**Segmenting private household energy consumption pattern automatically using QPSO and smart meter data**

Mengmeng Zhuang, Student Member, IEEE, Zuyi Li, Member,IEEE,

***Abstract***— Household electricity energy consumption pattern classficationing is very crucial for utility to build accurate load forecasting model and demand side management. The advent of Advanced Metering Infrasture (AMI) brings a fresh flood into the deployment of smart grid. Smart meters, as first step to smart grid, are accumulating mass customer’s data every day, however, much of data is collected, but not analyzed and potential value from these data is ignored. In order to automatically segment large scale customers’ energy consumption pattern in some power supply areas, a clustering analysis framework based on Quantum Particle swarm optimization (QPSO) algorithm is built. Clustering average radius, clustering average diameter and clustering average minimum distance are proposed to proposed to evaluate the effectiveness of clustering algorithm. The performance of different clustering algorithms was compared. Hourly smart meter data from Illinois state is utilized to analyze the daily electricity consumption curves of the whole customers, which can automatically identify the clustering numbers. Case study result shows the proposed approach is effective in automatically clustering large data set and segmenting energy consumption pattern.

***Index Terms***— Energy consumption pattern forecasting, Cluster analysis, QPSO algorithm

# **Introduction**

Mining and discovering large scale customer’s electricity energy behavior patterns is very crucial for utilities to well grasp electricity customer constitutes and electricity usage characteristics, and provide smart electricity usage service. Applying data mining [1] approach to extract knowledge from electricity customer behaviors pattern has become a hot topic recently. Cluster analysis [2-3] is proven to be a potential algorithm in data mining field. However, conventional data mining algorithms, such as k-means [4], k-medoids [5], self-organized maps (SOM) [6] and Fuzzy C-Means (FCM) [6] cannot prove to be effective in clustering performance and easy to fall into local optimization point [7]. Thus, it is necessary to optimize existing clustering algorithms.

Recently, an increasing number of research papers attempted to apply intelligent optimization algorithms to cluster algorithms [8-9]. However, most algorithms must be defined the cluster numbers in advance, however, in fact, the number of clustering is unknown. Many clustering approaches are difficult to meet the requirements in real applications. The reason can be summarized as: firstly, electricity indicators are adopted to classify electricity customer, the customer groups are unknown; secondly, electricity data are high dimensions, low dimension analyzing approach are not applicable in the high dimension data, no valid approaches are determined the group numbers; thirdly, “noise” is very common in customer electricity data, which will affect the clustering quality. Thus effective measure should be taken.

Particle swarm optimization (PSO) algorithm, developed from the original flocking system, is easily implemented, computationally inexpensive, and does not require gradient information of an objective function, but only its value [10]. In order to obtain faster convergence speed and high convergence accuracy when solving the clustering problems, most importantly, to obtain the best clustering number, an innovative QPSO[11] (Quantum Behaved particle swarm optimization) algorithms are proposed to solve the above clustering problems.

# **Data preparation**

## **Characters of customer electricity usage data**

Electricity index data is from a series of data acquisition system. The characters of these data can be summarized as:

Firstly, the customer data in electricity industry is various, a lot of electricity data form high dimension matrix. Secondly, customer electricity data have two key features: similarity and volatility. Thirdly, customer electricity data is high dimension and timing relevance. Fourth, the noise exists in the customer electricity data and cause electricity data non-smooth and have caused the bad performance of clustering. Finally, the data amount is very large, and customer electricity pattern is very rare in the data statics. From the figure.1, we can find the relationship between energy consumption through ambient temperature from the smart meter have close relationship.

## 

**Fig.1. Electricity consumption data based on sensitive temperature**

## **Data disposal and analyzing**

**Data collecting and processing:** Thus, we invested 125 smart meters with General Electric Model I-210+ through the meter supplied by Commonwealth Edison (ComEd). Besides, we download the ambient temperature from NOOA websiste: <http://www.ncdc.noaa.gov/>.Because the weather condition of the city nearby Chicago is similar with that of Chicago, we consider all the cities have the same temperature.

