Smart Meter Data-Driven Load Forecasting and Price Design in the Retail Market

Graduate Seminar @ KAUST

Yi Wang, Postdoc @ ETH
yiwang@eeh.ee.ethz.ch
Feb 21, 2021
Appointment
2019.2- Postdoc, ETH Zurich (Prof. Gabriela Hug)

Education
2014.9-2019.1 Ph.D., Tsinghua University (Prof. Chongqing Kang)
2017.3-2018.4 Visiting Student, University of Washington (Prof. Daniel Kirschen)
2010.9-2014.6 B.S., Huazhong University of Science and Technology

Research Interests
Data analytics for smart energy
Cyber-physical power and energy systems
Multi-energy systems integration
Acknowledgements

The works presented here are from the collaboration with the following professors and graduate students:

- **Prof. Gabriela Hug** and **Mr. Leandro Von Krannichfeldt** from ETH Zurich;
- **Prof. Chongqing Kang, Prof. Qixin Chen,** and **Mr. Cheng Feng** from Tsinghua University.


Outlines

- Backgrounds
- Aggregated Load Forecasting
- Personalized Retail Price Design
- Conclusions
Backgrounds

Supply the electricity with lower cost

Economy

Reliability

Sustainability

Keep the lights on

Accommodate more renewable energy

Trilemma

Hydro
Nuclear
Thermal
Grid
Load
Solar
Wind
Backgrounds

The power generation and consumption should be balanced in real-time.

Flexibility: Better ways of matching supply and demand over multiple time and spatio-scales.
Backgrounds

Various sensors and controllers will be installed in the power and energy systems.

Demand response in buildings, industry and transport could provide 185 GW of flexibility, and avoid USD 270 billion of investment in new electricity infrastructure.


https://www.iea.org/reports/digitalisation-and-energy
## Backgrounds

<table>
<thead>
<tr>
<th>No.</th>
<th>System/ Data</th>
<th>Data Source</th>
<th>Data Type</th>
<th>Frequency</th>
<th>Data Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Economic Information</td>
<td>Statistic Bureau</td>
<td>GDP, CPI, PMI (Purchasing Managers Index), Sales Value, Prosperity Index</td>
<td>Per Month</td>
<td>Non structural</td>
</tr>
<tr>
<td>2</td>
<td>Energy Consumption Data</td>
<td>Energy Efficiency Platform</td>
<td>Electrical Load, Output, Power Quality, Temperature</td>
<td>15Min</td>
<td>Non structural / Structural</td>
</tr>
<tr>
<td>3</td>
<td>Meteorological Data</td>
<td>Meteorological Bureau</td>
<td>Temperature, Humidity, Rainfall</td>
<td>Per Day</td>
<td>Structural</td>
</tr>
<tr>
<td>4</td>
<td>EV Charging Data</td>
<td>Charging-Pile RTU</td>
<td>Current, Voltage, Charging Rate, State of Charge</td>
<td>15Min</td>
<td>Structural</td>
</tr>
<tr>
<td>5</td>
<td>Customer Service Voice Data</td>
<td>Customer Service System</td>
<td>Customer Voice Data</td>
<td>Real Time</td>
<td>Non structural</td>
</tr>
</tbody>
</table>

**Variety**
10 million Smart Meters, 15min

**Value???
60GB per day, 21TB per year.

**Velocity**

**Volume**
Backgrounds

Participators and their businesses on the demand side

- Network Security
- Economic Operation

DSO

Retailer

Electricity Purchasing
- Price Design
- Personal Service
- Theft Detection

Aggregator

Demand Response
- Energy Efficiency

Consumer

Home Energy Management
- Transactive Energy

Smart meter data

Backgrounds

- Aggregated Load Forecasting
- Personalized Retail Price Design

Wholesale market

Retail market

Retailer

Consumers
Traditional load forecasting algorithms directly use historical data at the aggregation level.

With the prevalence of smart meters, fine-grained sub profiles reveal more information about the aggregated load and further help improve the forecasting accuracy.
Aggregated Load Forecasting

Introduction

Three strategies for aggregated load forecasting (ALF):
1) Top-down; 2) bottom-up; 3) clustering based.

Is it possible to utilize both ensemble techniques and fine-grained subprofiles to further improve the aggregated load forecasting accuracy?
Aggregated Load Forecasting

Introduction

Primary idea: instead of treating the aggregated load as a whole, partitioning consumers into several groups and making predictions might help improve load forecasting.

