Robust Data-driven Profile-based Pricing Schemes

Jingshi Cui, Haoxiang Wang, Chenye Wu, *and* Yang Yu

Abstract—To enable an efficient electricity market, a good pricing scheme is of vital importance. Among many practical schemes, customized pricing is commonly believed to be able to best exploit the flexibility in the demand side. However, due to the large volume of consumers in the electricity sector, such task is simply too overwhelming. In this paper, we first compare two data driven schemes: one based on load profile and the other based on user's marginal system cost. Vulnerability analysis shows that the former approach may lead to loopholes in the electricity market while the latter one is able to guarantee the robustness, which yields our robust data-driven pricing scheme. Although k-means clustering is in general NP-hard, surprisingly, by exploiting the structure of our problem, we design an efficient yet optimal k-means clustering algorithm to implement our proposed scheme.

Index Terms—Data Driven, Pricing Policy, Electricity Market, End-to-End Learning

I. INTRODUCTION

The legacy power grid has worked admirably for over a century under the notion that electricity access is a human right [1]. This notion makes the power system operators often overlook the opportunities brought by *extensively* exploiting the value of customers. The challenge also comes from the large volume of customers in the system. It is simply too overwhelming for the operator to design customized pricing schemes for all end users, especially with time varying energy consumption patterns. At its core, this is a big data issue.

Powered by the wide deployment of smart meters, the advanced data analytic methods and machine learning algorithms ease the process of customized pricing by providing promising solutions of using big data to inform better price decisions to the system operator.

To exemplify this solution for the electricity sector, in this paper, we study a microgrid scenario, where the microgrid operator sets the real time price according to the marginal generation cost. Inspired by Yu *et al.* in [2], we define a metric MCI, based on the individual load profiles, to reflect each user's value in the microgrid operation. This metric allows us to employ k-means clustering algorithm to design the customized pricing scheme. Unlike most previous studies, which conduct the k-means clustering based on user's load profile, we adopt the notion of end-to-end machine learning, which yields a robust data-driven pricing scheme. Figure 1 visualizes the structure of this work.

A. Literature Review

To the best of our knowledge, we are the first to design a robust data-driven profile-based pricing scheme for the

The authors are with the Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, Beijing, China, 100084. C. Wu is the correspondence author. Email: chenyewu@tsinghua.edu.cn.
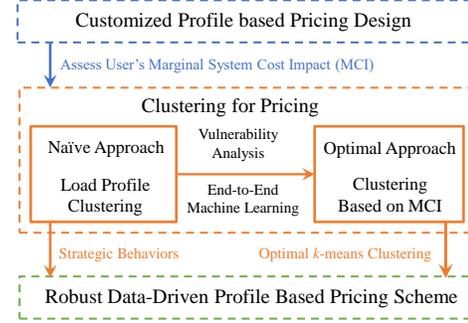

Fig. 1. Structure in Designing the Robust Pricing Scheme

electricity sector. We identify three groups of closely related literature.

The first related research direction investigates using machine learning techniques for better power system operation, mostly with a focus on load prediction. Just to name a few, Kong *et al.* exploit the temporal characteristics in residential load and propose an LSTM (Long Short-Term Memory) load forecasting method in [3]. Many deep learning models have also been employed for more accurate load prediction, including pooling-based deep recurrent neural network in [4], FFNN (Feed-Forward Neural Network) in [5], DBN (Deep Belief Network) in [6], etc. Chen *et al.* seek to integrate domain knowledge in different neural network building blocks in [7]. Researchers have also used machine learning to solve scheduling or dispatch problems for electricity market. For example, Mocanu *et al.* propose a deep reinforcement learning method to conceive an online optimization for building energy management systems in [8].

Another line of relevant research exploits opportunities in designing better pricing schemes for demand side. Based on utility maximization, Samadi *et al.* propose an optimal real-time pricing algorithm for demand side management in [9]. By combining hybrid particle swarm optimizer with mutation algorithm, Xu *et al.* design a data-driven pricing scheme to help the utility minimize peak demand in [10]. Exploiting price elasticity of electricity consumption, Yu *et al.* develop a parametric time-utility model to obtain the optimal real time price and maximize the social welfare in [11]. Qian *et al.* develop an SAPC (Simulated-Annealing-based Price Control) algorithm to reduce the peak-to-average load ratio via a two-stage optimization in [12].

Lastly, the robustness of clustering algorithms has also been well investigated. For example, Georgogiannis *et al.* use outlier detection to conduct the robustness analysis for quadratic k-means in [13]. Alok *et al.* overview the cluster validity measures and analyze the robustness of such indices

in [14]. Yan *et al.* investigate the consistency and robustness of two kernel-based clustering algorithms: semidefinite programming relaxation for kernel clustering and kernel singular value decomposition in [15]. Xu *et al.* design a robust and sparse fuzzy k-means clustering algorithm by incorporating a robust function to handle outliers in [16]. Li *et al.* introduce the concept of robust cluster objective function to enhance the robustness of clustering algorithms in [17].

B. Our Contributions

In pursuit of a robust data driven pricing scheme, the principal contributions of our work can be summarized as follows:

- *End-to-End Machine Learning*: We adopt the notion of end-to-end machine learning to design the optimal k-means clustering algorithm, which highlights that the vulnerability of many data-driven pricing schemes (e.g., the one in our recent work [18]) is due to the clustering criteria selection.
- *Characterize Robustness*: Following the definitions of strategic behaviors in [18], we propose a smoothness requirement to characterize robustness. This serves as the guideline to design robust data-driven pricing scheme. Surprisingly, although smoothness is a *global* concept, we prove that a *local* criteria for each cluster implies this *global* concept.
- *Optimal k-means Clustering*: Although k-means clustering in general is NP-hard [19], we exploit the structure of our problem and design a greedy algorithm to conduct k-means clustering with the smoothness requirement. We further prove that to achieve the smoothness requirement, the greedy algorithm is optimal in terms of minimizing the number of clusters.

