Düzce University
Journal of Science & Technology

Research Article

Two-Stage Clustering Approach for the Household Electricity Load Profiles by Fuzzy Logic and Neural Network Techniques

Uğur Buğra ETLİK [a], Yavuz EREN [a,*]

[a] *Department of Control and Automation Engineering, Faculty of Electrical & Electronics Engineering, Yıldız Technical University, Istanbul, TURKEY*
** Corresponding author's e-mail address: yeren@yildiz.edu.tr*
DOI: 10.29130/dubited.1009823

ABSTRACT

In this paper, household electricity load profile (LP) clustering problem is addressed. LP clustering analysis has been utilized as predicted end-user LPs for demand or supply management strategies to maintain the stability of the power systems. The consumption dynamics of the LPs are formed by the combinations of technical and social factors. Hence, discovering the dynamic patterns of the LPs has been a challenging problem. For this problem, we have offered successive applications of Sugeno fuzzy-logic (SFL) and self-organizing map neural network (SOMNN) techniques. Firstly, the data sets of the LPs are clustered by fuzzy logic approach by the reference models which are generated with the common family-types per persons. Then, considering the extra input of the weighted occupancy profiles, SOMNN is performed to improve the clustering result according to the dataset. The proposed strategy has been simulated by MATLAB® and the related results are presented.

Keywords: *Clustering, Sugeno Fuzzy Logic, Self-Organizing Map Neural Network, Household Load Profiles*

Evsel Elektriksel Yük Profilleri için Bulanık Mantık ve Yapay Sinir Ağları Teknikleri ile İki-Kademeli Kümeleme Yaklaşımı

ÖZ

Bu çalışmada, evsel elektriksel yük profili (YP) kümeleme problemi ele alınmıştır. YP kümeleme analizleri ile güç sistemlerinin kararlılığını sağlamada yararlanılan talep veya arz yönetimi stratejilerinin icrasında gerekli olan tahmini son kullanıcı YP tiplerinin elde edilmesi sağlanabilmektedir. YP tüketim dinamikleri hem teknik hem de sosyal unsurların etkileri ile şekillenmektedir. Bu bakımdan, YP dinamik davranışını anlamlandırmak zor bir problemdir. Bu çalışmada, bahsedilen bu problemin çözümü için iki ayrı aşamada sırasıyla Sugeno bulanık-mantık (SBM) ve öz-düzenleyici harita yapay sinir ağları (ÖDHYSA) tekniklerinin uygulandığı çözüm önerilmiştir. İlk olarak, YP veri seti ev halkı sayısı temelli aile tipleri üzerinden modellenen referans yük tipleri dikkate alınarak SBM tekniği ile sınıflandırılmıştır. Daha sonra, evde hâlihazırda bulunan hane halkının zaman bazlı ağırlıklandırılmış şekliyle de belirleyici bir giriş verisi olduğu düşünülerek ÖDHYSA tekniği uygulanıp kümeleme sonuçları iyileştirilmiştir. Önerilen stratejinin benzetim çalışması MATLAB® ortamında gerçekleştirilip ilgili sonuçlar sunulmuştur.

Anahtar kelimeler: *Kümeleme, Sugeno Bulanık Mantık, Öz-düzenleyici Harita Yapay Sinir Ağları, Evsel Yük Profilleri*

Received: 14/10/2021, Revised: 22/10/2021, Accepted: 15/11/2021

I. INTRODUCTION

Clustering can be defined as discovering the natural grouping of a set of patterns, points or objects [1]. Clustering is a widely applicable idea for engineering problems as well as social networks. Also, some problems which need to be addressed by the both of those fundamental fields become the center of attraction for clustering problems. In this manner, the clustering of electricity load profiles (LPs) has been a popular research interest since the common patterns of LPs have been formed by technical limitations and the socio-economic factors. To explore the consumption dynamics for LPs with respect to pre-determined or self-determined templates are roughly defined as end-user LP clustering. There have been addressed two main approaches for the LP clustering studies [2]: White box approach considers physical properties of the buildings, occupant behavior and periodic behavior of the electrical appliances. Besides that, black-box model focuses on prediction methods by using the criterions of demographics, meteorological database, pricing strategies etc. For evaluation the best performance, commonly used methods such as k-mean, k-medoid and self-organization maps have been used through dividing the LPs into the sub-segments to minimize the data contaminations [3].

