Change Detection in Electricity Consumption Patterns Utilizing Adaptive Information Theoretic Algorithms

Kojury-Naftchali, Mohsen; Fereidunian, Alireza; Savaghebi, Mehdi; Akhbari, Bahareh

Published in:
IEEE Systems Journal

DOI:
10.1109/JSYST.2020.3011313

Publication date:
2021

Document version:
Accepted manuscript

Citation for pulished version (APA):
Kojury-Naftchali, M., Fereidunian, A., Savaghebi, M., & Akhbari, B. (2021). Change Detection in Electricity Consumption Patterns Utilizing Adaptive Information Theoretic Algorithms. *IEEE Systems Journal*, *15*(2), 2369-2377. https://doi.org/10.1109/JSYST.2020.3011313

Change Detection in Electricity Consumption Patterns Utilizing Adaptive Information Theoretic Algorithms

Mohsen Kojury-Naftchali, Alireza Fereidunian, *SM, IEEE*, Mehdi Savaghebi, *SM, IEEE*, Bahareh Akhbari, *SM, IEEE*

***Abstract*—The high-resolution data on electricity consumptions, recorded by smart meters at customers' premises, are valuable sources of operational information and consumption patterns. In addition, customers' characterization plays an undeniable role in implementation of demand response (DR) programs, as any changes in the consumption patterns could affect DR programs. Therefore, an accurate algorithm for detecting changes in consumption patterns is very useful in not only implementation of DR, but also in other fields such as load forecasting and peak shaving. This paper proposes a reliable procedure for detecting changes in customers' consumption patterns. For this reason, an adaptive algorithm is introduced to improve the clustering quality of customers' consumption patterns by determining the optimum number of clusters, using locally weighted entropy-based segmentation (LWEBS). Moreover, considering customers' consumption records in different time slots as the features, another adaptive algorithm is introduced for feature selection, based on the mutual information (MI) concept. The proposed method is evaluated by applying a real dataset provided from the Irish Social Science Data Archive (ISSDA). Results corroborate the efficiency of the proposed procedure.**

***Index Terms*— Customer Characterization, Advanced Metering Infrastructure (AMI), demand response (DR), entropy-based segmentation, kernel function, mutual information**

NOMENCLATURE

$P(X)$	Probability distribution of random variable X
$H(X)$	Entropy of random variable X
$H(X,Y)$	Joint entropy of random variables X , Y
$MI(X,Y)$	Mutual information between random variables X , Y
$H(X\|Y)$	Conditional entropy of variable X given variable Y
$m_{n,c}$	Membership of customer n in cluster c
$MOIND_n$	Motion indicator corresponding to customer n
ξ	Predefined threshold for total error e_{tot}^2
$m_{n,\max}$	The maximum membership coefficient of customer n
C_c	The cluster (region) c
$x_{c,k,f}$	The value of feature f for k^{th} customer of cluster c
e_{tot}^2	Total error of customer assigning to the clusters
$cen_{c,f}$	The value of f^{th} feature in the c^{th} cluster center
$Ent_z(C_c)$	Entropy of feature z in cluster C_c
$x_{i,z}$	Value of feature z for customer i
Γ_c	Sum of $x_{i,z}$ in cluster C_c
K_c	The number of members in cluster C_c
N, F, C	The number of customers, features and clusters respectively
Ω	Accuracy of the model
$\Omega_{threshold}$	Threshold of the accuracy of the model
$Ent_{tot,z}$	Total entropy of feature z
f_z , f_f	The features with the largest entropy and MI respectively
Γ_{tot}	Sum of $x_{i,z}$ in all clusters
$x_{ACT_{i,t}}$, $x_{FOR_{i,t}}$	Actual and forecasted consumption of customer i in time slot t
$\overline{x}_{ACT_{i,t}}$	Mean of the actual consumption of customer i in time slot t
DD_i	Distance distribution associated to customer i
$CPS_{i,[t_0,t_1]}$	Consumption pattern string corresponding to customer i in time interval $[t_0, t_1]$
$SI_{i,a,b}$	Similarity index corresponding to customer i for two consumption time series a, b
$G(t_m)$	Gradient of time slot t_m
SD_i	The operator for calculating standard deviation of customer i
$x_{(t)}$	Value of time series in time slot t
$\gamma_{threshold}$	Threshold for FCA
$c_{n,old}$, $c_{n,new}$	The old and new cluster to which the customer n is assigned
Δt	Time interval
$G_{[t-1,t]}$	Gradient in time interval $[t-1, t]$
$x_{t_m,a}$, $x_{t_m,b}$	Value of time series a and b in time slot t_m
μ , σ	Average and standard deviation of the Gaussian kernels respectively

I. Introduction

THE energy consumption data repository is a valuable source of operational information for electricity distribution industry. Advanced metering infrastructure (AMI) provides a bidirectional communication infrastructure between the participants of the electricity market. Considering the emerging AMI facilities equipped with smart meters and the increasing need for high resolution consumption data, the volume of the data exchanged is drastically increased. Hence, customer characterization becomes more complicated. Besides, the granularity in the nature of the household consumption data is another point complicating customer characterization. Customer characterization is investigated in several studies for the reasons like participating in demand side management (DSM) programs [1], [2]. In [4], a demand response (DR) model is proposed for residential consumers to change their consumption pattern with the aim of maximizing their utility. In [5], residential customers are clustered in a flexible DR scheme for which customers' electricity consumption habits is analyzed utilizing historical data. In [6], energy consumption of customers is classified using smart meters data. A stochastic multi-stage planning model is proposed in [7] considering uncertainty of consumers' participation in demand response (DR) schemes. Besides, [8] and [9] utilize entropy for consumption analysis in the network and load forecasting, respectively. In addition, by great increase in the volume of data which is provided by smart meters, feature selection and trend analysis attracts more attention which facilitate decision making about the time series behavior. This is conducted in [10] using data mining techniques. Besides, [11] proposes an algorithm for trend detection in time series analysis considering granularity of time series. Moreover, different methods of dimension reduction for time series are reviewed in [12]. In [23], a method for clustering customers according to their load profiles is proposed for demand response program. In [25], a modified deep residual network is utilized and the proposed model is applied to probabilistic load forecasting using Monte-Carlo simulations. In [26] a forecasting model is proposed in which a kind of modified least square regression model is used. In [27], a residential load forecasting model based on a recurrent neural network is proposed using smart meter data. In [28], an ensemble of some individual probabilistic load forecasts are combined which is a constrained quantile regression averaging method. In [29], customers load consumption patterns is analyzed and similar ones are clustered by a method considering uncertainties of customers' loads. In [30], the probabilistic nature of residential appliance demand and short-term load forecasting is described using a conditional hidden semi-markov model.