The rest of the paper is organized as follows. We introduce our system model and identify each user's marginal system cost impact (MCI) in Section II. Then, Section III examines the economic loophole created by common k-means clustering. Inspired by such vulnerability analysis, in Section IV, we define robustness and propose that adopting the notion of end-to-end machine learning leads to a robust scheme: clustering based on MCI. We further design the greedy k-means clustering algorithm and prove its optimality in Section V. Simulation studies are conducted in Section VI for further insights into the proposed scheme. Concluding remarks and future directions are given in Section VII.

II. SYSTEM MODEL

We consider a microgrid of N users where the microgrid energy acquisition is maintained by an operator. Denote the total energy consumption in the microgrid at time t by L_t. Assume the cost to acquire energy is of a quadratic form, i.e.,

$$C(L_t) = \frac{1}{2}a \cdot L_t^2 + b \cdot L_t + c, \ \ \forall 1 \le t \le T,$$
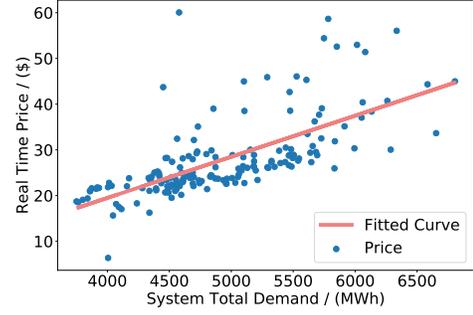$$s.t. \ \ L_t = \sum_{i=1}^{N} l_i^t, \tag{1}$$

Fig. 2. Relationship Between Price and Demand

where l_i^t denotes user i's energy consumption at time t. In such a model, the microgrid operator could simply set the electricity price $p(t)$ at time t according to the marginal cost:

$$p(t) = \frac{\partial C(L_t)}{\partial L_t} = aL_t + b, \ \ \forall 1 \le t \le T. \tag{2}$$

Remark: When N is large, the energy consumption in the microgrid forms a perfect competition (i.e. each user is a price taker) and the price determined by Eq. (2) is efficient. We adopt such a model in that it also allows us to understand the case when the price taker assumption does not hold, though out of the scope of this paper. Furthermore, the quadratic form of cost functions (i.e. linear form of prices) is a common assumption [20], which can be verified by real data. Figure 2 suggests that the real time price ($p(t)$) and the system total demand exhibit a linear relationship, using the PJM data [21].

A. Marginal System Cost Impact

We follow the metric, marginal system cost impact (denoted by MCI), first proposed by Yu *et al.* [2] to assess each individual user's system impact.

In the daily operation of the microgrid, the total cost for the microgrid operator can be obtained as follows:

$$C_T(\mathbf{L}) = \sum_{t=1}^{T} C(L_t), \tag{3}$$

where $\mathbf{L} = (L_1, ..., L_T)$. If the operator implements an hourly price, then T is simply 24 for the daily operation.

The MCI is characterized by each user's own load profile. For each user i, denote its daily load profile by $\mathbf{l}_i = (l_i^1, ..., l_i^T)$. Then, we have

$$
\begin{aligned}
MCI_i &= \lim_{\Delta \to 0} \frac{C_T\left(\mathbf{L} + \frac{\Delta \mathbf{l}_i}{||\mathbf{l}_i||_1}\right) - C_T(\mathbf{L})}{\Delta} \\
&= \lim_{\Delta \to 0} \frac{\sum_{t=1}^{T}(aL_t + b)\frac{\Delta l_i^t}{||\mathbf{l}_i||_1} + \left(\frac{\Delta l_i^t}{||\mathbf{l}_i||_1}\right)^2}{\Delta} \\
&= \sum_{t=1}^{T}(aL_t + b)\frac{l_i^t}{||\mathbf{l}_i||_1} \\
&= \sum_{t=1}^{T} p(t)\frac{l_i^t}{||\mathbf{l}_i||_1}.
\end{aligned}
\tag{4}
$$

Note that, by adopting MCI_i as the customized daily electricity rate for user i, its daily electricity cost is as follows:

$$MCI_i \cdot ||\mathbf{l}_t||_1 = \left(\sum_{t=1}^{T} p(t) \frac{l_i^t}{||\mathbf{l}_i||_1} \right) \cdot ||\mathbf{l}_t||_1 \qquad (5)$$
$$= \sum_{t=1}^{T} p(t) \cdot l_i^t.$$

This is a remarkable result! It shows that MCI_i, as a customized uniform price, maintains the same system efficiency as the prices determined by the marginal cost (i.e., $p(t)$'s). Moreover, MCI_i reflects the true value/cost of user i to the whole microgrid system.

However, we want to emphasize that our subsequent analysis does not rely on the quadratic cost functions. All of our conclusions hold as long as the first and second order derivatives of $C(L_t)$ exist, and its second order derivative is bounded.

This result again verifies the conclusion in [2] that each individual user's load profile should uniquely determine the user's electricity rate, which serves as the basis for data-driven pricing design.

B. Dataset Overview

Before diving into the details of discussion on the data-driven pricing, we first provide a brief overview on the datasets, which we will use throughout the paper.

We consider two types of demand profiles, residential users and commercial building users: one with more diversity and the other with an inherent pattern.

