This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2015.2479941, IEEE Transactions on Power Delivery

Hierarchical K-means Method for Clustering Large-Scale Advanced Metering Infrastructure Data

Tian-Shi Xu, Hsiao-Dong Chiang, *Fellow*, *IEEE*, Guang-Yi Liu, and Chin-Woo Tan

***Abstract*—Clustering of the load patterns from distribution network customers is of vital importance for several applications. However, the growing number of Advanced Metering Infrastructure (AMI) and a variety of customer behaviors make the clustering task quite challenging due to the increasing amount of load data. K-means is a widely used clustering algorithm in processing large dataset with acceptable computational efficiency, but suffers from local optimal solutions. To address this issue, this paper presents a hierarchical K-means (H-K-means) method for better clustering performance for big data problems. The proposed method is applied to a large-scale AMI dataset and its effectiveness is evaluated by benchmarking with several existing clustering methods in terms of five common adequacy indices, outliers detection, and computation time.**

***Index Terms*—Advanced Metering Infrastructure (AMI), big data problems, clustering, hierarchical K-means (H-K-means), load patterns.**

I. Introduction

WITH the proliferating of Advanced Metering Infrastructure (AMI), the clustering of the load patterns from distribution network customers has several applications. Firstly, for electricity providers, the variable consumption patterns make it quite challenging to provide satisfactory service, since the providers may have certain degrees of freedom in tariff design, a set of suitable tariffs with multiple rates can be developed to increase the profits of the market players after grouping the customers with similar electrical behaviors together [1]. Secondly, some abnormal consumption behaviors exist in AMI measurements due to several factors, such as the theft of electricity, improper installation of equipment, faulty metering, or statistical errors, which have led to nontechnical losses for distribution companies. To handle this problem at a low cost, clustering techniques have been applied in the task of detecting the outliers [2], [3]. Meanwhile, load pattern clustering has also been used to decompose unknown profiles into known profiles, which is desirable for distribution companies to influence the behaviors of the customers [4]. In addition, clustering load patterns with different consumption behaviors is also useful for distribution network state estimation [5], load forecasting [6], demand response [7], market strategies design [8], customer classification [9], etc.

Various methods have been applied to the clustering of load patterns, such as, self-organizing map (SOM) [10]-[15], hierarchical clustering method [12], [16]-[19], modified follow the leader [1], [10], [12], adaptive vector quantization [17], and Renyi entropy-based method [20], but the increasing number and the extending sampling period of AMI have enlarged the scale of the load dataset significantly, and most of the aforementioned methods suffer from excessive computational burden for large datasets. Apart of them, K-means [4], [5], [9], [11]-[18], [20]-[23], is a widely used clustering method with the simplest principle and fast convergence speed. K-means has been applied to the analysis of the large-scale datasets in [21]-[23]. However, since the classical K-means algorithm is sensitive to the initial centroids, the probability of finding appropriate initial centroids is especially low for large datasets so that local optimal solutions always appear in the final results. Hence, there is a pressing need to enhance the performance of classical K-means for clustering large dataset.

To meet this need, this paper proposes a hierarchical K-means (H-K-means) clustering method, which is different from the previously proposed “hierarchical K-means” methods. In previous studies, the “hierarchical” is based on the perspective of methodology, which usually refers to the aforementioned hierarchical clustering method. For example, hierarchical clustering method is combined with K-means in different ways in [24], [25]. While the “hierarchical” is explored from the perspective of data in this paper, which means the establishment of a hierarchical data structure before the process of K-means clustering. On this basis, the proposed H-K-means method is developed with the following goals:

1) It can significantly improve the quality of the clustering results given by classical K-means.
2) It can preserve the inherent speed advantage of classical K-means.
3) It is especially applicable to big data problems.

This work was supported by National High-Technology Research and Development Program (“863” Program) of China (2014AA051901).

T.-S. Xu is with the School of Electrical Engineering and Automation, Tianjin University Tianjin, China.

H.-D. Chiang is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA.

G.-Y. Liu is with the State Grid Corporation of China& SGRI North America, Beijing, China.

C.-W. Tan is with the Stanford Sustainability Systems Laboratory, Stanford University, Stanford, CA, USA.

4) It can effectively cluster large-scale load demand curves.

To evaluate its reliability and effectiveness, the proposed H-K-means method is applied to a large-scale AMI dataset. In addition, the quality of clustering results based on adequacy indices, outliers detection, and computation time are used to evaluate the clustering performance. As a result, outstanding performance has been achieved by the proposed H-K-means method after a comprehensive comparison with several existing K-means based clustering methods.

II. Classical K-means Algorithm

K-means [26] is a classical clustering algorithm which has been widely used in data mining. Its main function is to classify the patterns from a given dataset into a pre-specified number of clusters. In general, the quality of the clustering results given by classical K-means can be measured by an objective function such as, among others, the summation of the square of the Euclidean distance between each pattern and its centroid:

$$f = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left| x_i^k - \omega_k \right|^2 \tag{1}$$

where K is the number of clusters, ω_k is the centroid of the k_{th} cluster, n_k is the number of patterns belonging to the k_{th} cluster, x_i^k is the i_{th} pattern belonging to the k_{th} cluster.

To separate N patterns into K clusters, the basic procedure of the classical K-means algorithm is presented as follows:

Step 1. Select K patterns from the original dataset randomly as the initial centroids.

Step 2. For each pattern, calculate the distance between it and each centroid, and assign it to the nearest cluster.

Step 3. Update each centroid, say the k_{th} centroid, by:

$$\omega_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i^k \tag{2}$$

Step 4. If the convergence criterion is satisfied, terminate the algorithm with the current K clusters, otherwise, go to Step 2.

The convergence criterion in Step 4 can be set with a variety of forms, for example, a certain number of iterations, or the stabilization of the centroids, or an acceptable variation of the objective function value and etc.

