Electricity Market Customer Segmentation Based on DBSCAN and k-Means

—A Case on Yunnan Electricity Market

Xuejin Wang[1], Chongdong Zhou[1], Yijing Yang[1], Yueyong Yang[2*], Tianyao Ji[2]
Jifei Wang[1], Jingrui Chen[1], Ying Zheng[1]
[1]Kunming Power Exchange, Kunming, China
[2]School of Electric Power Engineering, South China University of Technology, Guangzhou, China
e-mail: Correspondence author: epyueyongyang@mail.scut.edu.cn

Abstract—**Customer segmentation is a proven and efficient method of differential management that has been widely used in various fields. By providing different types of customers with specific services to meet their heterogeneous needs, companies achieve profit growth and other objectives. Determined by the unique market background, especially the settlement mechanism, the trading behavior evaluation in use and the customer segmentation in plan together constitute the differential management system in Yunnan Electricity Market. They help Kunming Power Exchange to mine the information covered in participants' data, so that the exchange can be aware of their situation and better maintain the stable development of the market. A segmentation algorithm based on density-based spatial clustering of applications with noise (DBSCAN) and k-means method is designed to satisfy the requirement in Yunnan Electricity Market. A case analysis using the proposed algorithm is carried out, and the result shows that the segmentation of retailers using their actual operational data is effective and the mechanism can be considered reliable.**

Keywords-electricity market; differential management; customer segmentation meachnism; adaptive clustering algorithm

I. INTRODUCTION

Customer segmentation was first proposed by American scholar Wendell Smith in the mid-1950s and generally refers to a collection of customers based on customer features. The theoretical basis of customer segmentation lies in the heterogeneity of customer needs and the limited resource of enterprises in an effective market competition. In practice, companies categorize customers according to their attributes, behaviors, needs, preferences, and values in a specific market, and provide targeted products, services, and sales models to achieve greater profits.

Customer segmentation has been applied in many areas [1-3]. Banks are one of the earliest industries to segment customers. Through the basic attributes of customers such as age, gender, education, income, and purchasing attributes such as owned products, purchase channels and frequency, banks can obtain full-view information including customer's preferences for products, channels and affordable risk level, quickly generate and discover marketing leads, and provide differentiated products and service strategies for customers.

With the development of information technology, customer segmentation has been rapidly developed in the field of sales, especially in the online retail sector [4]. These companies are able to recommend products of interest to users based on their browsing history. They can even change the price of goods simultaneously, that is, increase the purchase cost of highly viscous users, and therefore squeeze profits with price discrimination.

Although the application of customer segmentation in the electricity market is still in research, its potential determines that it will become an important part of the operation and management of electricity markets. Taking Yunnan as an example, Kunming Power Exchange is responsible for the operation and management of Yunnan Electricity Market. As a non-profit-oriented exchange, the core of its market management is to regulate the order of market participants and promote healthy competition in the market. The evaluation of market participants' trading behavior has been carried out for many years in Yunnan Electricity Market, and the evaluation system has improved gradually, effectively distinguishing the market participants with different performances, thus achieving differentiated management [5]. For example, for those who perform poorly, because of a higher risk of default, their transaction types and trading volume are limited. Through this evaluation mechanism, Kunming Power Exchange manages to guide and promote trading behavior of market participants.

However, the role of the existing evaluation mechanism is limited. Its evaluation index tends to examine the ability and willingness of market participants in contract fulfillment, but it does not reflect the characteristics of market participants such as trading preferences and risk preferences. These characteristics are only a choice of the market participant, which should not be judged by the evaluation mechanism. Thus, it is more suitable to identify these characteristics by customer segmentation. Considering that the feature of clustering in the electricity market is that the amount of data is relatively small and there is little prior information, an adaptive algorithm is designed based on this situation. The algorithm combined the advantage of distance-based clustering and density-based clustering, so it can

978-1-7281-6782-4/20/$31.00 ©2020 IEEE

accurately segment samples while eliminating interference from outliers. The specific content will be introduced in the following sections.

The remainder of the paper is organized as follows. Section 2 describes the ideas and some details of the market management of Yunnan Electricity Market. Section 3 introduces the theory of customer segmentation and explores its potential application in Yunnan Electricity Market. Section 4 presents a case analysis based on the actual operational data. Finally, Section 5 concludes the paper.

