CSC4008: Techniques for Data Mining

Project Introduction

Jan. 26, 2021

Chenye Wu
wuchenye@cuhk.edu.cn

Data Mining in Power Systems

- **Motivation:**
  - To put the theory/toolbox into action
  - To give you a prototype research experience (find your own topic)
  - To have a concrete example during the rest of the course

- **Grades: 40 Points in total**
  - Course presentation: 10 Points
  - 1st report: 5 points
  - 2nd report: 25 points

Data Mining in Power Systems

- **Dataset:** Pecan Street Individual Users' Daily Energy Consumption Dataset (you may also find your own dataset).

- **Deliverables:**
  - Literature Review (1st Report, due on March 15, 2021)
  - During the Spring Festival, you are suggested to conduct the literature review to find your own topic, as well as conduct the pre-processing for the dataset.
  - 5-minute Course Presentation (in May, the last two weeks)
  - Final Report (2nd Report, tentatively due on May 26, 2021)

- **Example Directions**

The Dataset

- Collected from 400 users
- Location: Austin, TX
- Period: May 1 to Oct 30, 2015
- Granularity: 1 minute

- Need to ‘Preprocess’ the Dataset due to missing data

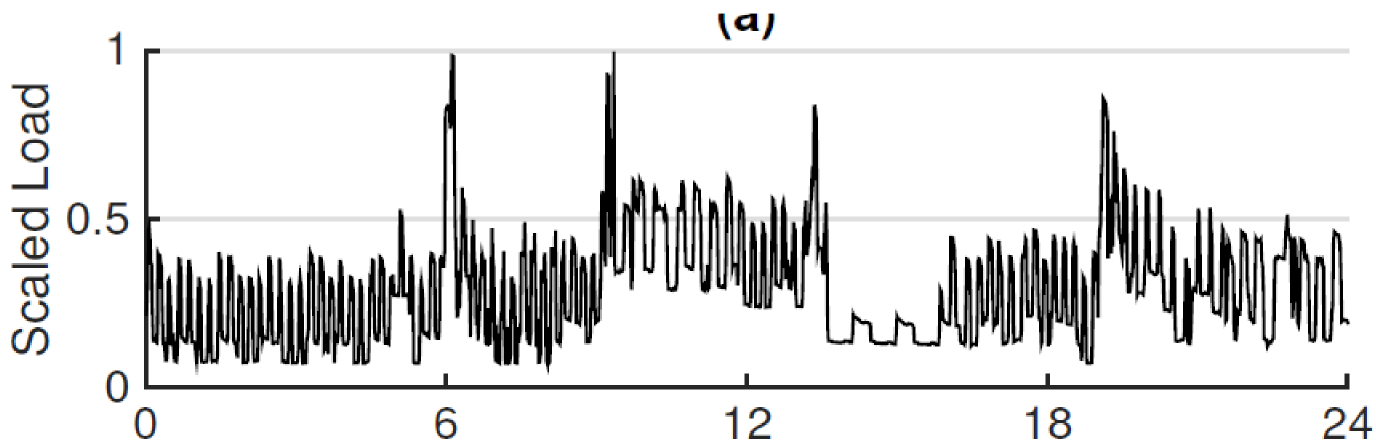
Example 1:

Prediction…

Chenye Wu, Wenyuan Tang, Kameshwar Poolla, Ram Rajagopal, " Predictability, Constancy and Contingency in Electric Load Profiles,"
In Proceedings of IEEE SmartGridComm 2016, Sydney, Australia, Nov, 2016.