A three-stage approach for aggregated load forecasting with smart meter data:

- **Clustering**: divide consumers into different groups
- **Forecasting**: develop forecasting model for each group
- **Aggregation**: sum forecasts of all groups
Aggregated Load Forecasting

Introduction

Go further steps by ensemble learning?
Deterministic Aggregated Load Forecasting

If there are different partitions of consumers, we can obtain different load forecasts.

\[
G(x) = \sum_{m=1}^{M} w_m f_m(x)
\]
Aggregated Load Forecasting

Deterministic Aggregated Load Forecasting

How much weight should be given to each method for the optimal combination?

\[
\min_w \quad L(y - G(x))
\]

\[
\sum_{m=1}^{M} w_m = 1, \quad w_m \geq 0, \quad m = 1, ..., M
\]

Real load

The \(n\)-th predicted load

\[
\hat{\omega} = \arg \min_{\omega} \sum_{t=1}^{T} \frac{1}{T} \left| \frac{L_{en,t} - \hat{L}_{en,t}}{L_{en,t}} \right|
\]

Minimize MAPE

\[
s.t. \quad \hat{L}_{en,t} = \sum_{n=1}^{N} \omega_n \hat{L}_{en,n,t}, \quad \sum_{n=1}^{N} \omega_n = 1, \quad \omega_n \geq 0.
\]

To determine the weights for the forecasts

It can be formulated as an LP problem.
Aggregated Load Forecasting

Deterministic Aggregated Load Forecasting

Weights, MAPE, and RMSE of different forecasts with different groups

<table>
<thead>
<tr>
<th>N</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>...</th>
<th>5237</th>
<th>Ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>ω</td>
<td>0.634</td>
<td>0</td>
<td>0</td>
<td>0.271</td>
<td>0</td>
<td>0</td>
<td>0.095</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>/</td>
</tr>
<tr>
<td>MAPE</td>
<td>4.25%</td>
<td>5.05%</td>
<td>5.29%</td>
<td>4.74%</td>
<td>5.55%</td>
<td>4.66%</td>
<td>4.79%</td>
<td>5.09%</td>
<td>5.59%</td>
<td>...</td>
<td>10.31%</td>
<td>4.05%</td>
</tr>
<tr>
<td>RMSE</td>
<td>210.95</td>
<td>229.73</td>
<td>228.01</td>
<td>217.68</td>
<td>244.9</td>
<td>217.64</td>
<td>227.36</td>
<td>232.61</td>
<td>250.27</td>
<td>...</td>
<td>441.33</td>
<td>202.88</td>
</tr>
</tbody>
</table>

The MAPE and RMSE of the proposed ensemble method are 4.05% and 202.88 which gain 4.71% and 3.83% improvements, respectively compared with the best individual forecast.

red line: actual load blue line: ensemble forecast
dashed lines: individual forecasts
Aggregated Load Forecasting

- Deterministic Aggregated Load Forecasting

Can we update the weights in a rolling window-based manner?

- **D_{Train}**
- **D_{Ensemble}**
- **D_{Test}**

- Clustering
- Train Individual | Test Individual | Test Individual
- Train Ensemble | Test Ensemble

- Round 1  
- Round 2
- \vdots
- Round W  

- Train Ensemble | Test
- Train Ensemble | Test
Aggregated Load Forecasting

- Deterministic Aggregated Load Forecasting

Benefits of window-based method

<table>
<thead>
<tr>
<th>Ensemble Method</th>
<th>Error Metrics</th>
<th>Window</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>COP_{MAPE}</td>
<td>MAPE</td>
<td>2.85%</td>
<td>3.13%</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>106.13</td>
<td>116.66</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>149.81</td>
<td>166.74</td>
</tr>
<tr>
<td>COP_{MSE}</td>
<td>MAPE</td>
<td>2.89%</td>
<td>3.15%</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>107.3</td>
<td>116.8</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>151.26</td>
<td>166.92</td>
</tr>
</tbody>
</table>

Ensemble weights over 17 weeks of the test set for all individual models.
Aggregated Load Forecasting

Deterministic Aggregated Load Forecasting

Combined model:

\[ G(x) = \sum_{m=1}^{M} w_m f_m(x) \]

Online Convex Optimization (OCO) is a unifying framework for the analysis and design of online algorithms.
Aggregated Load Forecasting

- **Deterministic Aggregated Load Forecasting**
  - **General formula**
    \[
    w_{t+1} = \arg \min_w \left[ d(w, w_t) + \eta_t \ell(y_t, w \cdot x_t) \right]
    \]
    - **Distance** \(d\)
    - Prevent information loss
    - **Loss** \(\ell\)
    - Integrate new sample

**Passive Aggressive Regression**

- **Aggressive:** weights change if losses are big enough
  \[
  w_{t+1} = \arg \min_w \left[ \|w - w_t\|_1 + \ell_\varepsilon(y_t, w \cdot f_t) + \lambda \|w\|_1 \right]
  \]
- **Passive:** weights do not change every time slot
  \[
  \ell_\varepsilon(y_t, w \cdot f) = \begin{cases} 
  0 & \text{if } |y - w \cdot f| \leq \varepsilon \\
  |y - w \cdot f| & \text{otherwise}
  \end{cases}
  \]
Aggregated Load Forecasting

