Identifying susceptible consumers for demand response and energy efficiency policies by time-series analysis and supplementary approaches

Mohsen Kojury-naftchali
School of electrical and computer engineering, university of tehran
Tehran, Iran
Email: kojury.savadkoohnaft@gmail.com

Alireza Fereidunian
Electrical engineering faculty, k.n.toosi university of technology
Tehran,Iran
Email: fereidunian@eetd.kntu.ac.ir

Hamid Lesani
School of electrical and computer engineering, university of tehran
Tehran, Iran
Email: lesani@ut.ac.ir

Abstract—**One of the most important concepts in present restructured market for utility companies, depends on having appropriate knowledge about customer consumption patterns. This information is necessary whereby, they will have a proper situation for better monitoring customers and optimum planning for future. In recent years, with advent of Advanced Metering Infrastructure (AMI), which created a proper communication infrastructure, the quality of monitoring of both suppliers and customers, had enhanced. The goal of this paper is acquiring more complete and accurate information about customer's behavior in consumption to Identifying susceptible consumers for demand response(DR) and energy efficiency(EE) policies. This goal will be achieved by mixing data-mining algorithms for time-series analyzing, with entropy index and apply them simultaneously. The susceptibility of customers for proposing tariff is analyzed too. For evaluation of the process, a real dataset, which relates to Irish social science data archive (ISSDA) is used.**

Keywords-component; Demand Response, Energy Efficiency, AMI, Entropy, Time-seies, Data-mining

I. INTRODUCTION

In present restructured electricity market, all of the utility companies, try to service customers with the best range of quality and the least cost payment for losses emerging in several aspects. It is known, there are always some constraints in generation and transmission of electricity energy produced by these companies, and the present competitive condition of market, complicates these constraints too. So, utility companies need applying substitute strategies minimizing these costs.

Energy efficiency(EE) and demand response(DR) are two famous policies, whose performance could show satisfactory and persuasive results, so, for this case, are of interest. Appropriate implementation of these programs necessitates sufficient knowledge about consumer's behavior in electricity energy usage, because they are end-users of This energy. For this reason, in recent years, technologies enabling capability of better monitoring of grid, has progressed quickly. One of this up to date technologies, is Advanced Metering Infrastructure(AMI). Bidirectional Communication capability, prepared by emersion of AMI, not only reduces a big amount of expenditure related to old gathering data methods and many other factors, but also reinforces communicational novel space of the utility companies and energy consumers in restructured electricity market. AMI data receive from several resources and time-interval between this records is short and near real-time, so this resolution of data, culminates in a worthwhile data-warehouse which is the best requirement for consumers profiling and being aware of their consumption patterns. Information concluded of analyzing these dataset, could help utility companies for more accurate load forecasting, expenditure reduction, short-term, mid-term and long-term planning and will be useful in the operation of system. because of creating a bilateral monitoring space between consumption and supplying side of the grid, AMI can give them more option for contribution in the market. As a result, the importance of this issue is obvious.

Research on customer profiling and segmentation, specially in recent years, represent the significance of analyzing customer's usage pattern. considerable percent of these papers, used data-mining algorithms to discover patterns which exist in data behavior. All of requirements of process related to knowledge discovery and data-mining is produced in [1]. [2] aims to segment users, and for this reason proposes to infer occupancy states from consumption time-series data using a Hidden Markov Model framework. [3] aims to investigate a household electricity segmentation methodology that uses an encoding system with a pre-processed load shape dictionary. In [4], a method for short-term load forecasting is discussed. This methodology is based on periodic time-series. Demand response is analyzed and scheduled by SCUC in [5]. Data-mining time-series data for smart meter is studied in [6]. In [7], the smart meter data used for segmenting residential consumers for energy efficiency and demand response program targeting is discussed. Discussion about short-term load forecasting and enhancing the accuracy of forecasting model for demand response by anthropologic structural variables is presented in[8]. Climate effects on consumption patterns is considered in [9] and data-mining techniques as tool of this analyze too. In [10],[11] clustering algorithms are applied and evaluated to customer segmentation. Determining the load profiles of consumers based with fuzzy logic and probability neural networks is discussed in[12]. In [13], classification of electricity consumer is investigated. In this research, indexes for classification and indexes for adequacy of clustering is proposed.

In this paper, authors aim to identify consumers which are susceptible for demand response and energy efficiency programs. For achieving better insight about consumer behavior, we used of data-mining algorithms and analyzing time-series features simultaneously. In addition, discussion on consumer consumption behavior and tariff proposing is presented too.

