helpful to peak shifting and valley filling, and finally obtain some economic significance

6. Conclusion

This paper presents a mining model for household electricity behavior based on system clustering and gray relational analysis. According to the analysis of a group of actual electricity data in a certain family, this model effectively excavates the household electricity consumption pattern of a certain period, as well as the electricity consumption behavior affects order of the corresponding model, and validates the predictive ability of the latest power consumption mode. This work will help to mine the users' potential electricity consumption habits for the power companies to develop the appropriate smart power strategy and improve the quality of power service.

References

- [1] Sun, G.Q., Li, Y.C., Wei, Z.N., Yang Y.B., Zang, H.X. and Bian, D. (2015) Discussion on Interactive Architecture of Smart Power Utilization. *Automation of Electric Power System*, **39**, 68-74.
- [2] Li, Y., Wang, B.B. and Li, F.X. (2015) Outlook and Thinking of Flexible and Interactive Utilization of Intelligent Power. *Automation of Electric Power System*, **39**, 2-9.
- [3] Wang, G.H. (2012) Practice and Prospect of China Intelligent power Utilization. *Electric Power*, **45**, 1-5.
- [4] Lin, H.Y., Zhang, J., Xu, K.P., Pi, X.J. (2012) Design of Interactive Service Platform for Smart Power Consumption. *Power System Technology*, **36**, 255-259.
- [5] He, Y.X., Wang, B.Xiong, W., Zhang, T. and Liu, Y.Y. (2012) Analysis of Residents' Smart Electricity Consumption Behavior Based on Fuzzy Synthetic Evaluation and the Design of Interactive Mechanism. *Power System Technology*, **36**, 247-252.
- [6] Sheng, W.X., Shi, C.K., Sun, J.P., Zhang, B. and Zhang, T.S. (2013) Characteristics and Research Framework of Automated Demand Response in Smart Utilization. *Automation of Electric Power System*, **37**, 1-7.
- [7] Yin, S.G., Zhang, Y., Bai, K.M. (2009) A Smart Power Utilization System Based on Real-Time Electricity Prices. *Power System Technology*, **33**, 11-16.
- [8] Zhang, X., Li, D.H. and Cheng, M. (2015) Study on Peak Load Shifting Management Based on the Big Data Technology. *Modern Electric Power*, **32**, 66-70.
- [9] Zhao, L., Hou, X.Z., Hu, J., Bo, H. and Sun, H.L. (2014) Improved K-Means Algorithm Based Analysis on Massive Data of Intelligent Power Utilization. *Power System Technology*, **38**, 2715-2720.
- [10] Peng, X.G., Lai, J.W. and Chen, Y. (2014) Application of Clustering Analysis in Typical Power Consumption Profile Analysis. *Power System Protection and Control*, **42**, 68-73.
- [11] Guo, X.L. and Yu, Y. (2015) A Residential Smart Power Utilization Strategy Based on Cloud Computing. *Automation of Electric Power System*, **39**, 114-119+133.
- [12] Zhang, S.X., Liu, J.M., Zhao, B.Z. and Cao, J.P. (2013) Cloud Computing-Based Analysis on Residential Electricity Consumption Behavior. *Power System Technology*, **37**, 1542-1546.
- [13] Han, J.W. (2012) Data Mining Concepts and Techniques. China Machine Press, Beijing.

- [14] Zhang,X.L., Hao, S.P., Li, J. and Jiang, C.R. (2015) Grey Correlation Based Analysis on Impacting Factors of Maximum Power Point Tracking Control of Wind Power Generating Unit. *Power System Technology*, **39**, 445-449.
- [15] Zhang, W.Y., Men, D.Y., Liang, J.F. and Wang W.Z. (2012) Monthly Load Forecasting Based on Grey Relational Degree and Least Squares Support Vector Machine. *Power System Technology*, **36**, 228-232.
- [16] Kong L.H., Jiao, Y.J. andDai, Z.H. (2014) A New Substation Area Protection Principle Based on Gray Correlation Degree. *Power System Technology*, **38**, 2274-2279.
- [17] (2014) UCI Machine Learning Repository [EB/OL]. University of California, Irvine (UCI). <http://archive.ics.uci.edu/ml/index.html>.



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact epe@scirp.org