Classification of Electricity Customer Groups Towards Individualized Price Scheme Design

Tao Chen*¶, Kun Qian†§, Antti Mutanen¶, Björn Schuller‡§, Pertti Järventausta¶ and Wencong Su*

* Department of Electrical and Computer Engineering, University of Michigan, Dearborn, Michigan, U.S.A.
† Department of Electrical and Computer Engineering, Technische Universität München, München, Germany
‡ Department of Computing, Imperial College London, London, U.K.
§ Chair of Complex and Intelligent Systems, University of Passau, Passau, Germany
¶ Faculty of Computing and Electrical Engineering, Tampere University of Technology, Tampere, Finland

Abstract—This paper introduces classification of electricity residential customers into different groups associated with individualized electricity price schemes, such as time-of-use (TOU) or critical peak pricing (CPP). We use an unsupervised learning method, K-means, assisted by a dimensionality reduction technique and an innovative supervised learning method, extreme learning machine (ELM), to cluster daily load profiles based on hourly AMI measurements. Then, the achieved typical daily load profiles are analyzed and utilized for the design of an electricity price scheme for every subgroup based on symbolic aggregate approximation (SAX). These carefully designed and customized retail price schemes can provide a potential tool for price-based and incentive-based demand response in the Smart Grid context.

I. Introduction

Smart grids have been revolutionizing electrical generation and consumption through a two-way flow of power and information. As an important information source from the demand side, advanced metering infrastructure (AMI) has gained increasing popularity worldwide. For example, in Nordic countries, the Finnish government passed a new act, which states that at least 80% of the customers of each distribution system operator (DSO) must have a smart meter by December 2013, and nowadays in 2017 almost every customer (98%) in Finland is supplied with a smart meter [1]. The abundant data set of electricity consumption of residential customers enables accurate load profiling and data analytic application [2] [3]. Usually, the load profiles refer to electricity consumption behaviors of customers over a specific period, e.g., one day, and can help utility companies understand how electricity is actually used for different customers and obtain the load patterns to provide better customized service.

Traditionally, the information or data set about an individual energy customer's load profile has been unavailable or incomplete. Consequently, research on retail electricity price design usually assumes that all energy customers have very similar electricity usage patterns [4]. Thus, the implemented retail price schemes are designed independent of energy customers' load profiles, even for some demand response (DR) projects. Nowadays, however, with the comprehensive data sets of individual load profiles having been made available, many researchers have found remarkable heterogeneity in energy customers' load profiles [5] [6].

In this paper, instead of focusing on the clustering of different load curves, we mainly focus on the design of individualized electricity price schemes, such as time-of-use (TOU) [7], for different types of residential electric customers based on their classification results. However, the clustering of load profiles still plays a vital role in the reasonable price design. So far, a large number of clustering techniques, including K-means [8], hierarchical clustering [9], self-organizing maps (SOM) [10] and support vector machines (SVM) [11], have already been widely applied in power systems. Most of these do not provide a concrete description of how to utilize the clustering results towards improving electricity services though. Here, we use the simple K-means method combined with a dimensionality reduction technique, principal component analysis (PCA), and a fast efficient supervised learning method, extreme learning machine (ELM) [12] [13], to make the classification of load profiles more reasonable. Then, the achieved typical daily load profiles in every group can serve better for the design of an individualized electricity price scheme, with the help of the symbolic aggregate approximation (SAX) method. At the high level of a distribution network, the proposed method aims to provide a potential tool for price-based coordinated control and future DR programs in a Smart Grid.

II. Framework and Methods

The proposed price scheme design mechanism can be separated into two parts: achieving the typical daily load profiles of every assigned group after classification, and matching a suitable price structure to every typical daily load profile.

A. Classification

1) Data normalization: Data preparation including data cleaning is not the subject of this paper, and will not be discussed. In order to focus on the relative consumption level of specific energy customers and make the load profiles comparable, the normalization process transforms the AMI data, y_{ij}, as shown in (1).

$$y_{ij} = \frac{y_{ij}^{old} - y_{i,min}}{y_{i,max} - y_{i,min}}, \tag{1}$$

where, y_{ij}^{old} denotes the actual electricity consumption for customer i at time j, and $y_{i,max}$ and $y_{i,min}$ denote the minimum and maximum consumption over T periods, respectively.