The computation speed of classical K-means is usually fast for large datasets. However, the quality of the clustering results is sensitive to the initial centroids and may suffer from the issue of local optimal solutions. To overcome these difficulties, a H-K-means method is proposed in the next section.

III. Hierarchical K-means Method for Big Data Clustering

The quality of given initial centroids plays a decisive role in the final results of classical K-means. Unfortunately, the larger number of the patterns included in the dataset, the lower probability the K-means is to find the high-quality initial centroids among them.

The main idea of our proposed H-K-means method is as follows.

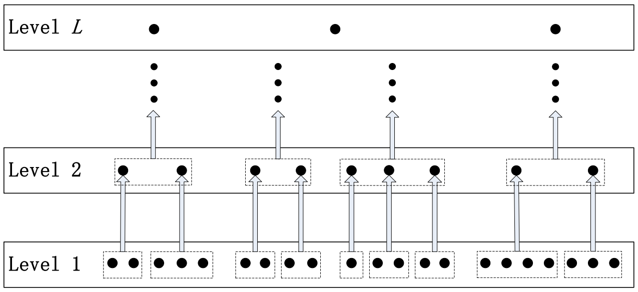

Fig. 1. Hierarchical structure based on the original dataset.

For a given dataset, if we treat a set of patterns which are close to each other as one representative-pattern via their centroid, then the entire dataset is represented as a smaller-size dataset consisting of several representative-patterns with a similar distribution as the original one. From this perspective, a hierarchical structure for the original dataset is built up as shown in Fig. 1, where the number of levels L is defined by the user or some criteria, the 1_{st} level is the original dataset, and each subsequent level is constituted by a smaller dataset of its previous level. Based on this multi-level structure, the H-K-means method can be described as the following steps:

Stage I: Establishing the hierarchical structure

Step 1. Set the original dataset as the 1_{st}-level dataset, start from $i = 2$.

Step 2. Establish the i_{th}-level dataset based on the $(i–1)_{\text{th}}$-level dataset (please see the later part for details).

Step 3. If $i = L$, go to Stage II, otherwise $i = i+1$ and go to Step 2.

Stage II: Weighted clustering

Step 4. If $i = L$, select K patterns among the i_{th}-level dataset randomly, otherwise, use the K obtained centroids in Step 5 instead.

Step 5. Implement K-means clustering for the i_{th}-level dataset and use the patterns given by Step 4 as the initial centroids. During each epoch of K-means, calculate the centroid of each cluster, say the k_{th} cluster, by:

$$\omega_k = \sum_{p=1}^{n_k} \left(r_p \cdot x_p \right) \Big/ \sum_{p=1}^{n_k} r_p \tag{3}$$

where n_k is the size of the k_{th} cluster in the i_{th}-level dataset, x_p is the p_{th} pattern of this cluster. If $i > 1$, r_p equals to the number of the $(i–1)_{\text{th}}$-level patterns represented by x_p, otherwise, $r_p = 1$.

Step 6. If $i = 1$, terminate the process and output the clustering results in Step 5 as the final results, otherwise, $i = i–1$ and go to Step 4.

Details for Step 2:

1) In this step, the following two requirements should be met to effectively reduce the scale of original clustering problem:

(i). Ensure a high similarity between the $(i–1)_{\text{th}}$-level dataset and the i_{th}-level dataset, in other words, each pattern in the $(i–1)_{\text{th}}$-level dataset should be close enough to the corresponding representative-pattern in the i_{th}-level dataset.

(ii). The size of the i_{th}-level dataset should be as small as possible.

```
Algorithm 1: Establishment of the $i_{th}$-level dataset
Input: the $(i–1)_{th}$-level dataset ($D_{i-1}$), the estimated size and
  the threshold $t$ for the $i_{th}$-level dataset ($M_i$ and $t_i$).
Output: the $i_{th}$-level dataset ($D_i$).
  Set $D_i = \{\Phi\}$.
  Cluster the patterns in $D_{i-1}$ into $M_i$ clusters with K-means.
  for $j \leftarrow 1$ to $M_i$ do
    Calculate the $\theta$ value for the $j_{th}$ cluster ($\theta_j$).
    if $\theta_j \le t_i$ then
      Add the centroid of the $j_{th}$ cluster into $D_i$.
    else
      Set $k = n_j$ ($n_j$ is the size of the $j_{th}$ cluster), $\theta_{\max} = 0$, all
        the $k$ patterns in the $j_{th}$ cluster as $C_{j1},\ldots,C_{jk}$, then $C_j =$
        $\{C_{j1},\ldots,C_{jk}\}$.
      while $\theta_{\max} \le t_i$ do
        $R_j = C_j$, then remove one of the two nearest patterns
          from $C_j$, $k = k–1$.
        Cluster the patterns in the $j_{th}$ cluster into $k$
          sub-clusters with K-means, using the patterns in $C_j$
          as the initial centroids.
        Update $C_j$ with the centroids of the $k$ obtained
          sub-clusters.
        Calculate the $\theta$ values for the $k$ obtained
          sub-clusters ($\theta_{j1},\ldots,\theta_{jk}$), set $\theta_{\max} = \max\{\theta_{j1},\ldots,\theta_{jk}\}$.
      end while
      Add the patterns in $R_j$ into $D_i$.
    end if
  end for
```

2) The basic procedure for building the i_{th}-level dataset is described in Algorithm 1. To meet the requirement (i), it firstly separates the patterns in the $(i–1)_{\text{th}}$-level dataset into M_i (an estimated size set by the user) clusters using K-means, then for each cluster, a θ indicator is calculated as follows to evaluate the similarity between all its members and the centroid:

$$\theta = \max_{q=1,\cdots,n}\left(\left|x_q - c\right|^2 \Big/ \left|c\right|^2\right) \qquad (4)$$

where n is the size of this cluster, x_q is the q_{th} pattern of this cluster, and c is the centroid of this cluster. A higher similarity between the members and the centroid can be expressed by obtaining a smaller θ since all the members are much closer to the centroid. Hence, for the members of each cluster, if the corresponding $\theta \le t_i$ (a threshold set by the user), we use the centroid as their representative-pattern, otherwise, we split this "bad" cluster into several sub-clusters and use the obtained centroids as the corresponding representative-patterns.