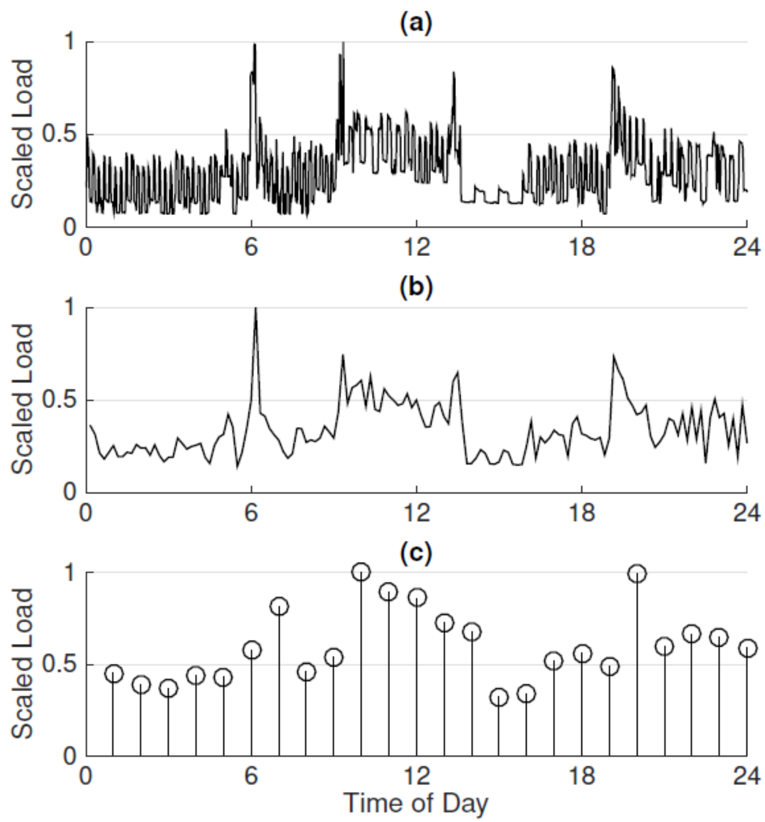
The most straightforward task

- The challenges:
  - It might be impossible to predict the time series sequence.



What we can do?

- Not all predictions are hard!!!



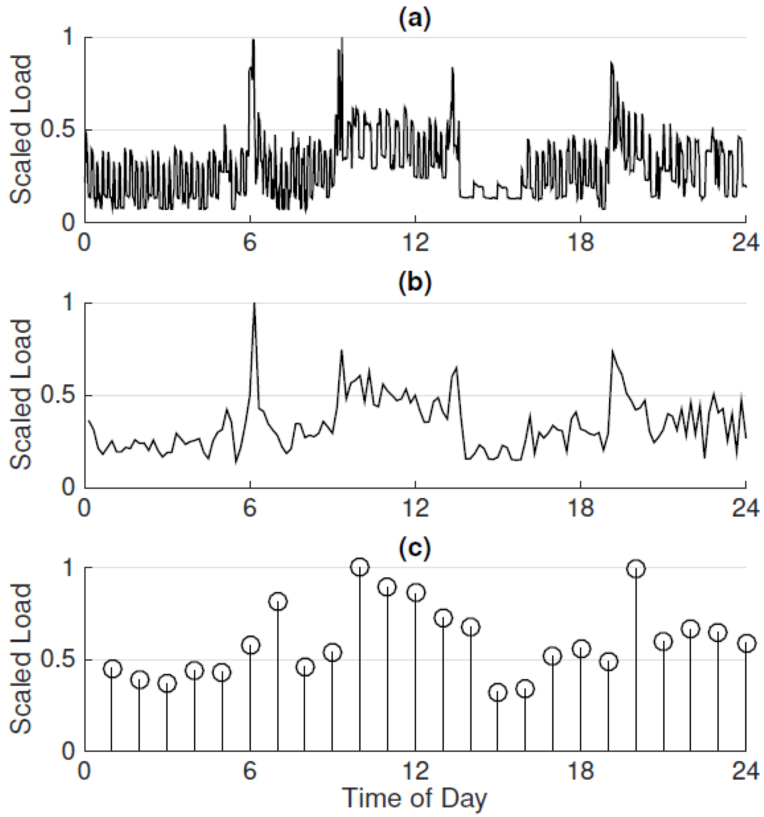
ig. 7: Individual scaled load profile with granularity of (a) one minute, (b) 10= minutes, (c) one hour.