II. BACKGROUND

A. Time series

Time-series consist of sequences of values or events, acquired over repeated measurement of time. This values are recorded at equal time intervals [1]. Some idioms define for time-series are as follows:

Trend: movements in up or down direction of time-series, after noticeable amount of time intervals. Trends show the incremental or decremental behavior of time-series. Fig.1 shows a sample of time-series which illustrates price versus time. In this figure, the trend is specified with dashed curve.

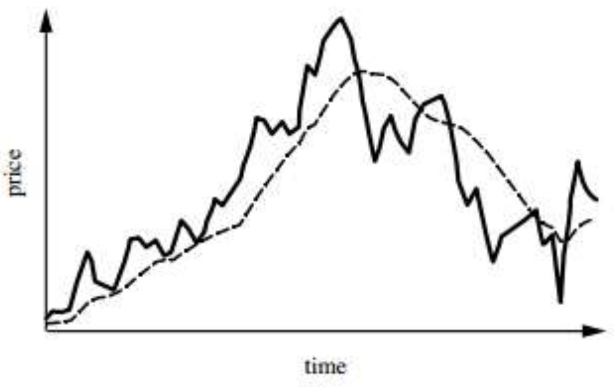


Figure 1. Trend in time-series[1]

Seasonality: seasonality or seasonal variations, occurs in alternative periods and it means, similar behavior of time-series in same times. This periods of time could be daily ,weekly ,monthly and so on.

Stationary time-series: if there is no up or down movement in time-series after noticeable amount of time intervals, the time-series are stationary. It means, in a long time and after several periods, the average of these kind of time-series is the same, and values of the profile, fluctuate around average value.

This paper aims to analyze time-series related to consumer's electricity energy, and by clustering them into the clusters, identify similar ones. Similarity between time-series means: data sequences that differ a little from the specified query sequences [1]. So, if time-series related to different individual consumers, show similar behavior in their movements and sequences, it means, the electricity usage patterns of these two sample is similar. In this process, when the cluster which an individual consumer is assigned to it, changes from one clustering phase to next one, it means, there is a variation in consumer energy usage behavior. So, repeating this process for several times, will help us to identify each individual consumer's variability in consumption patterns and assess the susceptibility of them for demand response and energy efficiency policies.

B. Demand response

According to the federal energy regulatory commission, demand response defines as such: changes in electric usage by end-use consumers from their normal consumption patterns, in response to changes in the price of electricity over time, or incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized. This definition shows, demand response programs are created to either decrease electricity consumption in peak times or evaluating consumer reaction and response to fluctuation of market prices. But it's major attention will be assigned to shift consumption of peak times into off-peak ones and the financial motivation could be appropriate interface. Demand response policies could be implemented in two ways, time-based demand response programs and incentive-based demand response programs [5]. Time-based demand response is related to changes in consumer's consumption patterns, responding to variation of market prices. It is divided into critical-peak pricing, real-time pricing and time of use tariffs. it is expected, such situation, induce consumers to reduce their electricity consumption in time of peak-load, or make changes in their timing of electricity consumption patterns. In incentive-based demand response programs, the consumers are offered to receive payments and decrease their consumption instead. This policy can be followed by satisfactory outcomes, especially when the reliability of grid is threated.

C. Energy efficiency

Energy efficiency means: reducing the amount of energy which is necessary for producing a product or service, without reduction in quality of service. for example, this policy aims to encourage residential consumers, using appliances with more efficient rate of energy, or industrial ones, substituting timeworn machines or equipments with new ones, and so on. Such actions in different parts, will reduce the total energy usage and restrain increasing of payment related to each individual consumer.

D. knowledge discovery steps

according to [1], The process of knowledge discovery, could be illustrated in following sequences:

1. data cleaning. Removing noise and inconsistent data
2. data integration. Combining data coming from multiple data sources
3. data selection. Selecting appropriate features

4. data transformation. Normalizing data for appropriate intended process

5. data mining. Extracting patterns which exist in dataset by using intelligent algorithms.

Fig.2 shows steps of knowledge discovery process Schematically.

