Machine Learning for Demand Forecasting in Smart Grid

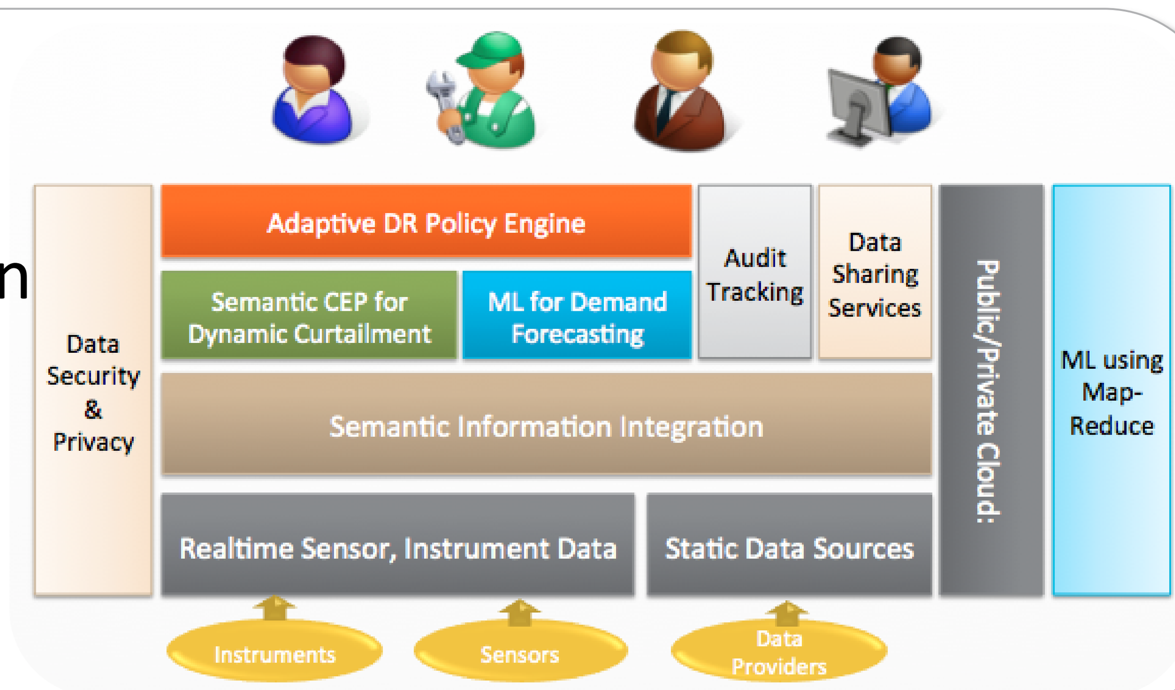
Saima Aman, Wei Yin, Yogesh Simmhan, and Viktor Prasanna
University of Southern California, Los Angeles, CA

Machine Learning Approach

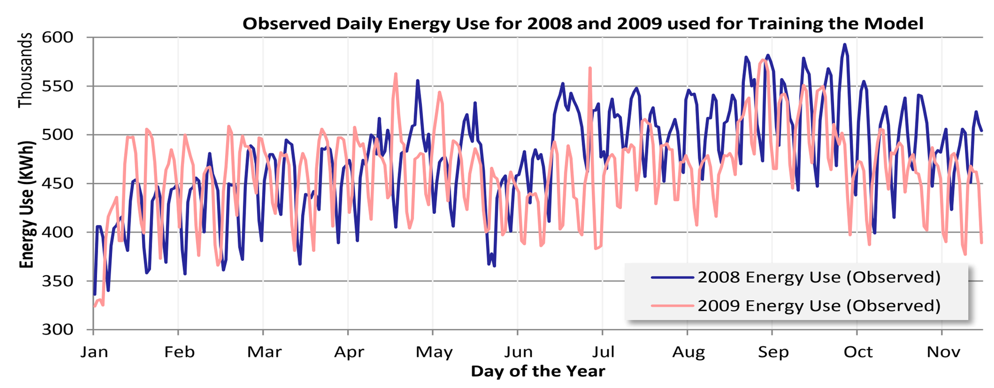
- An informatics approach for forecasting electricity demand
- Use direct and indirect indicators of power usage
- Use scalable Machine Learning for data intensive workloads
- Use Hadoop Map-Reduce on Clouds platforms for performance, scalability and reliability
- Experiment on the USC campus microgrid testbed

