

A Review on Clustering of Residential Electricity Customers and Its Applications

Amin Rajabi, Li Li, Jiangfeng Zhang, Jianguo Zhu, Sahand Ghavidel, and Mojtaba Jabbari Ghadi
Faculty of Engineering and Information Technology, University of Technology Sydney, Australia
Email: amin.rajabi@student.uts.edu.au

Abstract--Clustering is a well-recognized data mining technique which enables the determination of underlying patterns in datasets. In electric power systems, it has been traditionally utilized for different purposes like defining customer load profiles, tariff designs and improving load forecasting. Some surveys summarized different clustering techniques which were traditionally used for customer segmentation and load profiling. The recent changes in power system structure and introduction of new technologies necessitate the new investigation of applications and benefits of clustering methods for power systems. In this regard, this paper aims at reviewing the new research for clustering techniques for residential customers.

Index Terms-- Clustering, Customer segmentation, Data mining, Smart meters

I. INTRODUCTION

Data mining techniques have been extensively used in different areas of research to analyze underlying patterns of collected data. In power system sector, the implementation of these techniques was limited primarily due to the lack of access to fine grained electricity data. The recent changes in electricity system structure toward smart grids and the great utilization of advanced metering infrastructure (AMI) and smart meters have affected this trend greatly. The vast rollouts of smart meters around the world will provide power system operators with a huge amount of data especially from residential households which usually were ignored at individual level. These data are normally collected by smart meters in a time resolution of 30 or 60 minutes [1]. Use of even more fine grained measurements like every 15 or 1 minute is also possible depending on system requirements.

Beside the physical challenges and costs of installation of AMI and smart meters and providing communication system for bi-directional flow of data, the other major problem is the use of this invaluable data to increase the system efficiency and reliability. Application of data mining techniques is necessary as the millions of records must be processed. In this regard, clustering as an unsupervised data mining technique can help greatly to find similar patterns in data and to categorize electricity customers. Clustering of residential customers can benefit system operators in various ways especially through peak load reduction and introduction of demand response programs. So, in this paper a detailed review of different clustering techniques and their applications for residential customer's segmentation will be presented. It is not intended here to focus on definition and mathematical

formulations of clustering concepts as they have been extensively reported in the literature [2-5]. Instead, the new areas of research are presented in a well-ordered format and for interested readers the proper references of each concept are cited. The organization of the paper is as follows: Section II is dedicated to a brief review of clustering concepts and techniques. Section III reviews the recent literature related to residential clustering. Applications of residential customers' segmentation are discussed in Section IV. Finally, Section V summarizes the paper.

II. REVIEW OF CLUSTERING CONCEPTS AND METHODS

A. Clustering Techniques

The main problem of clustering is to partition a data set into a number of groups which are as similar as possible [6]. Clustering has numerous applications in different fields like biology, market analysis, and social-network studies for finding intrinsic patterns in datasets [7].

In power systems, clustering techniques can be applied to reduce the dimension of data and to find customers which show similar patterns. The consumption data of users are used for the clustering and customers are categorized to different clusters. The final goal is that the objects within a cluster be as similar as possible, while, two clusters can be completely differentiated from each other and have as less as possible similarity. In this regard, distance measures are used to find similarity or dissimilarity between any pair of objects. For numeric attributes like electricity consumption data, measures like Minkowski (L_p -norm), cosine similarity, and dynamic time warping (DTW) are usually used [4] [8, 9].

Various clustering techniques have been utilized for electricity customer segmentation. Table I reports some of these techniques with corresponding references. The best references for the concepts and mathematical definitions are also mentioned. Among these techniques K-means, hierarchical, self-organizing maps and modified follow the leader clustering techniques and different variations of them are the most popular ones in power system literature.

