2017

Identification and characterization of irregular consumptions of load data

Desh Deepak Sharma
Indian Institute of Technology, deshdeepak101@gmail.com

S. N. Singh
PJM Interconnection, snsingh@iitk.ac.in

Jeremy Lin
PJM Interconnection, Jeremy.Lin@pjm.com

Elham Foruzan
University of Nebraska-Lincoln, elham.foruzan@huskers.unl.edu

CrossMark

Identification and characterization of irregular consumptions of load data

Desh Deepak SHARMA[1], S. N. SINGH[2], Jeremy LIN[2], Elham FORUZAN[3]

MPCE

Abstract The historical information of loadings on substation helps in evaluation of size of photovoltaic (PV) generation and energy storages for peak shaving and distribution system upgrade deferral. A method, based on consumption data, is proposed to separate the unusual consumption and to form the clusters of similar regular consumption. The method does optimal partition of the load pattern data into core points and border points, high and less dense regions, respectively. The local outlier factor, which does not require fixed probability distribution of data and statistical measures, ranks the unusual consumptions on only the border points, which are a few percent of the complete data. The suggested method finds the optimal or close to optimal number of clusters of similar shape of load patterns to detect regular peak and valley load demands on different days. Furthermore, identification and characterization of features pertaining to

unusual consumptions in load pattern data have been done on border points only. The effectiveness of the proposed method and characterization is tested on two practical distribution systems.

1 Introduction

During the last few decades, there has been a major shift from the vertically integrated monopolistic system to the open power market system. The restructuring of electricity supply industry has created many new challenges in providing the secure, stable and economical electric power to the end users [1–3]. The electric prices vary significantly during the day due demand variations. To overcome the peaking problems, the demand response programs are suggested under the smart grid initiatives [4–6]. Under demand response scheme, customers reduce the electrical load demand during the peak-price period by rescheduling the demand for low-price periods [4–9]. Peak clipping, valley filling and load shifting are key tools of demand response [9].

Power operators are concerned about irregular behavior of electricity consumption in their decision making process. In the load profile data, abnormal consumptions may happen due to measurement error, undetected consumption, illegal electricity connection, improperly installed equipment, etc. [10–14]. Clustering of load profiles helps in developing working methodology for energy losses (technical and commercial) evaluation [10–13]. For peak shaving and distribution system upgrade, it is very essential

✉ Desh Deepak SHARMA
deshdeepak101@gmail.com; ddsharma@iitk.ac.in

S. N. SINGH
snsingh@iitk.ac.in

Jeremy LIN
Jeremy.Lin@pjm.com

Elham FORUZAN
elham.foruzan@huskers.unl.edu

[1] Indian Institute of Technology, Kanpur, Kanpur, India

[2] PJM Interconnection, Audubon, PA, USA

[3] Department of Electrical Engineering, University of Nebraska–Lincoln, Lincoln, NE, USA

 Springer

to know the changes in loading at the substations i.e the consumption behavior of customers. At the peak load, the power losses in different feeders and different transformers are to be estimated [15]. This will provide fair calculation of network pricing.

Data mining and artificial intelligence techniques such as support vector machines [11], fuzzy clustering [12], etc. are explored in the identification of irregularities in energy consumption. A comparison of a load profile is done with standard or average load profile to identify the abnormal consumption [12, 16]. Extensive experimental testing was carried out in [17] for selection of parameter values such as the sensitivity threshold to detect anomalous events, maximum cluster radius for the nearest neighbor cluster method and parameter used for fuzzy rule extraction based on identified clusters.

Different authors, in their research works, discussed various methods of classification of the electrical consumption data [14, 18–28]. These methods can facilitate development of different types of demand response strategies and improvement of grid reliability. For different customers, the representative load patterns (RLPs) are obtained and these are clustered on the basis of RLPs [26, 27]. The customers of each cluster will have same load pattern and thus, TLP (typical load profile) of each customer of a group is a centroid of that cluster [27]. Based on similar electrical consumption behavior, classical k-means [23, 27], fuzzy c-means [23, 27], hierarchical clustering, self-organizing feature maps (SOFM) [23, 27], principal component analysis (PCA) [23], curvilinear component analysis (CCA) algorithms [23], ant colony clustering (ACC) [28], support vector clustering (SVC) [26], etc. have been suggested for the classification of electrical load profile data. Different comparison methods such as clustering dispersion indicator, Davies-Bouldin indicator, stability index are utilized for cluster validity assessment [23, 27].

A data object is characterized by a set of similarity or dissimilarity measures which are described by distance function. Various clustering algorithms have been applied in separating the data object into different clusters while employing distance function. Major clustering methods which are applied in classification of data are partition based, hierarchical (agglomerative and divisive) clustering, neural network based, density based, grid based, model based, etc. [29–31]. Partitioning algorithms (k-means, fuzzy c-means, etc.) applied in clusters of load data need a number of clusters as input data. In the hierarchical clustering algorithm, dendrogram is created from the leaves up to the root (agglomerative approach) or from root down to leaves (divisive approach) with merge or divide operation in each iteration. A termination criterion is required to stop the iterations [23]. In an ant colony clustering (ACC)

concept, a specified number of clusters are required as input or number of clusters is defined in post-processing phase. In an iterative process of ACC, an initialization phase requires a number of clusters and number of ants in ending phase, a stopping criterion is to be defined [28]. In support vector clustering (SVC), the final clusters are obtained in post-processing phase, which is computationally intensive [26].

The ISODATA algorithm, which includes temperature dependency and outlier filtering, is proposed in [32] for customer classification. For the classification of load profiles, the Gaussian mixture model is used in assigning the labels, only, to the most recurrent load profiles [33]. Inter-cluster behaviour classification model and intra-cluster consumption volume prediction model are constructed using agglomerative hierarchical clustering algorithm [34]. In density based clustering algorithm, random initialization of any parameter is not required. Therefore, after setting the global parameters heuristically, similar results are obtained in each iteration and hence, consistency of the algorithm is preserved. There are different variants of density based algorithm available in the literature. Density based spatial clustering of applications with noise (DBSCAN) is one of the most popular density based algorithms being used in data mining [29–31].

Most of the clustering algorithms require an iterative control strategy to optimize the objective function and random initialization of some parameters. Thus, clustering results vary with different iterations. Selection of appropriate number of clusters is another tedious task in implementing these algorithms. The problem in implementation of DBSCAN is selection of global parameter while k-means and fuzzy c-means are based on iterative control scheme. Generally, statistical methods are used to identify the outliers and these methods are based on fixed probability distribution of data. However, the real time information is not fixed to any distribution. Further, all the irregular consumption detection methods work on whole load pattern data set. Outlier detection approaches based on k-means and fuzzy c-means approaches finds variation of data object from the centroid. The main problem with existing density based clustering algorithm is that intrinsic cluster structures cannot be detected by global density parameters. Different local densities are to be revealed to find local clusters in the data space with further partition [29, 35].

Motivated by aforementioned facts, in this paper, a new method, which is suitable for clustering and identifying the unusual electricity consumptions and their quantification according to the nature of irregularity, is proposed. The proposed method utilizes the concept of Local Outlier Factor (LOF) [36, 37] for ranking of unusual consumptions based on neighborhood densities i.e. k- nearest neighbors

(k-NNs) of these consumptions in the load pattern data. Clustering results are compared with k-means and fuzzy c-means with clustering validation using Davies Bouldin index and Silhouette coefficient.

