Machine Learning-Based Clustering of Load Profiling to Study the Impact of Electric Vehicles on Smart Meter Applications

Saeed Ahmed, Zafar Ali Khan
Mirpur University of Science & Technology, Pakistan
saeed.ahmed@must.edu.pk

Noor Gul
University of Peshawar, Pakistan
noor@uop.edu.pk

Junsu Kim, Su Min Kim
Korea Polytechnic University, Korea
suminkim@kpu.ac.kr

***Abstract*— The data collected from advanced metering infrastructure enables the electric utilities to develop a deep insight about the energy consumption behavior of the consumer. However, the load signature and consumption pattern varies due to addition of multiple types of new loads, such as electric vehicles (EVs). Therefore, it becomes imminent to further dig down these variations. To this end, this paper investigates the impacts of insertion of EV profiles in the household level smart meter data. The Irish CER dataset and EV data from the NREL residential PEV are utilized in this study to classify the users with and without EVs' loads. The results show that change in the cluster membership can help to separate the consumers with the EV load from the stand-alone consumers without the EV load.

Keywords*— *Data clustering; electric vehicles; load profiling; smart meter

I. Introduction

The smart meters deliver meticulous knowledge about the individual consumers' load patterns that can be further utilized to control the loads even at individual household [1, 2]. The challenges faced by the curse of dimensionality of the data can be managed by classifying the consumers into different classes to extract typical load profiles using machine learning-based techniques.

Extraction of load patterns from smart meter data is a cumbersome process and it can be tackled by supervised or unsupervised machine learning (ML) techniques [3]. Multiple studies have been carried out to classify the pattern in the unlabelled smart meter data, however, impact of integration of electric vehicles (EVs) on consumer classification is a promising area.

Information of EV charging may help the electric utilities to predict load and also to comprehend temporal and spatial aspects for: 1) load scheduling and 2) Evade distribution network renovations [4]. The consumers with EVs hide the purchase of EV usually from utility resulting in shift in their energy consumption pattern without the knowledge of utility, leading in wrong categorization of such consumers. Broadly, non-intrusive load monitoring is employed to disaggregate load for EV detection. However, it is an complicated technique that requires high granularity of data at frequency of seconds [5].

In this paper, we investigate the impact of inclusion of EVs at consumer level considering different diffusion levels of EV charging profiles. The smart meter data from Irish CER dataset [6] with 30 minutes resolution is interpolated to 10 minutes resolution to embed EV charging profiles with 10 minutes resolution. The profiles with and without EVs are clustered and changed in a cluster membership due to the inclusion of EVs. Accordingly, the impact of EVs are investigated in this paper.

The rest of the paper is arranged as follows. Section II explains the proposed scheme. The case studies and results with their analysis are presented in Section III. The paper is concluded in Section IV.

II. Methodology

This paper aims to investigate the clustering of the load profiles inclusive of EVs. The Irish CER smart meter dataset employed in this work [6] contains data snapshots with a frequency of 30 minutes for more than 5,000 residential and small business consumers for a period of 18 months. The EV charging data used in this case is from 2009 RECS data set provided by NREL [7]. 200 random customers are selected from the smart meter dataset and similarly, 30 EV charging profiles are selected from the NREL dataset for case studies. A flowchart of the proposed scheme adopted for the case studies is given in Figure 1.

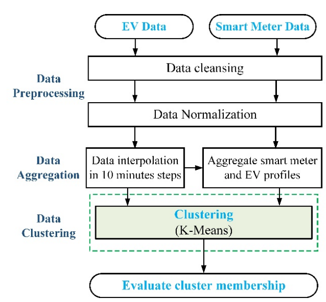


Figure 1: Flowchart Proposed scheme

The proposed scheme is explained as follows:

A. Data Pre-processing

In the data pre-processing stage, the first step is to ensure the quality of the data. To ensure high data quality, the outliers are removed, and data is cleansed by removing the erroneous values. Potential hardware failures in the first month can lead to zero kWh readings, therefore, all such readings are removed

[8]. The EV profiles contained only EV charging and data. For the rest of the time, the values are shown as zero which is considered normal in this case. Once the data is cleansed, 200 random customers are extracted from the smart meter data and the 30 random EV profiles are chosen from the EV data.