These methods differ based on different factors. Some algorithms like K-means require the number of clusters to be pre-determined before applying the clustering, whereas adaptive k-means, hierarchical and modified follow the leader do not need such criteria. Also, every distance measure is more suitable for each clustering technique, for instance, the most used measure for K-means is L_2

TABLE I
CLUSTERING TECHNIQUES USED FOR POWER SYSTEM

Clustering techniques	References	Definition
K-means (KM)	[2] [4] [7] [10] [11] [12] [13]	[6]
Variations of K-means including K-medoids, K-medium and Fuzzy C-means (FCM)	[4] [13] [9] [14] [15] [16] [17]	[14] [17]
Adaptive K-means	[18] [19] [20]	[20]
Hierarchical (H)	[2] [4] [7] [13] [15] [18] [21]	[6]
Self-organizing maps (SOM)	[4] [22] [11] [23] [15] [24]	[15] [24]
Modified follow the leader (FDL)	[2] [25] [26]	[25]
Expectation maximization (EM)	[27] [28] [16]	[27]
Online clustering/ parallel clustering/ divide and conquer	[19] [20]	--

norm while K-medoids minimizes L_1 norm. Furthermore, the complexity and computation time of these methods are also different.

B. Clustering Validity Indexes

Usually one or a few of clustering validity indexes (CVI) are used along with the clustering techniques. These indexes are used to evaluate the performance of clustering methods and hence, they can be utilized to determine the suitable number of clusters [29] [22], to compare the performance of different clustering techniques [2] and to evaluate the performance of clustering when some attributes (features) are added or removed [30]. Table II reports some of the most used CVIs along with the corresponding references.

C. Features

Application of clustering on load profiles can be performed in different ways based on the data set that is used. Generally clustering can be applied to [2] [5]:

- Raw consumption (time series) data of customers
- Features that are defined by the user based on the characteristics of load shapes and specific application
- Features that are extracted from load shapes by use of techniques such as frequency domain analysis
- A reduced data set which is suitably obtained from the original data for instance by use of techniques like principal component analysis (PCA)

Table III reports some of these techniques and references for each of them.

III. RESIDENTIAL LOAD PATTERN CLUSTERING

A. Use of Clustering for Power Systems

As a common trend, in the past, utilities usually divided their customers in industrial, commercial and residential sectors based on some fixed information like voltage level and nominal demand or based on the surveys which asked customers about various socio-demographic features [31-33]. Based on this approach a set of customer class load profiles were defined and each user was assigned to one of these classes [34]. For example, Finnish Electricity Association (FEA) published customer class load profiles for 46 different customer classes, 18 of which are for housing and the rest for agriculture, industry and services.

TABLE II
CLUSTERING VALIDITY INDEXES

Cluster validity indexes (CVI)	References
CDI (cluster dispersion indicator)	[12] [4] [26] [15] [30]
DBI (Davies-Bouldin indicator)	[12] [29] [4] [26] [15] [30]
MIA (mean index adequacy)	[22] [4] [26] [15] [30] [2]
SMI (similarity matrix indicator)	[4] [15] [2] [30]
DI (Dunn index) and MDI (modified Dunn Index)	[2] [12]
SIL (Silhouette index)	[12] [19]
WCBCR (the ratio of within cluster sum of squares to between cluster variation)	[9] [15] [2]
SI (scatter index)	[2]

TABLE III
CLUSTERING APPLICATION ON TIME SERIES, FEATURES, OR REDUCED DATA SET

Method	Examples	Reference
Raw consumption (time series) data	--	[12] [11]
Feature definition	Load factor: daily $P_{av,day}/P_{max,day}$ mean; standard deviation; skewness	[29] [28]
Feature extraction	Discrete Fourier transform (DFT); harmonics-based coefficients; discrete wavelet transform	[26] [35]
Data size reduction	Principal component analysis (PCA); Sammon map; symbolic aggregate approximation (SAX)	[20] [36]

Each customer was attached to specific load curve which is used as a base of billing and distribution planning [34] [29]. This method was satisfactory as in most cases only monthly kWh energy usages of customers were available.

