Usage Disaggregation of Smart Meter Data of Japanese Commercial Customers Using Machine Learning

August 26th, 2019

16th European Conference Ljubljana

Minao Watanabe, Kenta Ofuji

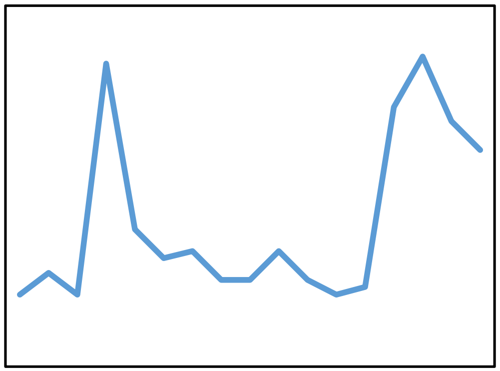
The University of Aizu

【Contents】

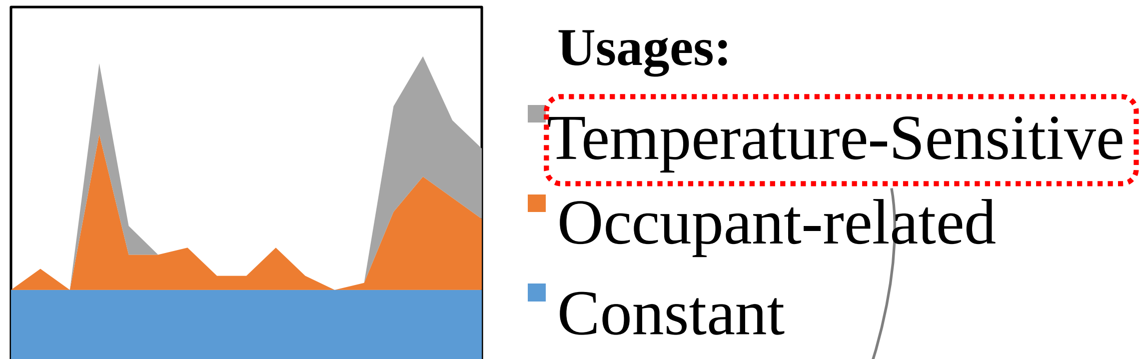
1. Approach and method of “Disaggregation”
2. Evaluation of intra-building by random forest regression
3. Evaluation of across-building by lightGBM

What is Disaggregation for?

- Disaggregate into several usage data from only whole-building smart meter data



Whole-building demand



Disaggregated by usages

Helps understand
- **peak occurrences** and
- **usage composition**

Data

- Japanese commercial buildings smart meter open data
  - Collected under Japanese government-aided energy efficiency projects
  - Covers 16 business types (Restaurants, Hospitals, Schools, ...), approx. **6,000** buildings across Japan
    - ✓ Use only "Supermarket" data
  - 1 hour granularity x 1 year (8,760 hours)
  - Data items:
    - ✓ Business Attributes

      E.g.: Industry, Location, Floor Area, …
    - ✓ **Six distinct usages** + Whole-building demand

- **Air Conditioning**
- **Freezing/Cooling Energy**
- Lighting, Power, ….

Disaggregate by :

- Random Forest regression (RFR)
- lightGBM algorithms

Motivation to disaggregate A/C versus Freezing/Cooling demand

- Both temperature-sensitive but with different underlying equipment
- **Of greatest magnitude** within gross building demand

Temperature-sensitive demand

Air Conditioning	Freezing / Cooling
e.g. Package air conditioning units	e.g. Freezer / Chiller

Forecasting model strategy

Explanatory Variables

Feature Matrix

X

RFR
lightGBM

Target Variable

Demand

y

- **Air Conditioning**
- **Freezing / Cooling**

<u>Feature Values used:</u>

1-1. Whole-building demand

1-2. 1-hour previous value of *y* → Auto Regressive(1) model

2-1. Hour of day (1 - 24)

2-2. Day of the week (Sun. – Sat.)