LP clustering have been utilized for the demand response (DR) programs, time of use tariff design, demand side management etc. Those applications have frequently addressed to improve the operational quality of power network as well as decreasing the investments to the new power sources. Regarding the recent clustering research, estimated or measured LPs have been referred to propose DR applications. DR applications are addressed to maintain the energy balance between supply and demand. The main objectives fulfilled in the DR programs are shifting the peak load demands to the off-peak periods or decreasing the cost of energy generation by flatten the LP segments which include frequent spikes or sole peak demand. By those modification based activities, end-users consumption have been manipulated according to the supply rate. Even the basic actions implemented at the background, the DR strategies have been proposed with incentive tariffs to shape the individual LD through the ideal options in the clustering sets. In this respect, clustering also directly helps to develop new marketing strategies, determining and updating the pricing policy [4].

For more accurate clustering or more efficient incentive based policies, it has been required periodic acquisition the consumption data of the each end-users. The data reveals the temporal variation in a day rather than the static billing data. Besides, installing the measuring device for each end-users is not realistic solution due to bringing extra costs. For fixing this problem and investigation the certain factors to build a comprehensive demand model, some other variables can be included to the LP estimation problem such as demographic details, environmental factors etc. Hence, instead of using large scale deployments of acquisition devices, the research activities have been directed to develop more efficient utilization of clustering techniques.

Considering the recent stream of the literature, the number of research activities on LP clustering analysis have been increased. Two-stage fuzzy C-means clustering techniques combining in searching for structural relationships with the minimum number of variables are discussed and experiments have been conducted with the composite users of domestic, commercial and small-scale industries to determine the typical LPs [5]. A concurrent k-means and spectral clustering method have been proposed for households [6]. By the mentioned method, the specified period divided into subintervals by the date and the diminished load patterns have been clustered by showing the advantageous of the mentioned technique in terms of computational complexity. Following the segmentation approach for short time forecasting, it has been mattered a unified methodology as handling the samples of 245 substations in Belgian National Grid Operator [7]. The related curves were segmented by extracting temperature and seasonal effects using the periodic auto regression model and then, the k-means clustering method was applied. In a similar fashion, the investigation the LPs of Queensland/Australia have been handled in [8]. Based on the wholesale load data, monthly primary load clusters are built, then the accuracy of those clusters have analyzed with clustering dispersion indicator to extract optimum number of clusters. As a novel perspective, optimization oriented future selection technique has been handled in as the customer profiles separated into load labels [9]. Moreover, the stochastic factors have been included to the profiles to investigate the adverse effects of uncertainty. In a similar manner, optimization based approach are used to cluster of the real LPs and five different cluster patterns as domestic, commercial, industry, hotels/restaurants and others are specified to analyze the set of profiles [10]. The type of heating ventilation and air-conditioning LPs for some higher educational buildings in Norway have been dealt [11]. A hierarchical cluster analysis focusing on load variation instead of the magnitude metric have presented with Pearson Correlation Coefficient based dissimilarity measure. In a similar fashion, heating load demand profiles of Danish residences with heat pumps have been presented and two types of clusters are offered by correlating the building characteristics such as home site, building year and also the socio economic variable of the existence of children [12]. Focusing on the capturing the load shape variability better than the classical algorithms, subspace projection method has been offered to discretize the load trends in the set of customers to obtain the appropriate cluster number [13]. As the proliferation of smart grid concept, the companies have took steps in the direction to develop incentive programs depending on their electricity consumption behaviors. To this end, companies internal database have utilized as input for clustering techniques. In this aim, it has addressed spectral clustering supporting the analysis with demographic and cartographic data types on the data over 6000 customers without auto meter appliances in Belgium [14]. For regional LP building, it has not been inconvenience to collect the detailed information from each of the home.

In this paper, we have dealt with household LP clustering problem with Sugeno fuzzy logic (SFL) and self-organizing map neural network (SOMNN) methods. At the first stage, we have addressed SFL technique to bring the corresponding LPs together which have the similarities at the most points through the day. Besides, the variation characteristics of the household LPs are strictly depended to the occupancy. Dealing this issue at the second stage, we have carried out the SOMNN method by using the outputs of SFL clustering, as well as the occupancy profiles (OPs). The reference profile set is defined by the profiles which have close relative similarities each other to reveal the merit of the proposed approach. In the MATLAB®environment, the simulation studies have been performed with the reference set and the adequacy of the approach is revealed.

II. METHODOLOGY

A. DESCRIPTION OF THE REFERENCE AND TEST DATA

Household LPs have been generated by the demand model presented in the reference of [15]. This tool provides random realistic electricity consumption data by optional date, household numbers and also the temporal OPs. In this study, we have specified four family types entitled as Type-1, Type-2, Type-3 and Type-4 corresponding to the household numbers of 2, 3, 4 and 5. Firstly, we have generated 100 different random load profiles per each family types to specify reference load profiles. To obtain the reference model for any family type, we have randomized the occupancy of the each LPs in the family type. Hence, the reference model represents the uncertainty in the household temporal occupancy in each family types. Besides, the resolution of the data is also another important factor to capture the dynamic behavior of the profile. The LPs are discretized by a certain sampling time and the corresponding value during the sampling period is assumed fixed. In this manner, Granell et al. have studied on the proper time intervals to distinguish the variation pattern of electricity LPs and they have specified that the load data must be sampled at most 30 minutes intervals [16]. Referring to this study, the profiles are refitted as taken the average consumption or average occupancy data for each 15 minutes in a day. Then, each 100 LPs in the corresponding family types have been averaged regarding the law of superposition through the timeframe.