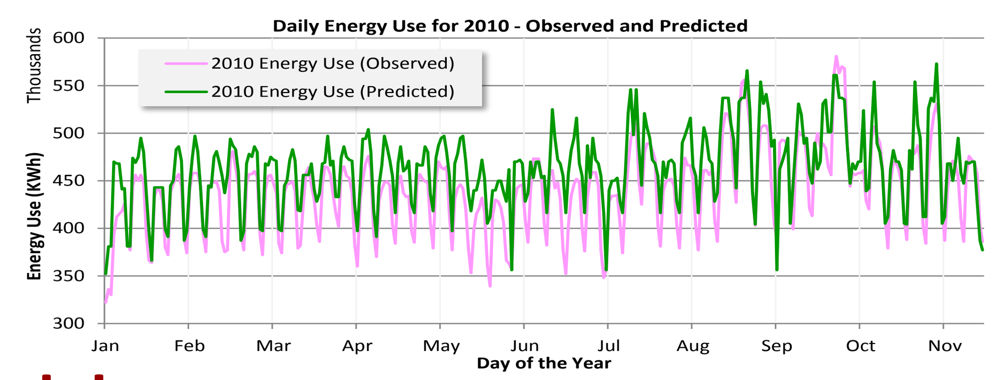
Scalable Forecasting

- AMI's collect TB of data from millions of customers @ 15min
- Several forecasting models
- New data arrives constantly
- Models operate on large data sets



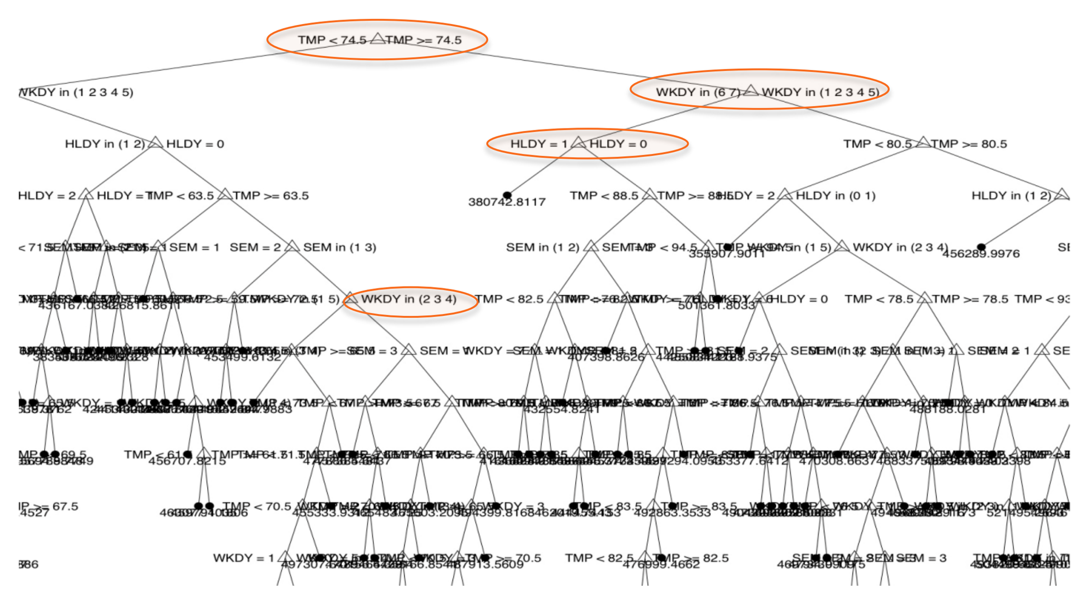
Demand Forecasting using Machine Learning†





Regression Tree Machine Learnt Models

- Use CV-RMSE to compare model accuracy – root mean square error between observed and predicted normalized to the mean of observed



Campus-level Daily Model

Model Used	CV-RMSE
Annual Mean	11.32%
Day of Week Mean	14.39%
Day of Year Mean	12.62%
Regression Tree	7.45%