The major contributions of this paper:

1) A method is proposed to obtain global density parameters in order to find an optimal partition of a data set into high and low density regions. The low density regions are known as border points which are a little part of whole load data and utilized to find irregular loading on distribution substations. Hence, computation work is highly reduced in identification of only irregular demand.

2) Micro clusters are obtained to reveal local clusters and, hence, further partition of the data set is avoided. Core points in load data help in analyzing the occurrence of a peak-valley in load pattern.

3) Furthermore, an approach to characterize and quantify the different features of unusual consumptions using feature irregularity factor (FIF) is introduced on only border points of load data. It identifies the irregularities in unusual consumption based on different irregularity features. This approach is scalable as different irregular features of unusual loadings on substations can be identified and added to decide FIF of different unusual consumptions. The suitability of the proposed method is demonstrated on two practical distribution systems.

2 Clustering methods

2.1 k-means

Classical k-means algorithm is a partition based clustering algorithm which separates a set of n data objects into k clusters based on similarity features. Given a set of n-number of observations, each observation is a d-dimensional real vector. This observation set is partitioned into k sets ($k < n$), while an objective function is minimized. Each set represents a cluster of data [29, 30].

2.2 Fuzzy c-means

In fuzzy c-means clustering, each data object is assigned to different clusters with different degrees of membership. Thus, membership of a data object is shared among different clusters. This algorithm tries to find the best partition of whole data while minimizing an objective function [29, 30].

2.3 Density based clustering

This algorithm separates high density and low density regions. A data point belongs to a cluster if its neighborhood density is high enough. Clusters get arbitrary shape while absorbing all the data points, those are in the neighborhood. Densities of all the clusters may be different. The classical density based spatial clustering of applications with noise (DBSCAN) forms clusters such that each data point in a cluster should consists of at least a minimum number of points (N_{minpts}) in its neighborhood defined by a given radius (r_{eps}). It means that the cardinality of the neighborhood has to exceed a threshold [29–31].

3 Local outlier factor (LOF)

LOF is density based outlier detection method [36, 37] in which the ratios between local density of data object and local density of the data objects' neighborhood are obtained. An outlier is defined based on the density of data objects existing in its neighborhood. A comparison of the density of each object with the density of its k-NNs is to be done. The local density of an outlier is relatively low compared to the local density of other data objects around its neighboring objects. In this approach, each data object can be represented by an outlying factor as per their nature of anomalies. If the value of LOF of a data object is higher, it means that there is a large change in densities of the object and its k-NNs. If the value of LOF of a data object is approximately equal to 1, the data object is close to dense region and not to an outlier [36, 37].

3.1 k-distance

Basically, it is the distance between an object under consideration and its k-th nearest neighbor. Let D is whole data set; $z \in D$ is the k-th nearest neighbor of $x \in D$ and $L_{dist}(x, z)$ is the distance of x to object z. The k-distance of x is written as

$$L_{dist,k}(x) = L_{dist}(x, z) \tag{1}$$

where D_x is the set of k-th closest objects to $x \in D$, then the distance of x to $o \in D_x$ is $L_{dist}(x, o) \leq L_{dist,k}(x)$ while $D_x \subseteq D$. Euclidean distance is considered for distance measurement.

3.2 k-distance neighborhood of x

The k-distance neighborhood of object x consists of k-th nearest neighbors i.e. objects whose distances from x are

less than or equal to k-distance of x. k-distance neighborhood of x is defined as

$$N_k(x) = \left\{ \forall o \in D_x | L_{\text{dist}}(x, o) \leq L_{\text{dist},k}(x), D_x \subseteq D \right\} \quad (2)$$

3.3 Reachability distance of x with respect to z

The reachability distance is an asymmetric measure. The reachability distance is used to find the density of k-nearest neighborhood of an object. The reachability distance of an object x with respect to object z is given as

$$L_{\text{dist},k}^{\text{reach}}(x, z) = \max\left\{ L_{\text{dist},k}(z), L_{\text{dist}}(x, z) \right\} \quad (3)$$

It maintains minimal distance between two objects x and z while object x is kept outside the neighborhood of z. If x is not close to z, then the reachability distance is simply the distance between x and z i.e. $L_{\text{dist}}(x, z)$. If x is very close to z then the reachability distance is k-distance of z i.e. $L_{\text{dist},k}(z)$.

3.4 Local reachability density of x

The local reachability density of x represents the density of its neighborhood. It is defined as the reciprocal of average reachability distance of k-distance neighborhood of x. If $|N_k(x)|$ is the number of objects in k-distance neighborhood of x, then the local reachability density of x is given as

$$R_{\text{lrd},k}(x) = \frac{|N_k(x)|}{\sum\limits_{z \in N_k(x)} L_{\text{dist},k}^{\text{reach}}(x, z)} \quad (4)$$

3.5 Local outlier factor of x

Basically, local outlier factor is the average of the ratio of local reachability densities of objects in k-distance neighborhood of x to the local reachability density of x itself and given as

$$LOF_k(x) = \frac{\sum\limits_{z \in N_k(x)} \frac{R_{\text{lrd},k}(z)}{R_{\text{lrd},k}(x)}}{|N_k(x)|} \quad (5)$$

The strength of reachability distance depends on positive integer k. The higher value of k ensures more stable results, but the burden of computation increases.

4 Outlier detection methods and problem assessments

"An outlier is an observation which deviates largely from the other observations as to arouse suspicions that it was generated by a different mechanism." Abnormalities, discordants, deviants, irregularities, or anomalies are the other terms used for outliers. Different basic models, such as extreme value analysis, probabilistic and statistical models, linear models, proximity-based models, information theoretical models, high dimensional outlier detection models, are used for detection of outliers in the data. These models are used depending on the type of the available data observation set. These algorithms are having pros and cons in the detection of outliers [38]. The objective of the outlier detection method is to identify data objects which are markedly different from or inconsistent with the normal set of data. The advantages and disadvantages of clustering based, nearest neighbour based, classification based and spectral anomaly detection techniques are discussed in [35]. It is shown that computational complexity is a big issue and most of the anomaly detection techniques are computationally expensive [35, 38]. These techniques work on the whole of the data observation set in the detection of anomalies. In this paper, a method is proposed for an optimal partition of load data into core points and border points. Irregular consumptions are part of border points.

Accurate selection of two global parameters $r_{\text{eps},o}, N_{\text{minpts},o}$ is to be done as per (6). Data point with LOF less than 1.0 is a part of the cluster. Possessing at least one LOF, of a data point, nearly equal to 1.0 but greater than 1.0, ensures that all less dense data points and outliers are included in border points. Thus, with this appropriate set of global parameters, it is ensured that all the high dense points are separated from the less dense points [36, 37], and clustering operation is performed on high dense points only and LOFs are computed of less dense points. Following equation is formulated for sub-optimal partition of whole load data into high and less dense regions.

$$(r_{\text{eps},o}, N_{\text{minpts},o}) = \begin{cases} (r_{\text{eps}}, N_{\text{minpts}}) | \min f_{\text{LOF}}, \\ \text{where } f_{\text{LOF}} = LOF(l_b^{lr}) - 1, \\ LOF(l_b) > 1.0, \ l_b, l_b^{lr} \in B, B \subseteq \Omega \end{cases}$$
$$(6)$$

where B is a set of border points l_b; Ω is the complete load pattern data; l_b^{lr} is the border point with lowest rank with LOF in B; $l \in \Omega$ is data point (a load profile) and $l_c \in \Omega$ is a core point in load data.