To have a better understanding of power system operation, some typical projects were performed to measure electricity consumption of a limited number of users in shorter time resolutions. Therefore, most of the research in the literature was focused on clustering of such consumption data to improve class load profiles or to present new load profiles [29] [37]. However, this is still a fundamental problem, and the procedures for dealing with customers' segmentation need to be revised greatly. Firstly, the consumption data of a great deal of customers (those who have installed smart meters which can amount to tens of thousands to millions of users) are now accessible. Secondly, the time period of measurement is not restricted and usage information for some successive years is available. These two factors affect the dimensionality of data which is not comparable with previously used data sets. Finally, as the data is continuously recorded, it can have possible applications for real-time operation and management of power systems. All of these factors emphasize the use of new clustering methods for electricity consumption characterization.

B. Residential Load Shape Characteristics

For the reasons that were mentioned in the previous section, application of clustering techniques for

residential households is still in its infancy. Individual residential load shapes are much different from the aggregated load shapes or industrial or commercial load patterns and the number of residential users is not comparable to the users in those sectors. The first obvious challenge arises from the nature of residential load patterns. They are much more variable, such that two households with similar characteristics can have completely different consumption patterns and the everyday usage can be different even for the same dwelling. This affects the application of any clustering technique. To overcome this variable pattern, many studies try to build a representative load pattern (RLP) for each residential user [2] [22] [10] [12] [26]. To this end, they combine the instant-by-instant data of daily load patterns based on a criteria (for instance, by averaging them) [2] [4] [38] and build a RLP for each home. Clustering can later be performed on these RLPs to categorize customers. However, this approach has advantages the useful information about the usage patterns will be missed. Another promising approach is to treat days of week separately and build a RLP for each day; however this method still suffers from the same mentioned problem.

Most of the studies in the literature try to form a number of clusters in which the customers that show similar usage patterns belong to the same cluster. A few other studies cluster daily loads of only one user to specify characteristics of just that customer. These researches can also be differentiated from each other based on the clustering algorithm, use of RLPs, use of features or data reduction methods, number of users, time period of study, and considering weekend and seasonality effects.

C. Literature Review

Ref [10] utilizes the data of 103 dwellings measured every 1 minute and creates a seasonal daily load profile for each household. Then, KM clustering is applied to these curves to partition the homes to two clusters for each season. Three clustering algorithms (KM, H, and Gaussian mixture model based clustering) are utilized in [7] to cluster the average monthly load profiles of each household. The clustering is in a way that each hour can be dedicated to a different cluster. Based on this procedure, they tried to find out the time of use prices for households. [11] evaluates the use of KM, K-medoids and SOM for clustering of daily load profiles of residential users. Using the DBI validity index, the most suitable clustering and the best number of clusters are decided.

The purpose of [15] is to cluster daily load curves of only one customer and present some patterns based on them. Several clustering methods (KM, FCM, SOM, and 7 different H) are applied to the measured daily consumption.

Three different attributes are used in [30] as the input of KM method to cluster residential customers. The attributes (flexibility measures) defined in a way to represent the flexibility of each household, hence, allowing applying demand response programs to those customers with highest possibility of changes in their

loads. These attributes are: total usage over evening period, the time of maximum usage, time of minimum usage. Four distinct time periods for electricity consumption are distinguished in [28] which are: overnight, breakfast, daytime and evening. Seven attributes are defined to obtain the main features of original data including the energy use in each of time periods, variability of energy use, and effect of seasonality and weekends. A clustering of these attributes is done by considering it as a finite mixture model of Gaussian multivariate distribution. To determine the number of clusters, Bayesian Information Criterion (BIC) is used.