3. Proxy temperature (of nearby city)

【Contents】

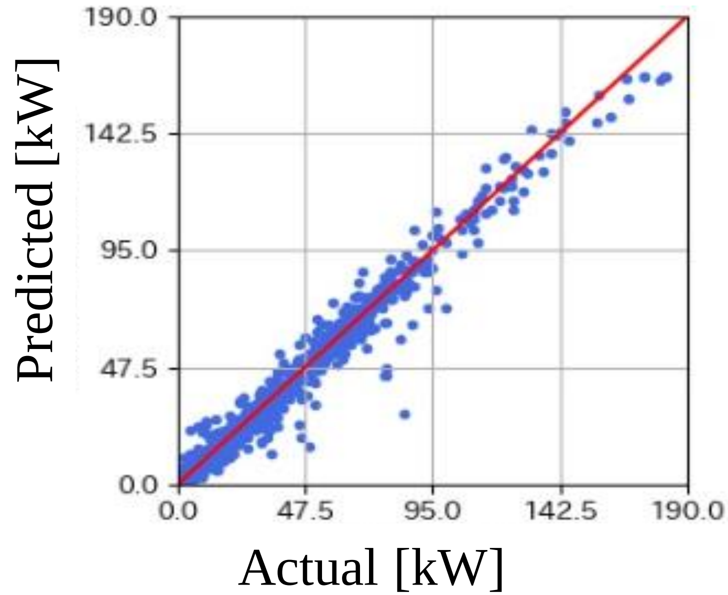
1. Approach and method of “Disaggregation”
2. Evaluation of intra-building by random forest regression
3. Evaluation of across-building by lightGBM

Results for one building : Air Conditioning , using RFR

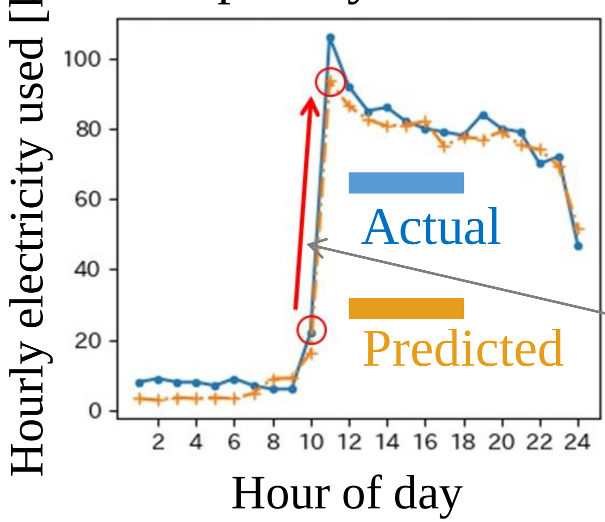
- Target: 30% in a year

  Mean Sq. Errors = 19.7 [%]

  (MSE)