5 Proposed method for identification of unusual consumptions and clustering

The proposed method, which acquires the basic concept of density based clustering, focuses on the core points for clustering purpose and border points for the identification of outlier. The LOFs are computed for only border points. So, all the border points are quantified with LOF according to their outlying nature. In the method, there is no

STATE GRID
STATE GRID ELECTRIC POWER RESEARCH INSTITUTE

consideration of any defined distribution of data to isolate the irregular consumption while assigning the degree of being irregular as LOF in load pattern data. Computation of LOF is done only on border points which are a few percent of whole load pattern data. No iterative control scheme is required for optimal or close to optimal clustering results obtained on a practical system [39, 40] by the proposed method. Although the method can find optimal clusters, but appropriate clusters are obtained in each zone in order to find distinguishable peaks and valleys for peak load clipping, load shifting. Heuristically, it is found that clustering, which produces distinguishable peak and valley, is validated as optimal clustering or close to optimal.

5.1 Distance matrix

Euclidean distance is considered to measure the closeness of data objects (load profiles). The distance between n-dimensional two data objects l_i and l_j is described as given below

$$L_{\text{dist}}(l_i, l_j) = \sum_{k=1}^{n} ||l_i^k - l_j^k|| \tag{7}$$

$$e_{ij} = L_{\text{dist}}(l_i, l_j) \tag{8}$$

A distance matrix represents the closeness of data objects and this matrix is a square matrix and its dimension is $N \times N$ where N is number of data objects. Diagonal elements such as e_{11}, e_{22}, …, e_{NN} are always zero. The scaling of the distance matrix is carried out by dividing all elements of distance matrix by a scaling factor if required.

5.2 Solution to obtain global parameters

The parameters $r_{\text{eps},o}, N_{\text{minpts},o}$ for a sub-optimal partition of load data can be obtained as given below:

1. Set arbitrarily $r_{\text{eps}}, N_{\text{minpts}}$ to generate set B;
2. Tune $r_{\text{eps}}, N_{\text{minpts}}$ to $r_{\text{eps},o}, N_{\text{minpts},o}$ to satisfy (6).

5.3 Generation of small clusters

Small cluster is formed from arbitrarily selected root core point and its direct density reachable core points at depth one. So, a small cluster, C_{sc}, is formed according to following theorem [31].

Theorem *If x_i is core point and $x_i \in C_{sc}$ then $x_j \in C_{sc}$ if x_j is core point and it is direct density reachable from x_i.*

5.4 Operation of merging small clusters

Two or more than two smaller clusters are merged into a single cluster such that the maximum deviation of averages

of these small clusters at any dimension is less than a threshold.

Consider $\{C_{sc}^1, C_{sc}^2, \ldots, C_{sc}^m\}$ is a set of small clusters of given n-dimensional data and $\{v_1, v_2, \ldots, v_m\}$ is set of averages (centroids) of these small clusters. Maximum deviation of two averages at any dimension is defined as given below.

$$\theta_{ij} = \max_{q=1:n} |v_i^q - v_j^q| \tag{9}$$

$\theta^{\max}$ and $\theta^{\min}$ are maximum and minimum values of θ_{ij} among all small clusters obtained as

$$\theta^{\max} = \max_{i,j=1:m, i \neq j} (\theta_{ij}) \tag{10}$$

$$\theta^{\min} = \min_{i,j=1:m, i \neq j} (\theta_{ij}) \tag{11}$$

Suppose K is the number of clusters as $\{c_1, c_2, \ldots, c_K\}$ after merging small clusters. $\theta = \theta_1$ can be set such that all small clusters are merged into single cluster i.e. $K = 1$.

$$\theta_1 = \{\theta^{\min} < \theta < \theta^{\max} | K = 1\} \tag{12}$$

$\theta = \theta_m$ can be set such that no small cluster is merged. In this case, the number of clusters K is equal to the number of small clusters m. Obviously, the number of clusters obtained, after merge operation, is less than number of small clusters i.e. $\forall K \leq m$.

$$\theta_m = \{\theta^{\min} < \theta < \theta^{\max} | K = m\} \tag{13}$$

θ_K can be set for K number of clusters as

$$\theta_K = \{\theta_1 < \theta < \theta_m | 1 \leq K \leq m\} \tag{14}$$

θ_K^o is the value of θ such that optimal number of clusters K^o is found. So θ_K^o is defined as given below.

$$\theta_K^o = \{\theta_1 < \theta < \theta_m | K = K^o\} \tag{15}$$

5.5 Assigning non-outliers to clusters

Border points which are having LOF approximately equal to 1.0 are located close to a homogeneous dense region and these may be part of any cluster through density reachable and density connected concepts. Higher values of LOF of points show that there is a large difference in the densities of these points and their k-nearest neighbors and hence, these points are considered to be outliers [36, 37].

A limiting value U_{LOF} is considered for LOF in order to define set of outliers, Ω_U, out of border points B as given below.

$$\Omega_U = \{l_b \in B | LOF(l_b) > U_{LOF}\} \tag{16}$$

The Ω_U is, obviously, set of unusual consumptions. Assume $\{c_1, c_2, \ldots, c_K\}$ is the set of clusters then the

border points which are not designated as outlier can be assigned to a cluster via following way.

$$l_b \in B := \left\{ l_b \in C_i \Big| \max_{i=1:K} N_{kNN}^{C_i}(l_b) \right\} \tag{17}$$

where $N_{kNN}^{C_i}(l_b)$ is the number of k-nearest neighbors of point l_b in cluster c_i [30].

5.6 Proposed method

Figure 1 shows the flow chart for finding the clusters and outlier with the proposed method. The steps of the proposed method are described as:

1) Get the database to find ranked outliers and clusters.
2) Select the proper distance function and obtain a distance matrix.
3) Find global parameters $r_{\text{eps},o}, N_{\text{minpts},o}$ as per section 5.2.
4) Construct small clusters with core points.

Fig. 1 Flow chart for finding clusters and LOF of unusual consumptions

5) Repeat the process of step 4 to obtain other small clusters until all of the remaining core points are visited.
6) Merge small clusters into clusters with variation in threshold θ to obtain the optimal number of clusters.
7) Compute LOF for each border point and consider a limiting value for LOF to isolate ranked outliers from border points.
8) Merge the non-outliers border points to clusters.

6 Proposed characterization of unusual consumptions

The electricity consumptions, which are different from regular electricity consumptions, are to be analyzed. Different types of peak demand, sudden large change and zero demand are some irregular consumption. These irregular consumption behaviors are defined below on only set Ω_U.