Possible Direction

- Define your own way of the ‘hardness’ for predicting the loads
- Examine the relationship between the hardness and other parameters, say granularity, aggregate level, etc.
- Preferably, with insights.
- What else we can do?

Re-examing the Load Profiles

- What would you expect to see if you conduct the FFT for the 3 time series data?



ig. 7: Individual scaled load profile with granularity of (a) one minute, (b) 10= minutes, (c) one hour.

How does NN work?

- Most predictors, during training, are expected to first capture
  - A) low frequency component
  - B) high frequency component
- Hence, will you be able to observe such phenomenon in predicting the load profiles?
- If yes, how can you make the phenomenon much easier to observe?

Example 2:

Good Consumers v.s. Bad Consumers

[1] Yu et al, Good consumer or bad consumer: Economic information revealed from demand profiles, IEEE Trans. On Smart Grid.

Common Belief

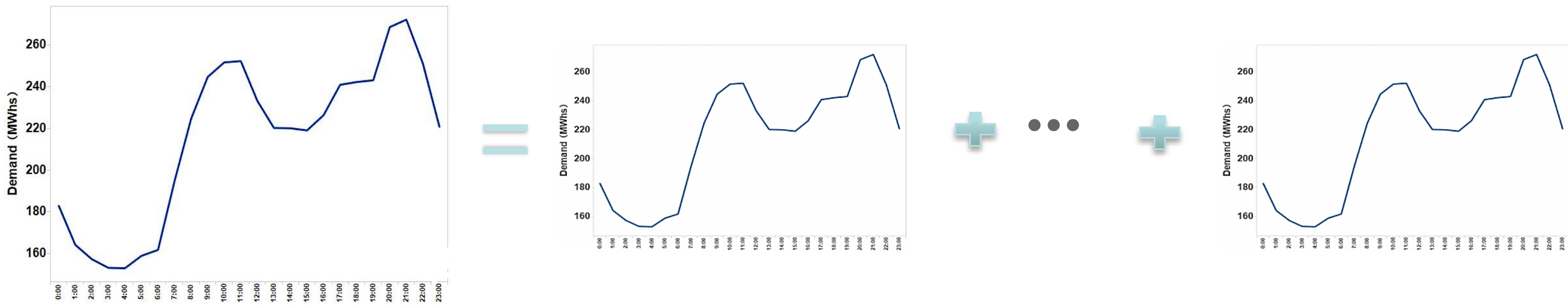
- In power system operation, the critical peak during the year drives the operational cost high.
- Hence, suppose the State Grid is asking you to distinguish bad consumers from good consumers, how would you suggest?

Common Belief

- People tend to believe those large consumers (consume large quantity of energy) are the bad consumers.
- What are the inherent assumptions behind this belief?

Underlying Assumption

- We thought that the energy consumption patterns are similar.