II. Market Management of Yunnan Electricity Market

A. Background of Market Management

As stated above, the core of market management in Yunnan Electricity Market is to regulate the order of market participants and promote healthy competition. It should be emphasized that due to the difference in market settlement, the focus of market management in Yunnan is not concentrated on margin management. Margin management is affected by the topology of the clearing network [6-8]. According to the settling mechanism explained in *Implementation Plan of Yunnan Power Marketization Transaction*, issued in 2018, Kunming Power Exchange is not the central counterparty of market transactions.

In most of the electricity markets in developed countries, retailers and producers settle through the exchange settlement department [9-11]. The user needs to pay the electricity fee to the retailer in accordance with their long-term retail contract signed beforehand, and usually the retail contract price is higher than the wholesale price or the bilateral contract price. Denote the retailer's transaction cost, i.e., the price paid to the producer to purchase electricity, by C, and denote the price difference between the fee paid by the user and the cost by revenue R. Therefore, the cash flow of the market is: the retailer collects C and R from the user, and then forwards C to the producer through the market settlement department.

In Yunnan Electricity Market, the cash flow is different. The retailer only needs to charge R from its proxy users, and C is directly transferred by the user to the producer through the settlement department. In Yunnan, China Southern Power Grid Corporation serves as the settlement department. Standing at the perspective of the power exchange, its only cash flow is the transaction service fee collected from the market participants.

From the overall consideration, the smooth cooperation between participants can promote the development of the market. If participants are not able to maintain normal market transactions, leading to risk event like default, it will bring a huge blow to the market.

Therefore, from the perspective of matching rights and obligations, the trading center should maintain market order and ensure the effective execution of market contracts by means other than trading margin. Evaluation of trading behavior and customer segmentation are very likely to be effective and efficient as the main methods of the market management.

B. Evaluation of Trading Behavior

The evaluation mechanism adopted in Yunnan is more like a credit management mechanism: they both evaluate market participants and get the corresponding grade based on the score, and then adopt different management strategies for different grades of market participants. Because it is to evaluate the contract fulfillment of the market entities instead of their credit status, the evaluation is carried out in three aspects: the willingness to perform, the ability to perform and the actual performance, so as to summarize the trading behavior of the market participants in recent period.

It is actually an auxiliary management measure rather than a mandatory regulation that must be obeyed. It indirectly achieves the goal of market management implicitly through the following means.

First of all, the evaluation mechanism and the risk prevention mechanism are organically combined. By limiting the low-grade participants' transaction types and trading volume to isolate the market with insufficient trading ability, poor service capability or trustworthiness participants, and reduce the overall risk of market operation.

Second, after obtaining the evaluation results of their own, market participants can make accurate adjustments. The evaluation mechanism has played a good role in stimulating and guiding.

Finally, the market entities spontaneously accept and apply the evaluation results. Some users began to use the rating result as an important reference for selecting retailers, which increased market transparency and promoted healthy competition in the market.

It should be emphasized that this kind of differentiated management is achieved by judging the merits and demerits of a market participant, and there is also a clear difference between the better and worse management methods adopted, which is in sharp contrast with the customer segmentation mechanism discussed below.

C. Customer Segmentation

With the completion of the trading platform, the market management department is able to grasp more comprehensive customer data. However, these data are not effectively used to improve efficiency. In fact, these data can provide a lot of information, including to describe the characteristics of market participants and to reflect their preferences [12]. However, the current evaluation mechanism can only extract a small part of the information. Therefore, a customer segmentation mechanism is required to make full use of it.

In Yunnan Electricity Market, the goal of customer segmentation may be applying differential management and improving management efficiency, which can be achieved through the extension of evaluation; that is, to expand classified management from the contract fulfillment of market participants to more aspects. Market players will be classified from various aspects, and each category may have its own characteristic labels, according to which different management methods will be adopted. For example, in terms of business structure, market entities take traditional business as their main income, and in terms of risk preference, they are labeled as risk-averse, so they are likely to be managed in the mode of traditional power enterprises.

Different from evaluation, the management department does not need to deliberately guide the market participants to make any change in preference. These results are mainly processed by the management department and need not to be disclosed to the whole market.

III. POTENTIAL APPLICATION OF CUSTOMER SEGMENTATION

A. Clustering Models for Segmentation

Clustering algorithm is a method to realize customer segmentation, it refers to the process that the whole set of objects is divided into multiple sub-sets, in which the objects in each sub-set are similar to one another but different from the objects in other sub-sets. For example, if we want to understand the risk preference of market participants in the electricity market, we can cluster the data of all participants related to this aspect, and get the segmentation of all the participants. For instance, they can be divided into three categories: conservative, steady and radical.