3) To meet the requirement (ii), the splitting process of each "bad" cluster continues with the decreasing number of sub-clusters as long as the corresponding $\theta_{\max} \le t_i$, and to improve the performance of the current splitting process (aims at k sub-clusters) via better initial centroids, k of the k+1 centroids obtained in the last splitting process will be used as the initial centroids instead of the random selection.

Remarks:

For large-scale datasets, the proposed H-K-means method is expected to achieve high-quality clustering results due to the following factors, which will be supported by the numerical studies in the next section:

1) The scale of the original clustering problem can be reduced gradually at Stage I, the probability for K-means to provide high-quality initial centroids is high at the top level.
2) The impact of high-quality initial centroids obtained at the top level influences the successive levels at Stage II, and finally promising clustering results of the original dataset will be obtained.
3) High-quality initial centroids will hopefully accelerate the convergence of K-means at each level, and for multiple executions, Stage I needs to be carried out only once. Hence, little extra computation time will be required for the proposed H-K-means method as compared with classical K-means.

IV. Application to AMI Data

A. Load clustering

Each load pattern is characterized by the consumption data of an electricity customer during a single day at successive 15-min or 1-hour time steps. To cluster all the load patterns accurately, small distances should be admitted between the load patterns with similar shapes or close peaking times. Hence, all the load patterns should be represented based on an unified scale. Then, for each load pattern, all of its measurements are normalized into the range of [0,1] by using its maximum value as the reference power.

For the purpose of clustering N daily consumption patterns into K clusters, the set to be clustered is defined as $X = \{\vec{x}_i : i = 1,\cdots,N\}$, and $\vec{x}_i = (x_{i1}, x_{i2}, \cdots, x_{id})^T$ stands for the i_{th} load pattern with d measurements. After the clustering process, we have the K non-overlapping clusters $W = \{\Omega_k$: $k = 1, \cdot\cdot;K\}$, where Ω_k represents the set of all the N_k patterns belonging to the k_{th} cluster. Each cluster, say the k_{th} cluster, can be represented by the centroid $\vec{\omega}_k = (\omega_{k1,}\omega_{k2,} \cdots, \omega_{kd})^T$, which is obtained by calculating the average vector of the N_k members of Ω_k. Note that in this application case, each variable x in (1)-(4) is replaced by the corresponding vector form $\vec{x}$, and so are the variables ω and c.

A set of AMI data consisting of the energy consumption values from 90 residential customers from a metropolitan area in the west coast of USA during a total of 8 months (roughly 220 days) is employed in this numerical study. For each customer, its smart meter recorded consumption data in every 15 minutes, and a large-scale AMI dataset containing 18662 daily patterns with d = 96 has been obtained.

B. Implementation of the H-K-means method

The proposed H-K-means method is applied to assign the load patterns to different clusters. At the Stage I of the method, a hierarchical structure is built from the original dataset. It can be observed from Algorithm 1, the establishment of the dataset at each level, say the i_{th} level, depends on the corresponding

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2015.2479941, IEEE Transactions on Power Delivery

estimated size M_i and the threshold t_i, hence, a parametric analysis is necessary.

Firstly, we set the 18662 patterns from the original dataset as the 1_{st}-level dataset, and the parametric analysis of the 2_{nd}-level dataset is shown in Fig. 2, in which various values of the parameters M_2 and t_2 have been tested respectively.

To "simplify" the 1_{st}-level dataset, the 2_{nd}-level datasets with smaller sizes are always desirable. However, although Fig. 2 shows that increasing t_2 can reduce the size of the 2_{nd}-level dataset effectively, t_2 cannot be too large since it limits the distances between the patterns from the two levels, namely, the similarity between the two levels. Hence, in the range shown in Fig. 2, only the smaller values will be considered further for t_2.

It can be observed from Fig. 2 that increasing M_2 when t_2 is relatively small, there is always a trend towards reducing the size of the dataset, but this trend becomes flat after about $M_2 = 1000$. Hence, M_2 is chosen as 1000 since the further increasing of M_2 will doubtless aggravate the computational burden.

Regarding the clustering effect as well as the computation efficiency, we observed that for several given numbers of levels, the hierarchical structure with a reduction of about 50% of the dataset size at each level (compared with its lower level) is the most suitable one. This observation (supported by numerical tests, omitted here due to the limitation of the space) leads to the selection of $M_2 = 1000$, $t_2 = 0.1$ (size: 8274) for the 2_{nd}-level dataset. Indeed, the choice of t_2 is not unique in a small range of [0.088,0.104], and the corresponding 2_{nd}-level dataset size is limited in [8110,9404], which is nearly 50% of 18662.

Then start from $i = 3$, the following steps are implemented:

Step 1: $M_i = M_{i-1}/2$.

Step 2: For the given M_i, keep increasing t_i from 0 until the size of the corresponding dataset is smaller than 50% of the $(i–1)_{th}$-level dataset, then use it as the i_{th}-level dataset.

Step 3: If the termination criterion (please see the later part for details) is satisfied, terminate this procedure and use the hierarchical structure with the $1_{st},\ldots,(i–1)_{th}$ levels for the Stage II of the method, otherwise, set $i = i+1$ and go to Step 1.

