# *USAGE DISAGGREGATION OF SMART METER DATA OF JAPANESE COMMERCIAL CUSTOMERS USING RANDOM FOREST REGRESSION*

Minao Watanabe and Kenta Ofuji, The University of Aizu, Phone +81-242-37-2577, o-fu@u-aizu.ac.jp

## Overview

Following European electricity industry, Japan also has started retail-level deregulation. In tandem with that, rapid deployment of smart meters for both residential and non-residential consumers is under way. Among various potential opportunities associated with smart meter data, usage disaggregation based on the whole-building power demand data is expected to promote energy efficiency, because it may discover important energy waste patterns. In this research, we take Japanese supermarket customer as an example (**Figure 1**), and establish a basic forecasting methodology to disaggregate the whole-building hourly smart meter data into several usages. Specifically, we used random forest regression (RFR) to predict demand for "refrigeration / cooling" across a year, which is the main temperature sensitive demand of this building. The remaining usage was then calculated by subtracting the predicted refrigeration / cooling usage from the entire demand. Lastly, we discuss the prediction accuracy by seasons.

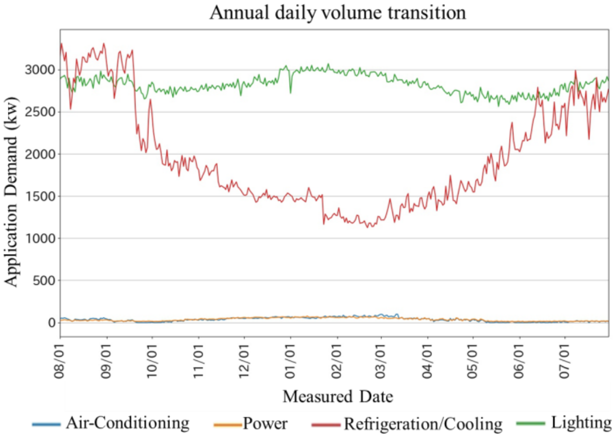


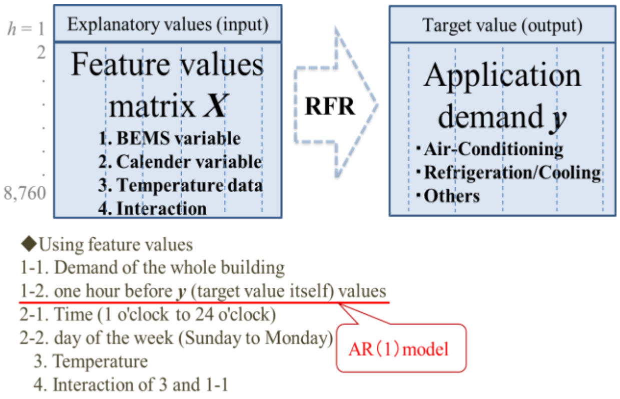
**Figure 1**. Annual daily volume transition [1] (Building ID: B110000181)

## Methods

**Figure 2** shows the methods for prediction. First, taking a consumer data of a supermarket located in the Tohoku (northern east) region of Japan [1] as an example, we built three models to predict usage-specific demand using linear regression, decision tree and RFR. We compared the model performance across the three algorithms. Since RFR performed the best, the results of RFR are mainly described in the following.

The used feature values are shown in **Figure 2**. As described in 1-2, the prediction model incorporated AR (1) (first-order auto-regression) concept by including the target variable itself one hour before. After considering various feature values for better prediction performance, we finally chose, based on the feature importance values, the interaction terms between the hour of day (1 o'clock to 24 o'clock) dummy variables and each of the hourly outside air temperature data of Sendai city [2] and the whole-building demand.

To divide the dataset into test data and training data, we extracted samples on a daily basis instead of random sampling, because it can consider the data’s continuous nature. As a result, we decided to choose 71 days from the year for the test data, so that the chosen days are the multiples of five each month. This also ensured uniformity across the year. The resultant number of the test data samples was 71 days × 24 hours = 1,704, and the training data was the remaining 7,056 hours. The ratio is about 2 (test): 8 (training).


$h = 1$
2
8,760

Explanatory values (input)

Feature values matrix $X$

1. BEMS variable
2. Calender variable
3. Temperature data
4. Interaction

RFR

Target value (output)

Application demand $y$

- ･Air-Conditioning
- ･Refrigeration/Cooling
- ･Others

◆Using feature values
1-1. Demand of the whole building
1-2. one hour before $y$ (target value itself) values
2-1. Time (1 o'clock to 24 o'clock)
2-2. day of the week (Sunday to Monday)
3. Temperature
4. Interaction of 3 and 1-1

AR (1) model

**Figure 2.** Outline of prediction method

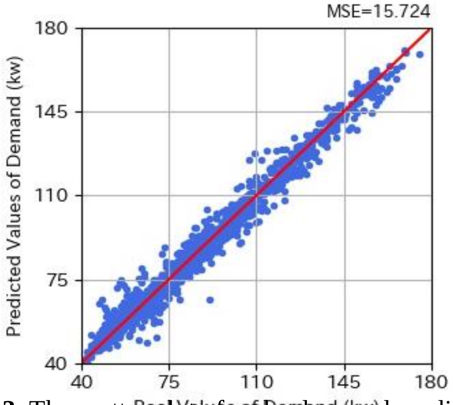


**Figure 3**. The scatter plot of real values and predicted values

## Results

**Figure 3** plots the measured (real) values on the horizontal axis and the predicted values on the vertical axis. The closer to the 45-degree line, the better the prediction. For prediction evaluation, Mean Squared Errors (*MSE*) was calculated using the measured values $y_{real}$, predicted values $y_{pred}$, and the sample size $n$. (Equation (1)). The *MSE* for this model was 15.724, implying reasonable accuracy across seasons.