6.1 Irregular peak unusual consumption

Irregular peak unusual consumption U_{irpeak} is defined as

$$U_{irpeak} = \left\{ l_b \in \Omega_U \big| \exists t : \Delta d_{irpeak}^t > \Delta d_{ref,a} \right\} \tag{18}$$

$$\Delta d_{irpeak}^t = d^t(l_b) - d_{ref}$$

$$\Delta d_{ref,a} = d_{peak,a} - d_{ref}$$

where $d^t(l_b)$ is the demand of a load data point (a load profile) l_b at time interval $t \in T$. d_{ref} is the reference demand and the demand which is more than d_{ref} is termed as peak demand. $d_{peak,a}$ is an acceptable peak demand in the system. $\Delta d_{ref,a}$ is a predefined acceptable change in demand more than d_{ref} to decide irregular consumption.

6.2 Broadest peak demand

Broadest peak demand U_{bpeak} is an unusual consumption as defined below. The demand in U_{bpeak} is more than d_{ref} for some consecutive time intervals τ_{peak} and n_{peak} is the cardinality of τ_{peak}.

$$U_{bpeak} = \left\{ l \in \Omega_U \big| (d^t(l) - d_{ref}) > 0, t \in \tau_{peak} \right\} \tag{19}$$

6.3 Sudden large gain unusual consumption

Sudden large gain unusual consumption, U_{sgain}, is the amount of increase in demand more than δ_a^g which is an acceptable gain in demand at any time interval $t \in T$.

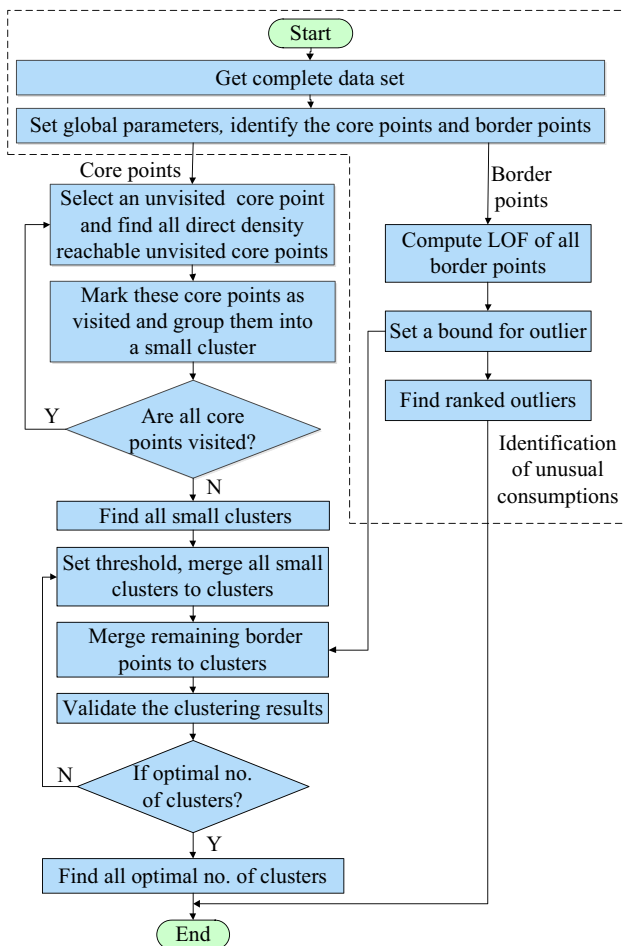$$U_{sgain} = \left\{ l \in \Omega_U \big| \exists t : \Delta d_{gain}^t > \delta_a^g \right\} \tag{20}$$

STATE GRID
STATE GRID ELECTRIC POWER RESEARCH INSTITUTE

$$\Delta d^t_{gain} = d^t(l) - d^{t-1}(l)$$

Similar to U_{sgain}, sudden large drop unusual consumption U_{sdrop}, is the amount of decrease in demand more than δ^d_a which is an acceptable drop in demand at any time interval $t \in T$.

$$U_{sdrop} = \left\{ l \in \Omega_U | \exists t : \Delta d^t_{drop} > \delta^d_a \right\} \quad (21)$$

$$\Delta d^t_{drop} = d^{t-1}(l) - d^t(l)$$

6.4 Nearly zero demand unusual consumption

Nearly zero electricity demand unusual consumption, U_{zero}, is the demand, which remains a very low value equal to zero at any time interval $t \in T$ or for some duration of time intervals.

$$U_{zero} = \{ l \in \Omega_U | d^t(l) = 0 \text{ and } t \in \tau_{zero} \} \quad (22)$$

where τ_{zero} is a set of time intervals on which demand is zero and n_{zero} is the cardinality of τ_{zero}. Based on aforementioned definitions, vector of features of unusual consumptions is defined as

$$Y_U = (\Delta d^t_{irpeak}, n_{peak}, \Delta d^t_{gain}, \Delta d^t_{drop}, n_{zero}) \quad (23)$$

$$\Delta d^t_{irpeak} > \Delta d_{ref}, \Delta d^t_{gain} > \delta^g_a, \Delta d^t_{drop} > \delta^d_a$$

To identify the degree of irregularity in unusual consumptions, feature irregularity factor (I_{FIF}) is introduced and defined below:

$$I_{FIF} = ||Y_U|| \quad (24)$$

Each feature of different unusual consumptions, in vector Y_U is normalized by min-max or z-score normalization method. In an unusual consumption, it is possible that more than one unusual characteristic may present. From a row of unusual consumption in Y_U, the most dominating unusual characteristics can be identified. Limiting values in I_{FIF} directly relate to real unusual behaviors of outliers. Once limiting values are decided, feature vector and hence, I_{FIF} of an unusual consumption are decided.

7 Case studies

The proposed method to identify unusual consumptions and to find clustering results for peak valley analysis is tested on the two practical systems. Regular peaks and valleys are identified with clustering results obtained from proposed approach in order to distinguish irregular peaks in the load pattern data. The proposed characterization of unusual consumptions has also been carried out. The 365 days are numbered as day 01 is Jan 01, similarly day 365 is

Dec 31 and so on. To validate the clustering of load pattern data, two most popular methods such as the Davies-Bouldin index (DBI) and Silhouette coefficient (SC) are used. Davies-Bouldin criterion depends on a ratio within the cluster and between cluster distances [25, 27, 30, 31]. The Silhouette coefficient criterion incorporates two approaches: cohesion and separation. Cohesion measures closeness of objects in a cluster and separation finds whether the clusters are well-separated or not [30, 31].

7.1 Case study-1

The effectiveness of the proposed method tests on a practical system of 20 zones [39, 40]. The data are annual hourly loaded (in kW) for US utility with 20 zones of year 2007. In most of the zones, the electricity consumption data are given in the range of thousands of kW. Therefore, the distance matrix is required to be scaled down. For each zone, the distance matrix is divided by the suitable divisor (scaling factor such as $10^3, 10^4$, etc.) so that elements of distance matrix are in the range of 10.