2) Principal component analysis: PCA is one of the most widely used dimension reduction techniques available. It aims to find a small set of orthogonal variables with manageable reduced dimensionality. These principal components are actually linear combinations of original variables, which represent the variance of the original data set in a low dimensional subspace [14]. The purpose of utilizing PCA is to speed up the convergence of the following clustering algorithm and make the result more robust [15]. More specifically, lower dimensional data will rationalize the clustering of time series based on the Euclidean distance.

3) K-means: The aim of classification is to divide a set of objects into different groups such that objects in the same group are more similar to each other than to those in other groups. K-means, as the most widely used and easily implemented clustering algorithm, will divide the input data set into K groups by their similarity [16]. Consider a data set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ consisting of N independent input vectors with D-dimension. The goal of the algorithm is to partition the data set into K groups. In order to obtain those groups, a set of vectors $\boldsymbol{\mu}_k$, with D-dimensionallity and $k = 1, \ldots, K$, is introduced to indicate centers (centroids) of K clusters. In other words, an assignment of data points to clusters is found, along with a set of vectors $\{\boldsymbol{\mu}_i\}$, to ensure that the sum of the squares of the distances of each data point to its closest vector $\boldsymbol{\mu}_k$ is minimized as in (2). In this paper, $\mathbf{x}_i$ stands for the PCA components of normalized AMI measurements, and $\mathbf{y}_i$, for an input vector.

$$J = \sum_{j=1}^{N} \sum_{i=1}^{K} r_{ij} \parallel \mathbf{x}_j - \boldsymbol{\mu}_i \parallel^2, \qquad (2)$$

$$r_{ij} = \begin{cases} 1 & \text{if } i = \operatorname{argmin}_i \parallel \mathbf{x}_j - \boldsymbol{\mu}_i \parallel^2 \\ 0 & \text{otherwise.} \end{cases} \qquad (3)$$

4) Extreme learning machine: The ELM is a comparably novel learning technology for working with generalized single hidden layer feed-forward neural networks (SLFN) [17]. An SLFN usually includes three layers, which are the input layer, hidden layer, and output layer, as shown in Figure 1. Given a training data set with N samples, the output function of the SLFN with L hidden nodes and the activation function θ is as shown in (4).

$$f_L(x_j) = \sum_{i=1}^{L} \beta_i \theta \left(\omega_i x_j + b_i \right) = t_j, \quad j = 1, 2, \ldots, N \quad (4)$$

ELM distinguishes itself from other conventional iterative learning algorithms because it randomly selects the biases and input weights for hidden nodes, ω and b. Besides, it usually calculates the output weights, θ, analytically by finding a least-square solution. In [17] and [18], the authors theoretically prove that the training error are usually minimized with better generalization performance and higher accuracy.
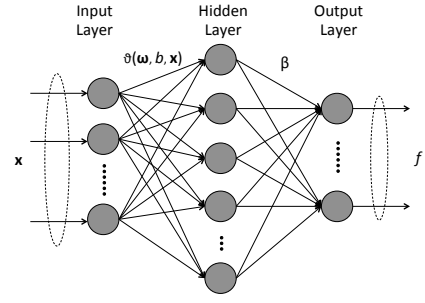

Fig. 1. Structure of an SLFN

B. Price Scheme design

1) Symbolic aggregate approximation: SAX mainly works as a powerful technique for the representation of time series data with lower bounding of the Euclidean distance [19]. Through the following two steps – transforming the load data into a piecewise aggregate approximation (PAA) representation and then symbolizing the PAA representation into a discrete string –, SAX can discretize a numeric time series into symbolic strings. As shown in (5), the intuitive idea of PAA is to use the mean values to represent the amplitude values that fall into the same time interval.

$$\bar{x}_i' = \frac{1}{k_i - k_{i-1}} \sum_{j=k_{i-1}+1}^{k_i} x_j', \qquad (5)$$

where j is the index of the normalized load data; i is the index of the transformed PAA load data; k_i is the ith time domain breakpoint; and $\bar{x}_i'$ is the average value of the ith segment [20]. In many applications, the averaging feature of the PAA can be utilized to smooth out short-duration, sudden and large 'spikes' of time series [2]. PAA has been proven to have all the pruning power of the Haar-based discrete wavelet transform (DWT) and can be defined with lower computation cost for arbitrary length queries [20].