The reference LPs for each family types are rendered in Fig. 1 and the OPs of the reference types are presented in Fig. 2. Root mean square (RMS) and average values of the reference set are given in terms of the load demand and occupancies by Table 1&2. Reference data set is utilized to check similarities on the test data. Therefore, we need to produce test data to give input to the clustering algorithm. For this aim, we have also generated main data set which has total of 1000 random LPs uniformly with the each family types. In a similar manner, the profiles are modified as averaging through the consecutive 15 minutes intervals. The LPs and OPs of the main data set are illustrated by Fig. 3&4, respectively.

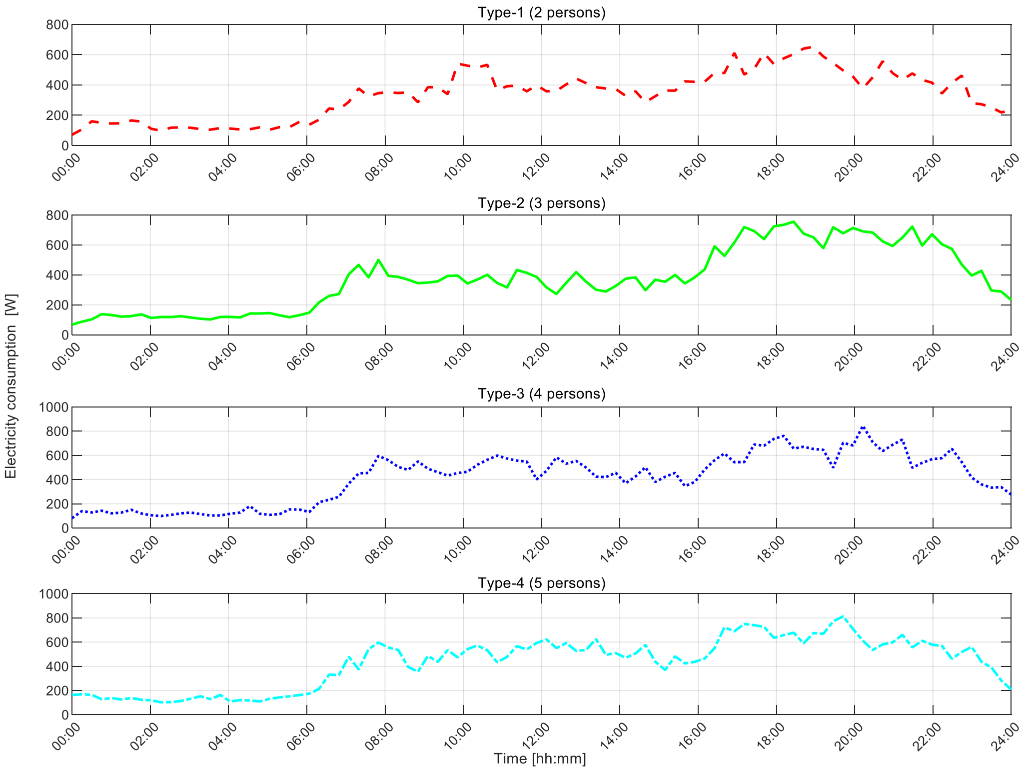


***Figure 1**. The reference load profiles for each family types.*

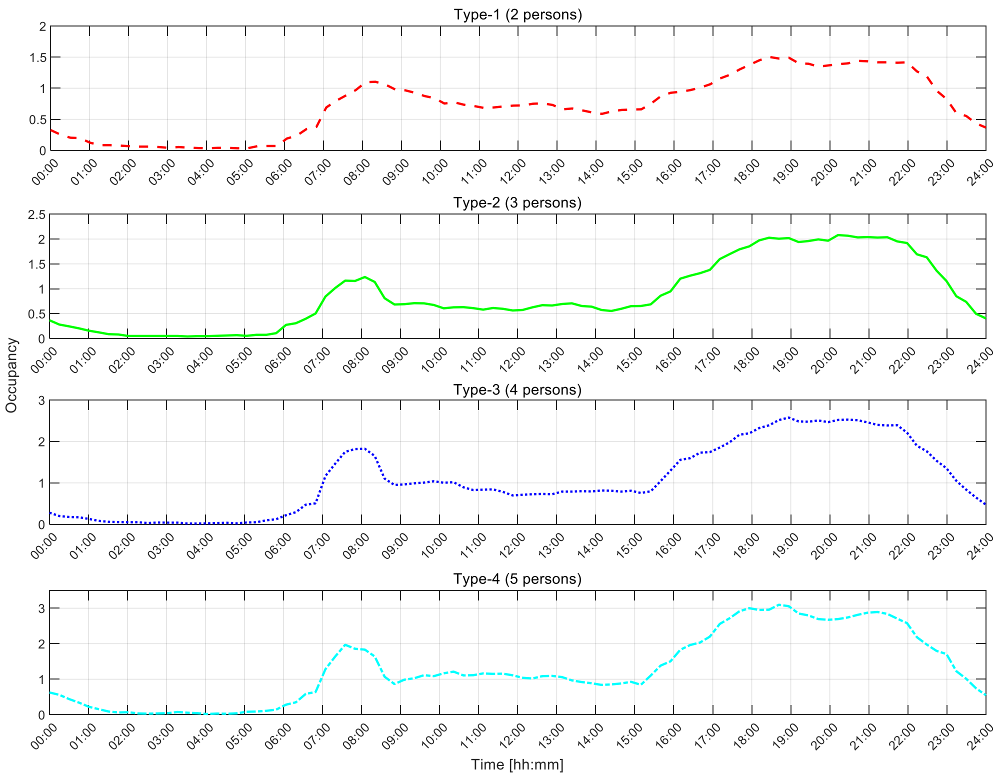


***Figure 2.** The reference occupancy profiles for each family types.*

Table 1. *The RMS values of the reference profiles.*

Family types	RMS value for the load	RMS value for the occupancy
2 persons	370.9369	0.8655
3 persons	425.6892	1.0944
4 persons	463.8757	1.3589
5 persons	478.1320	1.6072

Table 2. *The average values of the reference profiles.*

Family types	Average value for the load	Average value for the occupancy
2 persons	336.2161	0.7242
3 persons	375.2784	0.8627
4 persons	415.4393	1.0699
5 persons	431.7255	1.2801