The original dataset can be "simplified" with the increasing number of levels. However, if excessive levels are used, fewer patterns will reside at the top level dataset to reflect the distribution of the 18662 original patterns, thus some original patterns with different shapes will be represented by the same pattern at the top level, which may affect the final clustering performance. Hence, the following criterion is applied to choose the optimal number of levels (L_{opt}).

For the current hierarchical structure, all the original patterns represented by the same top level pattern are treated as a cluster, and this top level pattern is selected as the corresponding centroid. Then for each cluster, the θ value is calculated according to (4). Finally, the maximum value of all the obtained θ values ($\theta_{\max}$) is recorded. A large value of $\theta_{\max}$ will then arise if patterns with significant differences are assigned to the same cluster. Hence, the above 3 steps will be terminated with the optimal number of levels $L_{\text{opt}} = i–1$ when the current $\theta_{\max}$ exceeds a preset threshold, value p.

We set $p = 2$, and the detailed information for each level is summarized in Table I. It can be observed that the $\theta_{\max}$ keeps increasing slightly from $L = 2$ to 6, and becomes abnormally large at $L = 7$. According to this, the optimal number of levels for the Stage II of the H-K-means method is chosen as $L_{\text{opt}} = 6$.

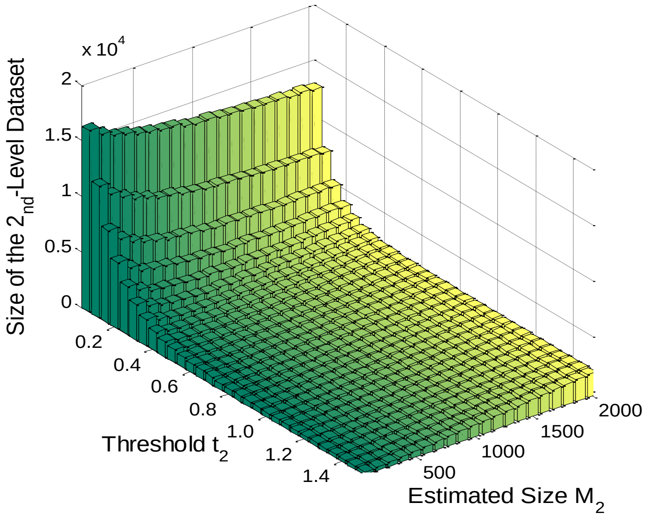

Fig. 2. Size of the 2_{nd}-level dataset variation for different M_2 and t_2.

TABLE I
INFORMATION FOR THE HIERARCHICAL STRUCTURE

Level index	Threshold t	Size	$\theta_{\max}$
1	-	18662	-
2	0.1	8274	0.1000
3	0.2	2750	0.4412
4	0.3	949	0.8264
5	0.4	437	1.1563
6	0.6	210	1.2961
7	1.0	83	9.0000

TABLE II
ADEQUACY INDICES FOR THE H-K-MEANS METHODS WITH DIFFERENT NUMBER OF HIERARCHICAL LEVELS AT $K = 40$

Adequacy indices	Number of hierarchical levels				
	3	4	5	6	7
MIA	0.148	0.146	0.146	0.144	**0.142**
SI	548.479	447.380	445.239	**404.716**	408.916
SMI	0.687	0.685	0.685	**0.681**	0.685
CDI	0.722	0.657	0.661	0.622	**0.592**
WCBCR	6.533	5.539	5.610	**5.137**	5.368

C. Performance comparisons

For comparison purposes, several K-means based methods available in the literature have also been applied to the dataset, including classical K-means [4], [5], [9], [11]-[18], [20], [21], fuzzy K-means [11]-[14], [17], [18], [27]-[29], K-means++ [30], [31], weighted fuzzy average K-means (WFA-K-means) [16], [29], [32], and developed K-means [17]. Moreover, the H-K-means methods with the other numbers of levels (L = 3, 4, 5, 7) are also included.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2015.2479941, IEEE Transactions on Power Delivery