Ref [37] investigates the customer classification for Finish electricity grid. Iterative self-organizing data-analysis technique algorithm is used for clustering and based on that, customer class load profiles are calculated. Optimum number of clusters is selected based on the knee-point criterion which is decided by using square sum of errors. [13] tries to generate the load profile of users without automatic meter reading (AMR) facilities based on the usage data of customers equipped with AMR. KM, FCM, and H clustering are used for this purpose. A framework for electricity consumer characterization is presented in [22]. A combination of SOM and KM algorithms is used in which firstly, SOM reduces the dimension of data set and secondly, KM method is utilized to perform the clustering. For each user, the RLP is built by averaging the measured load curves.

Authors in [23] use a PCA method to extract the main features from daily load profiles, and based on these features, consumption patterns are distinguished. Then the customers are clustered using SOM technique. Ref. [26] uses DFT to transform time-domain measurements to frequency domain. A set of features are then defined based on values in frequency domain which are used by a modified follow the leader algorithm to cluster customers. MIA, CDI, DBI and SI are used for evaluation and comparisons.

Authors in [18] evaluate a huge amount of consumption data (around 218 thousand customers and 66 million load shapes) with innovative techniques. Clustering is performed through an adaptive KM algorithm in which the total number of clusters can change. Later, a hierarchical algorithm is applied to decrease the number of clusters. As the clustering is performed on normalized daily load shapes, other measures are defined to capture the variability and quantity of usage. [16] divides electricity consumption into two parts: occupancy-related and weather-related consumption. A hidden Markov model (HMM) framework is utilized to infer the occupancy states from consumption data. Spectral clustering is used to segment collection of HMMs. Number of clusters is estimated by using the gap statistic technique, and K-medoid algorithm is applied as the last step in spectral clustering to identify representative users. A novel idea is proposed in [20] to cluster electricity customers based on the transitions between consumption levels instead of focusing on the shape of load profiles. After normalizing data of each

household, the dimension reduction is done using SAX technique. Then, a time-based Markov model is applied to capture the dynamics of load data, and a density-based clustering method is utilized to cluster the customers.

[19] proposes an online clustering method for high dimensional time series data. It utilizes an adaptive K-means algorithm and performs analysis of clustering based on an online algorithm. The principle behind this online time series clustering is a batch divide-and-conquer scheme in which the clustering is applied on chunks of data points and once the entire data set is scanned it combines the results to find the final clustering. The optimal number of clusters is determined using the SIL index. On the other hand, [39] introduces a “distributed method” based on an artificial neural network variant called deep-learning auto-encoder to cluster load profiles of 1000 customers.

IV. APPLICATIONS OF CLUSTERING

Among the possible applications of clustering for power systems are benefits for tariff design [25] [40], defining demand response programs [27] [18], improvement of load forecast [41] [42], identifying energy behavioral use patterns [28] [34] and classifying new or non-metered customers [21] [22]. Some of them are explained in the following.

A. Tariff Design

One of the most important applications of clustering of residential customers is to design suitable tariffs for different customers based on the classes that they belong to. As different customers show different load patterns, clustering can help to design cluster-specific tariff structures which can decrease peak load.

In [40] an electricity retailer clusters the customers in order to maximize the annual profits. For this purpose, the optimal selling price of each cluster is determined based on a profit function. Clustering is achieved through a weighted fuzzy average K-means method. [25] proposes a multiple rate tariff structure to replace the single rate structure. Firstly, the load diagrams of weekdays are built by averaging the load data of all the weekdays over the measurement period. Then, it introduces four features (called shape indicators) and clusters customers by using a FDL algorithm. To study new tariff structure, it is assumed that the total revenues under the new tariffs do not exceed the total revenues with the previous tariff. Furthermore, it suggests a procedure for identifying the customer's class and assigning an appropriate tariff to the customer. Clustering of a group of customers in Germany is performed in [43] where the authors use KM algorithm and DBI validity index to achieve an optimal number of clusters. Then, it proposes segment-specific rate design which sets a different rate for each segment of individual customers. For this purpose, a number of time zones based on the peaks and valleys of customer are defined and a time-variable rate is assigned to them.