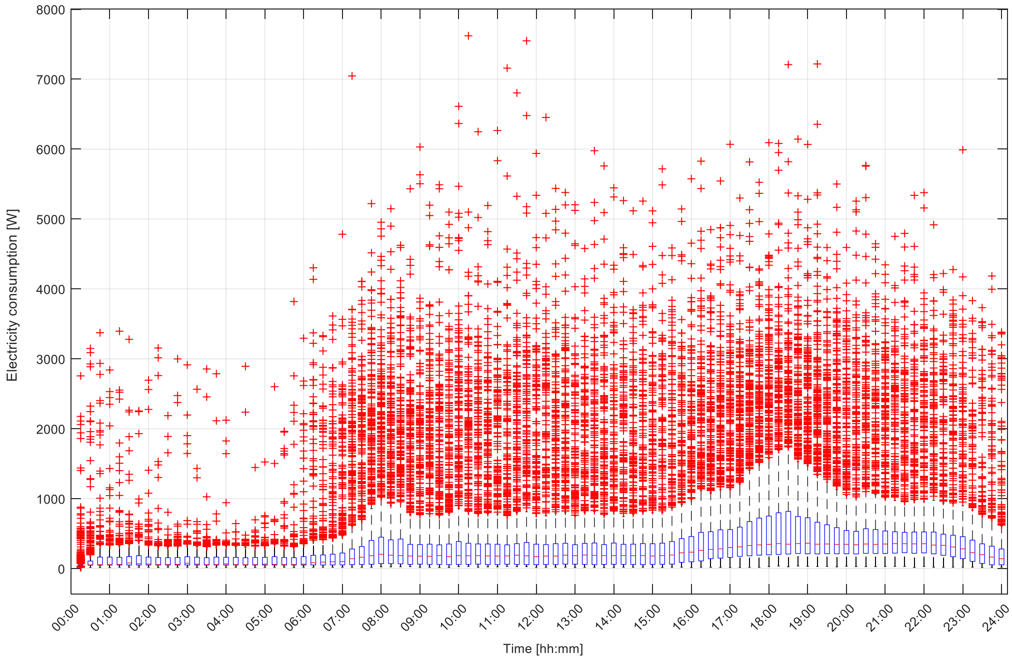


Figure 3*. Load profiles for main dataset.*

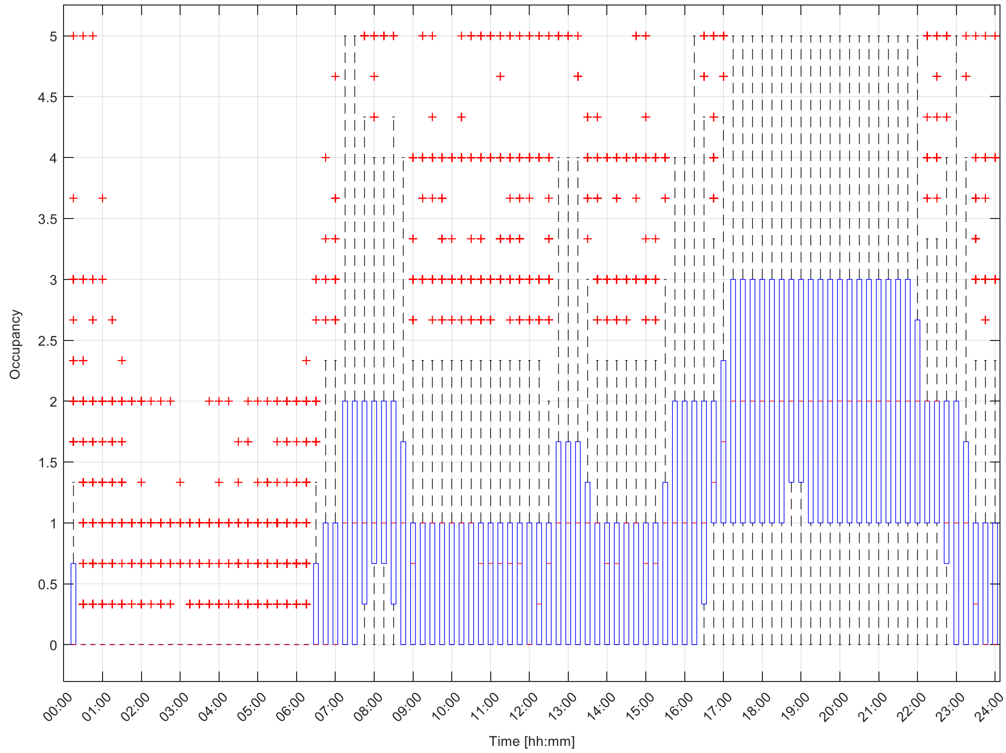


Figure 4. *Occupancy profiles for main dataset.*

B. CLUSTERING APPROACH

Recent research on the classification of LPs have been mainly referenced to the neural networks, data mining, fuzzy-logic decision, and character estimation techniques. From this perspective, we have adopted a cascaded approach with the techniques of SFL and SOMNN to aggregate household load profiles into specific groups according to their similarities. Fuzzy-logic technique represents the degree of uncertainties for the exact information or decisions [17]. Hence, this technique relaxes the restriction to assign a sample to a certain class which it is useful to make intermediate classification [18]. Instead, it helps to search more options to make a more correct correlation by the membership functions. Besides, SOMNN is also proper technique to detect and classify the dynamic behavior of the load profiles. As a distinct feature of SOMNN compared to the other neural network, it includes special reference vector to make connection between input and output layers [19].

Clustering of the 1000 LPs have been realized by cascaded applications of SFL and SOMNN. At the first stage, SFL technique is applied. In this manner, the LPs have been sampled with 15 minutes intervals for a typical day and 96 different periods of the day are discretized with 1000 elements to evaluate the temporal energy consumption data. Hence, all the data are represented with 96x1000 matrix which each row of the matrix corresponded to the different 15 minutes period of the day. For each row of