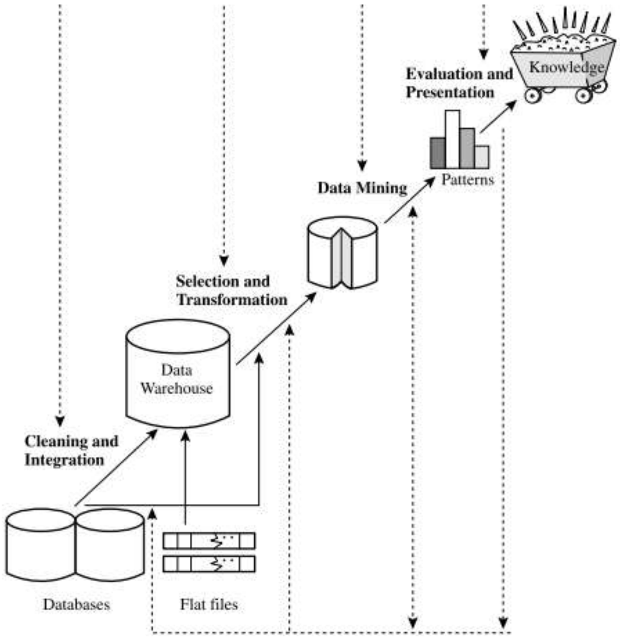


Figure 2. Steps of knowledge discovery, Data-Mining is of these steps[1]

In our purpose, which concerns about discovering patterns of consumer usage, classification and clustering are two popular practical instruments. So we should know more about them.

1) Classification

Classification is a supervised learning algorithm, allocating each sample to the classes which has defined before. By using data which has labeled before, the system learns(for example neural network which will be learned by samples that have labels) and makes a model which is able to dedicate new sample to one of these existing classes. Existence of labeled training data causes this process be supervised learning. It means there is a prior knowledge about data which are used for training system. For classification, algorithms such as decision tree and neural network could be appropriate choices.

2) Clustering

Unlike classification, clustering is an unsupervised learning algorithm. In this case, there is no prior knowledge about groups and labels of samples. In fact, clustering aims to divide existing space of samples to some subspace and clusters. For each individual cluster, we have a cluster center. Samples in each individual cluster should have the most similarity to each other and be dissimilar to samples assigned in other clusters. Similarity and dissimilarity of two samples, specifies by calculating distance between them. The higher distance implies more dissimilarity and lower distance implies more similarity. For this reason, there are some equations for calculating distances and a famous one is Euclidean distance. To define the Euclidean distance we can say: If we have two samples $X(x_1, x_2, ..., x_N), Y(y_1, y_2, ..., y_N)$, with N features , the Euclidean distance is as follows:

$$d(X,Y) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \quad (1)$$

The best clustering is performed, if samples of each individual cluster, have the minimum distance to their cluster center, which is in accordance with more compacted cluster, and should have the highest feasible distance to other cluster centers. Some of the more popular clustering algorithms are: k-means, hierarchical, SOM and so on.

III. MEHODOLOGY

The methodology for acquiring appropriate information about customer’s usage patterns and subsequent actions in accordance with this patterns, is simultaneous mixed analyzing of time-series consumption data and entropy index. So we will be able to have good insight about patterns of energy usage and make proper decisions for both issues; DR and EE. Fig.3 shows the sequences of process schematically.