B. Load Forecasting

Some research has used the clustering results to

improve the load forecasting. In [42] the authors aim to forecast household loads for two time periods: one hour and 24 hour ahead forecasting. They propose a new method called “cluster-based aggregate forecasting (CBAF)” and compare it with two other approaches, i.e., (1) to aggregate the energy consumption of all households into one time series (the aggregate consumption), then forecast the aggregate consumption, and (2) to forecast the energy consumption of each household separately, then aggregate the forecasts. They show that CBAF produces better results if a suitable number of clusters are selected and the size of customer base is large enough. [9] addresses load forecasting using dynamic time warping (DTW) distance. They try to predict energy usage based on a partial load curve (a few hours of actual energy use). So, the problem is like a classification problem in which clusters are treated as classes and partial load curves are assigned to them. The WCBCR is used for assessing the quality of clustering by K-mean (with L_2 similarity metric) and K-medoid (with DTW similarity metric) methods.

C. Demand Response

The final goal of demand response is to reduce the load of the network at peak hours (or when the system security is jeopardized) or to shift the load from a special time of day to another time periods. For this purpose, price-based or incentive-based demand response programs can be designed and offered to customers. Customers showing different patterns can be identified and different kinds of demand response programs can be offered to them.

[18] uses three factors i.e. shape of the load, quantity of usage, and variability to evaluate the consumption data of customers and to segment them for demand response purposes. Load curves are firstly normalized and a mixture of adaptive KM and H clusterings are performed on them. Through this approach, customers are clustered based on the consumption time. For instance, users with morning peak will be in the same cluster. Quantity and variability are also accounted for by use of other suitable measures. Ref [44] addresses the problem of segmentation of electricity users for the utilities by using consumption, demographics and previous program enrolment data. The final goal is to extract those users that are most probable to enroll in different energy efficiency or demand response programs and to target each group with efficient appropriate messages. A predictive segmentation is applied on a large population of about 1 million users.

To characterize the potential of demand response of wet appliances in Belgium, the EM clustering algorithm is utilized in [27] for clustering households based on their own load profiles. In addition, KM method is applied to perform a social segmentation based on the attitude of residents toward active demand reduction to form 4 groups. The results of these methods are then combined to estimate the potential of demand reduction.

D. Classification

Classification of new customers or those customers that are not equipped with smart meters is one of the

possible applications of clustering that is investigated in various works. To achieve this, firstly a number of customer classes are defined based on the data of customers with AMR facilities. Then each non-AMR user will be assigned to one of these classes based on the available information such as the monthly usage and other fixed information.

[13] clusters AMR users firstly and calculates a typical load profile (TLP) for each cluster as the mean of load patterns belonging to that cluster. Then the daily load profile of each non-AMR user is created by comparing different attributes of that customer with TLPs.

Ref. [22] presents a framework that includes two parts: the load profiling module and the classification module. In load profiling module, a load profile for each cluster is obtained by averaging all RLPs belonging to that cluster. In the classification stage, three load shape indexes (load factor, night impact, and lunch impact) and commercial indexes are used to classify customers.

V. CONCLUSIONS

This paper reviews the new research that has been conducted in recent years on clustering of residential load profiles. A brief review of the most used clustering algorithms for customer segmentation, an extensive literature review of current research and possible applications of clustering techniques for power systems were presented and discussed. Further research on new clustering techniques like online clustering for fast applications such as dynamic demand response needs to be investigated in future.