the matrix include simultaneous data of the test profiles, so SFL technique is separately carried to the each time slot of the day. So that, the reference profiles of the family types have been used to compare with each of 1000 random LPs. This process has been realized by the set of 16 rules which is given in Table 3.

Table 3. *Rules of SFL technique.*

Reference Energy Consumption	Household numbers			
	2 Persons	3 Persons	4 Persons	5 Persons
Type-1	0	0.33	0.33	0.66
Type-2	0	0.33	0.66	0.66
Type-3	0.33	0.33	0.66	1
Type-4	0.33	0.66	0.66	1

In this study, we also specialized the clustering analysis paying regard to the occupancy of the households. As we mentioned before, even though the number of households has relatively lower impact on the electricity consumption, the OPs of the households has major effect. This issue is explained as the vast of the appliances or the general energy sources such as lighting, heating, ventilation etc. have simultaneous use by the occupants. Therefore, the major factor on the electricity consumption is the OPs of the households. To include the OPs into the analysis of clustering, we referred self-organizing mapping technique. For this aim, we have weighted to the present output clusters which are obtained by SFL method through the temporal activity of the households. The profiles are weighted with the parameters of 0.05, 0.15 and 0.8 for the time zones of 22:00-06:00, 06:00-14:00 and 14:00-22:00, respectively. Consequently, we have attached priority to the time zone which the occupancy is relatively higher. Then, the profiles have been applied to SOMNN. All the steps of the general clustering procedure are given with Algorithm-1.