In the next step of the data pre-processing, the smart meter data with resolution of 30 minutes (48 daily readings) is converted to 10 minutes resolution. This is achieved by linearly interpolating the data using (1);

$$\frac{y-y_0}{x-x_0} = \frac{y_1-y_0}{x_1-x_0}. \quad (1)$$

With the linear interpolation, the daily reading of smart meter data become 1,440 which commensurate with EV data. The case studies and scenarios are detailed in the proceeding section. The finalized data is applied to the k-means clustering algorithm. Before application of k-means, the data is normalized to remove the impact of magnitude [9]. The normalization of a curve is carried out by dividing the curve by its maximum value. In this way all curves have values between 0 and 1 which results in clustering to focus on the shape of the profiles only.

B. K-means Clustering

k-means clustering is one of the most robust techniques with simple operating principle and is stable against the time resolution [10].

According to this approach after data pre-processing, the smart meter data is subjected to k-means clustering. Let $X = \{x_1, x_2, x_3, \ldots\ldots, x_m\}$ be the data set with m instances, and let $A_1, A_2, A3, \ldots, A_K$ be the k disjoint clusters of X [2].

i. Select k cluster centers randomly.

ii. Calculate the distance between each data point and cluster centers.

iii. Assign the data point to cluster with minimum distance to its center.

iv. Re-compute the new center with:

$A_i = (1/C_i)\sum_{j=1}^{C_i} x_i$ where C_i represents number of data points in the ith cluster.

v. Re-compute the distance between each data point and new cluster centers.

vi. If any data point member changes cluster membership, repeat the procedure from iii until no change in cluster membership occurs.

k-means clustering tries to minimize the error function which is given in (2):

$$E = \sum_{i=1}^{k} \sum_{x \in A_i} d(x, \mu(A_i)) \quad (2)$$

where $\mu(A_i)$ is the centroid of cluster, $A_i.d(x, \mu(A_i))$ denotes the distance between x data point and $\mu(A_i)$.

C. Load Profile Extraction

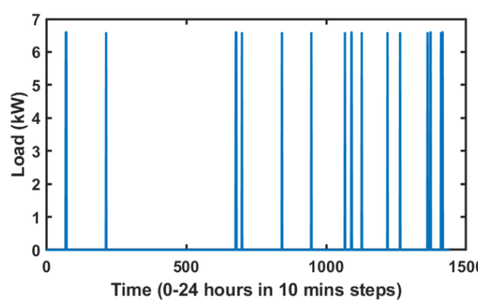
Once the k-means clustering on the data is performed on the dataset, the resultant clusters are used to extract typical load profiles. The typical load profiles of the clusters are extracted by taking arithmetic mean of the profiles in the cluster. The resultant profiles are considered as typical load profiles for the cluster which can be used for different power system application.

III. Numerical Application

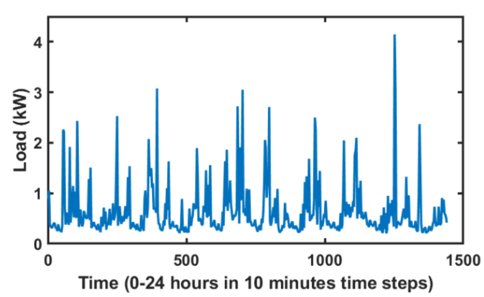
Distinction between EV profile and a normal consumer profile can be made from the Figure 2. From Figure 2, it can be observed that the EV profile tends to have higher magnitude and flat charging load, whereas the household load profile is highly volatile and variable. In total, 14 EV charging events took place on the particular day with different charging times.

Combining the two profiles provides aggregated profile which has EV profile embedded within the original household profile. Figure 3 shows an aggregated profile of profiles shown in Figure 2. The aggregated profile shows characteristics of the domestic load profile with its erratic behaviour and new peaks introduced due to the EV charging. Therefore, it is expected that due to change in shape of the cluster, the clustering results of the aggregated profile will be different from the household load profile without EV.