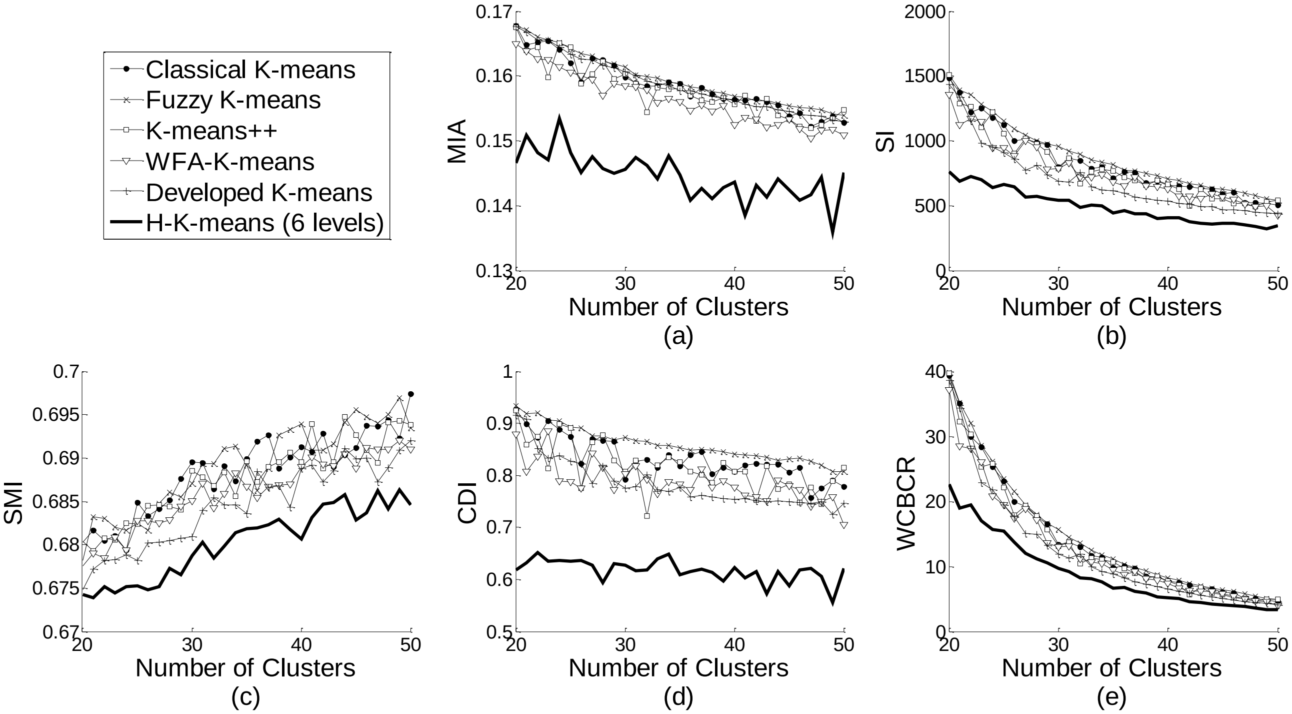

Fig. 3. Adequacy indices for the discussed clustering methods. (a) MIA. (b) SI. (c) SMI. (d) CDI. (e) WCBCR.

1) *Adequacy indices*

The following adequacy indices have been employed to evaluate the quality of clustering results: the Mean index adequacy (MIA) [1], the Scatter index (SI) [33], the Similarity matrix index (SMI) [13], the Clustering dispersion indicator (CDI) [1], and the Ratio of within cluster sum of squares to between cluster variation (WCBCR) [34].

The MIA index is used to evaluate the compactness of the clustering results based on the distance between each pattern and its centroid. It is especially suitable for the selection of the appropriate candidates for demand response programs. The SI and SMI indices are used to evaluate the distinctiveness of the clustering results based on the distances between different centroids, which can be applied to the analysis of the diversity of the customers or the detection of the abnormal patterns. Meanwhile, the CDI and WCBCR indices combine the evaluations of the compactness and the distinctiveness together, which can provide more thorough understanding of the clustering results.

As a common characteristic of the five adequacy indices, lower values will always correspond to better clustering results, namely, the clusters with better compactness or distinctiveness. A comparison of the adequacy indices will be conducted with the number of clusters ranging from 20 to 50, and a comparison between different methods is valid only for the same number of clusters. Considering the randomness of the discussed methods, 100 executions of each method are carried out for each number of clusters and the best result of each adequacy index is selected for comparison. Note that for the developed K-means method from [17], the steps of the variations of its two parameters have been adjusted in order to form the same number of executions.

As an example, Table II shows the adequacy indices for the H-K-means methods with different number of levels at *K* = 40. With the increasing of *L*, the improving of most indices terminates (or becomes less significant) after *L* = 6. Similar situations also hold with the other number of clusters. Hence, the case for *L* = 6 is used in the comparison with other methods.

Fig. 3 shows the comparison of all the discussed methods in terms of the five adequacy indices. As the number of clusters increases, all the methods generally, but not always, become more effective owing to the increasing of the resolution of the obtained clustering [10]. While in the presence of the same number of clusters, all these indices consistently show the distinct superiority of the H-K-means method. For each index, fuzzy K-means always gives the worst results. By contrast, although certain improvements can be achieved by K-means++, WFA-K-means, and developed K-means as compared with classical K-means, the H-K-means method still performs the best with the average improvements being 8.87% for MIA, 40.74% for SI, 1.18% for SMI, 25.90% for CDI, and 36.05% for WCBCR respectively.

2) *Outliers detection*

As noted in Section I, the outliers are to be isolated from the regular patterns during the clustering process, and generally they will appear as a series of small-size or singleton clusters with distinctive shapes in the clustering results. For the dataset in this study, the outliers can be characterized by severe fluctuations even with massive zero consumption readings.

To estimate how many outliers actually “exist” in the dataset, the following index is calculated for each pattern $\vec{x}$:

$$e=\sum_{i=1}^{d-1}\left|\vec{x}(i+1)-\vec{x}(i)\right| \tag{5}$$

Obviously, most of the normal patterns with flat shapes always have small values of *e*, while for the outliers, the values of *e*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2015.2479941, IEEE Transactions on Power Delivery