REFERENCES

- [1] C. Muscas, M. Pau, P. A. Pegoraro, and S. Sulis, "Smart electric energy measurements in power distribution grids," *IEEE Instrumentation & Measurement Magazine*, vol. 18, pp. 17-21, 2015.
- [2] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, pp. 68-80, 2012.
- [3] S.-I. Yang and C. Shen, "A review of electric load classification in smart grid environment," *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 103-110, 2013.
- [4] G. Chicco, R. Napoli, and F. Piglione, "Application of clustering algorithms and self organising maps to classify electricity customers," in *Power Tech Conference Proceedings, 2003 IEEE Bologna*, 2003, p. 7 pp. Vol. 1.
- [5] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Science and Technology*, vol. 20, pp. 117-129, 2015.
- [6] C. C. Aggarwal and C. K. Reddy, *Data clustering: algorithms and applications*. Chapman and Hall/CRC, 2013.
- [7] M. Crow, "Clustering-based methodology for optimal residential time of use design structure," in *North American Power Symposium (NAPS), 2014*, 2014, pp. 1-6.
- [8] I. Prahasano, D. King, and C. Ozveren, "A review of electricity load profile classification methods," in *Universities Power Engineering Conference, 2007. UPEC 2007. 42nd International*, 2007, pp. 1187-1191.
- [9] T. Teeraratkul, D. O'neilly, and S. Lallz, "Condensed representation and individual prediction of consumer demand," in *Smart Energy Grid Engineering (SEGE), 2016 IEEE*, 2016, pp. 11-16.
- [10] J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar, and M. E. Webber, "Clustering analysis of residential electricity demand profiles," *Applied Energy*, vol. 135, pp. 461-471, 2014.
- [11] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Applied energy*, vol. 141, pp. 190-199, 2015.
- [12] J. L. Viegas, S. M. Vieira, R. Melicio, V. Mendes, and J. M. Sousa, "Classification of new electricity customers based on surveys and smart metering data," *Energy*, vol. 107, pp. 804-817, 2016.
- [13] Y.-I. Kim, J.-M. Ko, and S.-H. Choi, "Methods for generating TLPs (typical load profiles) for smart grid-based energy programs," in *Computational Intelligence Applications In Smart Grid (CIASG), 2011 IEEE Symposium on*, 2011, pp. 1-6.
- [14] K. Zhou, S. Yang, and Z. Shao, "Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study," *Journal of Cleaner Production*, vol. 141, pp. 900-908, 2017.
- [15] G. Tsekouras, P. Kotoulas, C. Tsirekis, E. Dialynas, and N. Hatziaargyriou, "A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers," *Electric Power Systems Research*, vol. 78, pp. 1494-1510, 2008.
- [16] A. Albert and R. Rajagopal, "Smart meter driven segmentation: What your consumption says about you," *IEEE Transactions on Power Systems*, vol. 28, pp. 4019-4030, 2013.
- [17] I. Benítez, A. Quijano, J.-L. Díez, 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.
- [18] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, pp. 420-430, 2014.
- [19] R. Pal, C. Chelmiss, M. Frincu, and V. Prasanna, "Time Series Clustering for Demand Response, An Online Algorithmic Approach," *Online: www.cs.usc.edu/assets/007/93954.pdf*.
- [20] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of electricity consumption behavior