The case study considers different levels of the EV profiles i.e., 10, 20 and 30 EVs and each case study further considers four clustering scenarios i.e., clustering one day data, clustering one week, clustering one month data and clustering one year data. However, due to the limitations of the space EV profile with 20 and 30 EVs are not presented here. The results show a similar trend in all the cases.



a. EV charging profile



b. Domestic consumer load profile

Figure 2. EV charging and domestic consumer 24 hours load profiles

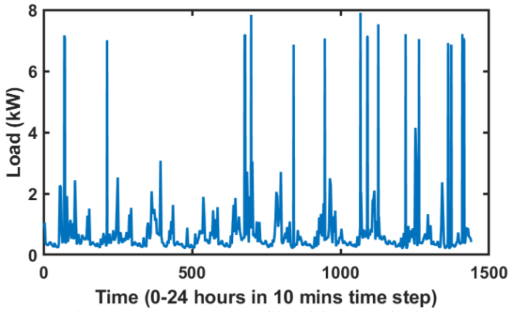


Figure 3. Aggregated profile with EV embedded

A. Case. 10 EVs

In the Case I, 10 random EV profiles for a year are selected. Initially, smart meter data of 200 customers is clustered for four different scenarios i.e., for one day only, for one week only, for one month only, and then finally for one year. Once this is achieved, aggregated profiles for all four scenarios using the 10 EV profiles are generated and then these aggregated profiles of the 10 consumers replace 10 consumers in the smart meter data. For ease of the comparison, the first 10 consumer profiles are taken out, their profiles are aggregated with EV and finally these aggregated profiles replace the original profiles.

Different number of clusters are simulated ranging from 2 to 8 and 4 is chosen as final cluster numbers. Having too high number of clusters can result in extraction of non-significant patterns and too low numbers can lead to loss of general patterns. Number of consumers in all four clusters remains almost the same in both smart meter and aggregated profiles.

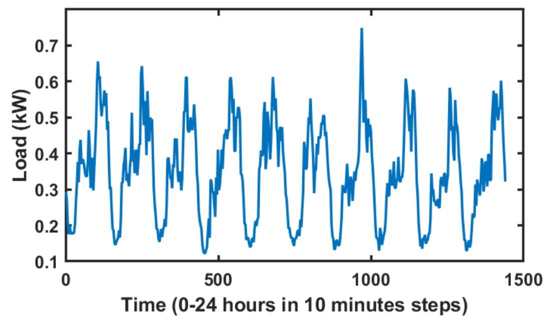
The change in cluster membership can be seen from the Table I, where clustering results for first 10 users show that these consumers belong three different clusters. Initially, the first 10 consumers without EV belong to clusters 1,3 and 4, but after aggregation of profiles with EV profiles, 80% EV customers belong to cluster 4 and rest of the 20% to cluster 2 (highlighted in red). It is important to note that the cluster number in both cases may not be the same clusters i.e. cluster number 2 in non-EV load may be marked as cluster number 4 in case of EV data. However, apart from the EV, the other consumers tend to retain their original cluster membership with acceptance of a few consumers. The change in the membership of these consumers is result of the change of centroid position due to the aggregation of profiles. The results for all scenarios i.e. for one week, one month and on year data show similar trends as can be seen from the Table I. Moreover, the load profiles of the cluster containing daily clustering before and after EV are given in Figure 4. Most of the EV consumers spread across different clusters tend to be in the same cluster.

B. Discussions

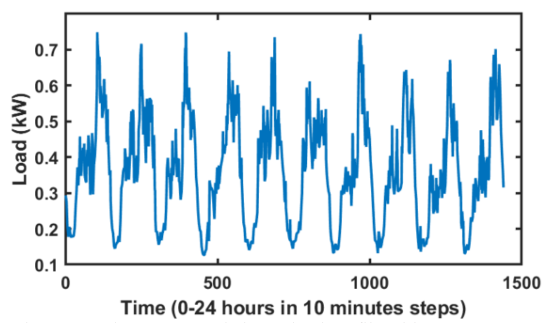
The study shows that sensitivity of EV penetration does not have significant impact on the clustering results and most of the EV consumers tend to remain in the same cluster. The daily, weekly, monthly or annual clustering exhibit somewhat similar outcome in terms of the EV classification, however, overall changes in number of consumers within a class are significant.