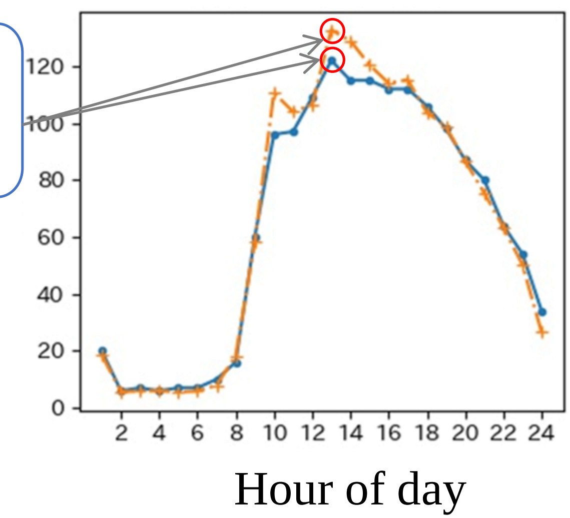
Sample day in **Winter**



Prediction error only10kW in peak hour

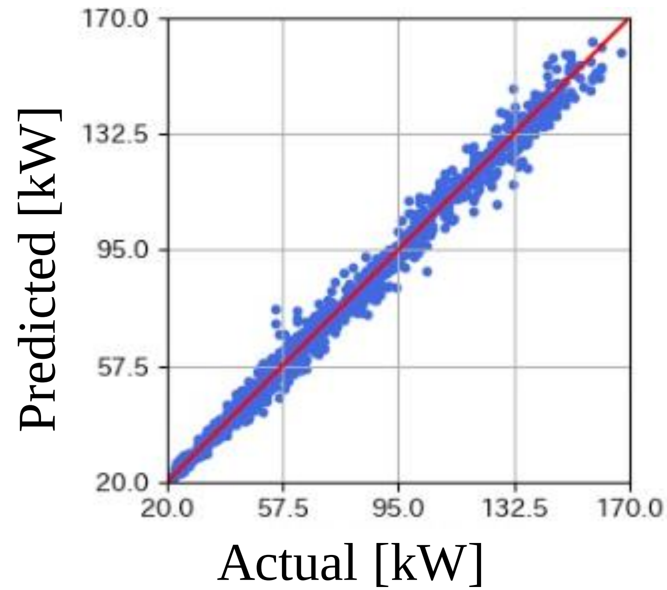
Cannot keep up with **sudden surge**

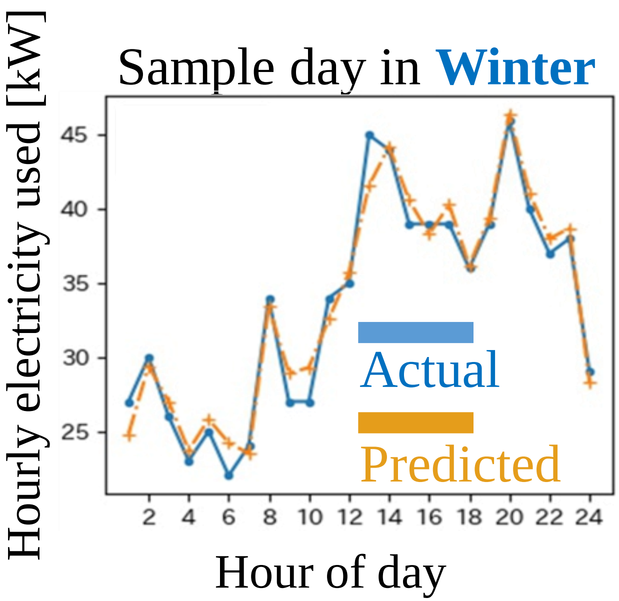
Sample day in **Summer**



Results for one building : Freezing / Cooling , using RFR

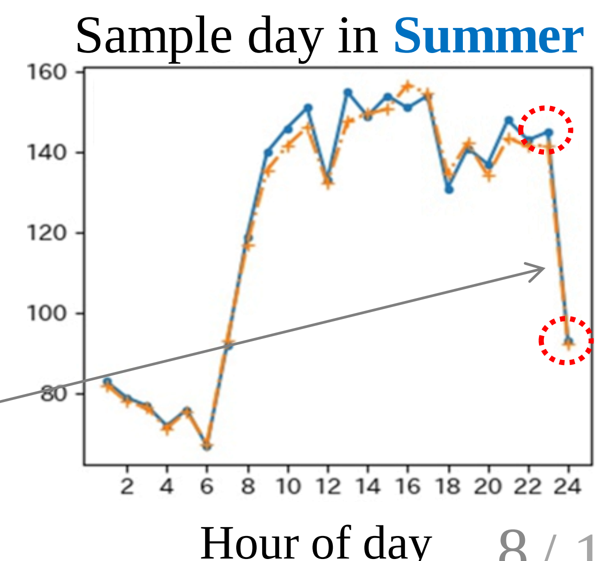
- Target: 30% in a year
  - MSE = 10.9 [%]
    - ➢Better prediction than A/C