$$MSE = \frac{1}{n}\sum(y_{pred-}y_{real})^2 \qquad ........ \quad (1)$$

A box plot of the prediction error $y_{\text{error}}$, calculated from Eq. (2), was drawn to understand the size and the variation of the prediction errors (**Figure 4**).

$$y_{error} = \frac{y_{pred} - y_{real}}{y_{real}} \qquad ........ \quad (2)$$

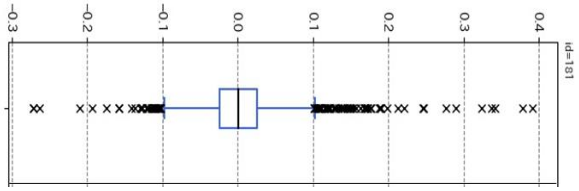


**Figure 4**. The box plot of $y_{\text{error}}$

The median is close to 0 and spans only within ± 0.1 between the first and third quartiles. This means the model can predict within ± 10% of the error in more than half of the cases.

Finally, typical daily measured values and predicted values in four different seasons are shown in **Figure 5**. The change in the amount used in summer (represented by 2012/8/15), on the right-top panel, is smaller than in any other seasons because the demand forecast is the closest to the real values.

In spring (2013/4/15) and autumn (2012/10/15), the by-hour demand varies greater than in summer, making the prediction for our AR(1)-based model more challenging, because the forecast cannot catch up the hourly variation quickly enough. However, in winter (2013/1/15), even with the large increase and decrease in the data, the overall forecast looks to be reasonable.

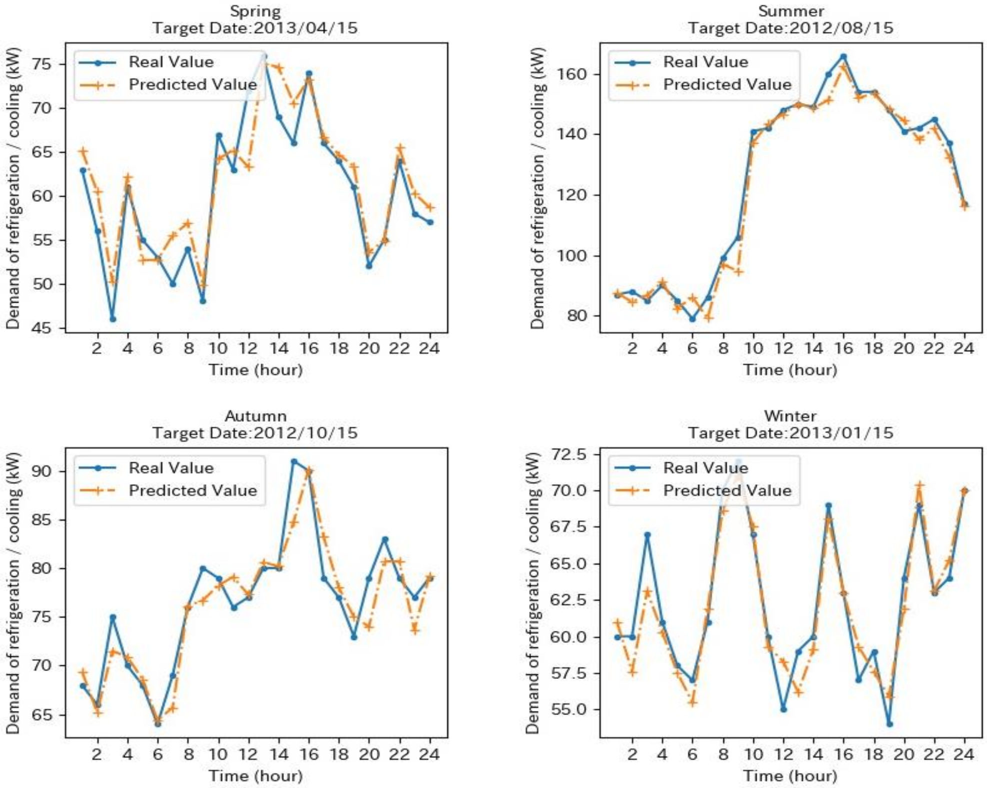


**Figure 5**. Forecast results for each four seasons, on representative dates

## Conclusions

Our current results are summarized as follows:

a) In the example supermarket taken this time, it was possible to predict with an error of less than ± 10% in more than half of the cases by the RFR model, using 20% of the 8,760 hours as training data.

b) However, in seasons when the data's rise and fall are steep, prediction can be difficult.

c) Temperature sensitive demand such as "refrigeration / cooling" can be predicted using the proposed model. Based on this, there is a possibility that simple disaggregation can be made.

d) Future works include validating the proposed method by applying it to a greater number of supermarkets and other commercial buildings. Establishing a general approach to choose effective feature values, especially the interaction terms, are among foremost challenges.

## References:

[1] EMS Open Data https://www.ems-opendata.jp/

[2] Japan Meteorological Agency, past weather data https://www.data.jma.go.jp/gmd/risk/obsdl/inde