Machine Learning for Demand Forecasting in Smart Grid

Saima Aman, Wei Yin, Yogesh Simmhan, and Viktor Prasanna
University of Southern California, Los Angeles, CA

We use Machine Learning methods for forecasting the energy consumption patterns in the USC campus microgrid, which can be used for energy use planning and conservation. These experiments are part of the Los Angeles Smart Grid Demonstration Project, supported by the DOE. We use *direct indicators* of power consumption, such as outside temperature and season, along with novel *indirect indicators* that are available on a university campus, like academic calendar, class schedules and occupancy data. We use *machine-learnt models* to predict campus power usage for coarse- (daily) and fine-grained (15-min) time intervals based on the direct and indirect features, utilizing 3 years of sensor data on power usage collected at 15-minute intervals from 170 smart power meters on USC campus.

Regression tree models generate a decision tree with nodes directing to the left or right child based on a feature condition and the leaves of the tree ending in a regression function that predicts the target variable's value (e.g. KWh). Our models perform better than the baselines that are based on the annual mean, time of week mean, and time of year mean. We evaluate the impact of each feature on the model accuracy by training and testing all combinations of the features and comparing their CV-RMSE. We observe that the day of the week feature is the most important feature for daily campus-level power usage prediction. For the 15-minute campus-level prediction, hourly outside temperature was the best predictor, while weekday was the next. This is understandable, given that temperature values through the day and are more relevant at a finer level of granularity.

The projected increase in the use of AMIs and data collection in a Smart Grid environment means that all applications, including demand forecasting, will be data intensive and require the use of scalable and reliable platforms for operations. For example, the Los Angeles Grid has over 1.4 million customers and will require huge capacity to store and analyze terabytes of data. The data could further grow if frequency of data collection is increased and newer sources of data are added. Building prediction models when new data comes in is compute and data intensive.

In order to motivate the need for scalability, we tested the problem of building forecasting models for daily predictions using a single computer. We found that for 25,000 buildings, we need 8 GB memory and 8 hours to build the model. For 15-min predictions, we would need 700 GB memory and 32 days. This is infeasible. We address the need for a scalable platform by leveraging the Hadoop Map Reduce platform to implement our machine learning models and are analyzing its performance.