Table I. Clustering membership with and without 10 EV Consumers

Cluster No.	One day without EV	One day with EV	One week without EV	One week with EV	One month without EV	One month with EV	One year without EV	One year with EV
	3	4	1	2	2	4	1	2
	4	2	2	4	1	1	2	1
	3	4	1	4	2	4	1	2
	1	4	1	4	4	4	1	2
	3	4	4	4	2	4	1	2
	1	2	1	4	4	4	1	2
	3	4	4	4	4	4	1	2
	1	4	1	2	4	4	1	2
	3	4	4	4	2	2	3	1
	1	4	1	2	4	4	1	2



a. One day domestic load profile without EV



b. One day aggregated cluster load profile with EV

Figure 4. One day Pre and Post EV cluster profiles: (a) without EV and (b) with EV

As the focus of this study is on the impact of EV penetration, these aspects are ignored in this study.

It is observed that, inclusion of EV results in cluster load profiles with higher magnitude of load and changes in patterns with sudden spikes. The EV charging profiles show that EV loads are high load and whenever they are installed at consumers premises, the consumer load patterns tend to change. This change can be observed in the form of high spikes of loads and overall increase in the maximum demand of the consumer. Although, stratification of the loads can be potentially used to isolate the EV consumers, presence of other high loads can lead to misclassification of consumers. This study shows that when a consumer buys new EV, their energy consumption pattern changes such that the new EV consumers tend change their membership into the same cluster. As the element of magnitude is already omitted by normalizing the data, the results are purely based on the shape of the load curves of energy consumption patterns/habits.

IV. Conclusion

This paper presents an investigation into the impact of EV penetration in load profiling of domestic consumers. The results suggest that presence of EV significantly changes energy consumption patterns of the domestic consumers. The consumers classified into different classes before EV tend to fall in the same class after incorporation of EV charging profile. Sensitivity analysis with different scenarios shows that the change in load patterns of the consumers due to EV makes the patterns of the consumers similar enough to be in the same class. Studies using the historic classification of consumer and present class change can potentially be used to identify the presence of EV.

Acknowledgment

This work was supported in part by the MSIT, Korea, under the ITRC support program (IITP-2021-2018-0-01426) supervised by the IITP and in part by the NRF funded by the MSIT (no. 2019R1F1A1059125).

References

[1] D. L. Donaldson and D. Jayaweera, "Effective solar prosumer identification using net smart meter data," *International Journal of Electrical Power & Energy Systems,* vol. 118, p. 105823, 2020.

[2] Z. Khan, D. Jayaweera, and H. Gunduz, "Smart meter data taxonomy for demand side management in smart grids," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2016, pp. 1-8.

[3] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, J. Zhang, and P. Siano, "A comparative study of clustering techniques for electrical load pattern segmentation," *Renewable and Sustainable Energy Reviews,* vol. 120, p. 109628, 2020.

[4] E. Apostolaki-Iosifidou, S. Woo, and T. Lipman, "Challenges and Opportunities for Electric Vehicle Charging Detection Using Utility Energy Consumption Data," 2019.

[5] A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, "Low Complexity Non-Intrusive Load Disaggregation of Air Conditioning Unit and Electric Vehicle Charging," in *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, 2019, pp. 2607-2612.

[6] *ISSDA, CER Smart Meter Customer Behaviour Trials Data, accessed via the Irish Social Science Data Archive, CER Electricity, Accessed via www.ucd.ie/issda (revised March 2012)*.

[7] N. R. E. Laboratory. (2019). *Impact of uncoordinated plug-in electric vehicle charging on residential power demand - supplementary data*. Available: https://catalog.data.gov/dataset/impact-of-uncoordinated-plug-in-electric-vehicle-charging-on-residential-power-demand-supp

[8] A. Aligholian, M. Farajollahi, and H. Mohsenian-Rad, "Unsupervised Learning for Online Abnormality Detection in Smart Meter Data," in *IEEE Power & Energy Society General Meeting*, 2019.

[9] G. Milligan and M. Cooper, "A study of standardization of variables in cluster analysis," *Journal of Classification,* vol. 5, pp. 181–204,, September 1988 1988.

[10] [10] R. Granell, C. J. Axon, and D. C. Wallom, "Impacts of raw data temporal resolution using selected clustering methods on residential electricity load profiles," *IEEE Transactions on Power Systems,* vol. 30, pp. 3217-3224, 2014.