Baseline Models

Campus-level 15-MinModel

Model Used	CV-RMSE
Annual 15-min Mean	17.37%
Time of Week Mean	16.00%
Time of Year Mean	15.07%
Regression Tree	13.70%

Baseline Models

Features *used in models were evaluated for their impact on prediction*

Weekday	Semester	Temperature	Holiday	CV-RMSE
♦	♦	♦	♦	7.40%
♦		♦	♦	7.60%
♦		♦		7.95%
♦	♦	♦		8.05%
♦	♦		♦	8.37%
♦	♦			8.54%
♦				8.86%
		♦		10.48%
	♦			11.05%
			♦	11.54%

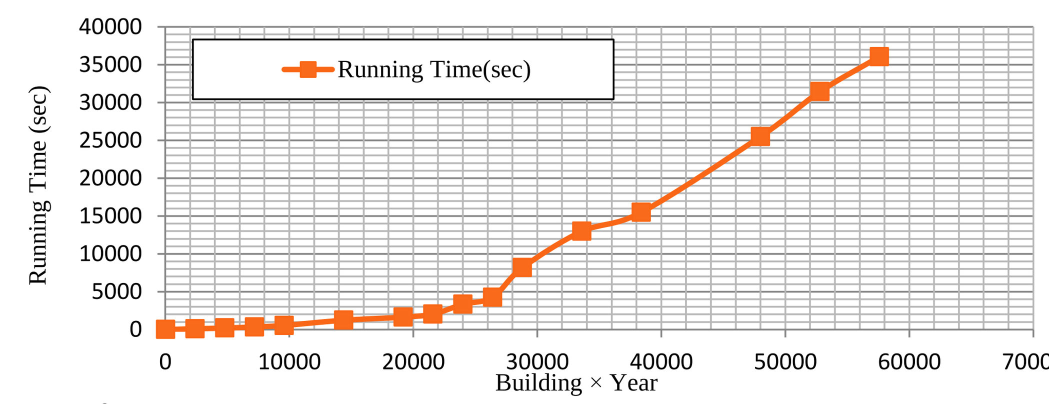
Campus & Building-level features		Building-level features
Weekday Semester Holiday	Max temp Avg. temp	Gross Building Area Net Area in Use Year of construction Type of building

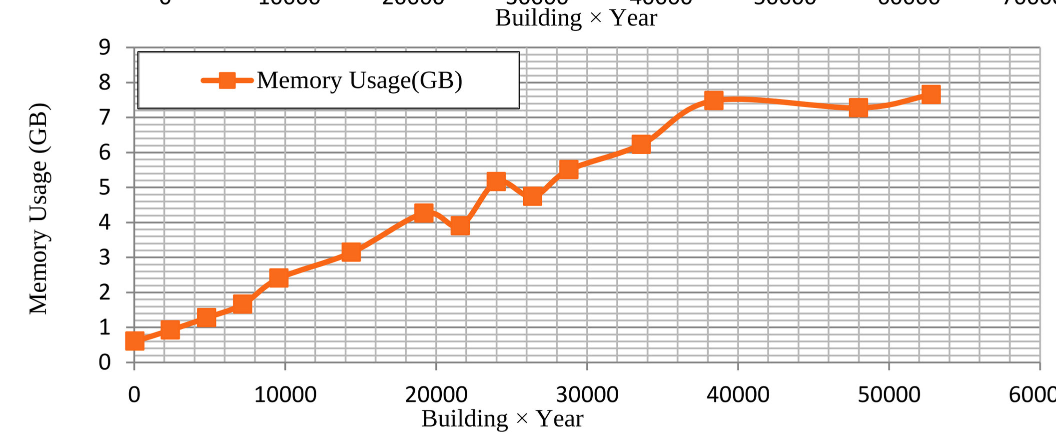
Ongoing Work

- Dynamic feature selection
- Moving windows of predictions
- Data feeds from information repository
- Effect of data granularity on prediction
- Multiple models for buildings/consumers
- From Medium term » Short term Models

†*Improving Energy Use Forecast for Campus Micro-grids using Indirect Indicators, Saima Aman, Yogesh Simmhan and Viktor K. Prasanna, DDDM, 2011*