- dynamics toward big data applications," *IEEE Transactions on Smart Grid*, vol. 7, pp. 2437-2447, 2016.
- [21] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of low voltage network templates—Part I: Substation clustering and classification," *IEEE Transactions on Power Systems*, vol. 30, pp. 3036-3044, 2015.
- [22] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Transactions on power systems*, vol. 20, pp. 596-602, 2005.
- [23] O. Motlagh, G. Foliente, and G. Grozev, "Knowledge-mining the Australian smart grid smart city data: A statistical-neural approach to demand-response analysis," in *Planning Support Systems and Smart Cities*, ed: Springer, 2015, pp. 189-207.
- [24] S. V. Verdú, M. O. Garcia, C. Senabre, A. G. Marín, and F. J. G. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *IEEE Transactions on Power Systems*, vol. 21, pp. 1672-1682, 2006.
- [25] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," *IEEE Transactions on Power Systems*, vol. 18, pp. 381-387, 2003.
- [26] E. Carpaneto, G. Chicco, R. Napoli, and M. Scutariu, "Electricity customer classification using frequency-domain load pattern data," *International Journal of Electrical Power & Energy Systems*, vol. 28, pp. 13-20, 2006.
- [27] W. Labeeuw, J. Stragier, and G. Deconinck, "Potential of active demand reduction with residential wet appliances: A case study for Belgium," *IEEE Transactions on Smart Grid*, vol. 6, pp. 315-323, 2015.
- [28] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE Transactions on Smart Grid*, vol. 7, pp. 136-144, 2016.
- [29] T. Räsänen and M. Kolehmainen, "Feature-based clustering for electricity use time series data," in *International Conference on Adaptive and Natural Computing Algorithms*, 2009, pp. 401-412.
- [30] I. Dent, T. Craig, U. Aickelin, and T. Rodden, "An approach for assessing clustering of households by electricity usage," 2012.
- [31] L. Lutzenhiser, "Behavioral assumptions underlying California residential energy efficiency programs," *CIEE Energy & Behavior Program*, 2009.
- [32] B. A. Smith, J. Wong, and R. Rajagopal, "A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting," in *ACEEE Summer Study on Energy Efficiency in Buildings*, 2012, pp. 374-386.
- [33] L. Dethman and D. Thomley, "Comparison of Segmentation Plans for Residential Customers," *Energy Trust*, 2009.
- [34] B. Stephen, A. J. Mutanen, S. Galloway, G. Burt, and P. Järventausta, "Enhanced load profiling for residential network customers," *IEEE Transactions on Power Delivery*, vol. 29, pp. 88-96, 2014.
- [35] S. Zhong and K.-S. Tam, "Hierarchical classification of load profiles based on their characteristic attributes in frequency domain," *IEEE Transactions on Power Systems*, vol. 30, pp. 2434-2441, 2015.
- [36] A. Notaristefano, G. Chicco, and F. Piglion, "Data size reduction with symbolic aggregate approximation for electrical load pattern grouping," *IET Generation, Transmission & Distribution*, vol. 7, pp. 108-117, 2013.
- [37] A. Mutanen, M. Ruska, S. Repo, and P. Jarventausta, "Customer classification and load profiling method for distribution systems," *IEEE Transactions on Power Delivery*, vol. 26, pp. 1755-1763, 2011.
- [38] G. Chicco, R. Napoli, F. Piglion, P. Postolache, M. Scutariu, and C. Toader, "Comparisons among clustering techniques for electricity customer classification," *IEEE TRANSACTIONS ON POWER SYSTEMS PWRS*, vol. 21, p. 933, 2006.
- [39] E. D. Varga, S. F. Beretka, C. Noce, and G. Sapienza, "Robust real-time load profile encoding and classification framework for efficient power systems operation," *IEEE Transactions on Power Systems*, vol. 30, pp. 1897-1904, 2015.
- [40] N. Mahmoudi-Kohan, M. P. Moghaddam, and M. Sheikh-El-Eslami, "An annual framework for clustering-based pricing for an electricity retailer," *Electric Power Systems Research*, vol. 80, pp. 1042-1048, 2010.
- [41] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Transactions on Smart Grid*, vol. 6, pp. 911-918, 2015.
- [42] T. K. Wijaya, M. Vasirani, S. Humeau, and K. Aberer, "Cluster-based aggregate forecasting for residential electricity demand using smart meter data," in *Big Data (Big Data), 2015 IEEE International Conference on*, 2015, pp. 879-887.
- [43] C. Flath, D. Nicolay, T. Conte, C. van Dinther, and L. Filipova-Neumann, "Cluster analysis of smart metering data," *Business & Information Systems Engineering*, vol. 4, pp. 31-39, 2012.
- [44] A. Albert and M. Maasoumy, "Predictive segmentation of energy consumers," *Applied Energy*, vol. 177, pp. 435-448, 2016.