Clustering has been widely used in power industry [13], especially in defining characteristics of end-users based on their load data. Appropriate processes may be applied to reduce the extreme dimensionality of load time series [14]. Considering the development of smart meter, an improved methodology which uses an encoding system [15] as well as a two-stage clustering algorithm [16] are studied independently to deal with more available data. However, there are few applications of subject classification in the electricity market field. An adaptive method is investigated, basing on two commonly used clustering algorithms, distance-based clustering and density-based clustering. The details are introduced below.

1) Distance-based clustering: The algorithm aims to choose centroids that minimise the distance within each cluster. The distance of a specific class can be expressed as:

$$d = \sum_{i=0}^{n} \min_{\mu_j \in C} \left(\left\| x_i - \mu_j \right\|^2 \right) \qquad (1)$$

where x_i represents point i in class j and μ_j represents the centroid of the points in class j. The first step of the algorithm is to identify how many classes the scattered points eventually cluster into, denoted by parameter k_{sample}. Then, select a few points as the initial center points, and iteratively relocate the points until the distance within and between the final classes meet the requirements. According to this principle, algorithms such as k-means are proposed [17]. The classification results of a randomly generated two-dimensional dataset are shown in Fig. 1.

2) Density-based clustering: Distance-based algorithm can not solve irregular shape clustering. Therefore, the density-based algorithm is developed to solve this problem. The density-based clustering algorithms generally assume that points of the same class are closely connected with each other. In other words, there must be points of the same class not far away from any point of the class. A cluster classification can be obtained by grouping closely related points into one class. By setting two parameters, the maximum distance between two samples, denoted by eps , and the smallest number of samples to form a category, denoted by min_{sample} , we can define the density of adjacent areas (the number of objects or data points), and thus the algorithm can return a final classification results. For the same set of data, the classification result using a density-based clustering algorithm is shown in Fig. 2.

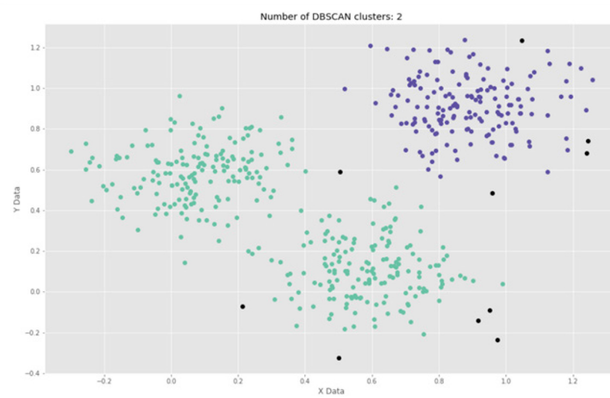

Figure 1. Example of k-means clustering.

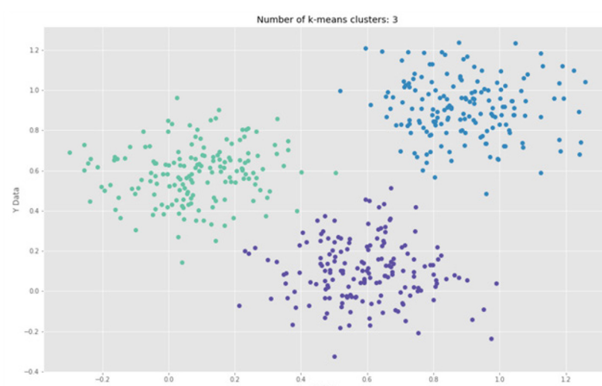

Figure 2. Example of DBSCAN clustering.

In contrast to the k-means method, the DBSCAN method can identify outlier points (i.e. black points). However, for the two classes of data in the lower left corner, as they are closely connected by boundary points, they can only be regarded as the same classification.

B. Adaptiv Algorithm Based on DBSCAN and k-Means

In Yunnan Electricity Market, the trading volume of producers and users is based on the actual electricity generated and maintains at a relatively stable level. However, the trading volume of retailers depends on its trading strategy and the users it currently agents, which is the object that needs to be focused on. This paper takes the retailers as an example to analyze the actual application process of customer segmentation in the electricity market. It is expected that through customer segmentation, the retailers' feature in each aspect are analyzed, so that they can be labeled and a comprehensive image of each retailer can be formed.

The design of the customer segmentation algorithm is based on the following premise.