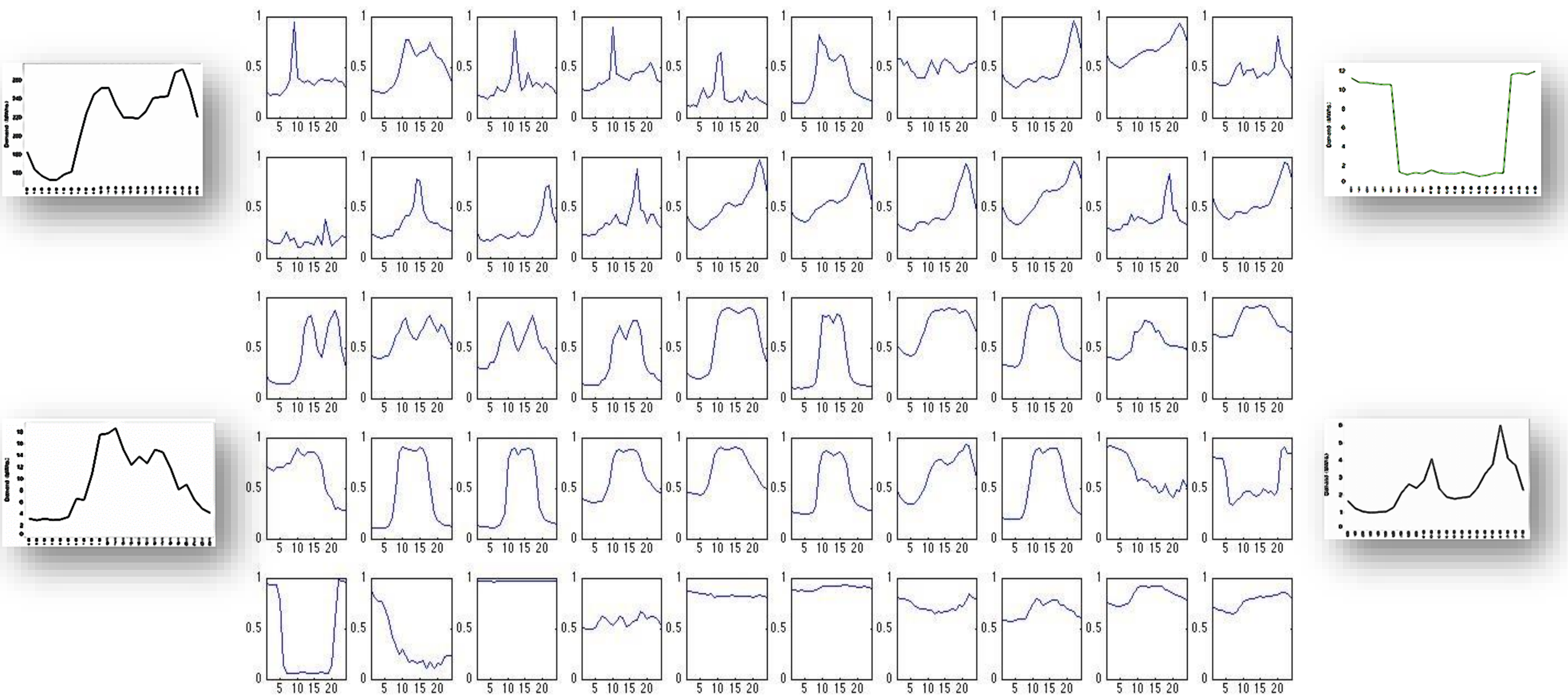
- Based on this assumption, of course, large consumers are bad consumers.

- Hence, we should charge them a higher electricity rate.

Is this assumption true?

How to validate your conclusion?

Ask the Data!

