**Machine Learning (EECE 6822)**

****

**Project Proposal**

**Mohammad Saber**

**(MUID: 005852837)**

**Due date of submission: 02/19/2015**

**Electricity Load Forecasting Using Decision Tree Method**

Electricity load forecasting is very important for utility service providing companies for planning ahead and running their business efficiently. Mainly, electric load forecasting research is categorized based on time duration of forecast [1]. Short-term load forecasting normally predict load for one or two weeks ahead of time. Long-term load forecasting is normally considered prediction of load for next 1 year or more. Forecasting load for next quarter of the year would be considered as mid-term electricity forecast.

Short-term load forecast [1-5] depends on weather data (temperature, wind, dew point etc.) mainly. Weather is a chaotic system and it’s very difficult to predict. Though short-term load forecast heavily depend on weather forecast, thus it’s very difficult to forecast electric load perfectly. However short-term electric load also depend on human behavior which is very difficult to measure. For example, some people may tolerate more hot temperature than others and need not to turn on air-cooler. These things are hard to measure.

Long-term load forecast depends on economic data (personal income, unemployment rate, electricity price, gross domestic product etc.) as well as weather data [6]. Long-term weather forecast is also more difficult than short-term weather forecast. After 1 year a country may go through economic turmoil, thus their electricity demand may decrease because of bad economy. Thus we have to consider economic factors as well as weather factor to predict long-term load which makes it more difficult than short-term load forecasting.

My target is to predict short-term electricity demand using decision tree method. My expected data source is GasDay Lab, Marquette University. My alternative data sources are US Energy Information Administration (EIA) and National Oceanic and Atmospheric Administration (NOAA). Considering availability of data I will forecast the electricity demand for long-term also. I am considering Bureau of labor statistics (BLS) and Bureau of Economic Analysis (BEA) as my possible source of collecting economic data.

Accuracy of load forecasting depends on frequency of data. If we use low-frequency data for forecasting then we can’t expect good forecast. Cooling degree day (CDD) is a popular weather factor used in electricity load forecasting. Normally average temperature of a day has been subtracted from 65F to find CDD. If the result is negative then CDD is considered zero. Considering any temperature over 65F would lead us to turn on the air-cooler, we calculate CDD which is a very important factor to predict electric load. But sometimes CDD is misleading if we have low-frequency data. Say, we have CDD = 0, that means temperature is below 65F all the time for a particular day. Consider we have high temperature of 80F and low temperature of 50F, then we have CDD of 0 (average temperature is 65F) for that day. May be we have 50% of time over 65F in that particular day, but we are considering that day as a cold day which is misleading. Thus I am planning to use hourly data to calculate CDD considering availability of data and name this new variable as heating degree hour (HDH).

In my long-term electricity load forecasting model, I will use population, GDP, unemployment rate, per capita income, cooling degree day, previous electric load, day of the week as factor to predict electricity load for next 1 year. For short-term load forecast I will use temperature, wind, dew point, previous load, cooling degree hour as variables to forecast electricity. I am also planning to use linear regression model along with decision tree method and compare their performance.

**References:**

[1] Xin Jin, Jie Wu, Yao Dong, Dezhong Chi, “Application of A Hybrid Model to Short-Term Load Forecasting”, International Conference of Information Science and Management Engineering, IEEE, 2010, pp. 492 – 497.

[2] Du Xin-hui, Tian Feng, and Tan Shao-qiong, “Study of Power System Short-term Load Forecast Based on Artificial Neural Network and Genetic Algorithm”, 2010 International Conference on Computational Aspects of Social Networks, IEEE, 2010, pp. 725 – 728.

[3] James W. Taylor, Lilian M. de Menezes, Patrick E. McSharry, “A comparison of univariate methods for forecasting electricity demand up to a day ahead”, International Journal of Forecasting 22, Science Direct, 2006, pp. 1-16.

[4] Du Xin-hui, Tian Feng , Tan Shao-qiong, “Study of Power System Short-term Load Forecast Based on Artificial Neural Network and Genetic Algorithm”, International Conference of Computational Aspects of Social Networks, IEEE, 2010, pp. 725-728.

# [5] Sérgio Ramos, João Soares, Zita Vale, Sandra Ramos, “Short-term Load Forecasting Based on Load Profiling” Power and Energy Society General Meeting (PES), IEEE, 2013.

# [6] Tao Hong, Jason Wilson, “Long Term Probabilistic Load Forecasting and Normalization with Hourly Information”, Smart Grid, Vol. 5, No. 1, IEEE, 2013, pp. 456-462.

[7] Qia Ding, “Long-Term Load Forecasting Using Decision Tree Method”, Power System Conference and Exposition (PSCE), IEEE, 2006, pp. 1541-1543.