Performance Limitations of Centralized Machine Learning





For 15-min interval data, we need 32 days & 700GB memory to train the model

Solution

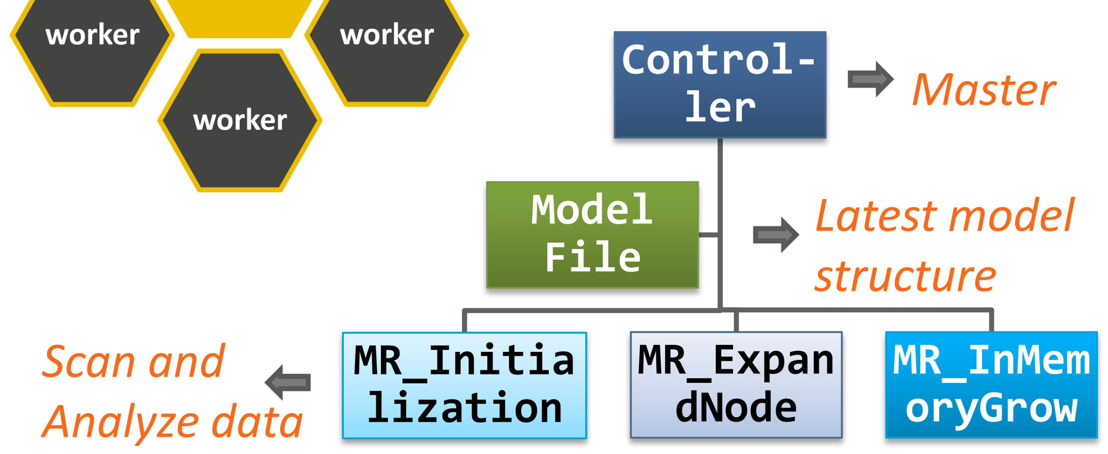
- Parallel processing on Cluster/ Cloud using Hadoop Map-Reduce

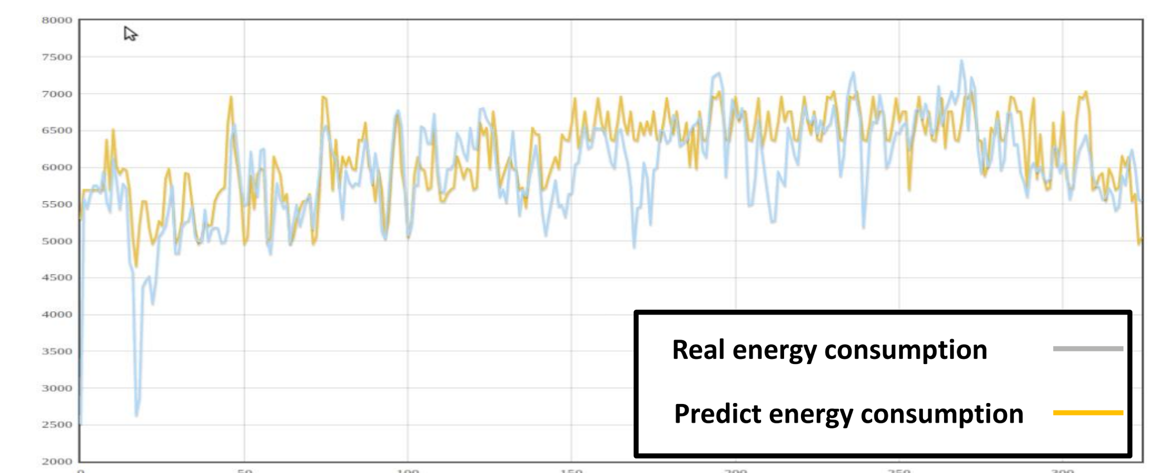
Benefits

- Divide and conquer: Scalable as the training data size increases
- High Performance, Lower time for data intensive modeling
- Reliable operation for continuous execution

PLANET Scalable Learning Algorithm

- Google's Regression Tree Algorithm
- Combine MR+Weka





Ongoing work

- Migrate prototype to Eucalyptus
- Automatic feature selection
- Software infrastructure for launching ensemble models

Acknowledgement *This material is based upon work supported by the Department of Energy under Award Number DE-OE0000192. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.*