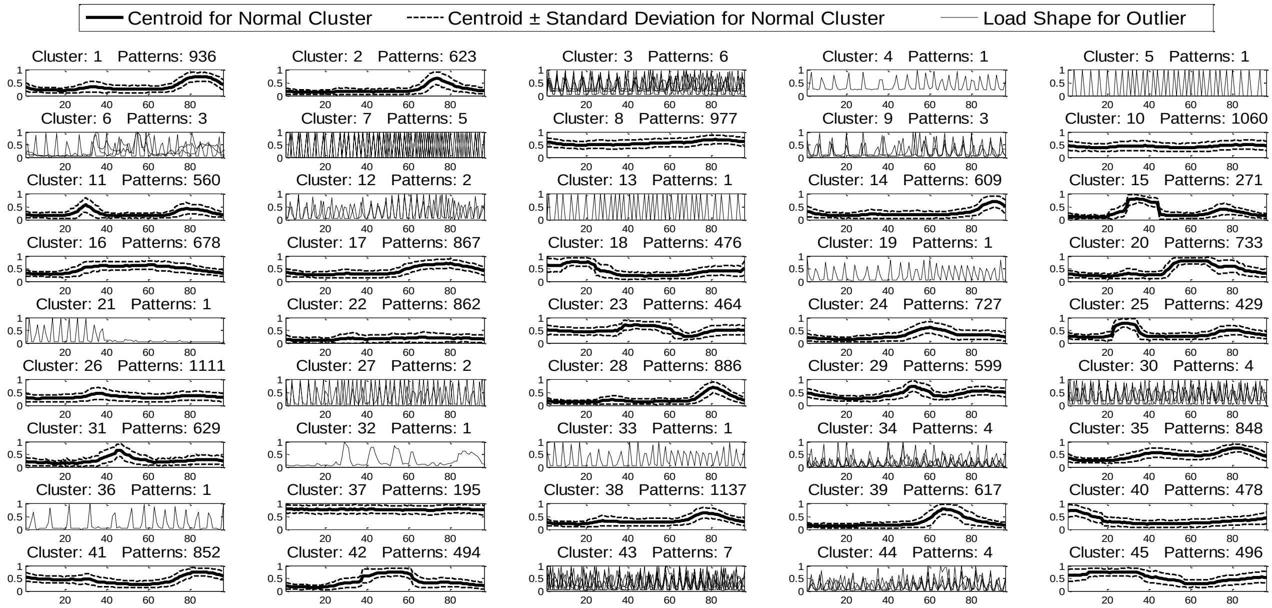

Fig. 4. Clustering results for H-K-means (*L* = 6) at *K* = 45 with the best detection of outliers. Horizontal axis: quarters of hour. Vertical axis: per-unit power.

TABLE III
NUMBER OF ISOLATED OUTLIERS FOR THE DISCUSSED METHODS

Method	Number of clusters															
	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Classical K-means	5	7	9	6	10	5	6	10	2	8	7	1	11	11	10	10
Fuzzy K-means	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K-means++	6	10	9	6	1	4	7	10	7	12	9	8	7	14	12	10
WFA- K-means	10	14	9	10	11	24	13	12	16	11	9	10	14	14	11	10
Developed K-means	5	2	5	5	5	5	5	2	6	2	2	10	3	5	2	5
H-K-means (3 levels)	22	27	20	22	24	23	23	24	25	28	21	23	31	24	32	28
H-K-means (4 levels)	26	35	31	33	36	32	38	35	34	34	40	36	40	39	33	44
H-K-means (5 levels)	36	**39**	38	37	41	**46**	38	39	41	41	43	**47**	**44**	39	**55**	**54**
H-K-means (6 levels)	**41**	**39**	**41**	**42**	**48**	40	**42**	40	**49**	**43**	**48**	46	43	**52**	47	48
H-K-means (7 levels)	38	38	34	35	37	37	39	**41**	37	39	37	39	**44**	38	39	43

usually become extremely large. As a result, 18588 *e* values distribute evenly in [0,25], while the other 74 ones disperse in [25,69]. However, some normal patterns with multiple peak periods may also have considerable *e* values. Hence, the real number of the outliers should be close but lower than 74.

As an example, Fig. 4 shows the clustering results of H-K-means at *K* = 45 with the best performance for the detection of the outliers, in which 48 isolated outliers have been shown directly in clusters #3-7, #9, #12, #13, #19, #21, #27, #30, #32-34, #36, #43, while the other normal clusters are represented by the centroids and the standard deviation curves.

Table III shows the performance of the discussed methods in terms of the number of isolated outliers from *K* = 35 to 50. The best detection rate of the H-K-means methods is 74.3% (55/74), compared to 32.4% by WFA-K-means, 18.9% by K-means++, 14.9% by classical K-means, 13.5% by Developed K-means, and 0% by fuzzy K-means. We observe that the H-K-means methods have consistently outperformed all the other methods especially for the case with 6 hierarchical levels. The reason for this improvement can be explained by the selection of the initial centroids. Since the total amount of the outliers is relatively small, it is usually difficult for traditional K-means based methods to detect them from the original dataset during the random selection process for the initial centroids, and thus producing the corresponding poor performance. Whereas the features of these outliers can be preserved during the data "simplification" process of the H-K-means method, a set of patterns with uncommon shapes also appear in the top-level dataset. Fig. 5 shows the patterns in the 6_{th}-level dataset, about 20 outliers exist in the 210 patterns, which would increase the probability of obtaining small or singleton clusters in the final results, and furthers the outliers' isolation.

3) Computation time

Table IV shows the comparison of all the discussed methods in terms of the average, maximum and minimum computation times for all the executions. Each value given in the table is the

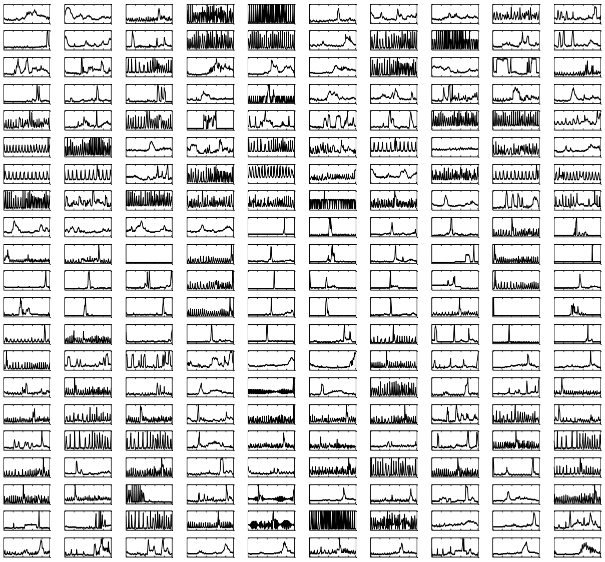

Fig. 5. Patterns in the 6_{th}-level dataset of the hierarchical structure.