Despite the valuable research work done, prior studies lack the following considerations in customer characterization: First, an accurate procedure for change detection in households' consumption patterns using smart meter data is scarcely investigated. In addition, accurate clustering for time series entails determining the efficient number of clusters, followed by proper assignment of the samples to the clusters. If the number of clusters are too big, the results are accompanied with too unnecessary scrupulosity, while customers' characterization is accompanied with intuitive uncertainty. Moreover, if the number of clusters is too small, the accuracy and similarity of assigned samples to each cluster decreases. In addition, the reliability of inferred points of this clustering decreases, too. Therefore, a reliable procedure is necessary for determining the efficient number of clusters, as the second less considered point in the aforementioned studies.

In addition, analyzing customers' consumption patterns is usually associated with implementation of programs such as demand response. Therefore, changes in these patterns is important, thus the ability to detect these changes is very useful in customers' characterization. For instance, [24] reveals that, the ability of detecting changes in consumption patterns (i.e. abnormalities) can improve the fraud detection performance. In general, detecting changes in customers' consumption patterns is important in all of the programs in which customers' behavior matters, like demand response, peak shaving as well as theft and fraud detection.

Motivated by the above discussion, two primary goals are pursued in this paper: segmentation of customers with an acceptable accuracy and introducing a structured procedure for change detection in customers' consumption patterns.

In this paper, a novel method for detecting changes in customers' consumption patterns is introduced as performed by the two following actions:

- Optimizing the number of clusters related to the customers' consumption patterns by utilizing locally-weighted entropy-based segmentation (LWEBS)
- Determining the more effective time slots in consumption time series by an adaptive feature selection algorithm using mutual information and locally weighted Gaussian kernel learning (LWGKL).

Here, features for each customer mean the consumption records in each time slot of the consumption time series. These recorded consumption values in different time slots are used directly to have a comprehensive insight about the variation of the consumption behavior of the customers.

As the first contribution of this paper, a new procedure is introduced to detect changes in customers' consumption patterns and trends of the consumption time series by using a subset of the features. Customer characterization is performed as reliable as the case with the complete feature set, if the reduced feature subset is selected accurately. Moreover, an adaptive method is introduced to determine the optimum number of clusters here as the second contribution, because of the influencing role of the number of clusters in customers' characterization. Besides, introducing an adaptive procedure for feature selection is the third contribution. As the fourth contribution, the similarity index (SI) is introduced to represent and forecast the shape of the consumption profile and scrutinize its changes. Applying the trend analysis tools on the selected time slots (features) considerably enhances the change detection performance. Therefore, smart selection of these time slots culminates in reduction in computational burden, while the results are still accurate enough. The real dataset provided by ISSDA [22] is used for analysis, considering the anonymity of the dataset and preserving the privacy of the customers.

The rest of the paper is organized as follows: In section (II), some prerequisites are described which are mainly used for features selection. The proposed methodology is analyzed in section (III), and its implementation is performed in section (IV). Discussion on the performance evaluation of the proposed method is conducted in section (V). Eventually, the paper is concluded in section (VI).

II. Feature Selection Basics

Feature selection (FS) is a popular process used in data analytics in which a subset of features that can show almost all key properties of the samples is selected. Algorithms such as principal component analysis (PCA), wavelet and independent component analysis (ICA) are usually of interest for this aim [3]. The FS procedure in this study entails some concepts which are described in this section.

A. *Entropy*

In the information theory literature, the entropy of a random variable or signal shows its uncertainty, where larger entropy means more uncertainty. Eq.(1) represents the formula of entropy ($H(X)$) for the random variable X in which $P(X)$ is the probability mass function [16]. By inspiration of the persuasive capabilities of entropy in image segmentation [21], in this paper, the entropy of each individual feature is calculated to determine the capability of each feature in customers' segregation. Indeed, entropy of each feature is regarded as a practical instrument for measuring distinguishing capability of the features for customers' segmentation. More detailed points about the performance of the entropy in FS are discussed in Section IV.

$$H(X) = -\sum P(X)\log_2(P(X)) \tag{1}$$

B. *Mutual information*

Mutual information (MI) is one of the most popular algorithms in FS which is based on the entropy concept. To introduce this algorithm, consider two random variables X, Y. MI shows the amount of information these two variables share [16]. Eq. 2 is the formula of MI between variables X, Y. In Eq.2, $H(X)$is the total uncertainty (entropy) of the random variable X and $H(X|Y)$is the conditional entropy which shows the amount of revealed information about variable X by observing variable Y. If X and $Y$are independent variables, MI equals zero and with dependency increase, MI increases.

$$MI(X,Y) = H(X) - H(X|Y) \tag{2}$$

Through clustering, each sample is identified by a cluster label. Hence, from the MI point of view; the aim of this paper is to identify a group of features maximizing the MI between cluster labels and the selected features. In short, the MI measures the reduction in uncertainty of the assigned labels to the samples (customers), caused by applying the (observing) features. Definitely, features with higher MI have more ability in distinguishing samples and composing clusters with more segregation accuracy. More detailed information about the performance of MI is discussed in Section IV.

III. Proposed Methodology

A. *Structure of the Proposed Method*

The proposed procedure used for changes detection by customer segmentation is briefly shown in Fig. 1. According to this flowchart, the initial dataset gathered by the AMI is preprocessed using steps indicated in [15]. Hereafter, the real load profiles (time series) of customers are used as the dataset to be processed. After that, by using entropy concept, the first phase of FS is performed. This phase concerns determining the optimum cluster number and assigning samples to them accurately. In the second phase of FS, the most influencing features which are time slots of a day are determined and selected by calculating MI. Then, training process is performed on the selected features using the locally weighted learning with Gaussian kernels to construct models for describing customers' consumption pattern. Eventually, the performed changes detection is evaluated using evaluating indices which are introduced in Section V.

The two described phases of FS are as follows:

B. *First phase of feature selection: optimization by entropy*

Determining the number of clusters and accurate assignment of customers to these clusters is addressed in by utilizing a two-step adaptive procedure for FS as portrayed in Fig.2; the first step for determining the accurate number of clusters and the second one for assigning samples to the clusters accurately. This procedure begins with calculating the entropies of each feature using Eq.5. Note that, customers are clustered using a fuzzy clustering (FC) algorithm. The output of this algorithm is a group of membership coefficients $m_{n,c}$ relating customer n to cluster C among the presumed clusters. In detail, for each individual sample, there are membership coefficients representing dependency of the indicated sample to each one of the clusters [17]. Moreover, in the image processing literature, pictures are divided into regions and the entropy is used to measure the uniformity of pixels within a region [21]. In this study, clusters are regarded to play the same role as regions in image segmentation. Therefore, by calculating the entropy of each feature, its ability in distinguishing customers and their effect on the uniformity of a cluster is determined.