Clustering result are presented with Table 4. We have considered 4 family types which has relatively similar numbers of households. Considering the household LPs, the electricity consumption does not have big difference just as adding or subtracting a few persons to the household numbers. If we consider the any family type in the set, it is differed by the nearest types with only one person in the household number. Therefore, the any type in the set has close similarities in the profile with nearest types. Hence this case makes the clustering procedure relatively difficult by those reference forms. Table 5 presents RMS values of each load and occupancy for each reference types. As seen from Table 5, the amount of the total energy consumption close to the each other. Therefore, the types should be paired with the neighboring ones. From this perspective, the proposed clustering approach has achieved 90% success as it is seen from the Table 4. Moreover, if we consider the exact detection of the types, the clustering carried out 50% success. In consideration of the significant resemblance of the profiles, this is acceptable performance.

Clustering results have also been visualized with Fig. 5. This figure present the clusters per each reference types and the proper profiles according to Algorithm-1. From Fig. 5, it can be inferred that the profiles are classified through the proper reference types because the average and the variation patterns are relatively corresponded to the types. Moreover, we have provided Fig. 6 to evaluate the results from general perspective. Fig. 6 projects the comparison of the reference types and the average of the each clusters. From this figure, it can be commented that overall trends of the each reference types and also the average energy consumptions show similarities.

Algorithm 1. *Clustering of the household load profiles.*

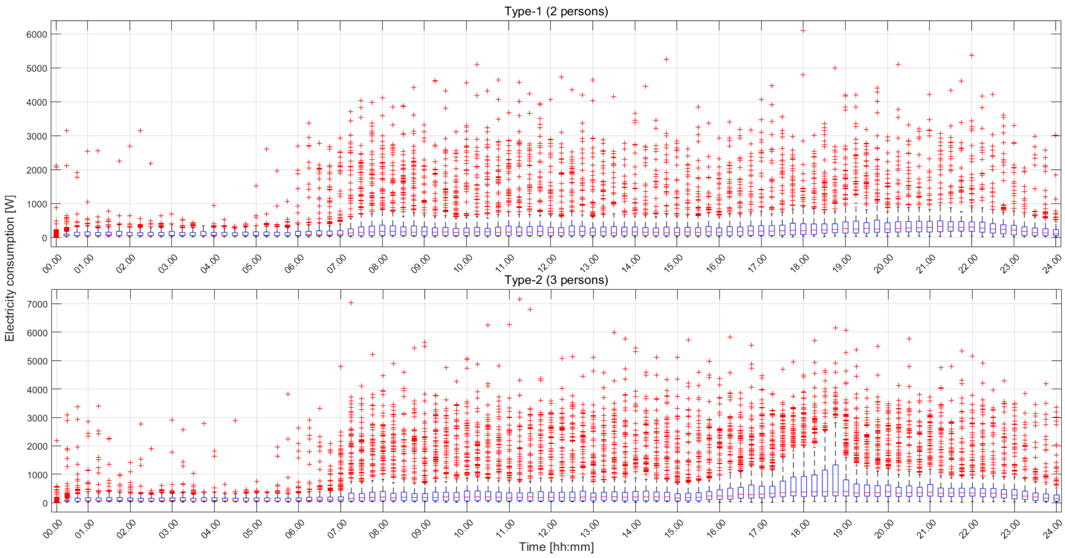
Step 1	:	Get the 100 load and occupancy profiles per each family types.
Step 2	:	Calculate the average energy consumption and the average occupancy for the reference profiles.
Step 3	:	Get the 1000 random load and occupancy test profiles.
Step 4	:	**foreach** 15-minutes time slots in a day
Step 5	:	Get time-based normalized form of the averaged reference profiles.
Step 6	:	Define four membership function per each normalized reference profiles.
Step 7	:	Define 4*4=16 fuzzy-logic rules per each membership functions.
Step 8	:	Get time-based normalized form of the test profiles.
Step 9	:	**foreach** test profiles
Step 10	:	Execute the SFL decision process.
		end foreach
	:	**end foreach**
Step 11	:	**foreach** output profiles of the SFL technique
Step 12	:	Execute the SOMNN decision process.
		end foreach