- **Deterministic Aggregated Load Forecasting**

  Update the weights online for a better performance

  Errors on test set after online learning

<table>
<thead>
<tr>
<th>Method</th>
<th>MAPE</th>
<th>SD</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGDR</td>
<td>2.43%</td>
<td>0.025</td>
<td>86.05</td>
<td>122.71</td>
</tr>
<tr>
<td>FTRLP</td>
<td>2.23%</td>
<td>0.021</td>
<td>81.09</td>
<td>113.87</td>
</tr>
<tr>
<td>OSELM</td>
<td>2.80%</td>
<td>0.029</td>
<td>106.03</td>
<td>155.03</td>
</tr>
<tr>
<td>Online Bagging</td>
<td>2.07%</td>
<td>0.021</td>
<td>74.33</td>
<td>106.23</td>
</tr>
<tr>
<td>PAR</td>
<td>1.67%</td>
<td>0.015</td>
<td>61.83</td>
<td>86.68</td>
</tr>
<tr>
<td>Proposed</td>
<td>1.62%</td>
<td>0.014</td>
<td>59.59</td>
<td>83.21</td>
</tr>
<tr>
<td>Best SVR</td>
<td>3.18%</td>
<td>0.032</td>
<td>117.54</td>
<td>171.72</td>
</tr>
<tr>
<td>Best RF</td>
<td>2.89%</td>
<td>0.029</td>
<td>108.25</td>
<td>156.84</td>
</tr>
<tr>
<td>Best GBRT</td>
<td>3.53%</td>
<td>0.032</td>
<td>127.81</td>
<td>175.78</td>
</tr>
<tr>
<td>Batch OPT</td>
<td>2.89%</td>
<td>0.028</td>
<td>107.55</td>
<td>154.88</td>
</tr>
<tr>
<td>Window OPT</td>
<td>2.85%</td>
<td>0.028</td>
<td>106.13</td>
<td>149.81</td>
</tr>
</tbody>
</table>

- All ensembles improve their forecasting performance through online learning.
- Nearly all ensembles outperform the benchmarks after online learning.
- The proposed method has the highest accuracy and stability among all examined ensembles.

SD: Standard deviation of the absolute percentage error
Aggregated Load Forecasting

- **Deterministic Aggregated Load Forecasting**

  Update the weights online for a better performance

  The hour of break-even for all ensembles

<table>
<thead>
<tr>
<th>Method</th>
<th>MAPE</th>
<th>SD</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGDR</td>
<td>39.5</td>
<td>86.5</td>
<td>41.0</td>
<td>64.0</td>
</tr>
<tr>
<td>FTRLP</td>
<td>66.5</td>
<td>87.0</td>
<td>64.0</td>
<td>60.5</td>
</tr>
<tr>
<td>PAR</td>
<td>17.5</td>
<td>9.0</td>
<td>19.5</td>
<td>17.5</td>
</tr>
<tr>
<td>OSELM</td>
<td>112.0</td>
<td>2.0</td>
<td>2833.5</td>
<td>no</td>
</tr>
<tr>
<td>Online Bagging</td>
<td>22.5</td>
<td>4.5</td>
<td>23.0</td>
<td>35.5</td>
</tr>
<tr>
<td>Proposed</td>
<td>1.5</td>
<td>2.0</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

MAPE over the course of the first day of online learning

- The proposed method has the earliest break-even after 2 hours for all metrics.
- The other ensembles have the break-even approximately within one or two days.
- An ensemble employing online learning is able to pay off at a relatively early point in time.
Aggregated Load Forecasting

Probabilistic Aggregated Load Forecasting

Compared with deterministic forecasting, probabilistic load forecasts provide comprehensive information about future uncertainties.
Aggregated Load Forecasting

- **Probabilistic Aggregated Load Forecasting**

Pinball loss (PL) and Winkler Score (WS) assess the calibration and sharpness simultaneously.

\[
\text{PL}(\hat{y}_{t,q}, y_t) = \begin{cases} 
(y_t - \hat{y}_{t,q})q & \hat{y}_{t,q} \leq y_t \\
(\hat{y}_{t,q} - y_t)(1-q) & \hat{y}_{t,q} > y_t
\end{cases}
\]

\[
\text{WS}(L_t, U_t, y_t) = \begin{cases} 
\delta_t + 2(L_t - y_t)/\alpha & y_t \leq L_t \\
\delta_t & L_t < y_t < U_t \\
\delta_t + 2(y_t - U_t)/\alpha & U_t \leq y_t
\end{cases}
\]

- **Performance of overall quantiles**
- **Performance of extreme quantiles**

**Average Coverage Error (ACE)** evaluate the reliability of the forecasts.

\[
ACE = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{y_i \in [L_i, U_i]\}} - (1 - \alpha)
\]

- **Performance of an certain interval**
Aggregated Load Forecasting

- **Probabilistic Aggregated Load Forecasting**

\[ x_i = [1, \hat{y}_{1,i}, ..., \hat{y}_{K,i}] \quad i \in [1, ..., n] \]

PCA

\[ \hat{w}_q = \arg \min_{w_q} \sum_{i=1}^{n} \rho_q(y_i - z_i w_q) \]

Factor Quantile Regression Averaging

\[ \hat{w}_q = \arg \min_{w_q} \sum_{i=1}^{n} \rho_q(y_i - x_i w_q) \]

Quantile regression averaging (QRA), a special form of quantile regression, is a kind of model averaging method.