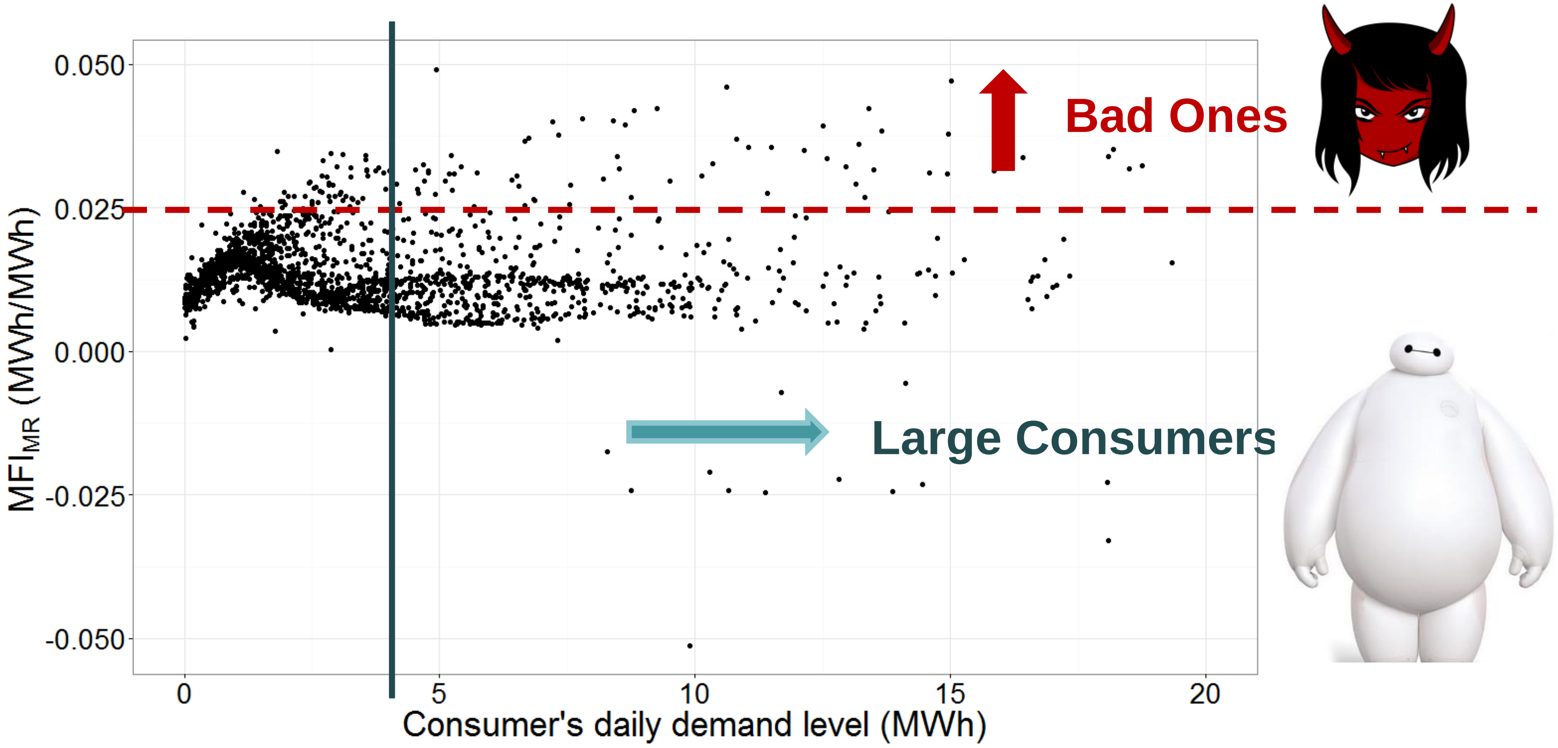


Good consumers v.s. Bad consumers

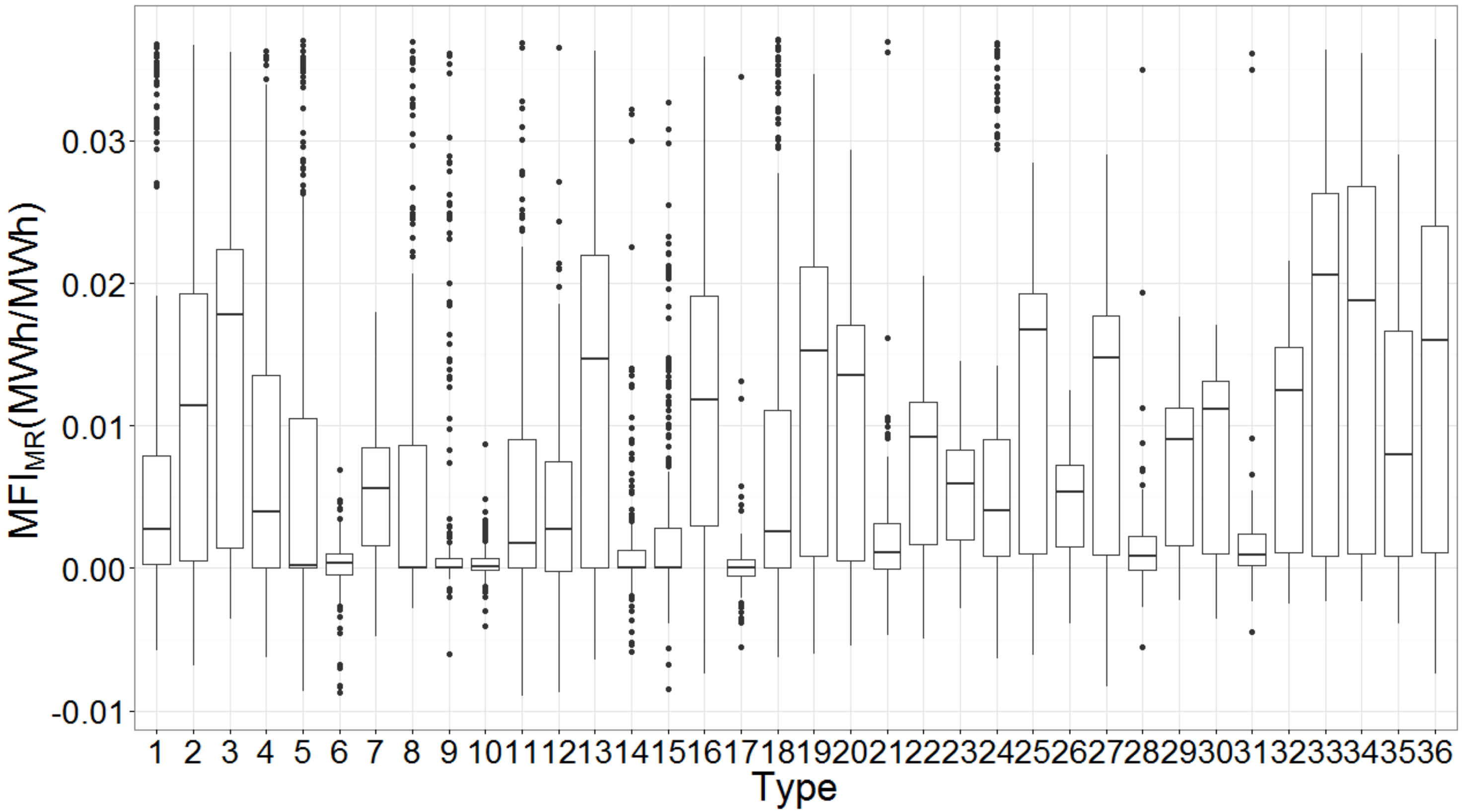
- To determine if one consumer is good or bad, we need to specify the system operator's marginal cost to serve the consumer.
- This marginal cost is uniquely determined by the consumer's load shape!
- Hence, instead of evaluation the consumer's total energy consumption, the system operator needs to evaluate its load profile!
- Yu et al. have developed such an index to use the load profile to determine the marginal serving cost in [1].

[1] Yu et al, Good consumer or bad consumer: Economic information revealed from demand profiles, IEEE Trans. On Smart Grid.