Empirical studies have shown that the load profiles across populations show similar patterns [22]. Hence, to enlarge the two datasets for more informative clustering results, we combine the data in different days. More precisely,

1) Residential Dataset: We use the Pecan Street dataset [23] with more than 300 residential load profiles from 2016.1.1 to 2016.2.10. We combine all the load profiles (excluding some profiles with missing data) into a single residential dataset of $7,699$ profiles.
2) Commercial Building Dataset: We use the Energy Plus dataset [24] with $4,820$ commercial building load profiles across US from 2007.1.2 to 2007.1.12. By combining all the data (again, excluding some profiles with missing data) to form a large dataset, we obtain our commercial building dataset of $48,200$ profiles.

As suggested by the definition of MCI, besides calculating the total load L_t, we normalize the profiles in two datasets to highlight the observation that user's load profile uniquely determines its electricity rate. Mathematically, for each user i's daily energy consumption $\mathbf{l}_i$, we define its normalized consumption $\mathbf{d}_i$ according to the l_1-norm:

$$\mathbf{d}_i = \frac{\mathbf{l}_i}{||\mathbf{l}_i||_1} = \frac{\mathbf{l}_i}{\sum_{t=1}^{T} l_i^t}. \qquad (6)$$

III. VULNERABILITY OF COMMON CLUSTERING

One straightforward way for clustering is to conduct the k-means clustering based on each individual's load profile.

However, since the users in the same cluster do not share exactly the same load profile, this data-driven pricing approach may create loopholes allowing certain users to bypass to other clusters for a better retail price.

In this section, we revisit the definition of disguising in [18], and provide more empirical evidence to support that strategic behaviors do exist under this pricing scheme.

A. Definition of Disguising

Users may strategically change their load profiles for a better price, especially for those who lie on the boundary of two clusters. They can conduct minimal load profile modifications to switch to another cluster, and potentially for a better price. We define such behaviors as disguising.

We denote the center (load profile) of cluster j by $\mathbf{c}_j$, and denote the cluster that user i belongs to by a function $u(i)$. To understand the strategic behavior, we need to examine the effort that user i needs to spend on moving to another cluster $n \neq u(i)$. This effort can be characterized by a scalar $\mu_{i,n}$. If user i can successfully switch to cluster n, the following condition holds:

$$||(1-\mu_{i,n})\mathbf{d}_i+\mu_{i,n}\mathbf{c}_n-\mathbf{c}_{u(i)}||_1 \geq ||(1-\mu_{i,n})(\mathbf{d}_i-\mathbf{c}_n)||_1. \quad (7)$$

Hence, the minimal effort that user i needs to spend for disguising can be solved by an optimization problem:

$$\min_{n \neq u(i)} \inf \mu_{i,n}$$
$$s.t. \ ||(1 - \mu_{i,n})\mathbf{d}_i + \mu_{i,n}\mathbf{c}_n - \mathbf{c}_{u(i)}||_1 \qquad (8)$$
$$\geq ||(1 - \mu_{i,n})(\mathbf{d}_i - \mathbf{c}_n)||_1,$$
$$p_n < p_{u(i)},$$

where p_n represents the price of cluster n, i.e., the MCI for cluster n's center profile $\mathbf{c}_n$.

Denote CR to be the minimal objective value to problem (8). We can formally characterize disguising as follows.

Definition 1: User i has the ability to disguise if

$$CR_i \leq \theta. \qquad (9)$$

Remark: In this *parametric* definition, the user's strategic behavior can be measured through a threshold θ. We also want to emphasize that by considering the load profile modification based on the normalized profile, we intentionally ignore user's possibility of disguising via reducing energy consumption, which allows us to focus on understanding the potential of shifting load profile for disguising.

B. Empirical Evidence: Clustering Results

We first conduct the k-means clustering for the two datasets. We choose a small k for better illustration of the subsequent analysis on strategic behaviors. Fig 3(a) and 3(b) visualize the clustering results for the two datasets, respectively.

The central profiles show interesting patterns in the two types of users. For example, cluster 8 in Fig. 3(a) and cluster 4 in Fig. 3(b) seem more like the classical load pattern with 2 peaks: one in the noon time and the other at night. We can observe that there are many more types of users, including

(a) Residential Dataset
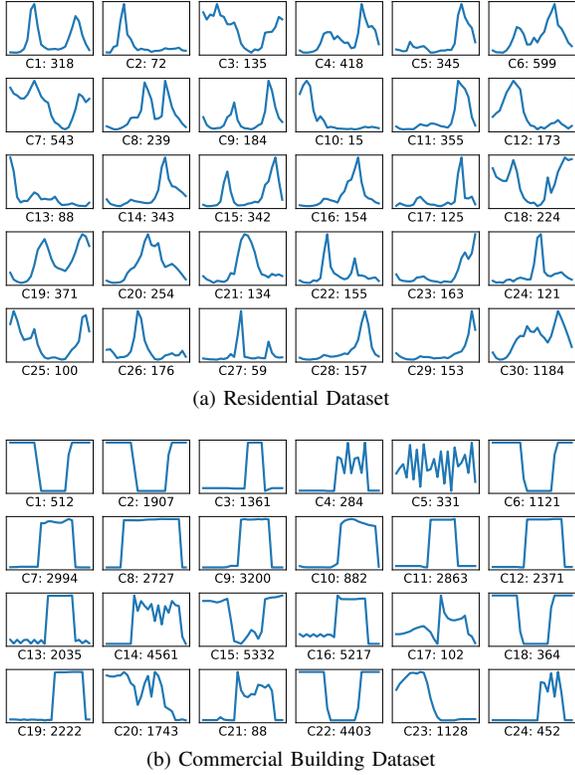

(b) Commercial Building Dataset

Fig. 3. Demand Profile of Each Cluster Center

those who are more active at night (e.g. cluster 11 in Fig. 3(a), and cluster 22 in Fig. 3(b)).