LASSO Quantile Regression Averaging

\[ \hat{w}_q = \arg \min_{w_q} \sum_{i=1}^{n} \rho_q(y_i - x_i w_q) + \lambda \| w_q \|_1 \]
Aggregated Load Forecasting

- Probabilistic Aggregated Load Forecasting

Similar to deterministic forecasting......

- Clustering
- Train Individual | Test Individual | Test Individual
- Train Ensemble | Test Ensemble

Round 1
Round 2
Round W
Aggregated Load Forecasting

- Probabilistic Aggregated Load Forecasting

Error metric comparison for all ensemble methods with a Prediction Interval of 90%.

<table>
<thead>
<tr>
<th>Ensemble Method</th>
<th>Error Metrics</th>
<th>Offline Ensemble</th>
<th>Benchmark 1</th>
<th>Rolling Window-based Ensemble</th>
<th>Benchmark 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>QRA</td>
<td>ACE</td>
<td>-1.73%</td>
<td>-1.85%</td>
<td>-0.56%</td>
<td>-0.92%</td>
</tr>
<tr>
<td></td>
<td>PBL</td>
<td>45.82</td>
<td>50.19</td>
<td>42.28</td>
<td>46.52</td>
</tr>
<tr>
<td></td>
<td>WKS</td>
<td>788.62</td>
<td>846.89</td>
<td>728.13</td>
<td>791.78</td>
</tr>
<tr>
<td>FQRA</td>
<td>ACE</td>
<td>-1.80%</td>
<td>-1.85%</td>
<td>-0.45%</td>
<td>-0.92%</td>
</tr>
<tr>
<td></td>
<td>PBL</td>
<td>45.82</td>
<td>50.19</td>
<td>42.26</td>
<td>46.52</td>
</tr>
<tr>
<td></td>
<td>WKS</td>
<td>787.26</td>
<td>846.89</td>
<td>727.24</td>
<td>791.77</td>
</tr>
<tr>
<td>LQRA</td>
<td>ACE</td>
<td>-1.71%</td>
<td>-1.83%</td>
<td>-0.63%</td>
<td>-0.98%</td>
</tr>
<tr>
<td></td>
<td>PBL</td>
<td>45.84</td>
<td>50.2</td>
<td>42.26</td>
<td>46.53</td>
</tr>
<tr>
<td></td>
<td>WKS</td>
<td>785.77</td>
<td>845.7</td>
<td>724.74</td>
<td>791.55</td>
</tr>
</tbody>
</table>

- The two naive benchmarks are obtained by directly forecasting the total loads without dimension reduction and clustering.
- Benchmark 2 updates the weights in a rolling window-based approach, while Benchmark 1 does not.
**Aggregated Load Forecasting**

- **Probabilistic Aggregated Load Forecasting**

**Combined model:**

$$G(x) = \sum_{m=1}^{M} w_m f_m(x)$$

- **General formula**

$$w_{t+1} = \arg \min_w \left[ d(w, w_t) + \eta_t \ell(y_t, w \cdot x_t) \right]$$

- **Distance** $d$
- **Loss** $\ell$
- **Prevent information loss**
- **Integrate new sample**
Aggregated Load Forecasting

Probabilistic Aggregated Load Forecasting

- General Formula

\[ w_{t+1} = \arg \min_w \left[ d(w, w_t) + \eta_t \ell(y_t, w \cdot x_t) \right] \]

- L₂-distance:

\[ d(\cdot) = \frac{1}{2} \| \cdot \|^2 \]

- \( \varepsilon \)-insensitive quantile loss:

\[
\ell_{\varepsilon, q}(w_q; x, y) = \begin{cases} 
q(y - w_q \cdot x + \varepsilon(q - 1)) & \text{if } y - w_q \cdot x > \varepsilon(1 - q) \\
0 & \text{if } -\varepsilon q < y - w_q \cdot x < \varepsilon(1 - q) \\
(q - 1)(y - w_q \cdot x + \varepsilon q) & \text{if } y - w_q \cdot x < -\varepsilon q
\end{cases}
\]

- Solving KKT conditions:

\[
w_{t+1} = w_t + \eta_t \text{sign}(y_t - w_t \cdot x_t) \tau_t x_t \quad \tau_t = \min \left\{ C, \frac{\ell_{\varepsilon, q}(y_t, w_t \cdot x_t)}{q \| x_t \|^2} \right\}
\]
Aggregated Load Forecasting

Probabilistic Aggregated Load Forecasting

- General Formula
  \[
  w_{t+1} = \arg \min_w \left[ d(w, w_t) + \eta_t \ell(y_t, w \cdot x_t) \right]
  \]

- L₂-distance:
  \[
  d(\cdot) = \frac{1}{2} \| \cdot \|^2
  \]

- \(\varepsilon\)-insensitive quantile loss:

- Solving KKT conditions:
  \[
  w_{t+1} = w_t + \eta_t \text{sign}(y_t - w_t \cdot x_t) \tau_t x_t
  \]
  \[
  \tau_t = \min \left\{ C, \frac{\ell_{\varepsilon,q}(y_t, w_t \cdot x_t)}{q \| x_t \|^2} \right\}
  \]
Aggregated Load Forecasting

Probabilistic Aggregated Load Forecasting

Mechanism of Quantile Passive Aggressive Regression

- Extension to probabilistic forecasting: $\varepsilon$-insensitive loss $\rightarrow$ $\varepsilon$-insensitive quantile loss
- $\varepsilon$-insensitive region: Preserve «quantile height» between $y_q$ and $y$

- Batch quantile regression
  - Access to whole data sequence
  - «Statistical height» implicitly given

- Online quantile regression
  - Only access to one sample per round
  - «Statistical height» collapses $\rightarrow$ Real value

- $\varepsilon$-insensitive quantile: Preserve «statistical height»

\[
\ell_q(y, \hat{y}_q) = \begin{cases} 
q(y - \hat{y}_q) & \text{if } y \geq \hat{y}_q \\
(q - 1)(y - \hat{y}_a) & \text{if } y < \hat{y}_a
\end{cases}
\]
Aggregated Load Forecasting

Probabilistic Aggregated Load Forecasting

The performance on Irish load data

<table>
<thead>
<tr>
<th>Method</th>
<th>ACE</th>
<th>PBL</th>
<th>WKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSGD</td>
<td>-0.92%</td>
<td>51.60</td>
<td>722.43</td>
</tr>
<tr>
<td>QPAR</td>
<td>2.23%</td>
<td>47.61</td>
<td>1075.02</td>
</tr>
<tr>
<td>QNN</td>
<td>-2.55%</td>
<td>54.94</td>
<td>776.86</td>
</tr>
<tr>
<td>Batch QRA</td>
<td>-5.25%</td>
<td>44.55</td>
<td>734.64</td>
</tr>
<tr>
<td>Window QRA</td>
<td>-1.90%</td>
<td>40.30</td>
<td>659.94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>ACE</th>
<th>PBL</th>
<th>WKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSGD</td>
<td>-0.02%</td>
<td>30.04</td>
<td>527.94</td>
</tr>
<tr>
<td>QPAR</td>
<td>-1.69%</td>
<td>29.47</td>
<td>484.59</td>
</tr>
<tr>
<td>QNN</td>
<td>-0.64%</td>
<td>56.10</td>
<td>930.23</td>
</tr>
<tr>
<td>Batch QRA</td>
<td>-5.25%</td>
<td>44.55</td>
<td>734.64</td>
</tr>
<tr>
<td>Window QRA</td>
<td>-1.90%</td>
<td>40.30</td>
<td>659.94</td>
</tr>
</tbody>
</table>

*QSGD: Quantile Stochastic Gradient Descent
*QPAR: Quantile Passive Aggressive Regression
*QNN: Quantile Neural Network
*Window OPT: window-based optimization

- All ensembles outperform the benchmarks after online learning except QNN
- The proposed method has the highest accuracy regarding pinball loss and winkler score
- A substantial performance improvement can be achieved by ensembles incorporating online learning.
Aggregated Load Forecasting

- **Probabilistic Aggregated Load Forecasting**

QSGD online forecast over one week

QPAR online forecast over one week
Aggregated Load Forecasting

Probabilistic Aggregated Load Forecasting

The performance on Irish load data

The hour of break-even for all ensembles

<table>
<thead>
<tr>
<th>Method</th>
<th>Break-Even ACE</th>
<th>Break-Even PBL</th>
<th>Break-Even WKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSGD</td>
<td>508.0 h</td>
<td>35.0 h</td>
<td>307.0 h</td>
</tr>
<tr>
<td>QPAR</td>
<td>2810.0 h</td>
<td>138.5 h</td>
<td>253.5 h</td>
</tr>
<tr>
<td>QNN</td>
<td>687.0 h</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

- The proposed QPAR has earliest WKS break-even
- QSGD has earliest Break-even for ACE and PBL
- Online learning enables to outperform batch approach within a month.
Aggregated Load Forecasting

Short Summary

Generate ensembles

Combine ensembles

Deterministic forecasting

Probabilistic forecasting

Clustering → Individual forecasting → Aggregation

Static ensemble → Rolling window-based ensemble → Online ensemble

Convex problem → Rolling window-based ensemble → PAR model

QRA → Rolling window-based QRA → Quantile PAR

Deterministic forecasting

Probabilistic forecasting

Generate ensembles

Combine ensembles

Clustering → Individual forecasting → Aggregation

Static ensemble → Rolling window-based ensemble → Online ensemble

Convex problem → Rolling window-based ensemble → PAR model

QRA → Rolling window-based QRA → Quantile PAR
Aggregated Load Forecasting

➢ **Short Summary**

- High quality point forecasting can be generated by making full use of the fine grained smart meter data;
- On this basis, we can utilize ensemble techniques to further improve the forecasting accuracy;
- Online learning can be a powerful tool in short-term load forecasting by integration new information and the proposed modified PAR model is very suitable in this context, especially as an online ensemble method;
- PAR model can be further extend to quantile PAR model using quantile regression averaging for probabilistic forecasting.
Personalized Retail Price Design

- **Introduction**
  - The opening of electricity retailing market
  - The need for diversified service
  - Consumers choose freely in market

How to provide **diversified services** for different consumers to enhance the competitiveness of the retailers?
Personalized Retail Price Design

Main Idea

- **Data-driven price design.** Smart meter data contains great value which may help retailing price design.
- **Respect self-selection.** Consumers’ willingness and rights to choose must be respected.