Absolute energy smaller in winter, due to lower outside air temperature

Can keep up with **sudden sag**



【Contents】

Across building forecast : Cross validation plan

215 Supermarkets *

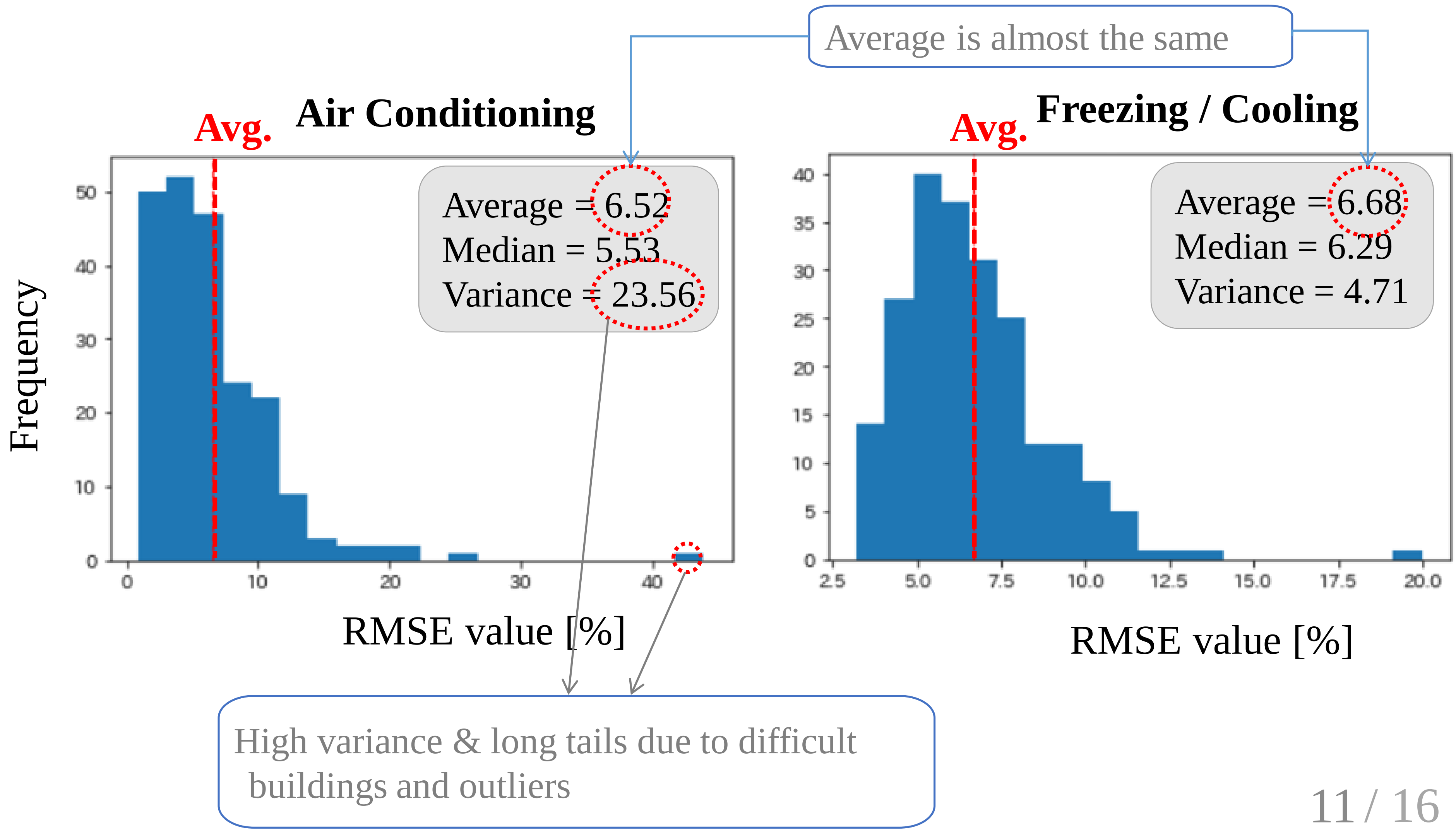
Number of CV iterations

i = 1	Evaluation 8760 x 43	Learning 1 8760 x 43	Learning 2 8760 x 43	Learning 3 8760 x 43	Learning 4 8760 x 43
i = 2	Learning 1 8760 x 43	Evaluation 8760 x 43	Learning 2 8760 x 43	Learning 3 8760 x 43	Learning 4 8760 x 43
i = 3	Learning 1 8760 x 43	Learning 2 8760 x 43	Evaluation 8760 x 43	Learning 3 8760 x 43	Learning 4 8760 x 43
i = 4	Learning 1 8760 x 43	Learning 2 8760 x 43	Learning 3 8760 x 43	Evaluation 8760 x 43	Learning 4 8760 x 43
i = 5	Learning 1 8760 x 43	Learning 2 8760 x 43	Learning 3 8760 x 43	Learning 4 8760 x 43	Evaluation 8760 x 43