TABLE IV
RELATIVE COMPUTATION TIME FOR THE DISCUSSED METHODS BASED ON CLASSICAL K-MEANS

Method	Relative indices for computation time		
	Average	Maximum	Minimum
Fuzzy K-means	8.6743	8.3744	6.6331
K-means++	8.6934	7.1561	6.7921
WFA- K-means	1.9204	1.8098	1.5920
Developed K-means	1.6367	1.6754	1.6060
H-K-means (3 levels)	1.5296	1.6395	1.3939
H-K-means (4 levels)	1.5129	1.7381	1.4468
H-K-means (5 levels)	**1.4220**	**1.3877**	1.4367
H-K-means (6 levels)	1.4561	1.4516	**1.1237**
H-K-means (7 levels)	1.5141	1.4194	1.7820

relative value which consists of the ratio of the average (or maximum, minimum) computation time for the method under test to the same factor for classical K-means. Notice that all the programs in this work were executed with Microsoft Windows 7, 3.10 GHz processor, 3.24 GB of RAM, and the actual computation times of classical K-means are 12.0260, 23.6230, and 6.2330 seconds for average, maximum and minimum values respectively. Obviously, the comparison shows that the H-K-means methods have preserved the speed advantage of classical K-means to the greatest extent over the other methods.

Overall, the H-K-means method with 6 hierarchical levels is considered to be the most suitable one for the analysis of the given AMI dataset by virtue of its promising performances in the aforementioned three respects.

V. CONCLUSIONS

The proliferation of AMI measurement has resulted in a huge increase in the amount of available load data in distribution systems, making the clustering task quite challenging. This paper presents a novel clustering method, the H-K-means method to enhance the performance of classical K-means for large dataset. For the task of clustering a large dataset, this method begins with "simplify" the given dataset to a manageable one, then restores it back to the original one gradually with the succession of high-quality initial centroids, and finally provides promising clustering results.

To evaluate the effectiveness of the proposed H-K-means method, we apply it to a large-scale AMI dataset consisting of 18662 daily patterns. The following advantages of the proposed H-K-means method can be observed based on the comparisons between it and several existing K-means based methods:

1) It achieves the best compactness and distinctiveness respectively in the corresponding clustering results.
2) It exhibits an outstanding capability for detecting the outliers.
3) It needs little extra computation time compared with classical K-means.

From a practical viewpoint, we believe several application areas will surely benefit from the promising clustering results, and our future study will focus on the following points:

1) Since the applied dataset has covered a relatively long period of time, the clustering results are applicable to the study of load prediction, demand response, tariff design. Of course, the scale of the applied dataset should be higher for reliable results if more customers are needed to be analyzed. We expect that the proposed H-K-means will then become more effective than the other K-means variants because of the increasing difficulty of finding high-quality initial centroids among the original dataset.
2) The customers' behaviors are changeable. To ensure the timeliness of the clustering results, new patterns can be assigned to the obtained clusters by an appropriate classification tool [9], and a re-clustering process will then be conducted once some performance indices such as the classification error (RMSE) is unacceptable.

REFERENCES

[1] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," *IEEE Trans. Power Syst.*, vol. 18, pp. 381–387, Feb. 2003.

[2] E. W. S. dos Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. de Souza, "Detection and Identification of Abnormalities in Customer Consumptions in Power Distribution Systems," *IEEE Trans. Power Del.*, vol. 26, no. 4, pp. 2436–2442, Oct. 2011.

[3] A. H. Nizar, Z. Y. Dong, and Y. Wang, "Power utility nontechnical loss analysis with extreme learning machine method," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 946–955, Aug. 2008.

[4] G. Nourbakhsh, G. Eden, D. McVeigh, and A. Ghosh, "Chronological Categorization and Decomposition of Customer Loads," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 2270–2277, Oct. 2012.

[5] A. Mutanen, M. Ruska, S. Repo, and P. Järventausta, "Customer Classification and Load Profiling Method for Distribution Systems," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1755–1763, Jul. 2011.

[6] H. L. Willis, A. E. Schauer, J. E.D. Northcote-Green, and T. D. Vismor, "Forecasting distribution system loads using curve shape clustering," *IEEE Trans. Power App. Syst.*, vol. PAS-102, no. 4, pp. 893–901, 1983.

[7] S. Valero, M. Ortiz, C. Senabre, C. Alvarez, F. J. G. Franco, and A. Gabaldon, "Methods for customer and demand response policies selection in new electricity markets," *IET Gen., Transm., Distrib.*, vol. 1, no. 1, pp. 104–110, 2007.

[8] R. F. Chang and C. N. Lu, "Load profiling and its applications in power market," in *Proc. IEEE Power Eng. Soc. General Meeting*, Jul. 13–17, 2003, vol. 2.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2015.2479941, IEEE Transactions on Power Delivery

[9] J.-H. Shin, B.-J. Yi, Y.-L. Kim, H.-G. Lee, and K.-H. Ryu, “Spatiotemporal Load-Analysis Model for Electric Power Distribution Facilities Using Consumer Meter-Reading Data,” *IEEE Trans. Power Del.*, vol. 26, no. 2, pp. 736–743, Apr. 2011.

[10] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, “Load pattern-based classification of electricity customers,” *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1232–1239, May 2004.

[11] G. Chicco and I. S. Ilie, “Support vector clustering of electrical load pattern data,” *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1619–1628, Aug. 2009.

[12] G. Chicco, R. Napoli, and F. Piglione, “Comparisons among clustering techniques for electricity customer classification,” *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 933–940, May 2006.

[13] G. Chicco, R. Napoli, and F. Piglione, “Application of clustering algorithms and self organising maps to classify electricity customers,” in *Proc. IEEE Power Tech Conf.*, Bologna, Italy, Jun. 23–26, 2003.

[14] T. F. Zhang, G.. Q. Zhang, J. Lu, X. P. Feng, and W. C. Yang, “A New Index and Classification Approach for Load Pattern Analysis of Large Electricity Customers,” *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 153–160, Feb. 2012.

[15] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, “An electric energy consumer characterization framework based on data mining techniques,” *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 596–602, May 2005.