2) Flowchart: The complete framework and method will follow the general principles of data analytics-type processing, including normalization, feature extraction (dimensionality reduction) and data post-processing of the clustering results.

Step 1: Pre-process the collected AMI data of regional energy customers, which includes removing of invalid data sets and normalizing of customers' daily load profiles.

Step 2: Implement the dimensionality reduction with the PCA technique to make daily load profiles more suitable and easier for classification.

Step 3: Cluster the PCA components of the analyzed daily load profiles into initial K typical groups of energy customers with the K-means classification algorithm.

Step 4: Check the clustering index and accuracy with ELM. If the training and testing accuracy are below a chosen threshold Th, the number of clusters will be decreased by $K' = K - N_c$, and *Step 3* will be repeated again.

Step 5: Obtain the typical daily load profiles for every energy customer group by averaging the grouped daily load profiles based on the clustering index.

Step 6: Use SAX to assign symbols to the segmentations of the obtained typical daily load profiles in every customer group.

Step 7: Match the symbols associated with every typical daily load profile to suitable energy price levels.

Step 8: Analyze the economic effect and explore the different potential demand response programs for all the energy customer groups.
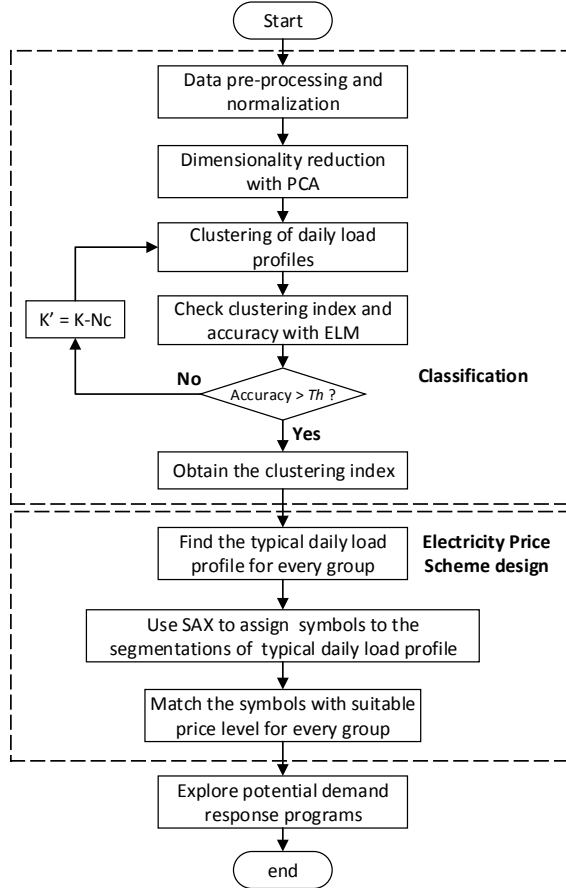

Fig. 2. Flowchart of the price scheme design process

III. RESULTS AND DISCUSSION

The following test cases include an AMI data set collected from a realistic Finnish distribution system operator (DSO), which includes 3,398 non-empty low voltage customers in a small region. We randomly picked 1,500 customers from them, and chose several typical normal dates (without special national holidays) to demonstrate the proposed framework.