Before experiments, we conduct data processing. **Firstly**, we conduct correlation analysis considering one day as an example. Data cleaning and disposal are very crucial. For some missing data, we adopt mean value between two value. Figure.2 shows the data characters for 125 smart meters for one day. Gauss filter algorithms was adopted to smooth the data to analyze electricity usage character of one meter.



**Fig.2. The electricity usage characters for one day in Chicago area**

# **Clustering analysis model**

## **QPSO**

Particle Swarm Optimization (PSO), a population-based evolutionary computation technique, is originally proposed by J. Kennedy and R. Eberhart [10] in 1995. In the standard PSO, each individual is treated as a particle in d-dimensional space, with the position and velocity of particle i represented as s a particle in d-dimensional space, with the velocity and position of particle i represented as:

（1）

 (2)

The state of particle update is as follow:  (3)

Where， and are all acceleration coefficients, ，is random number with even distribution between [0，1]；evolution generations；，represents location and velocity repectively；,  represents the personal best position of the particle I and global best position among all the particles in the swarm.

Quantum-behaved Particle Swarm Optimization (QPSO) [11], is developed from PSO. It not only has fewer parameters, but also has higher global search ability than PSO. In QPSO, the particles moves as follows:  (4)

Where ,is the mean value for

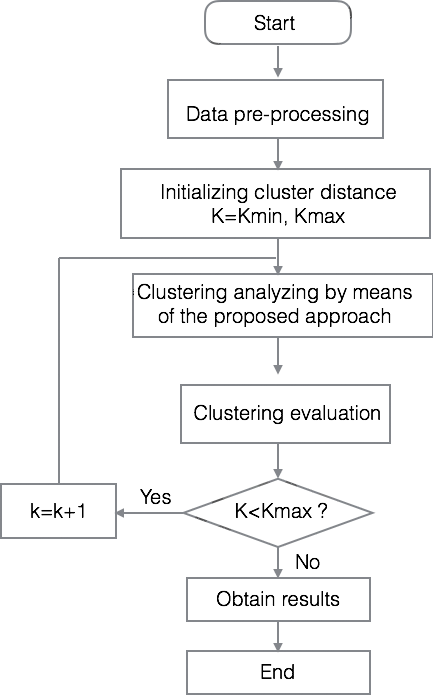
individual best optimization 

So we can get the following：

（5）

Where is slack coefficients，and can control convergence speed of algorithm. During particle iteration, ±is determined by ，when is less than 0.5 时，it can be considered as +，otherwise -.

## A QPSO-based data clustering technique was proposed in [12] with a particle representing the cluster centroid. The results showed that the QPSO could generate good results in clustering data vectors with tolerable time consumption. The flow chart of QPSO algorithms used for clustering analysis is in Figure.3.



**Fig.3. The Flow chart of QPSO model**

## **Clustering performance evaluating**

We define three kinds of evaluation index:

**Definition 1**: Clustering **average radius**

****

**Definition 2**: Clustering **average diameter**

**Definition 3**: Clustering **average minimize distance**

** is clustering number,  is** ** clustering center, ;** **** is **** classical clustering radius,  is ** clustering diameter;** **** is ****

**is the center of  clustering.**

In the experiment, the convergence processes of proposed algorithm can be seen in Figure.4.



**Fig.4. The** **convergence processes of QPSO algorithm**

In order to evaluate our proposed algorithm, we also compare the performance of different clustering approach.Different algorithms are make experiments on the IRIS data set.

The results are in table 1.

**Table.1 IRIS dataset clustering evaluation**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Algorithm items** | **Clusters** | **Experiment times** | **Best accurance** | **Worst accurance** | **Mean accurance** | **Time** |
| **Kmeans** | 3 | 20 | 74.00% | 63.50% | 70.70% | 11.5s |
| **Kernel-Kmeans** | 3 | 20 | 90.80% | 84.90% | 88.60% | 9.6s |
| **Rough Kernel -Kmeans** | 3 | 20 | 92.80% | 87.60% | 98.40% | 7.5s |
| **The proposed algorithm** | 3 | 20 | 100% | 98.50% | 99.20% | 23s |

# **Case study**

The performance of proposed algorithms is applied to one day data from 125 electricity customers in Illinois state. Two case studies are used to evaluate the performance of the proposed model and algorithms. This experiment simulation model are run on a laptop PC : Computer processor：1.4gGHz Intel core i5, Ram：4GB, 1600 MHz DDR3, operation system: Osx 10.9.5. The simulation models and design of experiment are implemented in MATLAB \_R2013a.