Challenges

- Diversified service
- Mine consumers’ inner need
- Satisfying consumers
- Self-selection in a real market
- Proper incentive

Data-driven price design

Compatible incentive design

Discover utility from data → Cluster load profiling data → Correlate preference with shape → Make centroids representatives → Form optimization problem
Personalized Retail Price Design

- A Stackelberg game

**Leader—Retailer**
- Design pricing schemes
- Predict consumer behaviors

**Followers—Consumers**
- Choose one pricing scheme
- Adapt electricity consumption

Diagram:
- Wholesale market
- Retail market
- Retailer
- Consumers
# Personalized Retail Price Design

## Problem formulation - Consumer

#### Consumer Utility
- Measure satisfaction
- Comparison between different plans
- Diminishing marginal utility

\[
F(p, q) = u(q) - \sum_{t=1}^{T} p_t q_t
\]

#### Consumer Strategy
- Strategic and rational consumers: 
  \[
  q^*(p) = \arg \max_{q} \{ F(p, q) \}
  \]

\[
U(p) = \max_{q} \{ F(p, q) \} = F(p, q^*(p))
\]

How can smart meter data be useful?

\[
F(p(0), q(0)) = 0 \quad \frac{\partial F(p(0), q(0))}{\partial q_t} = 0, \quad \forall t
\]

Original electricity consumption is the realization of Utility Maximization!
Personalized Retail Price Design

Problem formulation - incentive

**Individual rationality**

If the retailer wants consumer \( k \) to choose pricing scheme \( r \), the retailer **must guarantee** choosing \( r \) is consumer \( k \)’s dominant strategy

\[
U_k(p_r) \geq U_k(p') \quad \forall k
\]

**Compatible incentive**

If the retailer wants consumer \( k \) to choose new pricing scheme \( r \), the retailer **must guarantee** choosing \( r \) is at least as good as previous situation

\[
U_k(p_r) \geq U_k(p_0) \quad \forall k
\]
Personalized Retail Price Design

Problem formulation - Retailer

Where the retailer purchases electricity?
- Forward contracts
- Day-ahead market
- Real-time market

Balance predictable load
- Price uncertainty
- Risk loss measure — CVaR

Balance unpredictable load
- Price and load uncertainty

Which is considered more important?
Risk Weighting factor

Purchasing strategy
Personalized Retail Price Design

- Electric Reliability Council of Texas (ERCOT)
- Extreme Cold Scenarios
- Rotating Outages
- Extreme-high Price !!!
- Price Uncertainty
- CVaR

https://www.dallasnews.com/opinion/commentary/2021/02/20/dont-just-blame-ercot-what-caused-outages-is-our-competitive-electricity-market/
Personalized Retail Price Design

Problem formulation - Clustering

Different Clustering Methods

- Hierarchical Clustering
- K-means
- Fuzzy C-means
- Gaussian mixture

Clustering evaluation

Davies Bouldin Index
With-cluster compactness
Between-cluster separation

Centroid as representative

One method may not fit all data sets
Personalized Retail Price Design

Problem formulation – Optimization framework

Optimization framework – an MINLP model

- **Objective**: Retailing profit maximization
- **Constraints**:
  - Load balance
  - Consumer reaction
  - Compatible incentive
  - Risk measure CVaR
  - Price structure: Various choices

- **Price category**: CPP, RTP, ToU

Lower Risk
Less changes
3. Personalized Retail Price Design

Optimization framework – an MINLP model

- **Objective**: Retailing profit maximization
  \[ \max R = \sum_{r=1}^{R} \sum_{t=1}^{T} K_r \times p_{k,t} \times q_{k,t} - \sum_{t=1}^{T} \sum_{n=1}^{N_F} p_n^F \times L_n^F \times o_{n,t}^F \times o_n - \sum_{t=1}^{T} p_t^{D,est} \times L_t^D - \xi \times CVaR \]
  - Consumer payment
  - Forward contracts
  - DA
  - Risk Loss in DA & RT

- **Constraints**: Load balance
  \[ \sum_{r=1}^{R} K_r \times q_{r,t} = \sum_{n=1}^{N_F} L_n^F \times o_{n,t}^F \times o_n + L_t^D, \quad \forall t \]
  - Consumer load
  - Forward contracts
  - DA
  - DA=Day-ahead market
  - RT= Real-time market