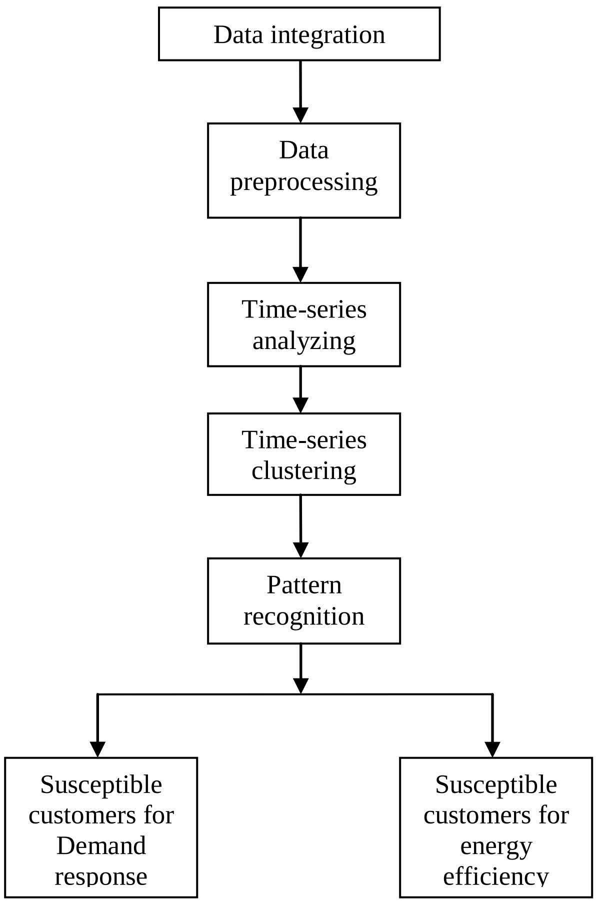


Figure 3. Sequenses of process of custumer’s identification

IV. CASE STUDY

To evaluate results of process which descripted in above sections, we tested it on a real dataset which is received from Irish social science data archive. This anonymized dataset is consumption data relating to 100 residential consumers for six months. These data are recorded with 30 minutes intervals. It means, for each customer, for each 24 hour, we have 48 consumption record. As said before, seasonality in time-series is a feature that might be seen in some cases, but if we restrict the period of data analyzing, we will be able to restrict and dedicate seasonality to a specific subsection of time, for example: weekly or monthly seasonality. In our study, the whole period of analyze is six months. for achieving pervasive results, we divide this time to six sections and apply our procedure to each one of these months respectively.

A. Process of time-series clustering

Dataset which used for knowledge discovery is consumption data time-series. As said before, for each 24 hour of a day, there is 48 record for each individual sample. So, our profile showing time-series, consists of 48 consumption record versus 24 hour. Here, each customer is displayed by a representative time-series. Clustering time-series with this dimension for all clustering algorithm is very hard to handle, So, it is necessary to reduce dimensions of each sample (customer). For this reason, the Principal Component Analysis (PCA) which uses Singular Value Decomposition (SVD) process is applied to do this. SVD presents eigenvalues and eigenvectors produced by this eigenvalues. By applying PCA algorithm, the dimension of data for each individual sample reduced to 10 numbers. By Using dimension reduction algorithms, because of reduction in whole size of dataset, not only speed of analyzing will increase and causes conservation in time, but also results of analyze and rules which is made by these results, are easier to be understood.

In clustering process, the number of clusters, usually is one of the challenging issues. So, some indexes are produced to facilitate and handle this problem. One of the well known indexes for this reason is Davise-Bouldin (DB) validity index [4].

This index is a function of the ratio of the sum of within-cluster scatter to between-cluster separation [4]. For clusters denoted $Q_i, i = 1, 2, ,..., N_c$ the DB index is

$$DB = \frac{1}{N_c} \sum_{J=1}^{N_C} \max_{l \neq j} \frac{S(Q_j) + S(Q_l)}{d(Q_j, Q_l)} \tag{2}$$

Where S(Q_K) is the (average) distance within cluster Q_K and $d(Q_j, Q_l)$ is the distance between clusters Q_j and Q_l . The optimal number of clusters N_c is the one for witch the DB validity index shows a minimum value [4].

In our study, each period of consumption which is analyzed, is one month and the whole testing period is six months, as a result, calculating BD index in six times will be applied. Fig.4 Shows the result of applying this index two our data set.

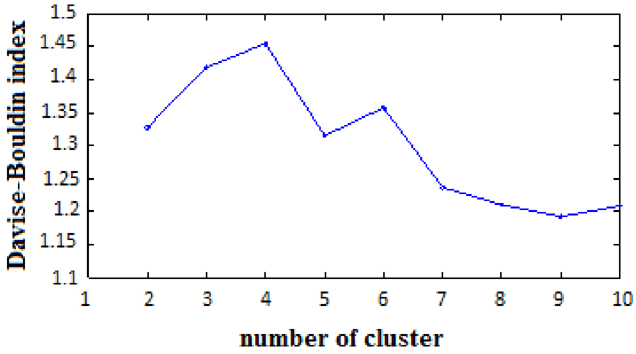


Figure 4. Davise-Bouldin index plot versus number of cluster

This figure shows, the best number of clusters for this dataset could be 9 clusters. Measures of DB index for this first clustering (first month of six) for different numbers of clusters are listed in tableI.