Table 4. Clustering of the load profiles by SFL&SOMNN techniques

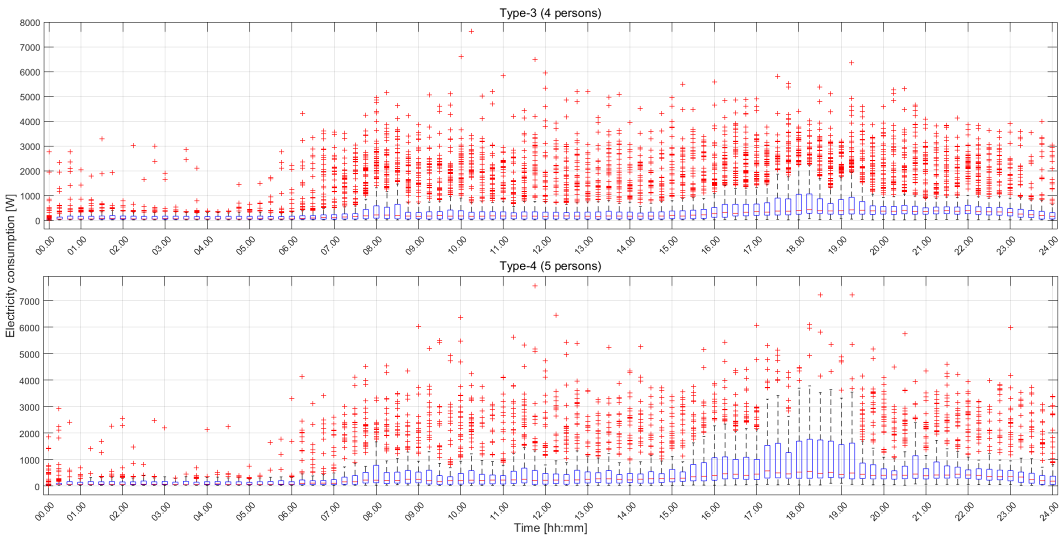
Reference Types	Clustering Results			
	Type-1	Type-2	Type-3	Type-4
Type-1	**165**	81	4	0
Type-2	60	**120**	66	4
Type-3	36	76	**111**	27
Type-4	11	50	90	**99**

Table 5. RMS values of load and occupancy for reference profiles

Reference Types	RMS	
	Energy Consumption	Occupancy
Type-1	370.9369	0.8655
Type-2	425.6892	1.0944
Type-3	463.8757	1.3589
Type-4	478.1320	1.6072



(a)



(b)

***Figure 5.** Clustering of the LPs corresponding to the family types*

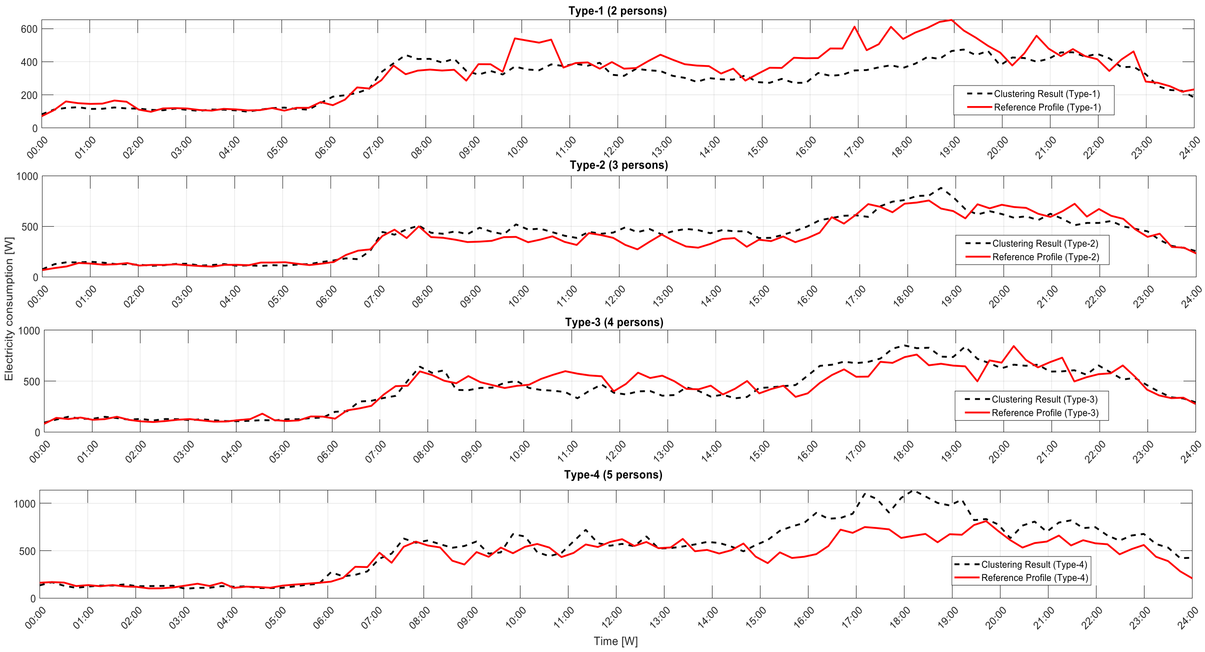


Figure 6. *Comparison of the reference profile types and the average of the each clusters*