Different notations are used in Table 1 as z_i is Zone-id; $\min^f_D, \min^k_D$ are minimum value of DBI with fuzzy c-means and k-means respectively; $\max^f_S, \max^k_S$ are maximum values of silhouette coefficient with fuzzy c-means and k-means respectively; $N^f_{o,D}, N^k_{o,D}$ are optimal numbers of

Table 1 Optimal number of clusters with fuzzy c-means and k-means

z_i	$\min^f_D$	$N^f_{o,D}$	$\max^f_S$	$N^f_{o,S}$	$\min^k_D$	$N^k_{o,D}$	$\max^k_S$	$N^k_{o,S}$
1	0.7370	3	0.6632	3	0.6787	5	0.6680	3
2	0.7456	5	0.6231	2	0.7523	3	0.6283	2
3	0.7672	3	0.6231	2	0.7364	3	0.6283	2
4	0.6826	2	0.7444	2	0.6804	2	0.7500	2
5	0.7286	4	0.6156	4	0.6571	4	0.6337	4
6	0.7506	3	0.6204	2	0.7138	5	0.6231	2
7	0.7731	5	0.6231	2	0.7883	5	0.6283	2
8	0.7517	3	0.6618	2	0.7998	2	0.6767	2
9	1.2000	2	0.4844	4	0.9411	3	0.5363	4
10	0.7628	3	0.6889	2	0.7627	5	0.6984	2
11	0.6843	4	0.6418	2	0.6802	5	0.6520	2
12	0.6844	4	0.6240	2	0.7024	5	0.6514	3
13	0.8385	4	0.6411	2	0.7734	4	0.6576	2
14	0.7513	5	0.6418	3	0.7953	3	0.6450	3
15	0.7988	4	0.6244	3	0.7619	5	0.6348	3
16	0.7710	4	0.6133	3	0.6975	5	0.6305	3
17	0.7547	4	0.6559	3	0.7515	5	0.6627	3
18	0.7700	5	0.6378	3	0.6856	5	0.6418	3
19	0.7554	5	0.6499	3	0.7710	4	0.6518	3
20	0.8099	5	0.5786	2	0.7670	5	0.5851	2

clusters with DBI and $N_{o,S}^f$, $N_{o,S}^k$ are optimal numbers of clusters with Silhouette coefficient using fuzzy c-means and k-means, respectively.

7.1.1 Results with fuzzy c-means and k-means

k-means and fuzzy c-means clustering algorithms are implemented to cluster the load pattern data of different zones with different number of clusters. Optimal numbers of clusters of each zone are identified with Davies-Bouldin index and Silhouette coefficient and results of all 20 zones are shown in Table 1.

7.1.2 Results with DBSCAN

While implementing DBSCAN, various combinations of N_{minpts}, r_{eps} are chosen, but no set of these global parameters is found to get clusters. Results at different values of parameters are given in Table 2, for Zone-1 and $N_{minpts} = 5$ only, with different values of r_{eps}. There are no cluster formations on complete days. Further partition of load data is needed to find regular and irregular consumptions.

7.1.3 Results with proposed method

Let φ_U is the percentage data used for unusual consumptions detection and defined as given below:

$$\varphi_U = \frac{C_B}{C_\Omega} \times 100 \tag{25}$$

where C_B and C_Ω are cardinalities of set of border points, B, and whole load data, Ω, respectively. Using (6), for optimal partition of load data, $N_{minpts,o}$ and $r_{eps,o}$ for different regions are obtained as shown in Table 3 and LOF are calculated on only φ_U. Thus, the computational work is highly reduced. With the proposed method, the irregular consumptions are identified in each zone and these are ranked using LOF as per the irregularity. Low to high anomalous levels of different unusual consumptions are identified with the assignment of LOF. For Zone-1,4,5, six irregular consumption days with their LOF are shown in Table 4.

In most of the zones, except Zone-4, the highest LOF is close to 2.0 so unusualness in electricity consumptions is not large in these zones. It is found that Zone-4 has most varied unusual consumptions (Fig. 2). In Zone-4, on days 152, 153, 350, 351(i.e. June 01, June 02 and Dec 16 and Dec 17, 2007), the LOFs are more than 3.0. It shows that on mentioned days, the electrical load consumption deviates in large amount compared to the normal load consumptions. In different zones, a limiting value for LOF can be set to isolate the outliers so that utilities can extract requisite information from outliers. Irregular consumptions of Zone-4 and 5 are shown in Figs. 3, 4, respectively.

Different irregularity features are obtained and shown in Table 5 only on border points of Zone-4 which consists most varied unusual consumptions. Minimum values in sudden drop and gain features are decided same as 100 kW for min-max normalization. The 1100 kW, heuristically, is assumed as an acceptable demand to decide the irregular peak unusual consumptions. In this zone in a year 2007, no day is found which has zero electricity demand. FIFs are calculated of different unusual consumptions to rank them as {350, 351, 152, 153, 37, 26, 36, 42, 103} based on irregularity features.

Table 2 Results with DBSCAN

r_{eps}	No. of clusters	No. of load pattern in clusters	No. of load patterns not in clusters
2.5	01	365	Zero
1.7–2.4	01	364	01
1.2–1.6	01	363	02
1.1	01	355	10
1.0	02	338, 06	21
0.9	01	325	40
0.8	01	312	53
0.7	03	276, 10, 08	71
0.6	04	8, 208, 5	139
0.5	05	117, 26, 9, 9, 10	194
0.4	05	63, 8, 10, 5, 5	274
0.3	04	05, 34, 9, 8	309
0.2	01	05	360
0.1	No cluster	–	365

Table 3 Optimal partition of load data in different zones

Zone-id	$r_{eps,o}$	$N_{minpts,o}$	φ_U	$LOF(l_b^{lr})$
1	1.5	20	13.15	1.03
2	0.8	20	13.25	1.00
3	1.0	25	10.41	1.03
4	5.5	20	3.01	1.00
5	0.8	25	10.14	1.03
6	0.9	25	15.06	1.00
7	0.9	26	17.53	1.00
8	2.5	25	19.72	1.03
9	0.9	25	21.64	1.02
10	1.6	25	19.72	1.02
11	0.9	28	12.05	1.01
12	1.1	28	18.63	1.01
13	1.4	30	16.98	1.01
14	2.2	30	15.98	1.00
15	0.5	28	13.42	1.02
16	3.0	25	12.60	1.00
17	2.0	25	20.82	1.00
18	1.9	25	11.50	1.00
19	0.7	25	16.98	1.01
20	0.5	22	18.35	1.02

Table 4 Irregular consumptions with LOF

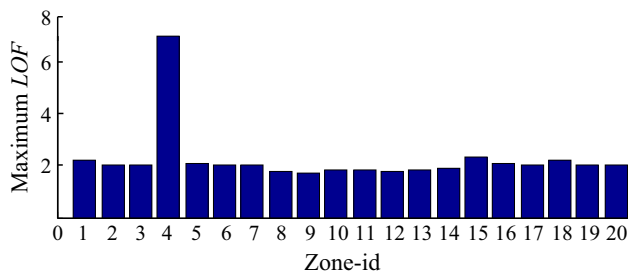| Zone-1		Zone-4		Zone-5	
Day	LOF	Day	LOF	Day	LOF
221	2.1890	152	6.9532	220	2.0630
220	1.9872	153	4.7295	20	1.8368
37	1.9807	350	4.2539	37	1.5186
222	1.7111	351	3.4112	39	1.3232
36	1.6461	26	2.2803	36	1.3221
237	1.5779	103	1.4808	21	1.2858