A. Individualized price scheme design

In the classification stage, 90% is chosen as a criteria for the explained variance in the PCA and Th. 16 energy customer groups are obtained (as shown in Figure 3) to stand for the typical energy consumption patterns extracted from the chosen 1,500 customers. In most groups, one or two peaks can be observed during a typical 24-hour time interval.
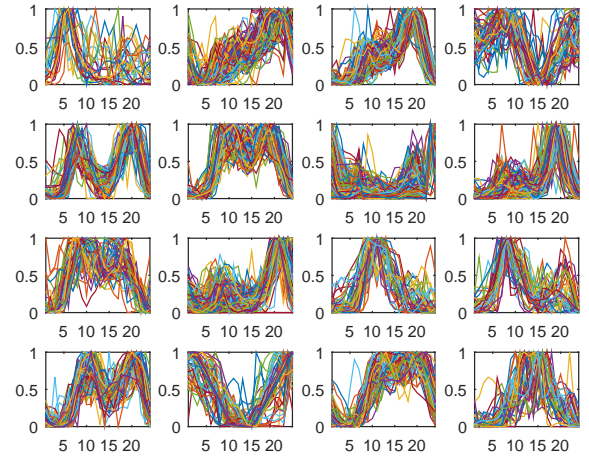

Fig. 3. Clustering of 1500 customers into 16 groups

The dynamic behavior of energy consumption for the whole group can be represented by SAX symbols as shown in Figure 4. It is noteworthy that the number of symbols forming a string can be very flexible. For simplification purposes and implementation easiness of the utility company (retailer), we just use three different symbols "**a**, **b**, **c**" in this test case. While, a larger number of symbol types will produce more accurate pricing for electricity products of different energy customers, it will also produce more complexity in terms of utility operation. The mapping between these SAX symbols and specific energy price levels should depend on a utility's historical operation experience and market analysis. A typical example of an electricity pricing level based on an existing global TOU pricing scheme is presented in Table I. Accordingly, the individualized price scheme designs for all 16 energy customer groups are shown in Figure 5.
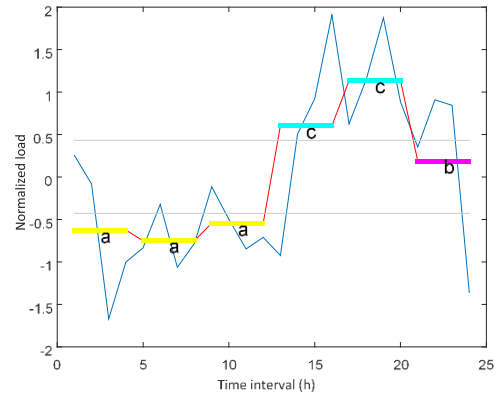

Fig. 4. The electricity price scheme design for a typical load profile with SAX

B. Economic analysis

In the retail electricity market, different energy customers are usually given different preferences and actually have cross-subsidy with each other [4]. By testing those 1,500 customers,

TABLE I
PRICE SCHEME DESIGN WITH SYMBOLS IN SAX

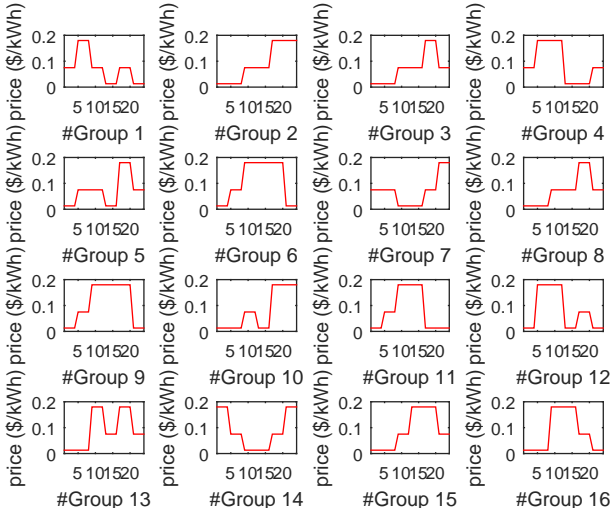| Symbols in SAX	Pricing Level	Price
a	Low level pricing	0.013 $ / kWh
b	Intermediate level pricing	0.075 $ / kWh
c	High level pricing	0.180 $ / kWh

Fig. 5. The design of an individualized price scheme structure for every group

we found that, as shown in Table II, the individualized TOU proposed in this paper mainly benefits retailers rather than energy customers in the short-term. However, if some demand response programs are introduced for customers, and their awareness of DR is reflected by some responsive rates, customers will still be able to achieve smart energy usage and economic benefit. In this way, retailers and customers will interact with each other more actively to commonly reach a better energy management and service.