* Used samples that had measurements of both A/C and Freezing/Cooling energy

Results for across-building: Root Mean Sq. Errors, by lightGBM

- Building-wise RMSE
  - ➢Many buildings fall between 0 and 10 [%] RMSE



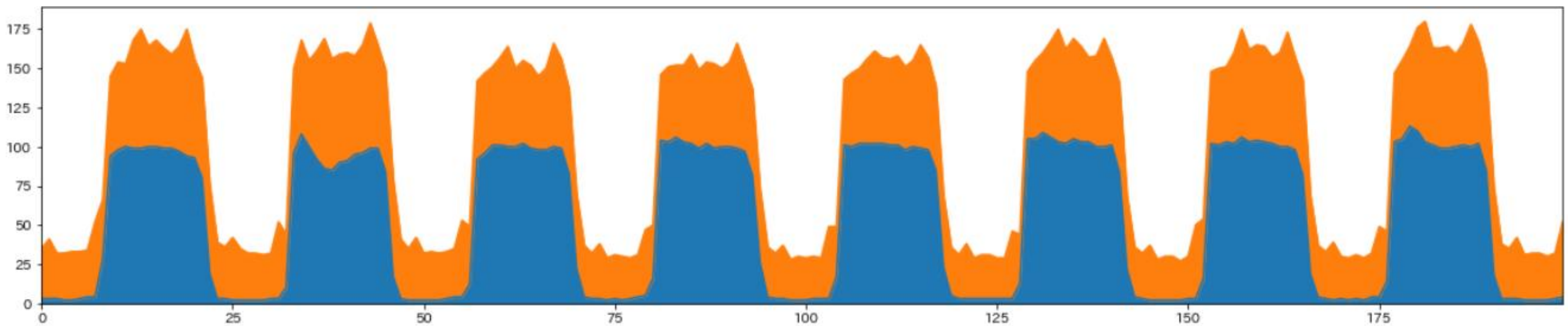
Example predicted time-series, across-building lightGBM

Air Conditioning

Freezing / Cooling

Actual

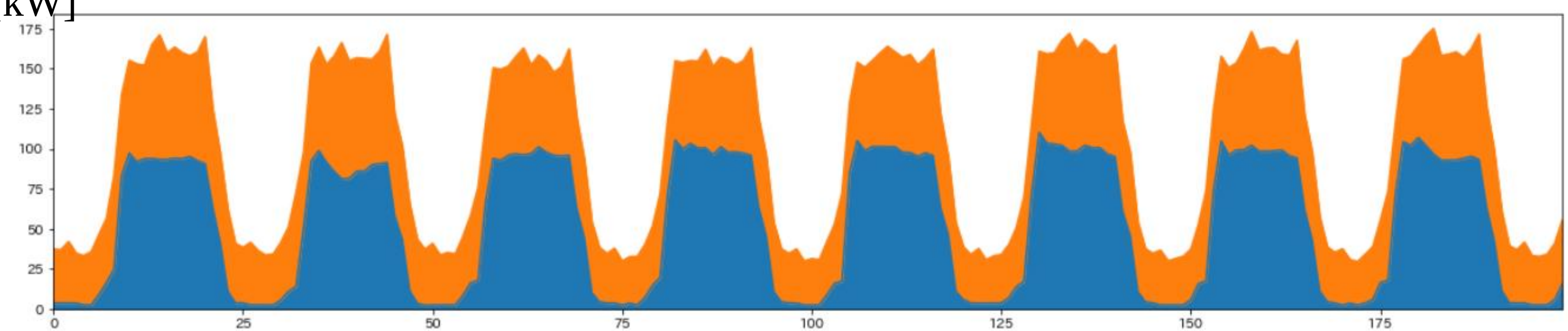
[kW]



Hour

Predicted

[kW]