- The number of retailer categories in the customer segmentation cannot be determined in advance, and only needs to satisfy the given range. If there are too few, the classification is meaningless; if too many, it will cause management confusion.
- Retailers within each category should have similar data and should be clustered based on distance. The result is better to be convex in shape.
- The characteristics of the retailers may not be known beforehand but are extracted after segmentation by summarizing the retailers' commonality in the same category.
- Retailers may not get their own labels in every aspect. Sometimes, they are just outliers that do not belong to any category.

In general, the expected algorithm should have two functions: identifying outliers and automatically selecting the optimal clustering number. The optimal number will be determined according to three indexes which evaluate the clustering effect, namely Silhouette coefficient, Calinski Harabasz score and Davies Bouldin score.

1) Silhouette coefficient: The index takes into account the average distance between samples in the same category and the average distance between samples in different categories. The closer the final value is to 1, the better the clustering effect will be.

2) Calinski Harabasz score: The index takes into account the covariance of samples in the same category and the covariance of samples in different categories. The larger the final value, the better the clustering effect.

3) Davies Bouldin score: The indicator takes into account the distance of the data and measures the maximum similarity of each categories. The smaller the final value, the better the clustering effect.

The algorithm designed in this paper integrates the advantages of DBSCAN and k-means. The DBSCAN algorithm is first applied to exclude the outliers, and then k-means is used to classify the remaining valid data. The detailed algorithm is shown in Algorithm 1.

Algorithm 1 Adaptive algorithm based on DBSCAN and k-means

Require: Standardized data of all retailers in a specific aspect x_i, threshold of two parameters in DBSCAN ($mul_{\text{eps}}, n_{\text{sample}}$), min and max number of parameters k_{sample} in k-means ($k_{\min}$, $k_{\max}$).

Set df = standardized data.

Calculate the average distance ave between all the data.

Run DBSCAN for df and with parameters ($eps = ave * mul_{\text{eps}}$, $min_{\text{sample}} = n_{\text{sample}}$). The outliers are identified and marked with label -1.

Set dfk = data whose label is not -1.

for i **in range** ($k_{\min}$, $k_{\max}$)**:**

> Run k-means for dfk and with parameters (k_{sample} =i). The clustering result are recorded in $labels[i]$. The data centers of each category are recorded in $centers[i]$.
>
> Calculate the three evaluation indexes $SC[i]$, $CHS[i]$ and $DBS[i]$.

Set k as the mode of the index values of each best result in the three sets: SC, CHS and DBS.

Return the clustering result $labels[\mathrm{k}]$ and summarize the attributes of each category according to $centers[k]$.

IV. Case Analyis

A. Selected Aspects of Segmentation

In view of the limitations of the article length, this paper selects the following aspect that is worth of attention in Yunnan Electricity Market.

1) Business Model: The second one is about the business model of the retailers. Currently, three indicators, number of proxy users, volume ratio and emerging business ratio, are used for analysis. They are calculated using the following formula.

- *Number of Proxy Users*: The total number of the proxy users of a retailer.
- *Volume Ratio*: The ratio of the actual electricity consumption of a retailer's users to the total electricity consumption of the whole market.
- *Emerging Business Ratio*: The proportion of a retailer's emerging business in its overall business, that is, the business other than being the purchasing agent in the electricity market.

B. Available Data

The data of 132 retailers from Yunnan Electricity Market are used in this analysis. Data for number of proxy users and emerging business ratio are taken from their basic situation data of 2019. Among them, 52 retailors didn't participate in trading activities, which means their volume ratios are 0. Considering that no transaction is also a basic attribute, they are included in the segmentation of business model. Therefore, there are still 132 samples of business model classification.

C. Empirical Results

The parameters of the algorithm are set to mul_{eps} =1/3, n_{sample} =4, k_{min} =3, k_{max} =8, i.e., the optimal clustering number ranges from three to eight, and a segmentation contains at least four samples.

The clustering results of retailer segmentation in Yunnan Electricity Market are shown as follows.

1) Business model: This issue is about the business model of the retailers. Table I reports the result of evaluation indexes, suggesting the clutering number should be 7, 5 and 5 respectively. As shown in Fig. 3, among the 132 retailers, except that five of them are considered to be outliers, the reamaining 127 retailers are segmented. Table II reports the data center and the number of retailers in each categoriy.