Note that, although the size of the building dataset is almost six times that of the residential user dataset, it leads to fewer clusters. Even within the 24 clusters, one may observe the inherent pattern of commercial buildings. This observation corresponds to our descriptions for the two datasets in Section II-B. Residential load profiles are more heterogeneous while commercial building profiles exhibit more homogeneity.

C. Empirical Evidence: Strategic Behaviors

To quantify the existence of strategic behaviors, we can define the number of users, having the ability to disguise, in each cluster n as follows [18]:

$$N_n(\theta) := \sum_{i, u(i) = n} I(CR_i \leq \theta), \qquad (10)$$

where $I(\cdot)$ is the indicator function.

Figure 4 plots how the percentage of potential strategic users varies with θ. Figure 4a(i) and 4b(i) demonstrate the statistical trends for the two datasets, which are quite intuitive. With a relaxed definition of disguising (larger θ), more users can become strategic. Yet, it is interesting to note that, in both datasets, users can disguise only if they change at least 0.8% of their load profiles.

We show more results in the subfigures of Fig. 4 to highlight the heterogeneity in clusters. There are rather stable clusters, e.g. cluster 11 in the commercial building dataset (as shown in Fig. 4b(ii)). And there are also rather unstable ones, e.g. cluster 28 in Fig. 4a(iv), and cluster 22 in Fig. 4b(iv).

IV. ROBUST DATA-DRIVEN PRICING: THEORY

To understand why strategic behaviors exist in the straightforward k-means clustering, we evaluate user's potential benefits of disguising.

Suppose user i has disguised itself to be a cluster n user, with a disguised profile $\widetilde{\mathbf{d}}_{i,n}$, where

$$\widetilde{\mathbf{d}}_{i,n} = ((1 - \mu_{i,n})\mathbf{d}_i + \mu_{i,n}\mathbf{c}_n). \qquad (11)$$

Then its potential benefit can be measured by the difference in electricity bills $b_{i,n}$:

$$b_{i,n}(\mu_{i,n}) = p_{u(i)}\|\mathbf{d}_i\|_1 - p_n\|\widetilde{\mathbf{d}}_{i,n}\|_1. \qquad (12)$$

Recall that p_n denotes the electricity rate for users in cluster n. Note that, since the load profiles have been normalized, $b_{i,n}$ measures the unit benefit of disguising. Mathematically, since $\|\mathbf{d}_i\|_1 = \|\widetilde{\mathbf{d}}_{i,n}\|_1 = 1$, we have

$$b_{i,n}(\mu_{i,n}) = p_{u(i)} - p_n. \qquad (13)$$

Hence, each user's maximal potential to disguise, denoted by b_i, can be defined as follows

$$b_{i,n}(\mu_{i,n}) = \max_{n \in \mathcal{K}_i, n \neq n(i)} \{p_{u(i)} - p_n\}, \qquad (14)$$

where $\mathcal{K}_i$ denotes the set of clusters that user i can successfully disguise itself with $CR_i \leq \theta$.

A. Definition of Smoothness

The potential benefits of disguising in Eq. (14) motivate us to define smoothness to ensure a robust data-driven scheme.

Definition 2: The k-means clustering is $\delta-$smooth if for any user i and its associated disguising set $\mathcal{K}_i$, the following condition holds:

$$|p_{u(i)} - p_n| \leq \delta, \forall n \in \mathcal{K}_i. \qquad (15)$$

Obviously, $\delta-$smoothness guarantees the bound of robustness. One straightforward way to ensure such smoothness is to design a supervised load profile based k-means clustering with the smoothness requirement.

B. k-means Clustering with Smoothness Guarantee

In this section, we want to show that, although smoothness is a global concept, it can be guaranteed by local properties. More precisely, if a k-means clustering could ensure that for each user i, its MCI_i is close to $p_{u(i)}$, then, this k-means clustering is smooth. Formally, we can prove the following theorem:

Theorem 1: Suppose a k-means clustering algorithm can guarantee that

$$|MCI_i - p_{u(i)}| \leq \rho, \forall i \in u(i), \qquad (16)$$

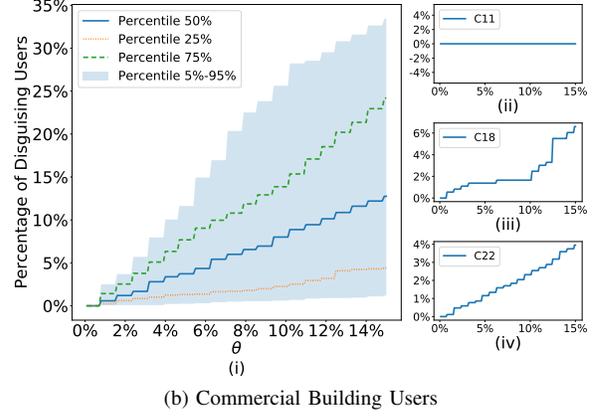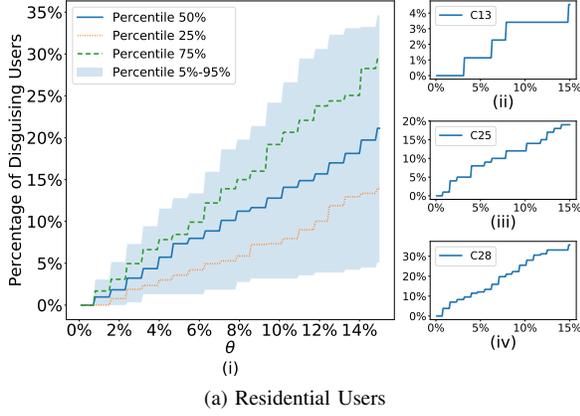then the clustering is $\rho(1 + \frac{1}{1-\theta})-$smooth.