- **Constraints**: Compatible incentive
  \[ U_r(p_r) \geq U_r(p') \quad \forall k \]  
  Choosing \( p_r \) is consumer \( k \)'s dominant strategy, \( k \) likes \( p_r \) than any other pricing schemes

  \[ U_r(p_r) \geq U_r(p_0) \quad \forall k \]  
  Choosing \( p_r \) is consumer \( k \)'s rational choice, \( k \) likes \( p_r \) than the old pricing schemes

* nonlinear terms are marked in red
3. Personalized Retail Price Design

Optimization framework – an MINLP model

**Constraints**: Utility and response

\[ q_t = \left( \frac{p_t}{p_{t(0)}} \right)^{\frac{1}{\alpha - 1}} \times q_{t(0)} \]

\[ U(p) = \sum_{t=1}^{T} \left( \frac{1}{\alpha} - 1 \right) \left[ \left( \frac{p_t}{p_{t(0)}} \right)^{\frac{\alpha}{\alpha - 1}} - 1 \right] \times q_{t(0)} p_{t(0)} \]

**Constraints**: Risk measure CVaR

\[ CVaR = \inf_{a \in R} \left\{ a + \frac{1}{(1 - \alpha)^{CVaR}} \cdot N_S \sum_{n_S=1}^{N_S} \left[ (- \Delta R^D - \Delta R^{RT}) - a \right]^+ \right\} \]

**Constraints**: Price structure

Price category: CPP  RTP  ToU

- Lower Risk
- Less changes

\[ \sum_{m=1}^{M} e_{r,t}^m = 1, \quad \sum_{t=1}^{T} e_{r,t}^m \geq D_{min}, \quad \forall m, r \]

\[ |e_{r,T}^m - e_{r,1}^m| + \sum_{t=2}^{T} |e_{r,t-1}^m - e_{r,t}^m| = 2, \quad \forall m, r \]

\[ p_{r,t} = \sum_{m=1}^{M} e_{r,t}^m \times p_r^m, \quad \forall t, r \quad m \text{ block ToU} \]

* nonlinear terms are marked in red

**MINLP model**  
Big M method  
Piecewise linear approximation  
**MILP model**
Personalized Retail Price Design

Solution method

Nonlinear model

- Power exponent \( \alpha \) \( \frac{\alpha}{p_{r,t}^{\alpha-1}}, \frac{1}{p_{r,t}^{\alpha-1}} \)
- Two variables’ product \( p_{r,t} \times q_{r,t} \)

Linear model

- Linear segment approximation
  Take \( p_{r,t} \times q_{r,t} \) as a whole
Personalized Retail Price Design

Solution method

Nonlinear model

- Binary variables times continuous variables
- Absolute value $|e_{r,t-1}^m - e_{r,t}^m|$
- CVaR

Linear model

- Add auxiliary variables
  - Conversed to linear equations

\[
\begin{align*}
\sigma_{r,t} &\leq M \times e_{t,r}^m \\
\sigma_{r,t} &\leq p_r^m \\
\sigma_{r,t} &\geq p_r^m - M \times (1 - e_{t,r}^m) \\
\sigma_{r,t} &\geq 0 \\
e_1 - e_2 &\leq A \leq e_1 - e_2 + 2 \times B \\
e_2 - e_1 &\leq A \leq e_2 - e_1 + 2 \times (1 - B) \\
CVaR &\geq a + \frac{1}{(1 - \alpha_{CVaR}) \cdot N} \sum_{n=1}^{N} W_{ns} \\
W_{ns} &\geq 0 \\
W_{ns} &\geq [(-\Delta R_{ns}^D - \Delta R_{ns}^{RT}) - a]
\end{align*}
\]

* new variables are marked in blue
Personalized Retail Price Design

Case Study

- 6435 consumers in Ireland.
- Data collected every 30 minutes.

Linear segment approximation (12 segments)
Personalized Retail Price Design

Case Study - clustering

DB index result

Clustering result
Personalized Retail Price Design

Case Study – prices and responses
**Personalized Retail Price Design**

Case Study – sensitivity analysis on elasticity

### Total load under different elasticity

<table>
<thead>
<tr>
<th>Elasticity</th>
<th>Original</th>
<th>-0.2</th>
<th>-0.3</th>
<th>-0.4</th>
<th>-0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retailing Profit($)</td>
<td>752</td>
<td>833</td>
<td>977</td>
<td>1186</td>
<td>1385</td>
</tr>
</tbody>
</table>

### Retailing profit under different elasticity

Elasticity ↓ 😞 Willingness to change ↓

---

**Notes:**
- **ToU under different elasticity**
- **Cluster1, Cluster2, Cluster3, Cluster4, Cluster5, Cluster6**
- **Power Systems Laboratory**
Personalized Retail Price Design

Case Study – sensitivity analysis on risk weighting factor

How CVaR, the quantity of power bought from day-ahead market and through forward contracts changes with the change of risk weighting factor?