TABLE I. DAVISE-BOULDIN INDEX FOR EACH NUMBER OF CLUSTER

Number of clusters	Davise-Bouldin index
2	1.327
3	1.4179
4	1.4542
5	1.315
6	1.3574
7	1.2369
8	1.2102
9	1.1921
10	1.2092

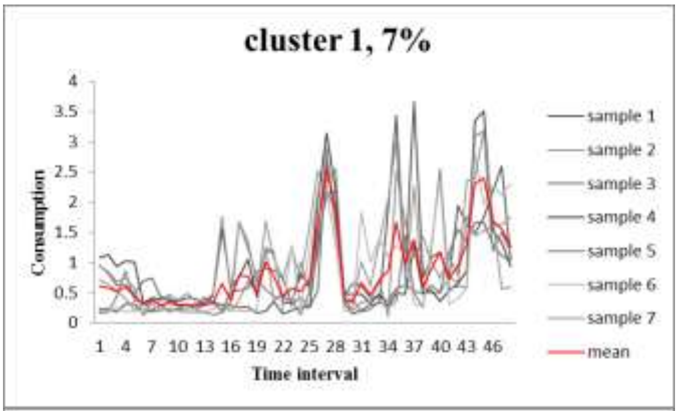
As such, For 5 subsequent clustering related to next 5 month, we have such a figure but with other optimum number of clusters. These cluster numbers are listed in table II.

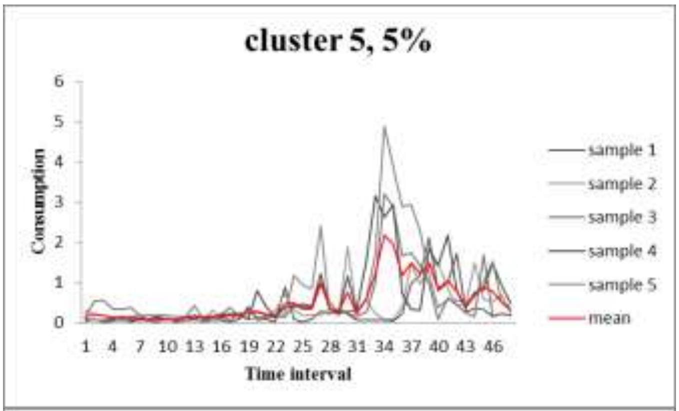
TABLE II. OPTIMUM NUMBER OF CLUSTERS FOR EACH CLUSTERING PHASE

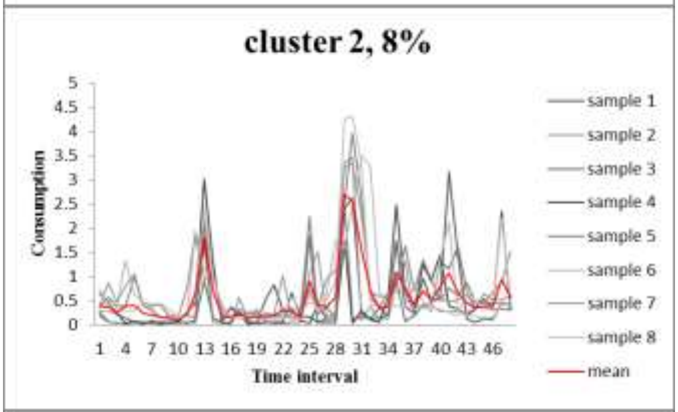
Clustering phase	Optimum number of clusters
2	9
3	9
4	10
5	9
6	8

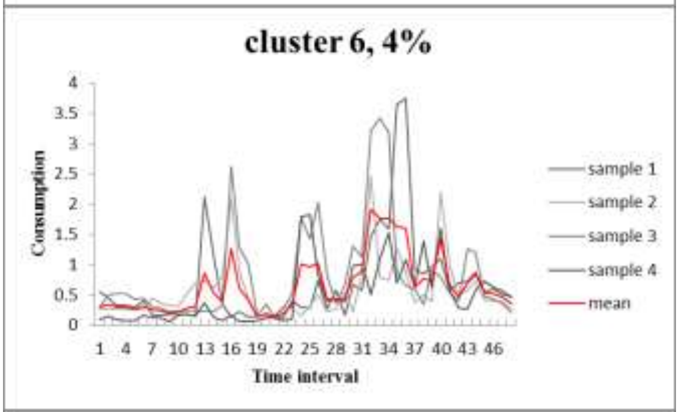
B. Clustering results and trend analysis

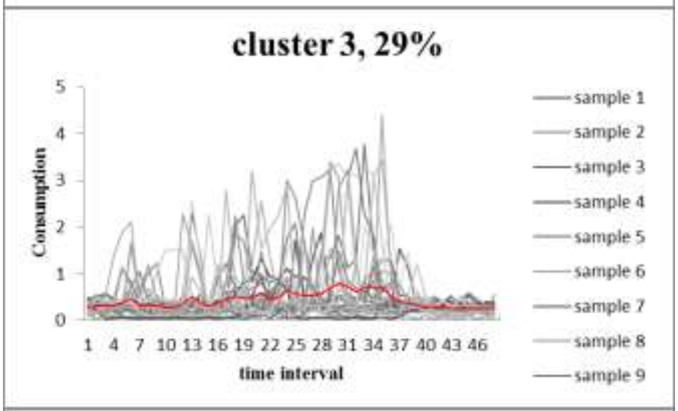
After dimension reduction in time-series data size, clustering process is implemented and shapes which are shown in Fig.5, are achieved. In each cluster, there are similar load profiles, and for each cluster, there is a representative for load profiles of that cluster which is specified by red color. The percent of members of each cluster is specified on the top of each figure.