(a) Residential Users

(b) Commercial Building Users

Fig. 4. Percentage of Strategic Users Evolving with θ

Proof: Suppose user i successfully disguises itself as cluster n user with a profile $\tilde{\mathbf{d}}_{i,n}$. The condition in the theorem implies that,

$$
\begin{aligned}
&\left| p_n - \sum_{t=1}^{T} p(t)((1 - \mu_{i,n})d_i^t + \mu_{i,n}c_n^t) \right| \\
&= \left| \sum_{t=1}^{T} p(t)(1 - \mu_{i,n})(c_n^t - d_i^t) \right| \\
&= |(1 - \mu_{i,n})(p_n - MCI_i)| \\
&\leq \rho.
\end{aligned}
\tag{17}
$$

Hence, for any cluster $n \in \mathcal{K}_i$, we have

$$
\begin{aligned}
|p_n - p_{u(i)}| &\leq |p_n - MCI_i| + |MCI_i - p_{u(i)}| \\
&\leq \frac{\rho}{1 - \mu_{i,n}} + \rho \\
&\leq \rho\left(1 + \frac{1}{1 - \theta}\right).
\end{aligned}
\tag{18}
$$

The second inequality is due to Eq. (17) and the last inequality is due to the parametric definition of disguising. Thus, we complete the proof. ■

The only remaining handle is to design an efficient k-means clustering algorithm with smoothness guarantee.

V. ROBUST DATA-DRIVEN PRICING: IMPLEMENTATION

In this section, we first introduce the most straightforward supervised k-means clustering implementation, and then propose our greedy algorithm. To our surprise, although the general k-means clustering is NP-hard, the proposed greedy algorithm is efficient and optimal.

A. Supervised k-Means Clustering

Given the outcome of profiled based k-means clustering (see the algorithm in [18] for an instance), we can modify the outcome (details are provided in Algorithm 1) to ensure the smoothness requirement. We term this algorithm as SkC for short.

This algorithm is commonly known as bisecting k-means clustering. Hence, SkC is a combination of load profile based clustering and bisecting clustering. While it obviously satisfies

---

Algorithm 1 Supervised k-means Clustering (SkC)

Input: User data set $C = \{i | i = 1, ..., n\}$, price data $p(t)$
Output: Clusters $C_1, ..., C_k$

1: $m = \min_i MCI_i$
2: $M = \max_i MCI_i$
3: **if** $M - m < 2\rho$ **then**
4:    **return** C as a single cluster
5: **else**
6:    $C_1 = \varnothing$, $C_2 = \varnothing$
7:    **for** $i = 1$ to n **do**
8:      **if** $|MCI_i - m| \leq |MCI_i - M|$ **then**
9:        $C_1 = C_1 \cup \{i\}$
10:      **else**
11:        $C_2 = C_2 \cup \{i\}$
12:      **end if**
13:    **end for**
14: **end if**
15: **if** $|\max_{i \in C_1} MCI_i - \min_{i \in C_1} MCI_i| > 2\rho$ **then**
16:    **if** $|\max_{i \in C_2} MCI_i - \min_{i \in C_2} MCI_i| > 2\rho$ **then**
17:      **return** $SkC(C_1)$ and $SkC(C_2)$
18:    **else**
19:      **return** $SkC(C_1)$ and (C_2 as a single cluster)
20:    **end if**
21: **else if** $|\max_{i \in C_2} MCI_i - \min_{i \in C_2} MCI_i| > 2\rho$ **then**
22:    **return** $SkC(C_2)$ and (C_1 as a single cluster)
23: **else**
24:    **return** C_1 and C_2 as two clusters
25: **end if**

---

the local property to ensure smoothness, it may produce significantly more clusters than that needed.

Remark: In fact, we don't need to design such a complicated supervised k-means clustering. We should re-consider what should be the clustering criteria.

We adopt the notion of the end-to-end machine learning. The ultimate goal of clustering is for pricing. Hence, instead of conducting the k-means clustering based on load profiles, we can directly conduct the clustering based on each individual's MCI.

The shift in selecting the clustering criteria is significant.

We can now directly guarantee the smoothness requirement with minimal room for market manipulation.

B. Greedy Algorithm

The idea is very simple. We first calculate the MCI for all users and sort the tuple (i, MCI_i) in the ascending order of MCI_i. For notational simplicity, we assume (i, MCI_i) is already ordered, and the index i simultaneously denotes user i and its rank with respect to MCI_i. Then, starting from the first user, we use the interval of length 2ρ to cover the neighboring users and form the cluster. This procedure can be repeated until all users have been clustered. We provide the detailed algorithm in Algorithm 2, and term this algorithm as greedy k-means clustering (or GkC in short).

Algorithm 2 Greedy k-means Clustering (GkC)

Input: Ordered tuple (i, MCI_i), $i = 1, ..., n$
Output: Clusters $C_1, ..., C_\kappa$
1: $i \leftarrow 1$
2: $k \leftarrow 1$
3: **while** $i \neq n$ **do**
4:    $j = \arg\max_m MCI_m$, $s.t. \; MCI_m \leq MCI_i + 2\rho$
5:    $C_k = \{i, ..., j\}$
6:    $k \leftarrow k + 1$
7:    $i \leftarrow j + 1$
8: **end while**
9: $\kappa = k$;
10: **return** $C_1, ..., C_\kappa$

C. Optimality of GkC

This greedy algorithm achieves the minimal κ to ensure the clustering criteria. Formally,

Theorem 2: The GkC selects the minimal number of clusters to ensure the clustering criteria:

$$|MCI_i - p_{u(i)}| \leq \rho, \forall i. \tag{19}$$

Proof: Suppose the optimal clustering result is with fewer number of clusters κ° (i.e., $\kappa^\circ < \kappa$), which also ensures the clustering criteria (19).