Equations (3)-(5) show the formulations for the entropy calculation. In Eq.3, Γ_c is the sum of values of feature z related to each sample $x_{i,z}$ in an individual cluster C.

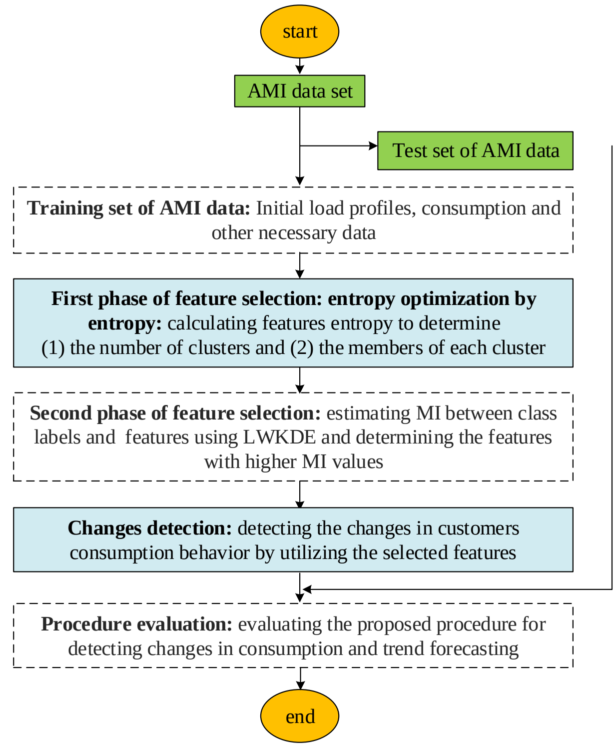

Fig.1. outline of the proposed methodology

In Eq.4, $Ent_z(C_c)$represents the entropy of feature z in cluster C. A cluster is composed of some samples with similar corresponding features. For example, suppose that, feature z is one of these features. Closer values of feature z which means less entropy of feature z for the samples of an individual cluster results in a large uniformity between these samples from this feature stand point. Similarly, Eq.5 represents the total entropy of each feature ($Ent_{tot,z}$) on the whole clusters. Furthermore, Γ_c/Γ_{tot} is considered as a weight for the entropy of each cluster representing the portion of that clusters intra-uniformity in the total uniformity of the sample space. Hence, as the size of a cluster increases, the effect of that cluster on the total uniformity increases too. Repeating this procedure, which is shown in Fig.2, culminates in the total entropy of each individual feature ($Ent_{tot,z}$). Then, the calculated $Ent_{tot,z}$ are regarded as the weights for each feature and FC is conducted by assuming a single cluster in the first step. Afterwards, the distances between each sample and the cluster center is calculated. If the condition in Eq.6 is not met, the number of clusters increases by one. This procedure continues until Eq.6 is met and when it is met, the corresponding number of clusters which is the last number is regarded as the optimal number of clusters.

Notice that, the threshold value of ξ is derived from $m_{n,max}$, therefore, selecting an efficient threshold is an important point which is derived by simulations and making trade-off by the assumption that each customer is assigned to a cluster with a specific percent of dependency. The value of ξ is determined in section (IV). Eventually and by satisfaction of Eq.6, C is derived as the optimal cluster number.

Up to now, the problem concerning the number of clusters is solved. The next problem is promoting the accuracy of customer assignment to the clusters which is conducted by applying an adaptive procedure. The calculated $Ent_{tot,z}$ values are considered in this step and the FC is performed using just the feature with the largest $Ent_{tot,z}$. In this step, motion indicator ($MOIND_n$) is introduced and used as a criteria for stopping the iterative procedure. $MOIND_n$ is defined as Eq.7 for each individual sample n using total entropy ($Ent_{tot,z}$) and membership coefficients ($m_{n,c}$). Eq.7 searches to find the cluster which maximizes the $m_{n,c}.Ent_{tot,z}$. Indeed, the aim of assigning each sample to the clusters accurately is realized by $MOIND_n$. If the cluster, to which a sample is assigned, does not change in an iteration, nothing occurs. Otherwise, the sample is assigned to the cluster with the largest corresponding $m_{n,c}.Ent_{tot,z}$ (Fig.2). Then, Fuzzy clustering adequacy (FCA) index is used as a metric to evaluate the accuracy of the constructed clusters [17]. According to the definition of FCA in Eq.8, less FCA corresponds to better clustering. If the condition on both FCA and $MOIND_n$ indices is met simultaneously, the clustering is terminated. Otherwise, the selected feature (f_z) is saved and neglected from the feature space and the procedure is repeated using the next feature with the largest entropy and continues as such. Notice that, feature selection reduces the dataset size and consequently the computational burden. Such capability is very practical especially when the volume of dataset is large.

$$\Gamma_c = -\sum_{\substack{i=1 \\ i\in R_c}}^{K_c} x_{i,z} \tag{3}$$

$$Ent(C_c) = -\sum_{\substack{i=1 \\ i\in R_c}}^{N} \frac{x_{i,z}}{\Gamma_c} \log \frac{x_{i,z}}{\Gamma_c} \tag{4}$$

$$Ent_{tot,z} = -\sum_{c=1}^{C} \frac{\Gamma_c}{\Gamma_{tot}} Ent_z(C_c) \tag{5}$$

$$e^2_{tot} = \sum_{c=1}^{C}\sum_{k=1}^{K_c}\sum_{f=1}^{F} (x_{c,k,f} - cen_{c,f})^2 \le \xi \tag{6}$$

$$MOIND_n = \arg\max_c (m_{n,c} \,.\, Ent_{tot,z}) \tag{7}$$

$$FCA = -\frac{1}{N}\sum_{n=1}^{N} m_{n,\max} \log(m_{n,\max}) \tag{8}$$

C. Second phase of feature selection: mutual information maximization

1) MI Maximization

This phase concerns determining the most influencing features to be selected for changes detection in customers consumption. Fig.3 portrays the steps of feature selection in this phase beginning with estimation of MI between every 48 features (time slot) and class labels as the variables. Then, an iterative process begins by selecting the time slot with the largest MI (f_f) as the first point of the customers' consumption time series. Then, locally weighted learning with Gaussian kernels is applied on this time slot. Checking the accuracy of the output with the predefined threshold ($\Omega_{threshold} = 95\%$) is the next step. This accuracy threshold is determined by considering the total error of 5% for predicting the ultimate changes detection in customers' consumption profiles (table (V)). If the intended accuracy is met, the process is terminated; otherwise, the procedure continues by neglecting the selected feature and the next feature with the largest MI is selected until the 95% accuracy is met.