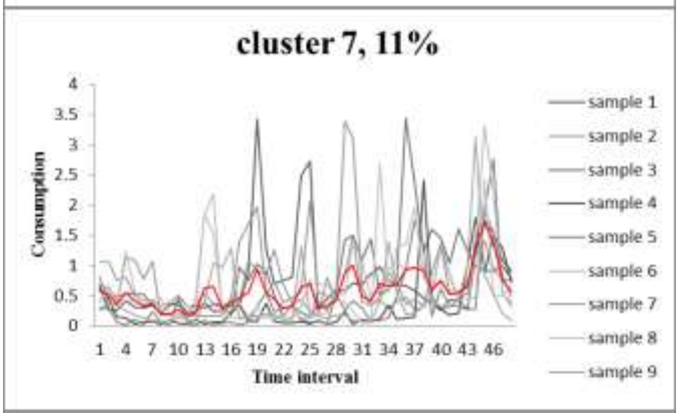

cluster 1, 7%
Consumption
Time interval
sample 1
sample 2
sample 3
sample 4
sample 5
sample 6
sample 7
mean

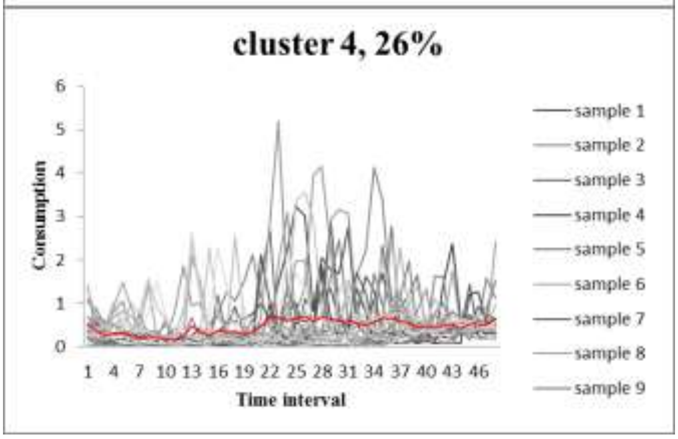

cluster 5, 5%
Consumption
Time interval
sample 1
sample 2
sample 3
sample 4
sample 5
mean

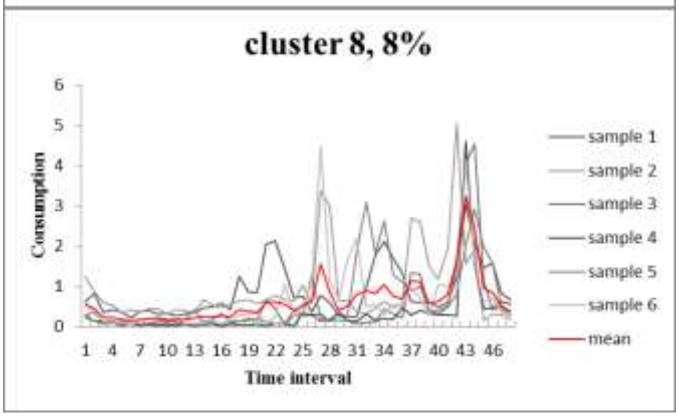

cluster 2, 8%
Consumption
Time interval
sample 1
sample 2
sample 3
sample 4
sample 5
sample 6
sample 7
sample 8
mean


cluster 6, 4%
Consumption
Time interval
sample 1
sample 2
sample 3
sample 4
mean


cluster 3, 29%
Consumption
time interval
sample 1
sample 2
sample 3
sample 4
sample 5
sample 6
sample 7
sample 8
sample 9


cluster 7, 11%
Consumption
Time interval
sample 1
sample 2
sample 3
sample 4
sample 5
sample 6
sample 7
sample 8
sample 9


cluster 4, 26%
Consumption
Time interval
sample 1
sample 2
sample 3
sample 4
sample 5
sample 6
sample 7
sample 8
sample 9


cluster 8, 8%
Consumption
Time interval
sample 1
sample 2
sample 3
sample 4
sample 5
sample 6
mean