Fig. 2 Maximum LOF in different zones

Type and occurrence of regular peak and valleys in clustering results are detected in different zones. Peak and valley as demand response opportunities of only Zone-4
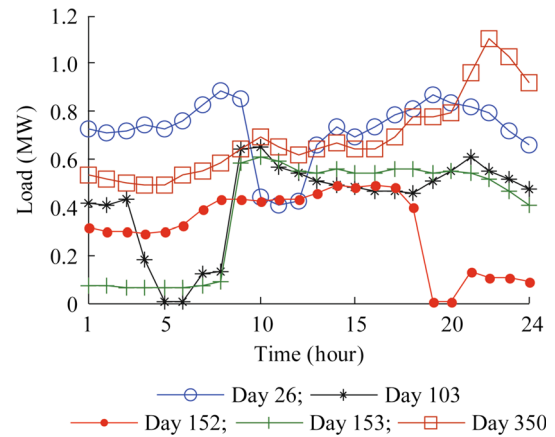
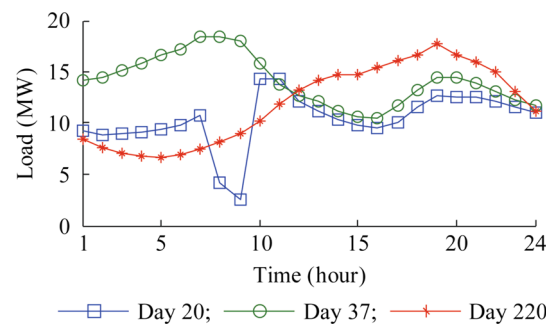

Fig. 3 Selected unusual consumptions in Zone-4

Fig. 4 Selected unusual consumptions in Zone-5

Table 5 Irregular consumptions features (Zone-4)

Day	Δd_{irpeak}^t	n_{peak}	Δd_{drop}^t	Δd_{gain}^t	n_{zero}	I_{FIF}
152	0	0	1	0.068	0	1.002
153	0	0	0	1	0	1
103	0	0	0.258	0.058	0	0.264
26	0	0.308	0.709	0.342	0	0.845
350	1	0.308	0	0.185	0	1.063
351	0	1	0.136	0.144	0	1.019
37	0	0.923	0	0	0	0.923
36	0	0.538	0	0	0	0.538
42	0	0.462	0	0	0	0.462

and 5 are shown in Table 6 and Figs. 5, 6. Morning peak (mp), evening peak (ep) and valley (v) are identified. In different zones, it is found with clustering results that 2 to 3 clusters are sufficient for peak-valley assessment and the numbers are optimal or close to optimal. Notations used in Table 7 are described as, min_D^p is minimum or close to minimum value of DBI and max_S^p is maximum or close to maximum value of Silhouette coefficient with proposed method; N_o^p is optimal or close to optimal number of clusters with proposed method.

Table 6 Peak-valley analysis

Zone-id	Cluster no.	DR opport.	Time (hour)	Demand (kW)
04	2	mp	9	1057
		v	15	737
		ep	20	1005
05	1	v	4	0.871×10^4
		ep	19	1.567×10^4
	2	mp	8	1.871×10^4
		v	14	1.354×10^4
		ep	20	1.616×10^4
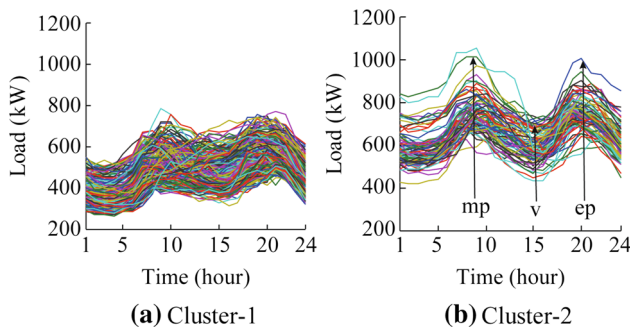
Fig. 5 Clustering results of Zone-4
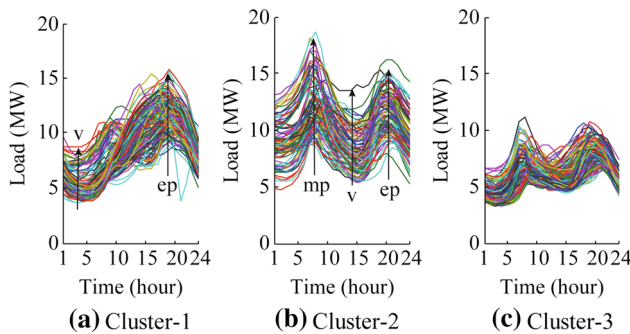
Fig. 6 Clustering results of Zone-5

Table 7 Clusters with proposed method

z_i	$\min_D^p$	$\max_S^p$	N_o^p	z_i	$\min_D^p$	$\max_S^p$	N_o^p
1	0.6677	0.7396	3	11	0.6261	0.7604	3
2	0.6316	0.7600	3	12	0.6471	0.7361	3
3	0.5677	0.7602	3	13	0.6292	0.7612	2
4	0.6587	0.7726	2	14	0.6521	0.7155	3
5	0.6339	0.7571	3	15	0.6416	0.7514	3
6	0.6644	0.7551	3	16	0.6826	0.7813	3
7	0.6923	0.7513	3	17	0.6899	0.7323	3
8	0.6568	0.7647	3	18	0.6313	0.7632	3
9	0.8738	0.6961	2	19	0.6706	0.7863	3
10	0.6098	0.7121	3	20	0.6039	0.7069	3

7.2 Case study-2

Indian Institute of Technology Kanpur (IITK) distribution system gets power supply from Panki power grid via 33 kV lines. One 10 MVA and two 5 MVA, 33 kV/11 kV transformers are installed in main substation [41]. The 10 MVA transformer (Tr-3) of main substation caters the major demand in IITK. Unusual consumptions along with regulars are identified and analyzed in hourly load data of year 2013 of 10 MVA, 33/11 kV transformer. Two optimal clustering are obtained and validated with Silhouette coefficient as 0.7865, 0.7832 and 0.7901 from k-means, fuzzy c-means and proposed method, respectively. Clustering results and unusual consumptions are shown in Figs. 7, 8, respectively. The ranked irregular consumptions with LOF are shown in Table 8.

The global parameters are set as $N_{\text{minpts},o} = 20$ and $r_{\text{eps},o} = 5$ according to (6). The number of border points is identified as 27 which is 7.40% of all 365 load patterns of year 2013. Unusual characteristics, with the proposed approach of characterization, are identified only in different border points. For these consumptions, the I_{FIF} are calculated while assuming limiting values of $d_{ref} = 325$A, $d_{peak,a} = 375$A, $\delta_a^g = 150$A, $\delta_a^d = 150$A. Day 198 is having
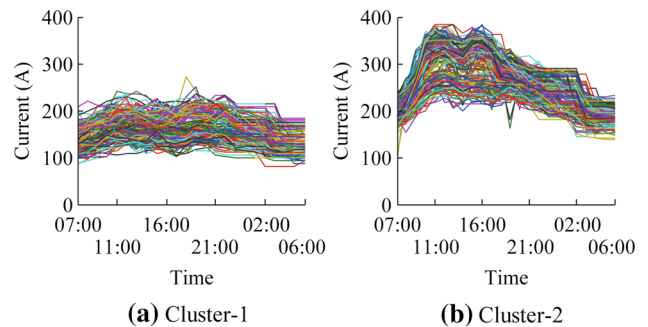