2) Locally weighted kernel density estimation (LWKDE) for MI estimation

An important issue in the second phase of the proposed feature selection methodology is the estimation of MI. Various methods are utilized for MI estimation [20]. In this paper, LWKDE is utilized to estimate MI persuasively. Determination of MI entails existence of the probability distribution function of the under-study variables.

In this algorithm, the histogram of the distributed data is constructed, and the ultimate probability density function is concluded by allotting sufficient kernels proportional to the shape of the histogram.

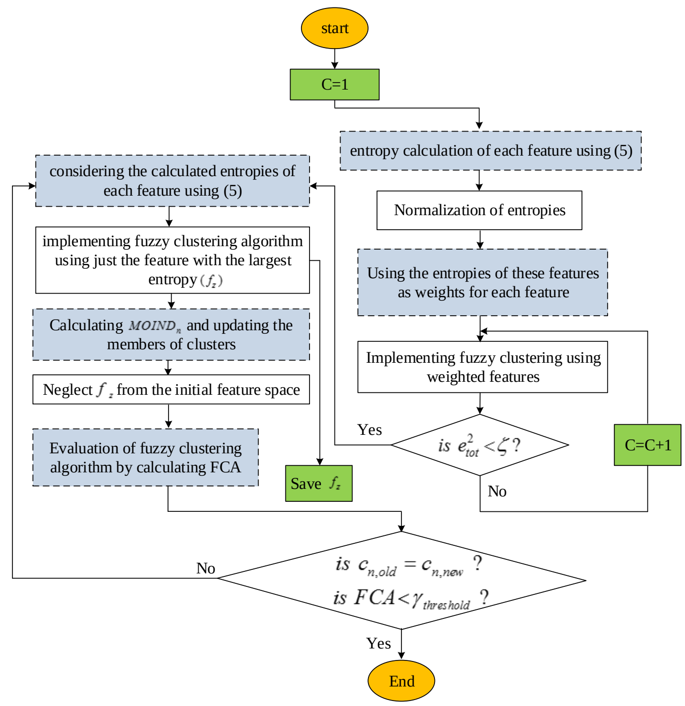

Fig. 2. Proposed Flowchart for the first phase of the feature selection.

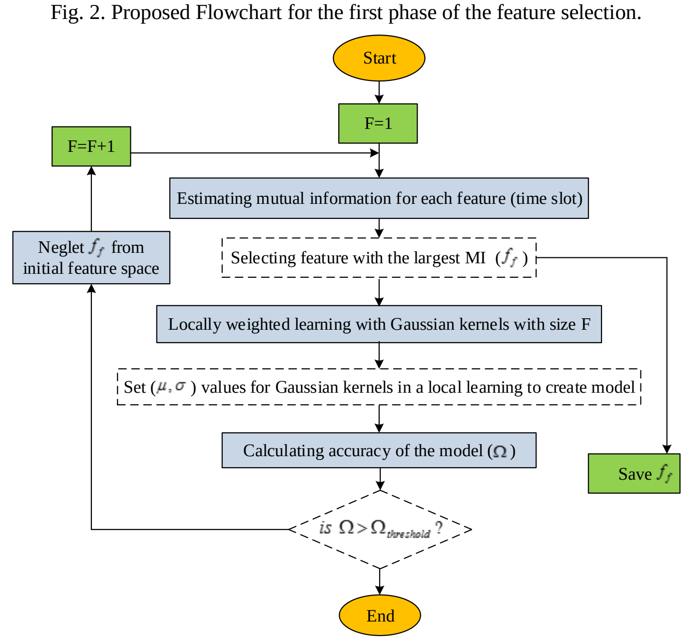

Fig.3. Proposed flowchart for the second phase of feature selection.

More detailed information about this algorithm is explained in [20] and neglected advertently for the sake of brevity. From another aspect, the customers' consumption pattern is undergone by season changes over the year. For this reason, identifying the most effective time slots is performed for two different periods. Differences of these periods are observable in the derived optimum time slots which is analyzed in section (V). In other words, time slots which are categorized as more weighty ones in the first period are not necessarily the same as the second period.

IV. Implementation and Results

Evaluation of the proposed procedures is performed using a real dataset related to 1200 anonymized Irish households [22]. These customers' consumption data are recorded in 30-minute intervals.

In the first phase of feature selection, the threshold value of $\xi = 0.05$ is considered after making trade-off by the assumption that each customer is assigned to a cluster with at least 90% dependency which means $m_{n,max} = 0.9$ in the FC literature. Fig.4 portrays the FCA index against different number of features in which $\gamma_{threshold} = 0.04$ is assumed as the threshold value of FCA. This threshold is the result of making trade-offs by the assumption that each customer is assigned to a cluster with at least 90% dependency like the prior step. By implementing the proposed procedure to the under-study dataset, the optimal number of clusters is derived 24 and the customers assigned to the clusters accurately using just 16 features not all 48 features, as shown in Fig.4.

In the second phase of feature selection, the MI values are estimated using the procedure explained in [20] and the normalized values of MI between all of the features and class labels are presented in Fig.5. These values are reported for two under-study periods ({spring, summer} and {fall, winter}). Moreover, Tables I and II report the time slots (features) with the largest MI with the size of 10 and 12 for two periods respectively and the correlation among these time slots. These correlation values corroborate the independencies among the selected time slots (features) which are expected to be the most effective ones for changes detection in customers' consumption. In such conditions, the intended aim in the second phase of feature selection is acquired and the needed features for analysis are reduced to 10 and 12 features instead of all 48 features in a 24 hours, as shown in Fig.5.

V. Discussions: Performance Evaluation

The authors believe that efficient change detection in consumption patterns is realized by analyzing two components of a times series; amount and trend. For this reason, two groups of indices are applied to measure the both of these components. The accuracy evaluation metrics for time series amount analysis and the similarity indices for trend analysis. More detailed information about these indices are described below.