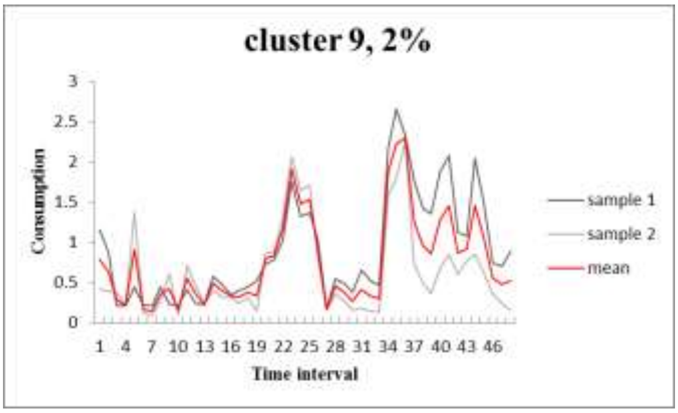


Figure 5. Customer's load profiles which assigned to each cluster

C. Discussion on entropy

The major purpose of this paper is detection and analyzing variability of consumers behavior in their electricity consumption patterns. In fact, the more accurate knowledge about changes in consumers behavior could show more persuasive results in implementation phase of demand response and energy efficiency programs.

Another concepts which can help us for this reason, is analyze of entropy in electricity consumption profiles. Entropy produces a criteria, showing changes in consumers consumption. According to [3], for each household *n*, the relative frequency $p_n(C_i)$ of each cluster center C_i .so the entropy of household is calculated by following equation:

$$S_n = -\sum_{i=1}^{k} p(C_i) \log p(C_i) \tag{3}$$

This parameter can be calculated for each clustering phase. the study period, for each subsection could be Daily, weekly and monthly during time. In our study, this period is monthly, it means we have six study phases. Consumers usually have regular energy usage in a week. As said before, entropy of electricity energy consumption related to an individual sample, shows customer's uncertainty for being assigned to a cluster with high confidence in a clustering process. If we look at the formula of entropy, it is clear that logarithmic phrase would be minimized, if $p(C_i)$ be bigger. In other words, customer whose entropy is low, in clustering process, will be assign to a cluster with a high confidence and it means, these samples are close to the cluster center which they are assigned to it. But this formula for another sample will be maximized, if this sample does not have stable behavior. It means, in clustering process, this sample is close to any one of the cluster centers noticeably, and this customer will be known as a sample with variable consumption pattern. If we map entropy related to each individual consumer, for the first clustering phase, the following distribution of entropy (Fig.6) is acquired.

According to This shape, the entropy of significant percent(82%) of under study samples, is higher than average entropy(0.2841). it means for the first phase of clustering, the samples showed reletively high rate of variability in their consumption.

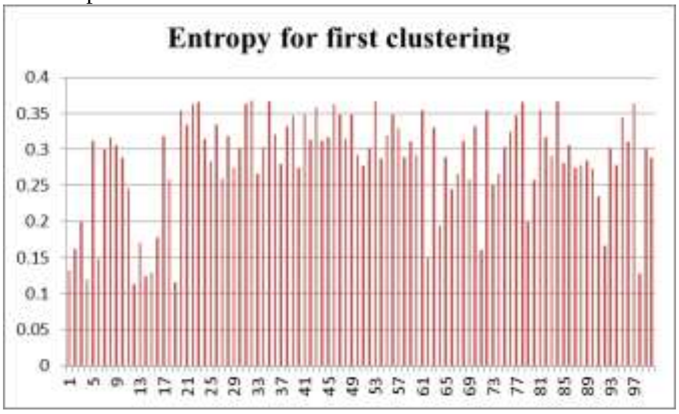


Figure 6. Entropy for under-study customers in the first phase of clustering

V. CONCLUSIONS

Two items for utility companies attract significant attentions, one of them is producing energy with high quality and sufficient enough for customers and another one is minimizing different costs existing in this process. One of the most effective capability which they need to handle both these two issues, is to have more complete information about their customer's behavior in electricity consumption. Because with this knowledge, making decision about other issues would be easier. In this paper we tried to achieve more correct information about consumer's consumption behavior by mixed analyzing time-series and entropy index for identifying proper customers for both DR and EE policies. Using time-series data related to consumption directly for clustering, not only culminates in better insight about patterns of energy usage, but also with the aid of this profiles, variability in time-series could be more tangible.

