White Paper for EEme,llc's Load Forecasting Service

Muchen Zhou

November 11th 2016

1 Background Information

Load forecasting is an integral part of utility operations for a large spectrum use cases ranging from optimal power generation, distribution network management to load balancing, shaving peak demand as well as integrated resource planning efforts. Utilities have historically relied on engineering models that typically leverage "top-down" aggregate data sets, e.g., substation level, or econometric models that account for economic attributes of a service territory. With the advent of mass AMI adoption collecting data from each building every 15 minutes unique analytic opportunities emerge. EEme, an enterprise Analytics-as-a-Service company that was spun out of Carnegie Mellon University, brings the scalable machine learning element to the Smart Grid industry by utilizing decades of artificial intelligence knowledge-base and existing AMI smart meter data. Specifically, EEme's proprietary machine learning platform infers appliance-level energy use insights without relying on additional hardware or end-user input. EEme's load forecasting technology, in turn, feeds from its load disaggregation engine to create more precise and accurate "bottom-up" predictions for power management starting at the appliance level. EEme's load forecasting technology enables utilities account for changes in appliance profiles, to discern between weather-sensitive and non-weather-sensitive end-uses and run probabilistic scenarios that incorporate energy efficiency impact on building energy profiles. EEme's "State-of-the-Art" load forecasting technology is now available via its cloud-based API in the form of Analytics-as-a-Service.

2 Data and Processing

In the validation tests, we will be focusing on two data sets. The first one is the 1-minute aggregated load usage of multiple industrial buildings in Foshan from May 2015 to July 2016. The second one is the individual 1-minute load usage of multiple residential buildings in Texas from March 2013 to March 2014.

There are two things that we need to discuss before the tests. The first one is time step of the data we feed into our tests. Although both datasets' original time-steps are 1 minute, it is not reasonable nor useful to do 1-minute forecast. It is not reasonable because the potential influencing factors(especially weather) won't change significantly in a 1-minute interval and there is a lack of small-interval weather data. Though we do not have detailed test over 1-minute data for this white paper's purpose, we have run the test during the development stage and the average MAPE is larger than 100%. Thus, to balance the performance and time-step, we will pick 60 minutes as our data's time step. The second thing we need to consider is if we should test against aggregated or individual building's data. During the validation test, we are using aggregated data for both Foshan and Texas since we are more focusing on the community demand for electricity rather than individual house's interest at current stage. And Sevlian et. al [1] mentions aggregation reduces the inherent variability in electricity consumption resulting in increasingly smooth load shapes

3 Methodology and Model

We will run two tests with different forecasting period to examine the short-term and long-term accuracy of EEme's model. The first test will forecast on a daily basis(24 ahead forecasting points) based on the previous week of data and the second test will forecast on a weekly basis(168 ahead forecasting points) based on the previous 30 days of data. We will evaluate EEme's model by measuring its MAPE, which is defined in next section, of each test's forecasting period through 330 days. In addition, we will also compare EEme's model with the conventional model.

There are multiple choices of models for load forecasting. Support Vector Machines(Regression)[2], Artificial Neural Network[3], Random Forest[4] and ARIMA[5] are some widely used methods in this filed. Support Vector Regression usually yield great accuracy and much more easier to set up, so we will be using Support Vector Regression model as our comparison model. We will be using local surface temperature, dew point and relative humidity as the input for Support Vector Regression model and EEme's hybrid model.

4 Error Rate

We used the following formula to calculate the error rate for individual hours:

$$ErrorRate(\%) = 100 \times \frac{|L_{actual} - L_{predict}|}{L_{actual}}$$

where $L_{actual}, L_{predict}$ represents the actual load and predicted load.

And we will use MAPE to calculate each period's mean error rate.

$$MAPE = \frac{\sum_{i=1}^{n} ErrorRate_i}{n}$$

where n is the length of testing points.