A. Accuracy Evaluation

The accuracy of the proposed method for changes detection in customers' consumption pattern is measured and compared using the following indices; mean absolute percentage error (MAPE), normalized mean square error (NMSE), and relative error percentage (REP) where their formulations are presented in (9)-(11).

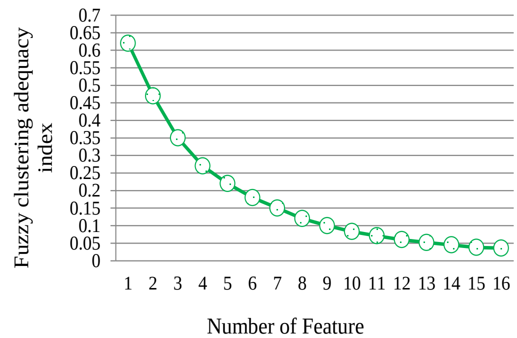

Fig.4. Fuzzy clustering adequacy index versus feature number.

Efficiency and advantage of the proposed method is analyzed using error indices and its performance in comparison with other methods is shown in Fig.6. These other methods are local support vector regression (SVR) [18] and locally weighted regression (LWR) [19]. SVR and LWR are two methods used in the case of load forecasting in which local weighting and regression are performed. For better insight and comparison, in addition to error indices, a new index (similarity index) is introduced in the next subsection and calculated for the proposed method in comparison with other methods. Figuare.9a portrays the total error in the form of box plot which is resulted by averaging the indicated error indices. Moreover, Fig.9b and 9c represent the results of changes detection and forecasted consumption trends for the same customer in two periods as a sample.

B. Similarity evaluation, using consumption pattern string (CPS) and distance distribution (DD)

The concentration of DR is mostly on the amount of energy which is consumed in some specific time intervals. Nonetheless, increase, decrease, i.e. the trend of this consumption from one time slot to another one is important, too. Actually, these increases and decreases compose a pattern which is basically rooted in customers' consumption behavior. To have a better insight, consider Figs. 7 and 8 exhibiting two groups of time series. The two time series of Fig.7 show dissimilar consumption patterns, neglecting the amount of consumption in each time interval. This condition is shown in Fig.8, yet with similar consumption patterns. The authors believe that customers' characterization without analyzing these patterns and trends

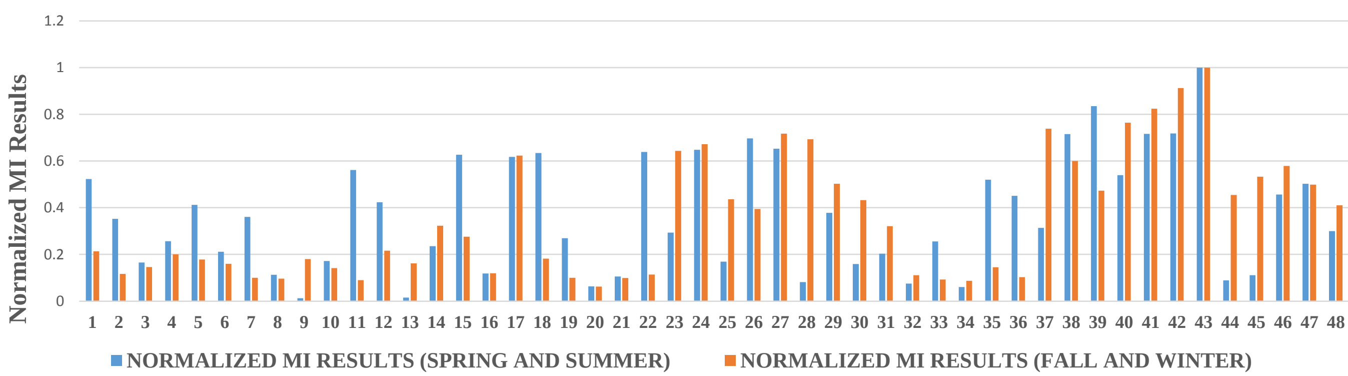

Fig.5. Normalized MI results for different seasons

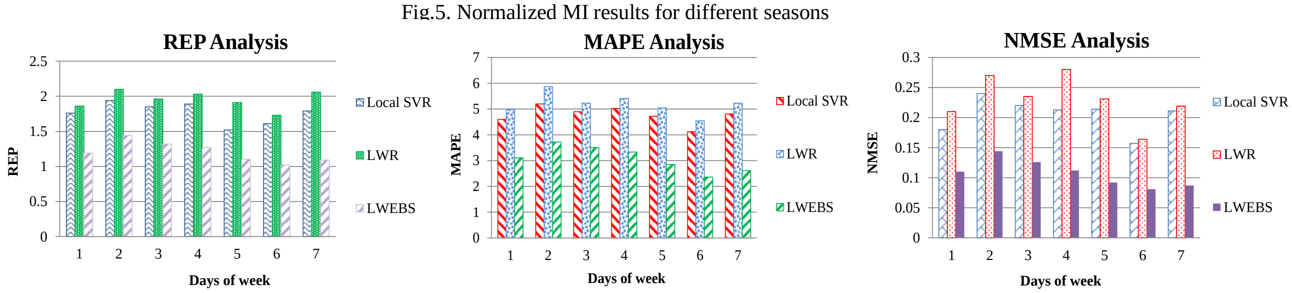

Fig.6. Error indices comparison for different methods

TABLE I

CORRELATION BETWEEN FEATURES (SPRING AND SUMMER)

Time interval	7-7:30 am	9:30-10 am	11-11:30 am	12-12:30 pm	13:30-14 pm	18-18:30 pm	19:30-20 pm	20-20:30 pm	20:30-21 pm	21-21:30 pm
7-7:30 am	1	0.0011	0.0009	0.0009	0.0014	0.0012	0.001	0.00303	0.00201	0.00202
9:30-10 am	0.0011	1	0.00105	0.00102	0.00301	0.0008	0.0008	0.0009	0.0006	0.001
11-11:30 am	0.0009	0.00105	1	0.003	0.0008	0.003	0.00203	0.00102	0.0008	0.0004
12-12:30 pm	0.0009	0.00102	0.003	1	0.00205	0.0016	0.0008	0.00202	0.004	0.00305
13:30-14 pm	0.0014	0.00301	0.0008	0.00205	1	0.0007	0.0007	0.0008	0.0009	0.00104
18-18:30 pm	0.0012	0.0008	0.003	0.0016	0.0007	1	0.00306	0.002	0.001	0.0009
19:30-20 pm	0.001	0.0008	0.00203	0.0008	0.0007	0.00306	1	0.0009	0.00203	0.0008
20-20:30 pm	0.00303	0.0009	0.00102	0.00202	0.0008	0.002	0.0009	1	0.00302	0.0012
20:30-21 pm	0.00201	0.0006	0.0008	0.004	0.0009	0.001	0.00203	0.00302	1	0.00105
21-21:30 pm	0.00202	0.001	0.0004	0.00305	0.00104	0.0009	0.0008	0.0012	0.00105	1