- risk weighting factor rises (↑)
- attach more importance to risk (↑)
- minimize CVaR (↓)
- buy less from day-ahead market (↓)
- buy more through forward contracts (↑)
## Personalized Retail Price Design

### Case Study - sensitivity analysis on clustering methods

<table>
<thead>
<tr>
<th>Method</th>
<th>RP</th>
<th>SW</th>
<th>AP</th>
<th>F/SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>752.03</td>
<td>0</td>
<td>0.2</td>
<td>-/-</td>
</tr>
<tr>
<td>HIA-COMP</td>
<td>1186.01</td>
<td>339.72</td>
<td>0.1947</td>
<td>65%/89%</td>
</tr>
<tr>
<td>HIA-WARD</td>
<td>1188.70</td>
<td>10.01</td>
<td>0.1971</td>
<td>33%/59%</td>
</tr>
<tr>
<td>KM-PLUS</td>
<td>1145.68</td>
<td>7.01</td>
<td>0.1973</td>
<td>9%/20%</td>
</tr>
<tr>
<td>KM-SAMPLE</td>
<td>1137.61</td>
<td>4.50</td>
<td>0.1975</td>
<td>22%/48%</td>
</tr>
<tr>
<td>KM-UNIFORM</td>
<td>1142.61</td>
<td>15.76</td>
<td>0.1973</td>
<td>11%/31%</td>
</tr>
<tr>
<td>FCM(m=1.1)</td>
<td>1150.43</td>
<td>9.43</td>
<td>0.1970</td>
<td>30%/47%</td>
</tr>
<tr>
<td>FCM(m=1.2)</td>
<td>1176.08</td>
<td>18.64</td>
<td>0.1968</td>
<td>19%/35%</td>
</tr>
<tr>
<td>FCM(m=1.3)</td>
<td>1208.06</td>
<td>0.64</td>
<td>0.1970</td>
<td>8%/20%</td>
</tr>
<tr>
<td>GMEM-PLUS</td>
<td>1145.82</td>
<td>36.01</td>
<td>0.1965</td>
<td>13%/28%</td>
</tr>
<tr>
<td>GMEM-RAND</td>
<td>1144.85</td>
<td>46.60</td>
<td>0.1967</td>
<td>10%/24%</td>
</tr>
</tbody>
</table>

- How much profit does the retailer get?
  - RP=Retailing Profit($)
- How much welfare do the consumers get?
  - SW=Social Welfare
  - AP=Average Price($/kWh)
- How well does clustering perform?
  - F/SC=First/Second Choice

- The most accurate prediction
- The most profitable for both retailer and consumers
Personalized Retail Price Design

- **Short Summary**

- The Stackelberg game between the retailer and the strategic consumers, an incentive-compatible market, and the retailer’s costs, risks and purchasing strategy are considered in this model.
- The ToU tariff can achieve the effects of peak shaving and valley filling, thereby simultaneously increasing the retailer’s profitability and ensuring consumers’ willingness and preferences.
- How elasticity of consumers and risk weighting factor of retailer influence the designed price is studied.
Conclusions

- How to make full use of fine-grained smart meter data?

- A better understanding of the consumer behavior helps to improve the accuracy/performance of aggregated load forecasting.

- A better understanding of the consumer behavior helps to make better decision for both retailer and consumers.

- Any other applications???
## Any other applications??

<table>
<thead>
<tr>
<th>No.</th>
<th>System/Data</th>
<th>Data Source</th>
<th>Data Type</th>
<th>Frequency</th>
<th>Data Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Economic Information</td>
<td>Statistic Bureau</td>
<td>GDP、CPI、PMI（Purchasing Managers Index）、Sales Value、Prosperity Index</td>
<td>Per Month</td>
<td>Non structural</td>
</tr>
<tr>
<td>2</td>
<td>Energy Consumption Data</td>
<td>Energy Efficiency Platform</td>
<td>Electrical Load、Output、Power Quality、Temperature</td>
<td>15Min</td>
<td>Non structural/Structural</td>
</tr>
<tr>
<td>3</td>
<td>Meteorological Data</td>
<td>Meteorological Bureau</td>
<td>Temperature、Humidity、Rainfall</td>
<td>Per Day</td>
<td>Structural</td>
</tr>
<tr>
<td>4</td>
<td>EV Charging Data</td>
<td>Charging-Pile RTU</td>
<td>Current、Voltage、Charging Rate、State of Charge</td>
<td>15Min</td>
<td>Structural</td>
</tr>
<tr>
<td>5</td>
<td>Customer Service Voice Data</td>
<td>Customer Service System</td>
<td>Customer Voice Data</td>
<td>Real Time</td>
<td>Non structural</td>
</tr>
</tbody>
</table>
This is a book worth reading; one will see how much insight can be gained from smart meter data alone.

Prof. Saifur Rahman
IEEE Fellow
President of the IEEE Power and Energy Society

Thank you for your attention

Yi Wang | yiwang@eeh.ee.ethz.ch | www.eeyiwang.com