VI. DISCUSSION

In this study, for identifying customers behavior in consumption, after applying data-mining techniques and clustering samples, analyzing trend of time-series and entropy is implemented. Trend analysis in our case study showed that, the direction of movements in time-series for each group of customers are different. But approximately clusters number 5,6,8 have similar trends, in which there is a slight positive slope in middle part of the profiles and then, returns with a negative one. Cluster number 1 in general has a slightly positive slope. Clusters number 2,3,4,7,9 have a slight positive slop in middle part of the profiles and then a negative one. As said before, high entropy for a sample means; this customer is not a stable one in consumption behavior. According to analysis which is done in [3], of aspect of demand response, customer whose consumption is regular, could be predictable easier and better choices for DR programs. Consumers with stable behavior in energy usage could be proper ones for DR programs. Samples, which showed analogous consumption behavior in the period of our study, are customers with low entropy, so, could be appropriate for demand response policies. In our study, 20% of samples are known as susceptible ones for demand response. Unlike this group, customers with high rate of variability in their consumption behavior or in other words customers with high entropy are proper ones for EE policies. From aspect of proposing tariffs for consumption, two points are important which is proposed in [13]. the first one is compactness of each individual clusters, and another one is segregation of clusters. These two points show quality of clustering. The more compactness of a cluster means, members of this cluster are more similar and more disaggregated clusters show more dissimilarity among these clusters. This situation will be achieved when a sample assigns to a cluster with high rate of confidence, and it means, low rate of entropy in electricity energy consumption. When a customer whose dependency to a cluster is more or this sample is closer (more similar) to the cluster center, as a result, for such consumer, proposing tariff will be more confident and easier. Because this sample with low entropy usually follows approximately analogous consumption pattern and has stable consumption behavior.

ACKNOWLEDGMENT

the data set used in the case study of this research, is prepared from Irish social science data archive(ISSDA), so the authors acknowledge ISSDA for providing this anonymized dataset. Authors acknowledge Dr. Mona Ghasemian of USB, for linking authors to ISSDA, too.

REFERENCES

[1] J. Han, M. Kamber (2006) “Data Mining concepts and techniques” 2nd edition, Morgan Kaufmann publisher

[2] A. Albert and R. Rajagopal “Smart Meter Driven Segmentation:What Your Consumption Says About You”, IEEE Transactions on power systems,Vol.28,Nov,2013

[3] J. Kwac,, J. Flora, and R. Rajagopal,”Household Energy Consumption Segmentation using houerly data”, , IEEE Transactions on Smart Grid, Vol.5,jan,2014

[4] M.spinoza, C.joie, R.belmanse, B.de moor, ‘’Short-term load forcasting, profile identification, and customer segmentation: a metodology based on time-series’’, , IEEE Transactions on Power systems, Vol.20. No.3, AUGUST 2005

[5] M.parvania, M. fotuhi-firuzabad, ‘’Demand response scheduling by stochastic SCUC’’, IEEE Transaction on smart gid, Vol.1, No.1, JUNE 2010

[6] A. lavin, D.klabjan, ‘’Clustering time-seies energy data from smart meters’’, DOI 10.1007/s 12053-014-9316-0

[7] B.Artthur smith, J.wong, R.rajagopal, ‘’ A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting’’, 2012 ACEEE

[8] F.javad, N.Arshad, F. Vallin, I.Vassileva, E.Dahlquist, ‘’Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short-term multiple loads forecasting’’, 2012, Elsevier

[9] V.Figueiredo,F.J.Duarte, F.Rodrigues, Z.Vale,and J.Gouveia et al., “Electric customer characterization by clustering”, in Proc. ISAP,Lemnos, Greece,Sep.2003

[10] S.Ramos, Z.Vale.,”Data Mining Techniques application in power Distribution utilities”,In Proc of the Trasmission and Distribution Conference and Exposition, Chicago, April, 2008

[11] V.Figueiredo,F.J.Duarte, F.Rodrigues, Z.Vale,and J.Gouveia, B.,”An Electric Energy Charactrization Framework based on Data Mining Techniques”. IEEE Transaction on Power Systems, Vol.18,596-602,May,2005

[12] 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.,Vol.151.,May,2004 .

[13] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, C. Toader, ‘’ Customer Characterization Options for Improving the Tariff Offer’’. IEEE Transaction on Power Systems, Vol.18, No.1, FEBRUARY 2003.