5 Results

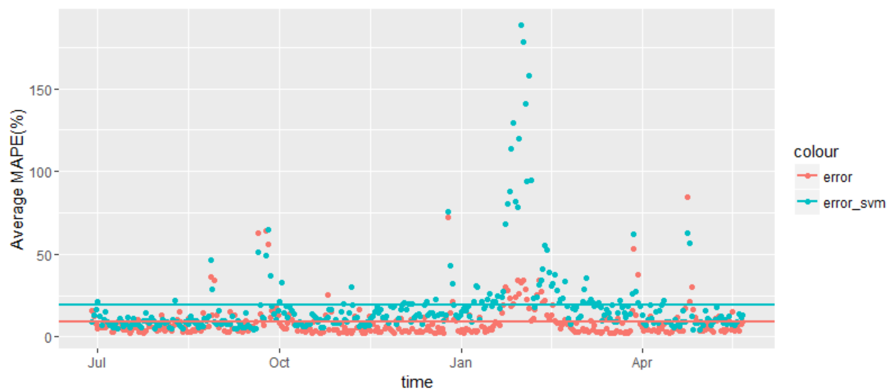

Figure 1: Foshan's Daily Forecast

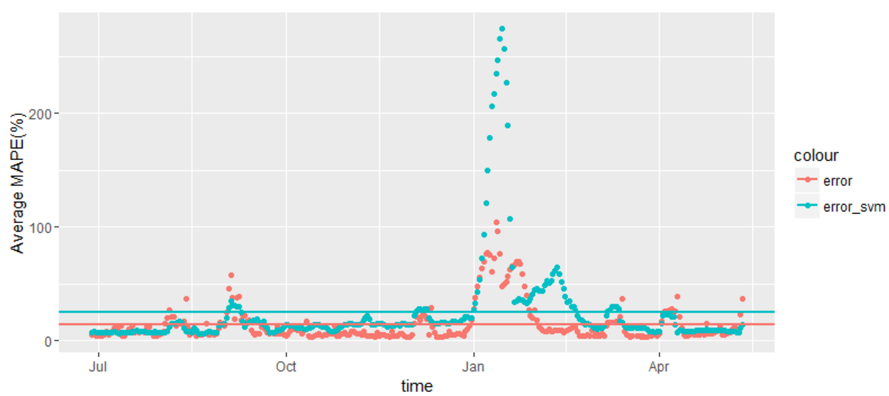

Figure 2: Foshan's Weekly Forecast

Figure 1 and 2 above demonstrates the MAPE for both EEme's hybrid Model and the conventional Support Vector Regression model in Foshan's case, where we run the test for daily and weekly forecast in Figure 1 and 2 respectively. In both plots, the red dots represents EEme's hybrid model, green dots represents SVR model, red horizontal line represents average MAPE through the 330-days period using EEme's hybrid model and green horizontal line represents average MAPE through the testing period using SVR model. The detailed Statistics are in 1, where the change denotes the the average change using EEme's hybrid model compared with SVR model. We can see clearly the MAPE increases

	EEme's Hybrid Model	SVR Model	Change
Daily Forecast	9.46%	18.87%	-49.87%
Weekly Forecast	14.80%	24.62%	-39.89%

Table 1: Detailed Result on Foshan's Dataset

significantly around the time of late January and early February, which is the Spring Festival in China. Although EEme's hybrid model produces significantly smaller MAPE compared with conventional model, the gap between Spring Festival period and normal days is still huge. The possible explanation for such insufficiency is that since the building is an industrial planet, the load usage during the Spring Festival can be varied a lot(does not follow the normal week's pattern) with some unpredicted features. There is no special variable in either model indicating the Spring Festival thus either model will not be sensitive to such sudden change of load usage in Spring Festival.

Figure 3 and 4 is the same tests in Texas and the detailed statistics are in table 2. We can clearly observe that EEme's hybrid model has a lower MAPE rate compared with SVR model in both daily forecast and weekly forecast. However, in weekly forecast, we can see the MAPE raised significantly around late and early summer. The possible explanation is that due to the sudden raise or drop of temperature, the usage of HVAC appliance can be hard to predict especially in long term(weekly).

Also we can notice the difference of average MAPE between Foshan and Texas. The reason behind this is the ...