TABLE I. RESULT OF EVALUATION INDEXES

Category	Result		
	Silhouette coefficient	***Calinski Harabasz score***	***Davies Bouldin score***
3	0.5993	117.23	1.1252
4	0.6499	155.78	0.9311
5	0.6742	194.11	0.7845
6	0.6799	186.17	0.8255
7	0.6877	179.19	0.8214
8	0.6537	174.29	0.8118

TABLE II. CLUSTERING RESULT OF BUSINESS MODEL

Category	Data Center			Number of retailers
	Number of proxy users	***Volume ratio***	***Emerging business ratio***	
0	0.0158	0.1312	0.8907	4
1	0.1319	0.3409	0.0697	24
2	0.0121	0.0201	0.0253	72
3	0.0797	0.9775	0.0901	23
4	0.7988	0.2903	0	4
Outlier				5

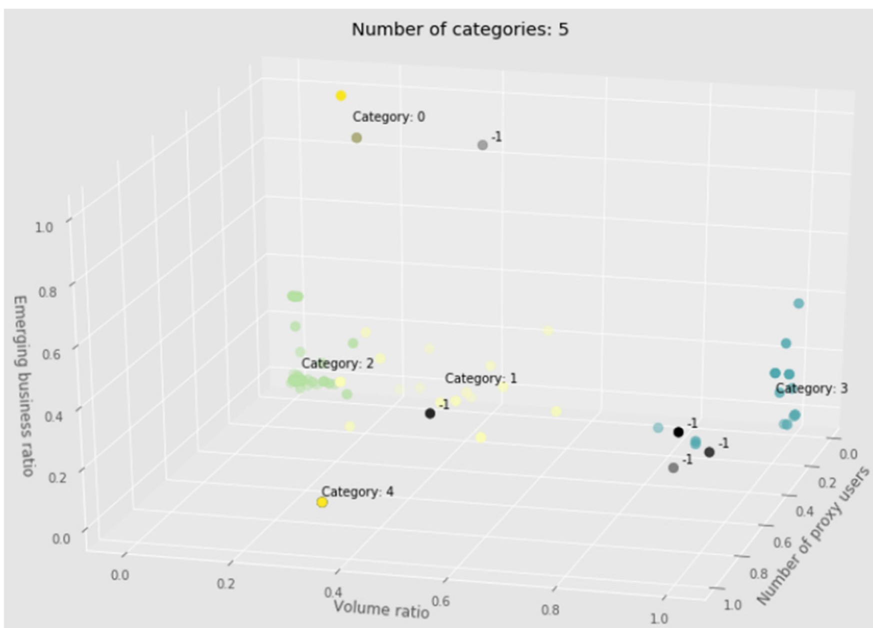

Figure 3. Clustering result of business model of retailers in Yunnan Electricity Market.

The attributes of each category can be summarized according to the data center.

- *Category 0*: The retailers in this category are the only minority who have a high ratio in emerging business.
- *Category 1*: The retailers in this category have a small number of proxy users and a medium trading volume with little emerging business. They may be independent retailers whose business relies on industrial and commercial users.
- *Category 2:* The retailers in this category have the least proxy users and the least trading volume. They are micro entities in the market.

- *Category 3*: The retailers in this category have a small number of proxy users but a huge trading volume. They are likely to be the retailer established by large power companies themselves, or old-fashioned entities with quality resources.
- *Category 4:* The retailers in this category have a huge number of proxy users but a medium trading volume also with little emerging business. They may be recently developed who mainly serve small business users.

A management strategy for the retailers of the above five categories can then be obtained. Category 0 is an excellent representative of the business transformation of a retailer and should be promoted to form a demonstration role. For category 1 and category 4, they are the indispensable backbone of the market. For category 2, they are the key to the market's non-discriminatory opening and can provide corresponding preferential and safeguard management when necessary. Finally, for category 3, it is important to supervise them not to mis-use their market power to disrupt the order of the market.

However, there is still some defects in this clustering algorithm. Some of these samples are actually similar to the features of a certain category, but are classified to be outliers because of the distance. Fortunately, the data of these points are likely to be extreme, and hence their conditions can be identified and judged by other means.

V. CONCLUSION

This paper has demonstrated the theory of market management in Yunnan Electricity Market, including the trading behavior evaluation mechanism and the customer segmentation strategy, and a case analysis of customer segmentation based on actual operational is presented.. Through these mechanisms, the market management department can implement differential management to improve effectiveness and efficiency.

The clustering algorithm is the key to the customer segmentation. A potential algorithm adapted to market characteristic is introduced. With the combination of the DBSCAN algorithm and the k-means algorithm, the outliers and the optimal clustering number are found automatically.