Denote this clustering result by $\pi_\circ$ and the GkC clustering result by π^*. And denote for each cluster $k_\circ \in \pi_\circ$ ($k^* \in \pi^*$), the element with minimal MCI by $s^{k_\circ}$ (s^{k^*}), and the element with maximal MCI by $e^{k_\circ}$ (e^{k^*}).

Since $\pi_\circ$ is optimal, there don't exist clusters $i_\circ$, $j_\circ$, such that $s^{i_\circ} < s^{j_\circ} < e^{j_\circ} < e^{i_\circ}$ or $s^{i_\circ} < s^{j_\circ} < e^{i_\circ} < e^{j_\circ}$. That is, the clusters in $\pi_\circ$ won't overlap with each other. This can be straightforwardly proved by contradiction. We omit the proof due to space limit.

Hence, the clusters in $\pi_\circ$, just as the case for π^*, are also disjoint intervals.

Next, we examine the clusters in $\pi_\circ$ and π^* along the axis of MCI. Since they are the clustering results for the same dataset, the first clusters in $\pi_\circ$ and π^* share the same starting point, i.e., $s^{1_\circ} = s^{1^*}$.
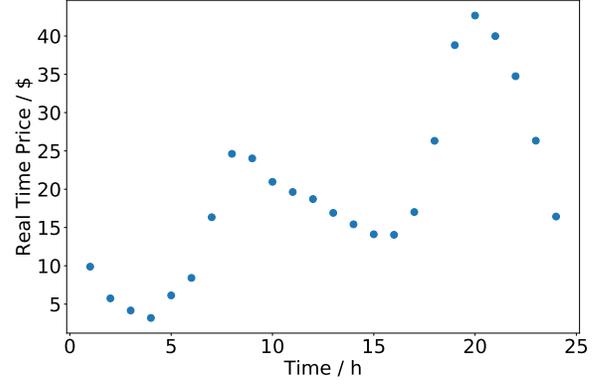

Fig. 5. Real Time Price Over a Day

Then, we observe the first difference between $\pi_\circ$ and π^*. We claim the first difference cannot be the starting point of some cluster. Otherwise, some profiles are left unclustered.

Suppose the first difference happens at cluster $k_\circ$ and cluster k^*, then there are two possible conditions.

(1) If $e^{k_\circ} > e^{k^*}$, then due to the construction process of GkC, $|s^{k_\circ} - e^{k_\circ}| > 2\rho$, which won't satisfy the clustering requirement.

(2) If $e^{k_\circ} < e^{k^*}$, then we can simply set $e^{k_\circ} = e^{k_*}$, $s^{k_\circ+1} = s^{k^*+1}$.

This forms a new clustering result with the same number of clusters in $\pi_\circ$. However, after the adjustment of (2) along the axis of MCI, $\pi_\circ$ becomes exactly the same clustering result as π^*! This contradicts with our assumption that $\kappa^\circ < \kappa$. ∎

Remark: The time complexity of GkC is $O(n \log n)$, which makes our data-driven pricing scheme even more practical. Since an algorithm with such a time complexity can be run even in real time, GkC solves the major concern of deploying our scheme: it requires the full knowledge of $p(t)$ to determine MCI. This implies that, the consumers are only aware of their electricity rates after the daily energy consumption. With the efficient GkC algorithm, based on advanced load forecasting methods, the microgrid operator can inform each user's estimated daily electricity rate in real time.

VI. SIMULATION STUDIES

In this section, we first compare the performance of SkC and GkC to demonstrate the optimality of GkC. Then, we analyze how different parameters (e.g. ρ, a, etc) will affect the clustering results. This inspires us to further exploit more insights into the GkC results.

Using the fitted price model, we select $a = 0.00012$ and $b = -37.38$. According to the total load in system, we obtain the real time price of the day, which helps us to conduct the clustering. We visualize the real time price determined by Eq. (2) in Fig. 5.

A. SkC versus GkC

While both can guarantee the smoothness requirement, we compare the running time as well as the number of clusters returned by SkC and GkC. Obviously, due to the heuristic
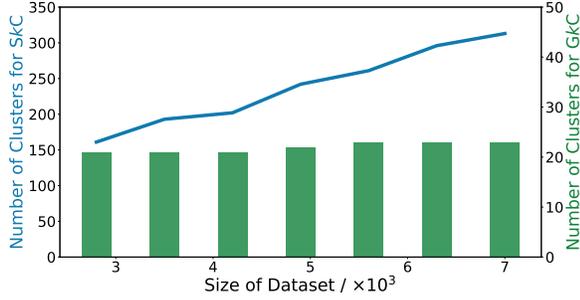
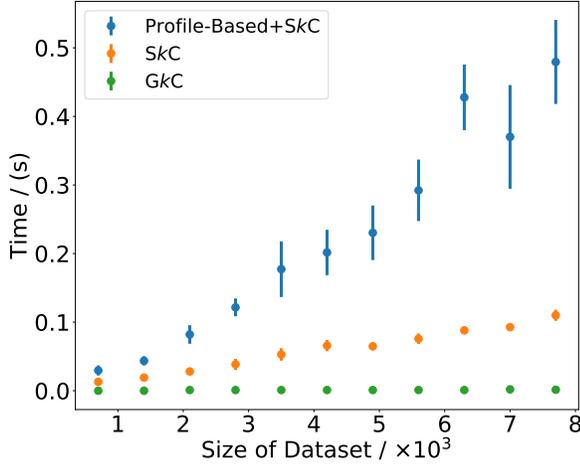Fig. 6. Efficiency Comparison between SkC and GkC