**Case 1: Clustering one household data up to almost two years**

In this case, we applied the proposed model to one household electricity usage pattern during almost two years. The energy consumption data covers from Jan.1th 2013to Oct.30th in 2014 in Chicago. We firstly obtain the results before QPSO algorithm optimizing in Figure.5. From Figure.5, we can find that the cluster result is not too obvious. This is because the proposed model is not optimized well.



**Fig.5. The clustering result before QPSO optimization**

The clustering numbers are set from 3 to 8. We try on 20 times and get the best optimization results in Figure.6. From Figure.6, we can find the proposed algorithm have a good performance. The clustering result can be very accurance.



**Fig.6. The clustering result after QPSO optimization**

**Case 2: Clustering one area for un-normal electricity energy usage**

In our experiment, we conduct experiments for 125 same model meters by ignoring the feature difference between different meters ID. When the network is prepared, the testing data is used to obtain 125 smart meters’ data clustering results, which can be seen in the Figure.7. In this case, we only consider to apply these areas into common and uncommon electricity energy pattern.



**Fig. 7. The results of the proposed model for one area**

# **Conclusion**

This paper compared several different clustering approach and proposed an innovative QPSO optimized clustering model to automatically identify and determine the electricity energy usage pattern. The technology roadmap in this paper is based on the processing of data mining[14]. According to the case study, the proposed model and algorithms can easily realize clustering analyzing functions automatically.

The proposed algorithm shows good performance in some smart meter. However, for large scale data set clustering, we need to improve the proposed model and algorithms. It is expected to apply these innovative clustering algorithms into practice.

# **Acknowledgment**

The data in this study is based on the data center of Comed, Inc. The author wishes to take this opportunity to express her sincere appreciation for the support from Comed, Inc.

# **References**

1. HAN J, KAMBER M. Data mining: concepts and

techniques[M]. Morgan Kaufmann Publisher, 2006.

1. A pattern recognition methodology for evaluation of load

profiles and typical days of large electricity customers[J].

Electric Power Systems Research, 2008, 78(9):1494-1510

1. Conejo A J，Morales J M，Baringo L．Real-time demand response model[J]．IEEE Transactions on Smart Grid，2010,1(3)：236-242
2. Anuar N, Zakaria Z. Electricity load profile determination by using fuzzy C-Means and probability neural network[J]. Energy Procedia, 2012, 14:1861-1869.Franklin L. Quilumba,

[5] RÄSÄNEN T, Voukantsis D, Niska H, et al. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data[J]. Applied Energy, 2010,87(11): 3538-3545.

[6] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping[J]. Energy, 2012, 42(1): 68-80.

[7] E. Yasunori, H. Yukihiro, Y. Makito, M. Sadaaki, On semi-supervised fuzzy c-means clustering, IEEE International Conference on Digital Object Identifier, 2009, 1119-1124

[8] Rui Xu ,and Donald Wunsch, "Survey of Clustering Algorithms", IEEE Transactions On Neural Networks, Vol. 16, No. 3, MAY 2005

[9] Petra Kudov́a ,"Clustering Genetic Algorithm," IEEE, DOI 10.1109/DEXA.2007.65, 2007

[10] Kennedy J, Eberhart RC. Particle Swarm Optimization. Proc.IEEE International Conference on Neural Networks, IV [C].Piscataway, NJ:IEEE Service Center,1995.1942-1948.

[11] J. Sun, B. Feng, W. B. Xu, Particle swarm optimization with particles having quantum behavior, Proceedings of 2004 Congress on Evolutionary Computation, 2004, 325-331

[12] Jun Sun, Wenbo Xu, Bin Ye, Quantum-behaved particle swarm optimization clustering algorithm, Lecture Notes in Computer Science, 2006, 340-347

[13]WITTEN Ian H , FRANK Eibe. Data Mining : Practical Machine Learning Tools and Techniques[M] . Elsevier Inc , 2005 :11 - 17.