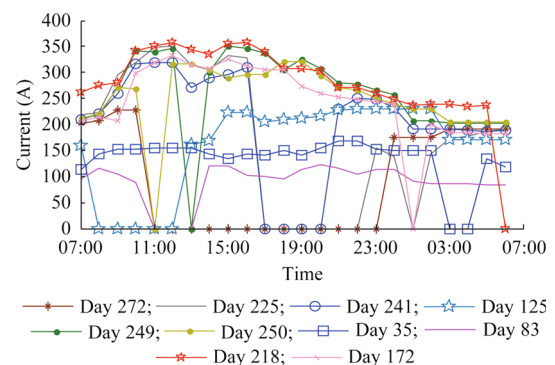**Fig. 7** Clustering results of electricity demand on Tr-3 (on a phase) of IITK

Fig. 8 Unusual consumptions of electricity demand on Tr-3 (on a phase) of IITK

Table 8 Irregular consumptions at IITK load pattern data

Day	LOF	Day	LOF	Day	LOF
249	7.10	35	2.40	176	1.37
250	6.03	83	2.28	263	1.31
272	5.01	172	2.19	106	1.29
218	4.89	97	2.15	231	1.26
241	3.72	76	2.05	178	1.22
198	3.27	119	1.76	88	1.20
151	2.90	86	1.61	288	1.19
225	2.80	15	1.58	89	1.16
125	2.54	224	1.48	348	1.14

Table 9 Features of unusual consumptions in IITK load data

Day	Δd_{irpeak}^t	n_{peak}	Δd_{drop}^t	Δd_{gain}^t	n_{zero}	I_{FIF}
249	0	0.38	1	0.9	0.08	1.40
225	0	0.25	0.91	0.03	0.54	1.088
272	0	0	0.40	0.15	1	1.087
241	0	0	0.81	0.48	0.31	0.99
250	0	0	0.61	1	0.08	1.170
198	1	0.625	0	0	0	1.179
172	0	0.125	0.46	0.3	0.08	0.569
218	0	1	0.44	0	0.08	1.095
263	0.059	0.5	0	0	0	0.504
125	0	0	0.05	0.06	0.38	0.39
35	0	0	0	0	0.15	0.15
83	0	0	0	0	0.23	0.23
224	0	0.5	0	0	0	0.5
151	0	0.875	0	0	0	0.875

392 A showing an irregular peak demand at 20:00 whereas Day 218 is having the broadest peak demand more than 325 A for maximum consecutive 8 hours (from 10:00 to 17:00). On Day 249, the demand drops sharply, a maximum drop in load pattern data, from 346 A to 0 A between 12:00 to 13:00. On the Day 250, the demand increases, sharply, from 0 A to 317 A between 11:00 to 12:00. On Day 272, the demand remains zero for 13 from 11:00 to 23:00.

Each column of Table 9 is normalized with min-max normalization method. Min values in sudden drop and gain features are decided same as 150 A, zero values for n_{peak} and n_{zero}, and 375 A in irregular peak while max values, in respective columns, are used for normalization. Thus, unusual consumptions are compared with one other and I_{FIF} are calculated. I_{FIF} is composed of irregularity features present in the unusual consumption and the features which are dominating and others which have less effect can be identified. The ranking of unusual consumptions with I_{FIF}
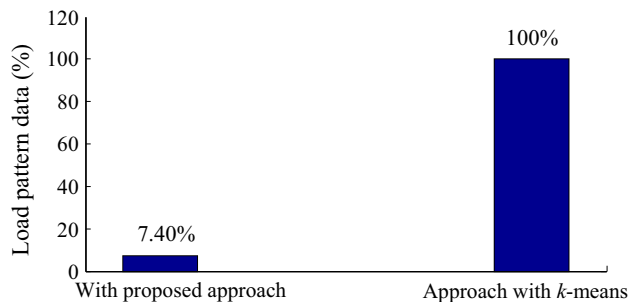

Fig. 9 Pertencage load data used for identification of broadest peak demand

is obtained as {249, 198, 250, 218, 225, 272, 241, 151, 224, 125, 83, 35}.

8 Size of energy storage

Based on the analysis of loading on a 10 MVA transformer at 33 kV/11 kV substation of IITK, in 2013, authors have identified a critical load profile using k-means algorithm [41] while utilizing complete load pattern data. This profile decides possible size of energy storage, without PV generation, for peak shaving operation. The broadest peak demand, defined in (20), is basically a critical load profile and helps in deciding the size of energy storage for peak shaving. To decide the critical load profile, the proposed approach of this paper works only on 7.40% of the load pattern data as that shown in Fig. 9. The profile of Day 218 shows the broadest peak demand.

9 Conclusion

In this paper, the unusual consumptions are obtained by the proposed method, using the local outlier factor (LOF), on only a few percent of whole load pattern data. Different, unusual loadings, and occurrence and type of peak-valley demand on substations are identified. The different features of unusual consumptions have been analyzed with proposed characterization on only border points of two practical test systems. Test results reveal that the proposed method is very effective in finding the irregular consumption, such as different types of unusual peak demand, sudden large change and zero demand. Regular peaks-valleys are identified with clustering results obtained from proposed approach in order to distinguish irregular peaks in the load pattern data. To validate the clustering of load pattern data, two most popular methods such as the Davies-Bouldin index (DBI) and Silhouette coefficient (SC) are used.

References

[1] http://www.pjm.com/markets-and-operations/energy.aspx

[2] https://energy.gov/fe/articles/developing-new-primer-nation-s-electricity-markets

[3] Ni YX, Zhong J, Liu HM (2005) Deregulation of power systems in Asia: special considerations in developing countries. In: Proceedings of the 2005 IEEE Power Engineering Society general meeting, vol 3, San Francisco, CA, USA, 12–16 Jun 2005, pp 2876–2881

[4] U.S. Department of Energy (2006) Benefits of demand response in electricity markets and recommendations for achieving them: a report to the United States Congress. Washington, DC, USA. Pursuant to Section 1252 of the Energy Policy Act of 2005

[5] Alhadi MH, El-Saadany EF (2008) A summary of demand response in electricity markets. Elect Power Syst Res 78(11):1989–1996

[6] Saele H, Grande OS (2011) Demand response from household customers: experiences from a pilot study in Norway. IEEE Trans Smart Grid 2(1):102–109

[7] Mathieu JL, Price PN, Kiliccote S et al (2011) Quantifying changes in building electricity use, with application to demand response. IEEE Trans Smart Grid 2(3):507–518

[8] Huang D, Billington R (2012) Effects of load sector demand side management applications in generating adequacy assessment. IEEE Trans Power Syst 27(1):335–343

[9] Logenthiran T, Srinivasan D, Shun TZ (2012) Demand side management in smart grid using heuristic optimization. IEEE Trans Smart Grid 3(3):1244–1252

[10] Nizar AH, Dong ZY, Zhang P (2008) Detection rules for non technical losses analysis in power utilities. In: Proceedings of the 2008 IEEE Power and Energy Society general meeting: conversion and delivery of electrical energy in the 21st century, Pittsburgh, PA, USA, 20–24 Jul 2008, 8 pp

[11] Nagi J, Yap KS, Tiong SK et al (2010) Nontechnical loss detection for metered customers in power utility using support vector machines. IEEE Trans Power Deliv 25(2):1162–1171