TABLE II
ECONOMIC ANALYSIS OF THE REVENUE AND PAYMENT

Pricing strategy	Customers (payment)	Retailer (cost)	Retailer (revenue)
Global TOU	$1458.33	$1178.12	$280.21
Individualized TOU	$1525.34	$1178.12	$347.22
Individualized TOU with DR	$1429.42	$1148.89	$280.53

IV. CONCLUSION

In this paper, we proposed an individualized electricity price scheme design mechanism for various types of customers based on SAX and some combined classification methods, namely K-means and ELM. The final goal is that the utility company can make better use of the collected smart meter data and provide customized service to end-users. The customers can also reach more awareness of the possible energy usage strategy.

In the future, some more accurate and computationally efficient classification method should be studied for other related applications involving large-scale distribution networks with an industrialized big-data platform. More innovative business models of demand response programs based on the designed individualized electricity price scheme may also be discussed in the Smart Grid context.

REFERENCES

[1] A. Mutanen, H. Niska, and P. Järventausta, "Ming smart meter data - case finland," in *The 23th International Conference & Exhibition on Electricity Distribution (CIRED) Workshop*, Helsinki, Finland, 2016, pp. 120–124.

[2] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of electricity consumption behavior dynamics toward big data applications," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2437– 2447, 2016.

[3] T. Chen, A. Mutanen, P. Järventausta, and H. Koivisto, "Change detection of electric customer behavior based on amr measurements," in *PowerTech, 2015 IEEE Eindhoven*, Eindhoven, Netherlands, 2015, pp. 1–6.

[4] Y. Yu, G. Liu, W. Zhu, and et al., "Good consumer or bad consumer: Economic information revealed from demand profiles," *IEEE Transactions on Smart Grid*, 2017.

[5] F. Quilumba and et al., "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, 2015.

[6] C. Beckel and et al., "Revealing household characteristics from smart meter data," *Energy*, vol. 78, pp. 397–410, 2014.

[7] E. Celebi and J. D. Fuller, "Time-of-use pricing in electricity markets under different market structures," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1170–1181, 2012.

[8] R. Li and et al., "Development of low voltage network templatespart i: substation clustering and classification," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3036–3044, 2015.

[9] T. Räsänen and et al., "Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data," *Applied Energy*, vol. 87, no. 11, pp. 3538–3545, 2010.

[10] S. V. Verdú and et al., "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *IEEE Trans. on Power Systems*, vol. 21, no. 4, pp. 1672–1682, 2006.

[11] G. Chicco and I.-S. Ilie, "Support vector clustering of electrical load pattern data," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1619–1628, 2009.

[12] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2012.

[13] K. Qian, Z. Zhang, F. Ringeval *et al.*, "Bird sounds classification by large scale acoustic features and extreme learning machine," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, USA, 2015, pp. 1317–1321.

[14] R. O. Duda, P. E. Hart, D. G. Stork *et al.*, *Pattern classification*. Wiley New York, 1973, vol. 2.

[15] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proceedings of the twenty-first international conference on Machine learning*. Alberta, Canada: ACM, 2004, p. 29.

[16] C. M. Bishop, "Pattern recognition," *Machine Learning*, vol. 128, pp. 1–58, 2006.

[17] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[18] R. Zhang and et al., "Short-term load forecasting of australian national electricity market by an ensemble model of extreme learning machine," *IET Gen., Trans. & Distribution*, vol. 7, no. 4, pp. 391–397, 2013.

[19] J. Lin and et al., "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, San Diego, CA, USA, 2003, pp. 2–11.

[20] J. Sevcech and M. Bielikova, "Symbolic time series representation for stream data processing," in *Trustcom/BigDataSE/ISPA, 2015 IEEE*, vol. 2, Helsinki, Finland, 2015, pp. 217–222.