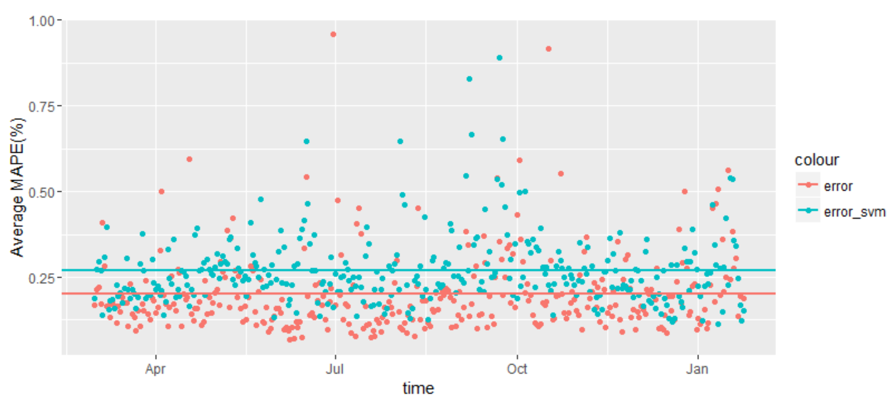

Figure 3: Texas's Daily Forecast

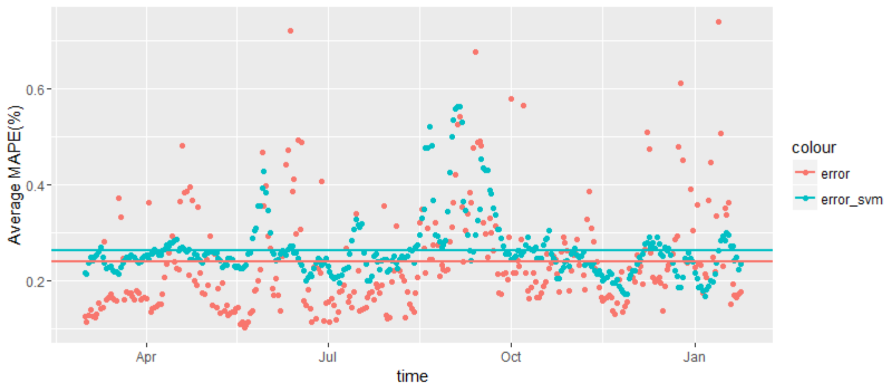

Figure 4: Texas's Weekly Forecast

	EEme's Hybrid Model	SVR Model	Change
Daily Forecast	20.13%	26.94%	-25.28%
Weekly Forecast	23.93%	26.40%	-9.35%

Table 2: Detailed Result on Foshan's Dataset

6 Summary

Through the test, we can see EEme's hybrid model has better performance compared with conventional model in both short-term(daily) and long-term(weekly) load forecasting though less significantly in long-term forecasting.

As for the insufficiency in EEme's model(prediction gap between Spring Festival and other time), there may be some possible ways to improve the model. Assigning an ordinal variable to represent the influence of big holidays such like Spring Festival in China and Christmas in United States may improve the model's sensitivity towards holiday. Training the model during Spring Festival separately from normal time may also be a viable option to improve the performance.

References

[1] R. A. Sevlian and R. Rajagopal, "A model for the effect of aggregation on short term load forecasting," *2014 IEEE PES General Meeting — Conference Exposition*, 2014.

[2] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on eunite competition 2001," *IEEE Transactions on Power Systems*, vol. 19, no. 4, p. 1821–1830, 2004.

[3] H. Hippert, C. Pedreira, and R. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, p. 44–55, 2001.

[4] A. Lahouar and J. B. H. Slama, "Day-ahead load forecast using random forest and expert input selection," *Energy Conversion and Management*, vol. 103, p. 1040–1051, 2015.

[5] G. Juberias, R. Yunta, J. G. Moreno, and C. Mendivil, "A new arima model for hourly load forecasting," *1999 IEEE Transmission and Distribution Conference (Cat. No. 99CH36333)*, 1999.