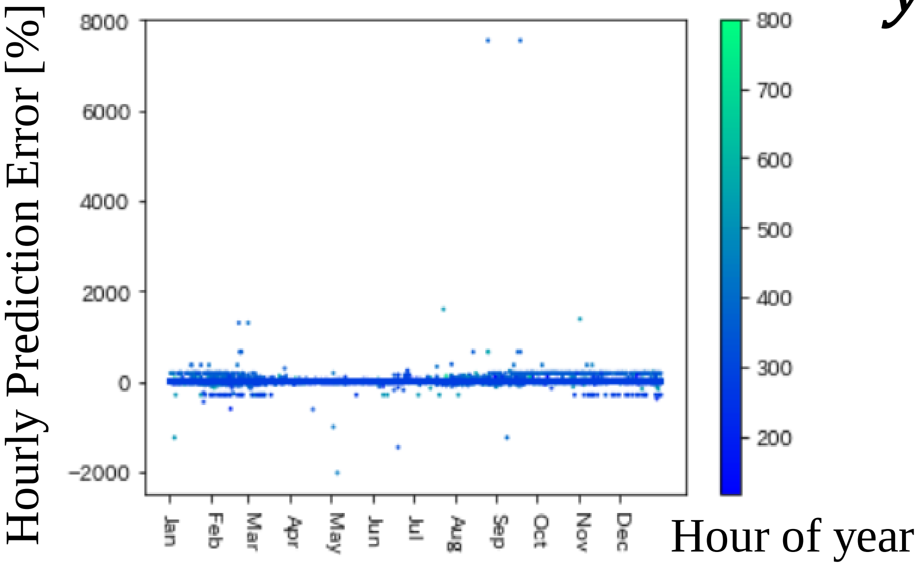


Hour

Prediction errors, across-building lightGBM

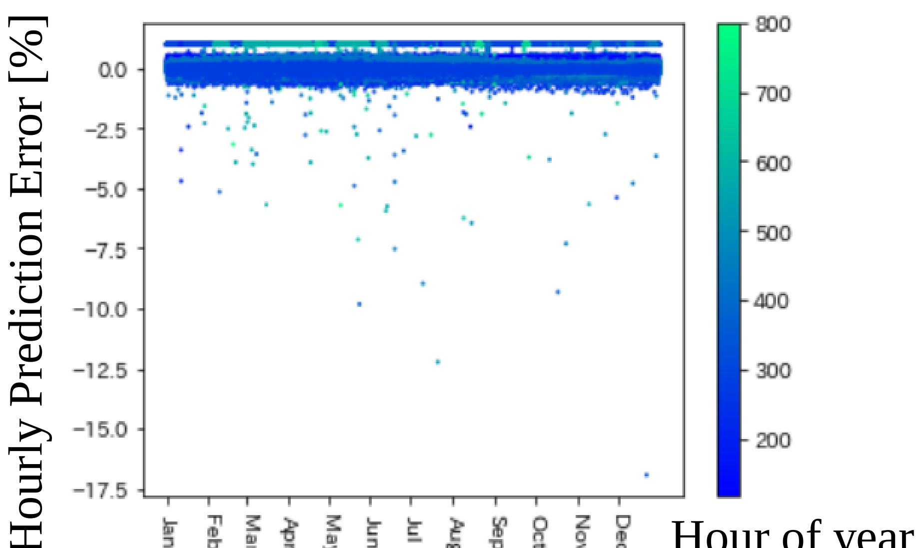
$$y_{error} = \frac{y_{real} - y_{pred}}{y_{pred}}$$