TABLE II

CORRELATION BETWEEN FEATURES (FALL AND WINTER)

Time interval	9-9:30 am	10-10:30 am	10:30-11 am	11:30-12 am	12:30-13 pm	13-13:30 pm	13:30-14 pm	18:30-19 pm	19-19:30 pm	20-20:30	20:30-21 pm	21-21:30 pm
9-9:30 am	1	0.0003	0.00204	0.0005	0.002	0.0007	0.00102	0.0006	0.00107	0.0005	0.0003	0.0105
10-10:30 am	0.0003	1	0.0021	0.0017	0.0008	0.00102	0.00103	0.0005	0.00107	0.0052	0.0017	0.0032
10:30-11 am	0.00204	0.0021	1	0.0002	0.0012	0.00406	0.007	0.00504	0.0004	0.00307	0.00102	0.00502
11:30-12 am	0.0005	0.017	0.0002	1	0.00309	0.0007	0.00202	0.0006	0.00604	0.0007	0.0025	0.00301
12:30-13 pm	0.02	0.0008	0.012	0.00309	1	0.0003	0.0003	0.0011	0.0009	0.0001	0.00206	0.001
13-13:30 pm	0.0007	0.00102	0.00406	0.0007	0.0003	1	0.0009	0.0006	0.0013	0.0045	0.00104	0.0019
13:30-14 pm	0.00102	0.0103	0.0007	0.00202	0.0003	0.0009	1	0.0002	0.0009	0.0008	0.00201	0.0001
18:30-19 pm	0.0006	0.0005	0.00504	0.0006	0.0011	0.0006	0.0002	1	0.00204	0.0007	0.0011	0.00203
19-19:30 pm	0.00107	0.00107	0.0004	0.00604	0.0009	0.0013	0.0009	0.0204	1	0.0002	0.0051	0.00108
20-20:30 pm	0.0005	0.0052	0.00307	0.0007	0.0001	0.0045	0.0008	0.0007	0.0002	1	0.00209	0.00101
20:30-21 pm	0.0003	0.0017	0.00102	0.0025	0.00206	0.00104	0.00201	0.0011	0.0051	0.0209	1	0.0036
21-21:30 pm	0.0015	0.0032	0.00502	0.00301	0.001	0.0019	0.0001	0.00203	0.00108	0.00101	0.0036	1

is rarely possible. Therefore, to scrutinize this issue and by inspiration of the points noted in [13], two indices are introduced, namely distance distribution (DD_i) and consumption pattern string (CPS_i) for each individual customer i .

$$MAPE=\frac{1}{N}\sum_{i=1}^{N}\frac{\left|x_{ACT_{i,t}}-x_{FOR_{i,t}}\right|}{x_{ACT_{i,t}}}\times 100 \tag{9}$$

$$NMSE=\frac{1}{\Phi^2 N}\sum_{i=1}^{N}(x_{ACT_{i,t}}-x_{FOR_{i,t}})^2$$
$$\Phi^2=\frac{1}{N}\sum_{i=1}^{N}(x_{ACT_{i,t}}-\overline{x}_{ACT_{i,t}})^2 \tag{10}$$

$$REP=\sqrt{\frac{\sum_{i=1}^{N}(x_{ACT_{i,t}}-x_{FOR_{i,t}})^2}{\sum_{i=1}^{N}x^2{}_{ACT_{i,t}}}}\times 100 \tag{11}$$

The DD_i index is a set of values composed of distances between two time series corresponding to an individual customer i in the break points. The break points of time series are shown in Figs. 7 and 8 by BP . Figs. 7b and 8b show the bar curves of distances between the two time series which results in DD_iaccording to (14). In Eq.14, SD is the operator for calculating standard deviation. Notice that the less DD means that customers' consumption pattern is approximately kept constant and with small changes. By using (12) and (13), consumption pattern string ($CPS_{i,[t_0,\dots,t_T]}$) is constructed in the time interval $[t_0,t_T]$. By calculating $G(t_m)$, 1 and -1 are respectively assumed for positive and negative signs of $G(t_m)$,. As such, a string composed of 48 records of 1 and -1 is provided for each customer i in 24 hours. Comparing the CPS_i of each customer i in two different time periods facilitates changes detection in customers' consumption pattern from one period to another. Tables III and IV report the derived CPS by comparing the time series in Fig.7a and 8a. Calculating CPS and DD simplify the customers characterization, however, the ultimate statement about changes detection is made utilizing similarity index (SI) which is defined as in Eq.15 using CPS and DD quantities. In Eq.15, $SI_{i.a,b}$ is the similarity index between the times series $a, b,$ both of them are related to the same customer. In other words, as an example, two time series for an individual customer is supposed to clarify the practicality of the similarity index. Notice to Figs.7 and 8, two groups of time series and corresponding CPS and DD values are reported. Numerical results show that DD values in Figs. 7 and 8 are 0.0122 and 0.0038, respectively. These DD values represent irregularities and regularities in the customers' consumption pattern, respectively. Actually, the smaller DD is more suitable, since it indicates more regularity in consumption behaviors. For more accurate changes detection in consumption, CPS should be intervened. Analyzing the reported CPS in the third row of Tables III and IV indicates that the percentage of changes in consumption behavior corresponding to the time series in Figs. 7 and 8 are 100% and 0%, respectively. CPS is derived by enumerating 0 and 1 records in the constructed string, and report that what percent of these records equals 1.

C. *Practical Insight*

Detecting changes in customers' consumption patterns performed and evaluated using the introduced procedures and indices. Table V presents a summarized report on the performance of the proposed procedure in comparison with the other methods for the same problem and the same data of this manuscript. According to the definition of the error and similarity indices, table (V) proves the advantage and the efficiency of the proposed method which confirms the main contribution of this manuscript which is detecting and forecasting changes in consumption patterns. Detecting changes in consumption patterns is the result of efficient customer characterization which is visualized as consumption forecasting. An important approach analyzed in this paper is detecting changes by two separated indices for measuring differences in values and trends of the time series. Such capability provides a flexible condition for grids analyzers to track the effect of these changes in trends and values on the grid separately. In general, the concept of detecting changes in customers' consumption patterns with the proposed procedure is a practical ability especially for the cases in the electricity network in which customers' role is embossed.