Good consumers v.s. Bad consumers



Good consumers v.s. Bad consumers



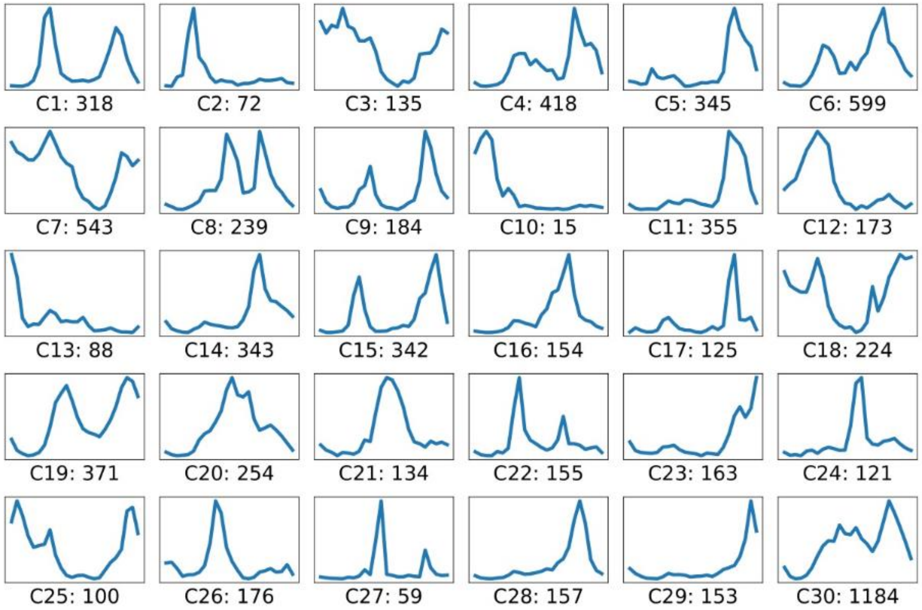
Example 3:

Vulnerability of Data-driven Pricing

Jingshi Cui, Haoxiang Wang,Chenye Wu, Yang Yu, “Robust Data-driven Profile-based Pricing Schemes", In Proceedings of IEEE ISGT NA 2021, Washington D.C. USA

Clustering Results

- Using Pecan Street Dataset
  - K-means clustering using l_1-norm!



Why choose l_1-norm?

- Cost function
  - $C_T(\boldsymbol{L}) = \sum_{t=1}^{T} C(L_t) = \sum_{t=1}^{T} \frac{1}{2} a \cdot {L_t}^2 + b \cdot L_t + c, \forall\, 1 \le t \le T, \quad s.t.\ L_t = \sum_{i=1}^{N} l_i^t$
- Real time price (real time marginal cost)
  - $p(t) = \frac{\partial C(L_t)}{\partial L_t} = a \cdot L_t + b, \forall\, 1 \le t \le T$
- Marginal System Cost Impact (MCI)
  - $MCI_i = \lim_{\Delta \to 0} \frac{C_T\left(\boldsymbol{L} + \Delta \frac{\boldsymbol{l}_i}{\left\|\boldsymbol{l}_i\right\|_1}\right) - C(\boldsymbol{l})}{\Delta} = \sum_{t=1}^{T} \frac{l_i^t}{\sum_{m=1}^{T} l_i^m} p(t)$

If we implement MCI as a uniform price, why it is good?

What does MCI tell us?

- If we choose to implement a uniform price, then we should choose MCI.
- And MCI is uniquely determined by user's load profiles.
- One straightforward implementation can be to conduct the k-means clustering based on load profiles.
- Users in the same cluster share the same price.

Clustering for Pricing

- **Clustering Challenges**
  - Problem: users in the same cluster do not share exactly the same load profile.
  - Loophole: certain users may bypass to other clusters for a better retail price.

Our Contributions

- **Two Questions**
  - What are the possible strategic behaviors?
  - How many users can conduct price manipulation?
- **Our Work**
  - Defining disguising ⇒ to identify strategic behavior.
  - Characterizing sensitive zone ⇒ to observe the impacts of different parameters.
  - Based on sensitive zone characterization and cost benefit analysis ⇒ vulnerability analysis.