Fig. 7. Running Time Comparison between SkC and GkC (SkC only accounts for the time after obtaining the profile-based clustering results; profile-based+SkC measures the total time for the supervised k-means clustering).
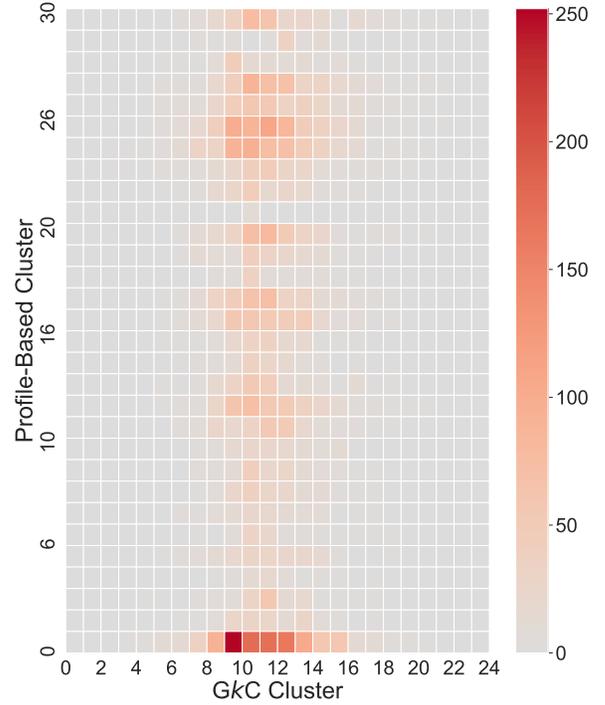

Fig. 8. Relationship Between GkC and Profile-Based Clustering when $\rho = 0.5$ (the darker the rectangle, the more users in the rectangle).

$$\sigma_k = \frac{\sum_{i \in C_k} ||\mathbf{d}_i - \mathbf{c}_k||_1}{|C_k|}, k = 1, ..K, \qquad (20)$$

where $\mathbf{c}_k$ denotes the center profile of cluster k and C_k denotes the set of users in cluster k, i.e., $C_k = \{i | u(i) = k\}$.

We can observe in Fig. 9 that the size of cluster seems like a binomial distribution: most users are around the medium MCI in the dataset (the medium user is in cluster 12). However, the diversity in each cluster has two peaks (cluster 6 and cluster 19). This is somewhat surprising. We conduct the profile-based k-means clustering for the two peaks for more insights.

Figure 10 shows that the demand profiles in cluster 6 are quite heterogeneous. In contrast, the demand profiles in cluster 19 all have a peak in the evening (corresponding to the peak price in Fig. 5), as suggested by Fig. 11. This implies that for high MCI users, since the high MCI is caused by the peak price, their profiles tend to be similar. For these users, the results obtained by clustering based on load profiles and clustering based on MCI are more likely to be aligned with each other. For other MCI users, clustering based on MCI is much more efficient!

C. Sensitivity Analysis

We conclude this section by understanding the impacts of ρ and a on the clustering results. As suggested by Fig. 12, the number of clusters produced by GkC diminishes quickly with the increasing of ρ. This also corresponds to our observation in Fig. 9, that the size of cluster distribution seems like a binomial distribution. Hence, the initial increase of ρ leads to a more dramatic decrease in the number of clusters.
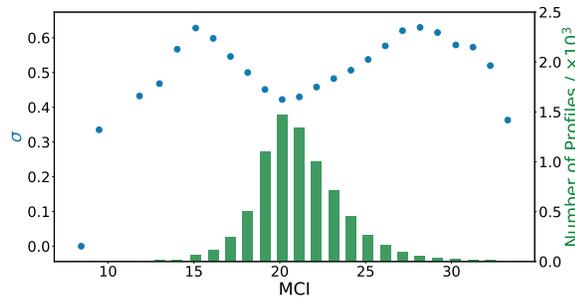
implementation of k-means clustering based on load profiles, GkC outperforms SkC in both aspects.

In Fig. 6, with the increasing size of dataset, GkC produces a rather stable number of clustering, around 20. In contrast, SkC produces hundreds of clusters and the number of clusters returned by SkC increases dramatically with the increasing size of dataset. Figure 7 compares the running time between SkC and GkC. Even if we don't take the time consumed by profile based clustering into account for SkC evaluation, GkC still runs faster than SkC.

In fact, as the dataset contains millions of profiles, GkC's running time is still $O(n \log n)$ yet SkC may suffer from the curse of dimensionality.

To understand why SkC produces more clusters than GkC, we use Fig. 8 for more insights. Since the initial clustering is conducted based on load profiles, most clusters need to be further decomposed into two or more clusters to ensure the clustering criteria (19). Hence, in Fig. 8, while the GkC only produces 24 clusters, SkC needs to produce at least the same amount of clusters as the number of red colored rectangles.