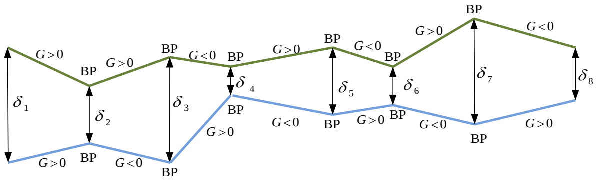

Fig.7a. Time series with dissimilar consumption patterns

TABLE III

THE DERIVED CPS BY COMPARING CONSUMPTION TIME SERIES

Upper curve	-1	1	-1	1	-1	1	-1
Down curve	1	-1	1	-1	1	-1	1
CPS	1	1	1	1	1	1	1

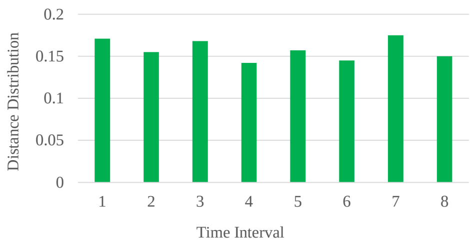

Fig.7b. Distances between two time series

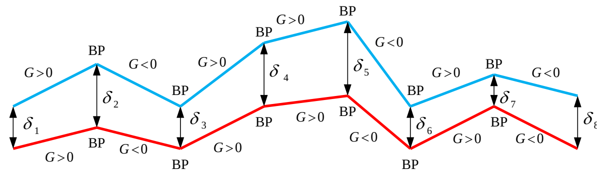

Fig.8a. Time series with similar consumption patterns

TABLE IV

THE DERIVED CPS BY COMPARING THE CONSUMPTION TIME SERIES

Upper curve	1	-1	1	1	-1	1	-1
Under curve	1	-1	1	1	-1	1	-1
CPS	0	0	0	0	0	0	0

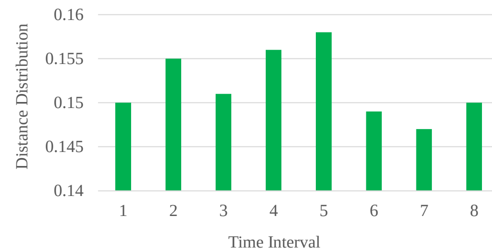

Fig.8b. Distances between two time series

$$G_{[t-1,t]} = \frac{x_{(t)} - x_{(t-1)}}{\Delta t} \tag{12}$$

$$CPS_{i,[t_0,\dots,t_T]} = [G(t_0),\dots,G(t_m),\dots,G(t_T)]$$
$$,G(t_m) = \begin{cases} 1 & if\ G(t_m) \ge 0 \\ -1 & if\ G(t_m) < 0 \end{cases} \tag{13}$$

$$DD_i = SD_i[(x_{t_{0,a}} - x_{t_{0,b}}),\dots\dots,(x_{t_{m,a}} - x_{t_{m,b}}),\dots\dots\dots,(x_{t_{T,a}} - x_{t_{T,b}})] \tag{14}$$

$$SI_{i,a,b} = {DD_i}/{CPS_i} \tag{15}$$

VI. Conclusion

This paper proposes an efficient procedure to detect changes in customers' consumption behaviors by utilizing an efficient subset of features (time slots). An adaptive algorithm is introduced for feature selection using LWEBS and MI with LWKDE algorithms. The accuracy of the forecasting and the similarity indices are recognized as the best metrics for evaluation of the proposed method, since the paper focuses on detecting and forecasting changes in the amount and the trends in customers' consumption patterns. All of the points reported in Fig. 9 justify the efficiency of the proposed method for changes detection. This conclusion is the result of a comparison shown in Table V by which the main contribution of the paper is justified. The results of Table V confirms the higher performance of the proposed method (LWEBS) comparing to the others. This comparison is observed in both accuracy (error indices) and trend detection (similarity index). In fact, the ability of the proposed algorithm in selecting the most efficient subset of features and optimum number of clusters leads to a relative advantage. The method for clustering customers is extended here to assign customers to the clusters accurately which results in better customers' characterization. From another aspect, introducing and applying the similarity indices representing the general shape of the consumption profiles helps with detecting changes be scrutinized and forecasted suitably. Comparing the values of the total error and normalized similarity indices corresponding to the proposed method and the other methods confirm the performance of the proposed method in detection of changes in the customers' consumption behavior. Extending other algorithms for detecting changes in customers' consumption patterns and analyzing its effects in the grid are considered as the future works.

Acknowledgement

The dataset used in the case study of this research, is prepared by Irish Social Science Data Archive (ISSDA), introduced by Dr. Mona Ghasemian of SBU, who is thankfully acknowledged.

TABLE V

GLOBAL COMPARISON OF DIFFERENT METHODS

Evaluation indices	LWR	Local SVR	LWEBS
Normalized similarity index	0.641	0.732	0.956
Total error	12%	10%	3.6%

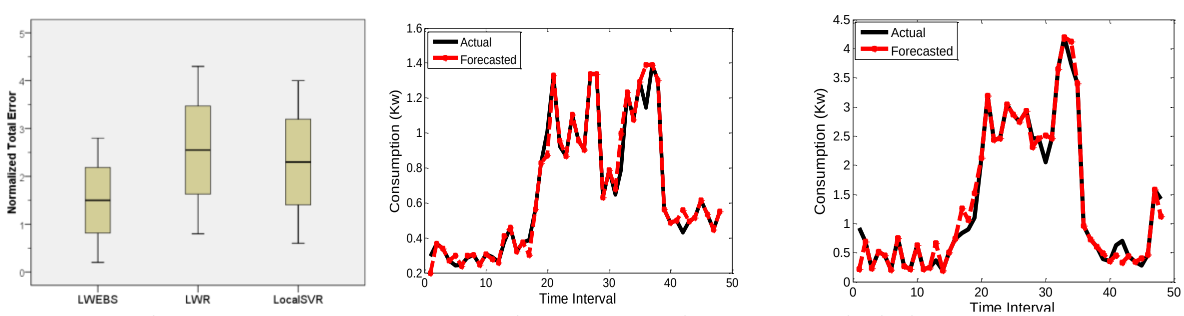

Fig.9. Distribution of normalized total error (9a), comparison of the actual and the forecasted load profile for fall and winter (9b), spring and summer (9c)