[12] Dos-Angelos EW, Saavedra OR, Cortés OAC et al (2011) Detection and identification of abnormalities in customer consumptions in power distribution systems. IEEE Trans Power Deliv 26(5):2436–2442

[13] Depuru SSSR, Wang LF, Devabhaktuni V (2012) Enhanced encoding technique for identifying abnormal energy usage pattern. In: Proceedings of the North American power symposium (NAPS'12), Champaign, IL, USA, 9–11 Sept 2012, 6 pp

[14] Willis HL, Schauer AE, Northcote-green JED et al (1983) Forecasting distribution system loads using curve shape clustering. IEEE Trans Power Appl Syst 102(4):893–901

[15] Grigoras G, Cartina G, Bobric EC (2010) An improved fuzzy method for energy losses evaluation in distribution networks. In: Proceedings of the 15th IEEE Mediterranean electrotechnical conference (MELECON'10), Valletta, Malta, 25–28 Apr 2010, pp 131–135

[16] Zhou G, Zhao W, Lü XJ, et al (2014) A novel load profiling method for detecting abnormalities of electricity customer. In: Proceedings of the 2014 IEEE Power and Energy Society General Meeting, Washington, DC, USA, 27–31 Jul 2014, 5 pp

[17] Wijayasekara D, Linda O, Manic M et al (2014) Mining building energy management system data using fuzzy anomaly detection and linguistic descriptions. IEEE Trans Ind Inf 10(3):1829–1840

[18] Chen CS, Kang MS, Hwang JC et al (2000) Synthesis of power system load profiles by class load study. Elect Power Energy Syst 22(5):325–330

[19] Chicco G, Napoli R, Postolache P et al (2003) Customer characterization options for improving the tariff offer. IEEE Trans Power Syst 18(1):381–387

[20] Gerbec D, Gasperic S, Smon I et al (2005) Allocation of the load profiles to consumers using probabilistic neural networks. IEEE Trans Power Syst 20(2):548–555

[21] Espinoza M, Joye C, Belmans R et al (2005) Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. IEEE Trans Power Syst 20(3):1622–1630

[22] Nizar AH, Dong ZY, Zhao JH (2006) Load profiling and data mining techniques in electricity deregulated market. In: Proceedings of the 2006 IEEE Power Engineering Society general meeting, Montreal, Canada, 18–22 Jun 2006, 7 pp

[23] Chicoo G, Napoli R, Piglione F (2006) Comparisons among clustering techniques for electricity customer classification. IEEE Trans Power Syst 21(2):933–940

[24] Verdu SV, Garcia MO, Senabre C et al (2006) Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. IEEE Trans Power Syst 21(4):1672–1682

[25] Tsekours GJ, Kotoulas PB, Tsirekis CD et al (2008) A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers. Elect Power Syst Res 78(9):1494–1510

[26] Chicco G, Ilie I-S (2009) Support vector clustering of electrical load pattern data. IEEE Trans. on Power Systems 24(3):1619–1628

[27] Zhang T, Zhang G, Lu J et al (2012) A new index and classification approach for load pattern analysis of large electricity customers. IEEE Trans Power Syst 27(1):153–160

[28] Chicoo G, Ionel O-M, Porumb R (2013) Electrical load pattern grouping based on centroid model with ant colony clustering. IEEE Trans Power Syst 28(2):706–1715

[29] Jain A, Murty M, Flynn P (1999) Data clustering: a review. ACM Comput Surv 31(3):264–323

[30] Xu R, Wunsch D (2005) Survey of clustering algorithm. IEEE Trans Neural Netw 16(3):645–678

[31] Patwary MMA, Palsetia D, Agarwal A, et al (2012) A new scalable parallel DBSCAN algorithm using the disjoint-set data structure. In: Proceedings of the international conference on high performance computing, networking, storage and analysis (SC'12), Salt Lake City, UT, USA, 10–16 Nov 2012, 11 pp

[32] Mutanen A, Ruska M, Repo S et al (2011) Customer classification and load profiling method for distribution systems. IEEE Trans Power Deliv 26(3):1755–1763

[33] Stephen B, Mutanen AJ, Galloway S et al (2014) Enhanced load profiling for residential network customers. IEEE Trans Power Deliv 29(1):88–95

[34] Hsiao YH (2015) Household electricity demand forecast based on context information and user daily schedule analysis from meter data. IEEE Trans Ind Inf 11(1):33–43

[35] Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv 41(3):1–58

[36] Breunig MM, Kriegel HP, Ng RT et al (2000) LOF: identifying density-based local outliers. ACM SIGMOD Rec 29(2):93–104

[37] Schubert E, Zimek A, Kriegel HP (2014) Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection. Springer Data Mining Knowl Discov 28(1):190–237

[38] Aggarwal CC (2013) Outlier Analysis. Springer, New York, NY, USA

[39] Global energy forecasting competition 2012-load forecasting—a hierarchical load forecasting problem: backcasting and forecasting hourly loads (in kW) for a US utility with 20 zones. Kaggle, San Francisco, CA, USA

[40] Hong T, Pinson P, Fan S (2014) Global energy forecasting competition 2012. Int J Forecast 30(2):357–363

[41] Sharma DD, Singh SN, Rajpurohit BS, et al (2015) Critical load profile estimation for sizing of energy storage system. In: Proceedings of the 2015 IEEE Power and Energy Society General Meeting, Denver CO, USA, 26–30 Jul 2015, 5 pp

Desh Deepak SHARMA pursued Ph.D. degree in Electrical Engineering at Indian Institute of Technology, Kanpur, Kanpur under QIP scheme. He is employed as Associate Professor in Department of Electrical Engineering, M.J.P. Rohilkhand University, Bareilly, India. His research interests include demand side management, application of data mining techniques in load profiling and distribution system; distributed control schemes for distributed generation and energy storages, cyber-security issues in distributed control system.

S. N. SINGH received the M.Tech. And Ph.D. degrees from Indian Institute of Technology, Kanpur, Kanpur, India, in 1989 and 1995 respectively. Presently, he is working as Professor in the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur. His research interests include power system restructuring, power system optimization and control, voltage security and stability analysis, power system planning, and ANN application to power system problems. Prof. Singh is a fellow of IE(India), a fellow of IETE(India), and a fellow IET(UK).

Jeremy LIN received his M.S.E.E in power and energy system from University of Illinois at Urbana-Champaign in 1999, M.B.A in finance from Villanova University in 2010 and a PhD from Drexel University. For more than a decade, he has been working in the electric power industry in various technical assignments, previously at Mid-America Interconnected Network (MAIN) Inc., GE Energy/Energy Consulting, and currently at PJM Interconnection, USA. His work included power system security operation and engineering support, power flow analysis, and calculation of available transfer capability. His recent assignments included economic valuation and reliability analysis of major transmission projects in PJM and neighboring regions. He is also leading some research projects at PJM, funded by DOE and other research consortium.

Elham FORUZAN received the master's degree in electrical engineering from the university of Tehran. Currently, she is a dual degree student pursuing a Ph.D. in electrical and computer engineering, and a master's degree in computer science, at the University of Nebraska-Lincoln. Her research interests include smart grid, artificial intelligence and machine learning, multi-agent systems, and power system cyber security.