B. Dive into GkC

We first examine the diversity of the load profiles in each GkC cluster. We define σ_k to evaluate the mean distance to the cluster center for each cluster k:

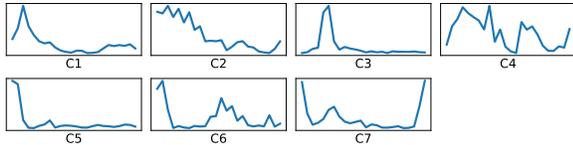Fig. 9. Number of Clusters Based on Profiles and σ

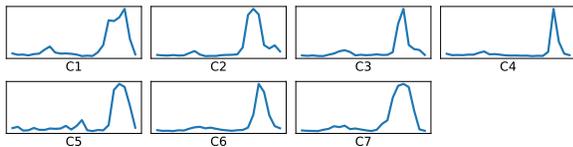Fig. 10. Clustered Demand Profile of Cluster 6 Based on MCI

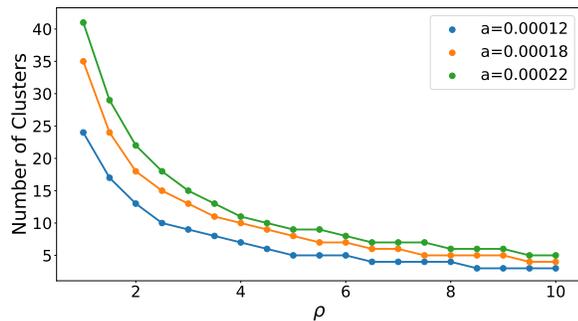Fig. 11. Clustered Demand Profile of Cluster 19 Based on MCI

Fig. 12. Number of Clusters Evolving with ρ for Different a

On the other hand, a larger a, which implies an even higher uniform electricity rate, leads to more clusters. Unlike the nonlinear impact of ρ, the impact of a exhibits linearity: to twice the value of a almost doubles the number of clusters.

VII. CONCLUSION

By adopting the notion of End-to-End machine learning, we identify the most efficient clustering criteria to design the data-driven profile-based pricing scheme. We further exploit the structure of the problem and the definition of smoothness, which inspire us to design an optimal and efficient greedy k-means clustering to implement the proposed pricing scheme.

This work can be extended in many ways. As we have mentioned in the remark in Section II, it will be interesting to examine the possibility to relax the price taker assumption using a game theoretic analysis. We also intend to examine the temporal characteristics of user load profiles over a longer time span for a more detailed user MCI profile characterization. This can enable the annual pricing plan design for the system operator.

REFERENCES

[1] World Bank, "State of electricity access report 2017," 2017.
[2] Y. Yang, G. Liu, W. Zhu, W. Fei, B. Shu, Z. Kai, N. Astier, and R. Rajagopal, "Good consumer or bad consumer: Economic information revealed from demand profiles," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.
[3] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on lstm recurrent neural network," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.
[4] H. Shi, M. Xu, and L. Ran, "Deep learning for household load forecasting a novel pooling deep rnn," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2018.
[5] Y. Wei, L. Xia, S. Pan, J. Wu, and Q. Li, "Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks," *Applied Energy*, vol. 240, pp. 276–294, 2019.
[6] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, "Deep belief network based electricity load forecasting: An analysis of macedonian case," *Energy*, vol. 115, pp. 1688–1700, 2016.
[7] K. Chen, K. Chen, W. Qin, Z. He, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2018.
[8] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. G. Slootweg, "On-line building energy optimization using deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.
[9] P. Samadi, A.-H. Mohsenian-Rad, R. Schober, V. W. S. Wong, and J. Jatskevich, "Optimal real-time pricing algorithm based on utility maximization for smart grid," in *IEEE International Conference on Smart Grid Communications*, 2010.
[10] Z. Xu, T. Deng, Z. Hu, Y. Song, and J. Wang, "Data-driven pricing strategy for demand-side resource aggregators," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 57–66, 2017.
[11] R. Yu, W. Yang, and S. Rahardja, "A statistical demand-price model with its application in optimal real-time price," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1734–1742, 2012.
[12] P. Q. Li, J. A. Z. Ying, J. Huang, and W. Yuan, "Demand response management via real-time electricity price control in smart grids," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1268–1280, 2013.
[13] A. Georgogiannis, "Robust k-means: a theoretical revisit," in *Advances in Neural Information Processing Systems*, pp. 2891–2899, 2016.
[14] P. Alok, S. Chaudhuri, and R. Chatterjee, "Cluster validity measures and their robustness," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, pp. 444–448, 06 2017.
[15] B. Yan and P. Sarkar, "On robustness of kernel clustering," in *Advances in Neural Information Processing Systems*, pp. 3098–3106, 2016.
[16] J. Xu, J. Han, K. Xiong, and F. Nie, "Robust and sparse fuzzy k-means clustering.," in *IJCAI*, pp. 2224–2230, 2016.
[17] J. Li, S. Song, Y. Zhang, and Z. Zhou, "Robust k-median and k-means clustering algorithms for incomplete data," *Mathematical Problems in Engineering*, vol. 2016, 2016.
[18] J. Cui, H. Wang, C. Wu, and Y. Yu, "Vulnerability analysis for data driven pricing schemes." https://arxiv.org/abs/1911.07453, 2019.
[19] S. Dasgupta, *The hardness of k-means clustering*. Department of Computer Science and Engineering, University of California , 2008.
[20] E. Martínez-Budría, S. Jara-Díaz, and F. J. Ramos-Real, "Adapting productivity theory to the quadratic cost function. an application to the spanish electric sector," *Journal of Productivity Analysis*, vol. 20, no. 2, pp. 213–229, 2003.
[21] PJM, "Pjm data." https://dataminer2.pjm.com/list.
[22] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, "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.
[23] Pecan Street INC., "Pecan street data." http://www.pecanstreet.org.
[24] U.S. Department of Energy, "Using energyplus for compliance." https://energyplus.net/sites/all/modules/custom/nrel_custom/pdfs/pdfs_v8.5.0/Using_EnergyPlus_For_Compliance.pdf.