REFERENCES

[1] J. Kwac, J. Flora, and R. Rajagopal, "Household Energy Consumption Segmentation Using Hourly Data", IEEE Trans. on Smart Grid, vol. 5, Jan. 2014
[2] M.K.Naftchali, A.Fereidunian, H.Lesani, "Identifying Susceptible Consumers for Demand Response and Energy Efficiency Policies by Time-Series Analysis and Supplementary Approaches", In Proc. of the Iranian Conference on Electrical Engineering (ICEE), Shiraz, Iran, 2016.
[3] A. Baheti, D. Toshniwal, "Trend Analysis of Time Series Data Using Data Mining Techniques", International Congress on Big Data, Pages: 430-437, 2014.
[4] U. Mori, A. Mendiburu, and J. A.Lozano, "Similarity Measure Selection for Clustering Time Series Databaes", IEEE Transactions on Knowledge and Data Engineering, Vol. 28, pp. 181-195, Jan. 2016.
[5] B. Dai, R. Wang, K. Zhu, J. Hao, P. Wang, "A Demand Response Scheme in Smart Grid with Clustering of Residential Customers", IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Beijing, China, 21-23 Oct. 2019.
[6] Z. A. Khan, D. Jayaweera, "Approach for Smart Meter Load Profiling in Monte Carlo Simulation Application", IET Journals, Feb. 2017.
[7] S. Giannelos, I. Konstantelos, and G. Strbac, "Option Value of Demand-Side Response Schemes Under Decision-Dependent Uncertainty", IEEE Transaction on Power Systems, Jan. 2018.
[8] N. Amjady, F. Keynia, "Day-Ahead Price Forecasting of Electricity Markets by Mutual Information Technique and Cascaded Neuro-Evolutionary Algorithm". IEEE Trans, Power Systems, pp 306 – 318, Vol, 24, Feb. 2009.
[9] W. K. Lai, I. M. Khan, G. S. Poh, "Weighted Entropy-Based Measure for Image Segmentation", international symposium on robotics and intelligent sensors, pp. 1261-1267, 2012.
[10] X.Shi, and N.Celik, "Relative Entropy-Based Density Selection in Particle Filtering for Load Demand Forecast", IEEE Trans. Automation science and engineering, Vol. 14, pp. 946-954, 2017.
[11] W.Lao, Y.Wang, C.Peng, C.Y, and Y.Zhang, "Time Series Forecasting via Weighted Combination of Trend and Seasonality Respectively with Linearly Declining Increments and Multiple Sine Functions", Neural Networks (IJCNN), 2014 International Joint Conference on, Beijing, China, July 2014.
[12] M.Suntinger, H.Obweger, G.Raidle, "Trend-Based Similarity Search in Time-Series Data", International Conference on Advances in Databases Knowledge and Data Applications (DBKDA), 2010 Second, Menuires, France, Apr. 2010.
[13] G. A. Van, F. Staals, M. Loffler, J. Dykes, B. Speckmann, "Multi-Granular Trend Detection for Time Series Analysis", IEEE trans. visualization and computer graphics, pp. 661-670, Vol. 23, 2017.
[14] E. W. S. dos Angelos, O. R. Saavedra, O. A. C. Cortés and A. N. DeSouza, "Detection and Identification of Abnormalities in Customer Consumptions in Power Distribution Systems," IEEE Trans. on PWRD., vol. 26, no. 4, pp. 2436-2442, Oct. 2011.
[15] J. Han, M. Kamber, "Data Mining Concepts and Techniques" 2nd edition, Morgan Kaufmann publisher. 2006
[16] T. M. Cover, J. A. Thomas, "Elements of Information Theory", 2nd edition, Wiley, September, 2006
[17] K. L. Wu, M. S. Yang, "A Cluster Validity Index for Fuzzy Clustering", pattern recognition letters, pp. 1275-291, 2005
[18] E. E. El-Attar, J. Y. Goulermas, and Q. H. Wu, "Forecasting Electricity Daily Load Based on Local Prediction", in proc. IEEE Power Eng. Soc. Gen. Meeting (PESGM), Calgary, Canada, Jul.26-30, pp. 1-6, 2009.
[19] W. S. Cleveland and S. J. Devlin, "Locally Weighted Learning", J. Amer, statist. Assoc, vol. 83, no. 403, pp. 596-610, Sep. 1988.
[20] S. Gao, G. V. Steeg, A. Galstyan, "Estimating Mutual Information by Local Gaussian Approximation", 31th conference on uncertainty in artificial intelligence (UAI), 2015.
[21] H. Zhang, J. E. Fritts, and S. A. Goldman, "An Entropy-based Objective Evaluation Method for Image Segmentation", in Proc of the SPIE, Vol. 5307, pp. 38-49, 2003
[22] Irish Social Science Data Archive (ISSDA), retrieved in November 2017 from: http://www.ucd.ie/issda/data/
[23] S. Lin, F. Li, E. Tian, Y. Fu, D. Li, "Clustering Load Profiles for Demand Response Applications", IEEE Transactions on Smart Grid, V. 10, March 2019.
[24] E. W. S. Angelos, O. R. Saavedra, O. A. C Cortés ; A N. d Souza, "detection and identification of abnormalities in customer consumptions in power distribution systems", IEEE Transactions on Power Delivery, V.26, Oct. 2011.
[25] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, J. Hu, "Short-Term Load Forecasting with Deep Residual Networks", IEEE Transactions on Smart Grid, V. 10, July 2019.
[26] J.Luo, T. Hong, S-C. Wang, "Robust Regression Models for Load Forecasting", IEEE Transactions on Smart Grid, V. 10, Sept. 2019.
[27] W. Kong, Z. Y. Dong, Y. Iia, D. J. Hill, Y.Xu, Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network", IEEE Transactions on Smart Grid, V. 10, Jan. 2019.
[28] Y.Wang, N. Zhang, Y. Tan, T. Hong, D. S. Kirschen, C. Kang, "Combining Probabilistic Load Forecasts", IEEE Transactions on Smart Grid, V.10 , PP. 3664 - 3674 , July 2019.
[29] M. Charwand, M. Gitizadeh, P. Siano, G. Chicco, Z. Moshavash, "Clustering of Electrical Load Patterns and Time Periods Using Uncertainty-Based Multi-Level Amplitude Thresholding", Elsevier, Electrical Power & Energy Systems, V. 117, May 2020.
[30] Y.Ji, E. Buechler, R. Rajagopal, "Data-Driven Load Modeling and Forecasting of Residential Appliances", IEEE Transactions on Smart Grid, December 2019.