Business model of retailers is chosen in case analysis of customer segmentation, and the data used are obtained from the operational data of 132 retailers collected in 2019. Retailers are clustered into five categories in this chosen aspect, and the retailers segmented into the same category show unique characteristics. Therefore, the customer segmentation mechanism can be considered reliable in this case.

ACKNOWLEDGMENT

The research is supported by Kunming Power Exchange under project *Construction of the Credit Evaluation and Management System of Yunnan Electricity Market.*

REFERENCES

[1] A. Albert and M. Maasoumy, “Predictive segmentation of energy consumers”, Applied Energy, Vol 177, Issue 1, pp. 435-448, 2016.

[2] I. Benitez, A. Quijano, J. Diez and I. Delgado, “Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers”, International Journal of Electrical Power & Energy Systems, Vol 55, pp. 437-448, 2014.

[3] H.W. Shin and S.Y. Sohnmeranz, “Segmentation of stock trading customers according to potential value”, Expert Systems with Applications, Vol 27, pp. 27-33, 2004.

[4] M. Muller, B. Pommeranz, J. Weisser and K. Voigt, “Digital, Social Media, and Mobile Marketing in industrial buying: Still in need of customer segmentation? Empirical evidence from Poland and Germany”, Industrial Marketing Management, Vol 73, pp. 70-83, 2018.

[5] Credit evaluation mechanism of power sales companies in Yunnan Power Market, Yunnan Energy Regulatory Office of National Energy Administration, 2017.

[6] D. Duffie, M. Scheicher, and G. Vuillemey, “Central clearing and collateral demand”, Journal of Financial Economics, Vol 116, Issue 2, pp.237-256, 2015.

[7] M. Galbiati and K. Soramäki, “Clearing networks”, Journal of Economic Behavior & Organization, Vol 83, Issue 3, pp.609-626, 2012.

[8] L. Tian, B. Y. Gan, Q. Sun, J. S. Sheng, and Q. Jie, “The Analysis of Credit Management in Guangdong Electricity Market”, in Proc. 2018 International Conference on Power System Technology Conf., pp. 740-746, doi: 10.1109/POWERCON.2018.8601592

[9] PJM Open Access Transmission Tariff: Attachment Q – PJM Credit Policy, PJM, 2014, [Online]. Available: http://www.pjm.com/-/media/documents/ag-reements/pjm-credit-overvie w.ashx?la=en.

[10] NYISO Market Administration and Control Area Service Tariff: Attachment K – Creditworthiness Requirements for Customer. New York Independent System Operator, 2018, [Online]. Available: http://www.nyiso.com/public/markets_operations/documents/tariffvie we r/index.jsp.

[11] X. D. Chen, L. Tian, B. Y. Gan, T. Y. Ji, and L. X. Xu, “Analysis of Credit Limit Quantification mechanism in Electricity Wholesale Market of the United States”, Automation of Electric Power Systems, vol. 42, Issue 19, pp. 98-105, 2018

[12] O. Dzobo, K. Alvehag, C.T. Gaunt and R. Herman, “Multi-dimensional customer segmentation model for power system reliability-worth analysis”, International Journal of Electrical Power & Energy Systems, Vol 62, pp. 532-539, 2014.

[13] J. Lopez, A. Aguado, F. Martin, F. Munoz, A Rodriguez and E. Ruiz, “Hopfield–K-Means clustering algorithm: A proposal for the segmentation of electricity customers”, Electric Power Systems Research, Vol 81, Issue 2, pp. 716-724, 2011.

[14] O. Motlagh, A. Berry, and L. O’Neil, “Clustering of residential electricity customers using load time series”, Applied Energy, Vol 237, Issue 1, pp.11-24, 2019.

[15] J.-S. Kwac, J. Flora, R. Rajagopal, "Household energy consumption segmentation using hourly data", IEEE Trans. Smart Grid, vol. 5, no. 1, pp. 420-430, Jan. 2014

[16] K. Mets, F. Depuydt and C. Develder, "Two-Stage Load Pattern Clustering Using Fast Wavelet Transformation," in IEEE Transactions on Smart Grid, vol. 7, no. 5, pp. 2250-2259, Sept. 2016.

[17] J. MacQueen, "Some methods for classification and analysis of multivariate observations", Proc. 5th Berkeley Symp. Math. Stat. Prob., pp. 281-297, 1967.