[16] N. M. Kohan, M. P. Moghaddam, S. M. Bidaki, and G. R. Yousefi,“Comparison of modified k-means and hierarchical algorithms in customers load curves clustering for designing suitable tariffs in electricity market,” in *Proc. 43rd Int. Universities Power Engineering Conf.*, Padova, Italy, Sep. 1–4, 2008, pp. 1–5.

[17] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, “Two-stage pattern recognition of load curves for classification of electricity customers,” *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1120–1128, Aug. 2007.

[18] D. Gerbec, S. Gasperic, and F. Gubina, “Determination and allocation of typical load profiles to the eligible consumers,” in *Proc. IEEE Power Tech Conf.*, Bologna, Italy, Jun. 23–26, 2003.

[19] A. Notaristefano, G. Chicco, F. Piglione, “Data size reduction with symbolic aggregate approximation for electrical load pattern grouping,” *IET Gen., Transm., Distrib.*, vol. 7, no. 2, pp. 108–117, 2013.

[20] G. Chicco, G. Chicco, J. Akilimali, “Renyi entropy-based classification of daily electrical load patterns,” *IET Gen., Transm., Distrib.*, vol. 4, no. 6, pp. 736–745, 2010.

[21] J.-S. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Trans. Smart Grid,* vol. 5, no. 1, pp. 420-430, Jan. 2014.

[22] M. Romero, L. Gallego and A. Pavas, “Fault Zones Location on Distribution Systems Based on Clustering of Voltage Sags Patterns,” in *Proc. 15th International Conference on Harmonics and Quality of Power*, 2012.

[23] M. Romero, L. Gallego and A. Pavas, “Estimation of voltage sags patterns with k-means algorithm and clustering of fault zones in high and medium voltage grids,” *Ingenier ú e Investigaci ón*, vol. 31, 2011.

[24] B. Chen, P.-C. Tai, R. Harrison, and Y. Pan, “Novel Hybrid Hierarchical-K-means Clustering Method (H-K-means) for Microarray Analysis”, in *Proc. IEEE Computational Systems Bioinformatics Conf.*, pp. 105–108, 2005.

[25] Y.-C. F. Wang and D. Casasent, “Hierarchical K-means Clustering Using New Support Vector Machines for Multi-class Classification”, in *Proc. International Joint Conference on Neural Networks*, Vancouver, BC, Canada, 2006.

[26] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, 1967, pp. 281–297.

[27] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, “Allocation of the load profiles to consumers using probabilistic neural networks,” *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 548–555, May 2005.

[28] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, “Determining the load profiles of consumers based on fuzzy logic and probability neural networks,” *Proc. Inst. Elect. Eng., Gen., Transm., Distrib.*, vol. 151, pp. 395–400, May. 2004.

[29] S. Nasser, R. Alkhaldi and G. Vert, “A Modified Fuzzy K-means Clustering using Expectation Maximization,” in *Proc. 2006 IEEE International Conf.*, pp. 231-235.

[30] I. P. Panapakidis, T. A. Papadopoulos, G. C. Christoforidis, G. K. Papagiannis, “Pattern recognition algorithms for electricity load curve analysis of buildings,” *Energy Build*. vol. 73, pp. 137–145, 2014.

[31] D. Arthur, S. Vassilvitskii, “k-means++: the advantages of careful seeding,” in *Proc. of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, NewOrleans, Louisiana, USA, 2013, pp. 1027–1035.

[32] N. Mahmoudi-Kohan, M. P. Moghaddam, M. K. Sheikh-El-Eslami and S. M. Bidaki, “Improving WFA K-means Technique for Demand Response Programs Applications,” in *Proc. 2009 IEEE Power and Energy Society General Meeting*, 2009.

[33] B. D. Pitt and D. S. Kirschen, “Application of data mining techniques to load profiling,” in *Proc. IEEE PICA*, Santa Clara, CA, May 16–21, 1999, pp. 131–136.

[34] D. Hand, H. Manilla, and P. Smyth, *Principles of Data Mining*. Cambridge, MA: MIT Press, 2001.

Tian-Shi Xu received the B.Sc. degree in electrical engineering and the automation from Tianjin University, Tianjin. China, in 2013. He is pursuing the M.Sc. degree at the School of Electrical Engineering and the Automation in Tianjin University. His areas of interest include power system and distribution system analysis, load management, computational techniques and artificial intelligence applications.

Hsiao-Dong Chiang (F’97) is currently Professor of Electrical and Computer Engineering at Cornell University, Ithaca, NY, USA. He and his colleagues have published more than 350 referred papers and have been awarded 14 patents arising from their research and development work, both in the United States and internationally. He was an associate editor IEEE Transactions on Circuits and Systems (1990-91), (1993-1995). He holds 17 U.S. and oversea patents and several consultant positions. He is Author of the book “Direct Methods for Power System Stability Analysis: Theoretical Foundation, BCU Methodology and Applications”, John Wiley & Sons, 2011 and of the book “Stability region of nonlinear dynamical system: theory, optimal estimation and applications”, Cambridge Press, 2015.

Guang-Yi Liu is a Senior Engineer and Deputy Head of Energy Management Systems Group at Electrical Power Research Institute, China. His current interests are the development of a new generation of EMS for use in a restructured electricity industry.

Chin-Woo Tan is currently Director of Stanford Smart Grid Lab. He has research and management experience in a wide range of engineering applications in intelligent sensing systems, including electric power systems, automated vehicles, intelligent transportation, and supply chain management. His current research focuses on developing data-driven methodologies for analysing energy consumption behaviour, and seeking ways to more efficiently manage consumption and integrate distributed energy resources into the grid. Dr. Tan was a Technical Lead for the LADWP Smart Grid Regional Demonstration Project, and a project manager with the PATH Program at UC Berkeley for 10 years working on intelligent transportation systems. Also he was an Associate Professor with the Electrical Engineering Department at California Baptist University. He holds a PhD and BS in Electrical Engineering, and a MA in Mathematics, all from UC Berkeley.