- Air Conditioning



- A little percent prediction points are anomaly error

- Freezing / Cooling



- Over 99 percent prediction points fall within +/-5% range

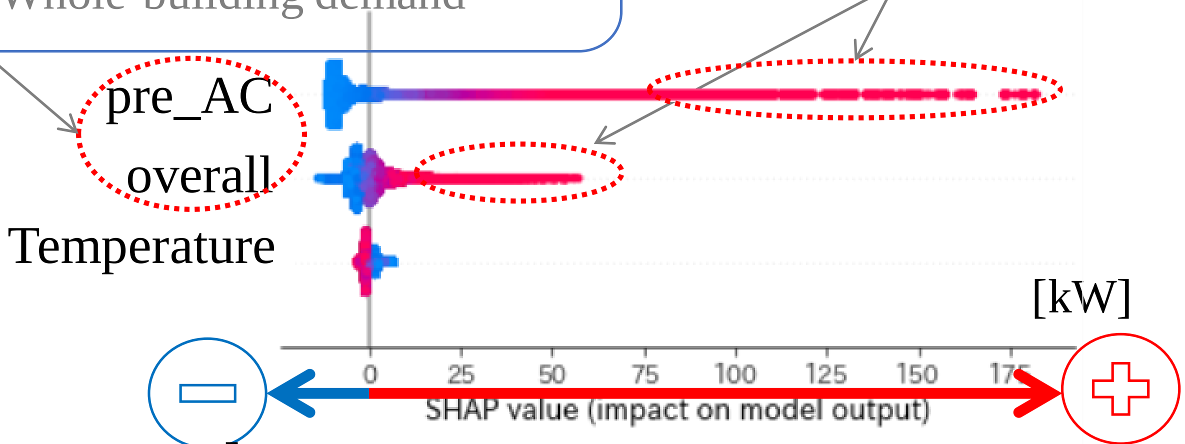
- only 10 points out of +/-5% range

SHAP figure : Feature importances, across-building lightGBM

- Air Conditioning

Highest impact on forecasted values from :
- pre_AC: 1-hour-before target value
- overall: Whole-building demand

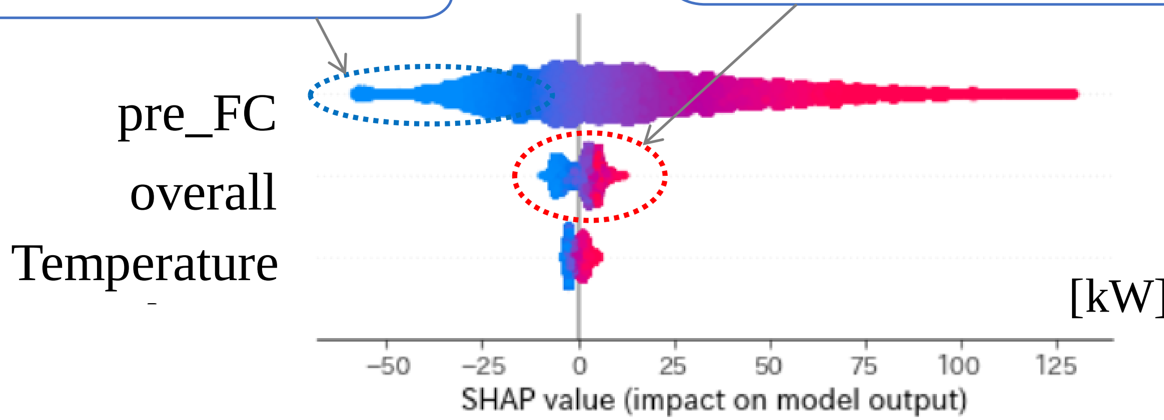
If these values are large:
- positive impacts to predicted values
- the greater these values, the greater the effect

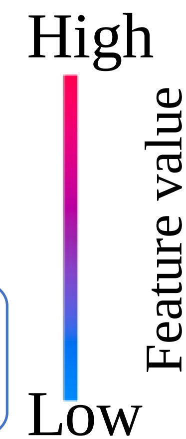


- Freezing / Cooling

Stronger negative impact

Less prominent impact of “overall”





Conclusions

- Successful disaggregation:
  - Temperature-sensitive demand was disaggregated between Air Conditioning and Freezing/Cooling
  - Made possible by lightGBM model trained with different buildings, capturing differences between A/C and Freezing/Cooling

- Possible application:
  - Demand side
    - ✓ potentially detect equipment fault
    - ✓ find peak timing in a year
  - Supplier side
    - ✓ Make energy without excess

Future work

- Problems and possible improvements
  - ➢ Air conditioning resulted to relatively less accurate prediction due to some difficult buildings and outliers
  - ➢ Better model tuning
    - ✓ Better parameter choice
    - ✓ Proper choice of training data

- Next steps
  - ➢ Want to extend to other types of industry
  - ➢ Try to predict with non-AR(1) model