Disguising: Strategic Behaviors

- Assumption
  - All users know the global information (central profile of each cluster and its corresponding price).
- Minimal Effort for Disguising

$$min_{n \neq u(i)} \, inf \lambda_{i,n}$$

user i's effort to move to cluster n

$$s.t. \ ||(1-\lambda_{i,n})\boldsymbol{d}_i + \lambda_{i,n}\boldsymbol{c}_n - \boldsymbol{c}_{u(i)}||_1 \geq ||(1-\lambda_{i,n})(\boldsymbol{d}_i - \boldsymbol{c}_n)||_1$$

$$p_n < p_{u(i)}$$

successfully switch to cluster n & lower price

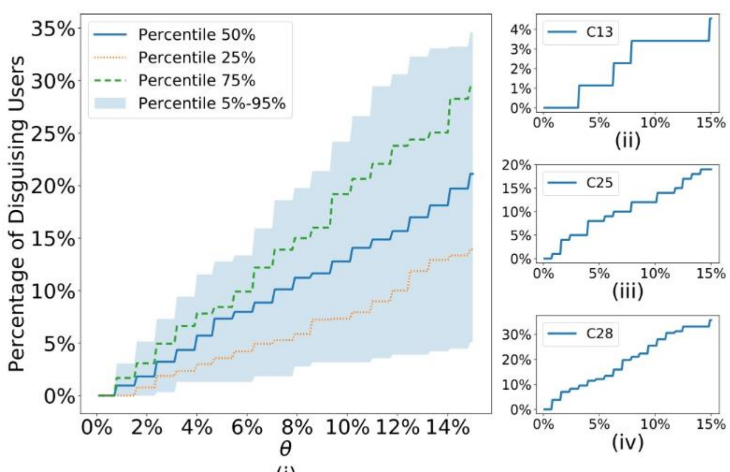
- Index to Differentiate Disguising
  - $CR_i = min_{n \in \{1,\ldots,k\}, n \neq u(i)} \, inf \lambda_{i,n}$

$u(i)$ is the cluster that user i belongs to.

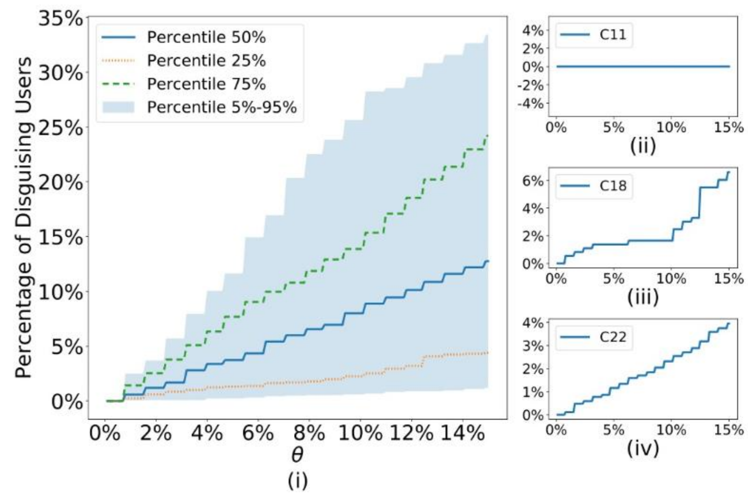
- Parametric Definition of Disguising
  - $CR_i \leq \theta$

Empirical Evidence: Strategic Behaviors

- Index to Quantify the Existence of Strategic Behaviors
  - $N_n(\theta) = \sum_{i,u(i)=n} I(CR_i \leq \theta)$
    - To define the number of users, having the ability to disguise, in each cluster n.



(a) Residential Users



(b) Commercial Building Users

Why Such Strategic Behaviors Exist?

- From End-to-End machine learning perspective…
  - It's because we select the wrong clustering criteria.
- We conduct k-means clustering for pricing. Hence, we should directly conduct the k-means clustering based on MCI.

- We define smoothness to guarantee robustness.
- ***(Definition)*** The *k*−means clustering is δ−smooth if for any user i and its associated disguising set $\mathcal{K}_i$, the following condition holds:

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

Local properties for smoothness!

- ***k*−means Clustering with Smoothness Guarantee**
  - ***(Theorem)*** Suppose a *k*−means clustering algorithm can guarantee that

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

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