III. CONCLUSION

More feasible electricity tariffs for each part of energy market can be obtained by utilizing the clustering techniques for demand or supply side energy management policies. Also, the end-users can be promoted to reform their consumption pattern through the economic tariffs. For this aim, we have provided two-stage approach with SFL and SOMNN techniques. With this approach, we have achieved reasonable clustering success to match the profiles in the dataset to the corresponding reference profiles. In this manner, we have reached the rate of 90% matching paying regard to the narrow-set of reference models. Besides, considering the reference models separately the success rate is about 50%. Taking accounting the fact that, the reference sets cannot be undifferentiated considerably by increasing or decreasing one person in the number of households for each family type. Hence the narrow-cluster types of reference models make the problem challenging as their profile dynamics have close similarities. For the future studies, we shall to consider different types of LPs to cluster. Also, referring to the inferences from the clustering analysis, we shall provide DR applications to maintain the stability of the grid.

IV. REFERENCES

[1] A. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no.8, pp. 651-666, 2010.

[2] Z. Wang and T. Hong, "Generating realistic building electrical load profiles through the Generative Adversarial Network (GAN)," *Energy and Buildings,* vol. 224, no. 110299, pp. 1-15, 2020.

[3] F. McLoughlin, D. Aidan and M. Conlon,"A clustering approach to domestic electricity load profile characterisation using smart metering data" *Applied Energy,* vol. 141, pp.190-199, 2015.

[4] J. Aghaei and M. I. Alizadehand, "Demand response in smart electricity grids equipped with renewable energy sources: A review," *Renewable and Sustainable Energy Reviews,* vol. 18, pp. 64-72, 2013.

[5] Z. Zakaria and K. L. Lo, "Two-stage fuzzy clustering approach for load profiling," presented at the *44th IEEE UPEC*, Glasgow, UK, 2009.

[6] L. Sun, K. Zhou and S. Yang, "An ensemble clustering based framework for household load profiling and driven factors identification," *Sustainable Cities and Society,* vol. 53, no. 101958, pp. 1-11, 2020.

[7] M. Espinoza, C. Joye, R. Belmans, and B. D. Moor, "Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series," *IEEE Transactions on Power Systems,* vol. 20, no. 3, pp. 1622-1630, 2015.

[8] D. Colley, N. Mahmoudi, D. Eghbal, and T.K. Saha, "Queensland load profiling by using clustering techniques," *IEEE AUPEC*, Perth, Australia, 2014.

[9] K. A. Choksi, J. Sonal and N. M. Pindoriya, "Feature based clustering technique for investigation of domestic load profiles and probabilistic variation assessment: Smart meter dataset," *Sustainable Energy, Grids and Networks*, vol. 22, no. 100346, pp. 1-11, 2020.

[10] J. N. Fidalgo, M. A. Matos and L. Ribeiro, "A new clustering algorithm for load profiling based on billing data," *Electric Power Systems Research*, vol. 82, no. 1, pp. 27-33, 2012.

[11] Z. Ma, R. Yan and N. Nord. "A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings," *Energy*, vol. 134, no. 1, pp. 90-102, 2017.

[12] C.M.R. do Carmo and T. H. Christensen. "Cluster analysis of residential heat load profiles and the role of technical and household characteristics," *Energy and Buildings,* vol. 125, no. 1, pp. 171-180, 2016.

[13] M. Piao, H. S. Shon, J. Y. Lee and K. H. Ryu "Subspace projection method based clustering analysis in load profiling," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2628-2635, 2014.

[14] D. Vercamer, B. Steurtewagen, D. V. Poel, and F. Vermeulen,"Predicting consumer load profiles using commercial and open data," *IEEE Transactions on Power Systems,* vol. 3, no. 1.5, pp. 3693-3701, 2015.

[15] I. Richardson, M. Thomson, D. Infield and C. Clifford, “Domestic electricity use: A high-resolution energy demand model,” *Energy and Buildings*, vol. 42, no. 10, pp. 1878-1887, 2010.

[16] R. Granell, C. J. Axon and D. C. H. Wallom, "Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles’,’*IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3217-3224, 2015.

[17] L. A. Zadeh.‘’Fuzzy sets’’,*Information and Control*, vol. 8, pp. 338-353, 1965.

[18] R. Xu and D.C. Wunsch, ’’Clustering,’’ *Wiley-IEEE*, 2009.