Multivariate Short Term Load Forecasting using Machine Learning Methods

Sumedh Sankhe

1 Introduction

Load forecasting is an important electric utility task for planning resources in an electric grid. This function also aids in predicting the behavior of energy systems in reducing dynamic uncertainties. National Utility agencies have the resources to apply computationally expensive and rich in data modelling techniques to forecast for future loads in a host of scenarios, but in markets where Demand Response is a very hotly contested commodity having accurate forecast using the least amount of time and not always having the cushion of tonnes of data is paramount. Enter machine learning models or complex time series modelling to predict acceptable forecasts. I have applied Multivariate Regression Models, Ensemble Learning and made some headway into applying Neural Networks to help predict better forecasts.

2 Assumptions

Before we get into the technicalities of the different methods here are a few assumptions that are made.

- During the process or data cleaning, load values for any customers that were missing or not reported for some reason were removed from our to experiment to calculated aggregated loads.
- Weather data received from NCDC(hourly interval) was used without delving deeper into how it was measured etc. Wet Bulb temperature and Relative Humidity are the two other predictors used.
- For predicting the load values, it reasonable to accept the weather forecast values provided by the Weather services.

3 Modelling Methods

- The aim of multiple linear regression is to model a dependent variable (load) by independent variables (weather, etc.) it can be formally written as

$$y_i = \beta_1 d_1 + \beta_2 d_2 + \ldots.\beta_N d_N + \varepsilon_i$$

Where y is the load to be forecasted, $\beta_1 \ldots \beta_N$ are the regression coefficients and the $d_1 \ldots d_N$ are the independent variables. Further to better our forecast we explore the effect of interaction between the independent variables on our forecast

- Ensembled learning methods applied here are the random forests, there were some difficulties setting up the random forest algorithm to work with multivariate time series data, hence a univariate time series was considered for this method forgoing the weather additions.
- Neural networks, the original plan to setting up a shallow LSTM proved to be difficult to accomplish in the given time frame, a prototype of the LSTM was developed but more testing/ validation is required before actually using it to forecast values.

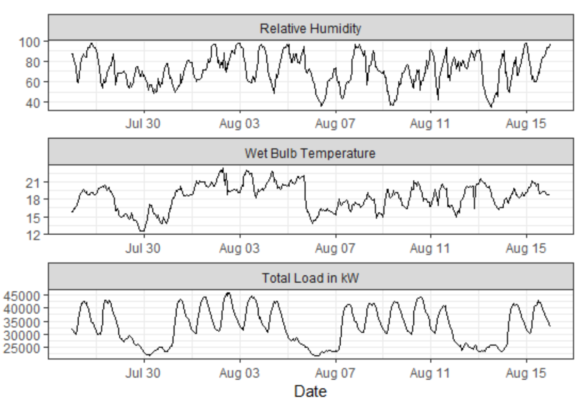

Figure 1. Visualizing our Predictors

4 Evaluations

Thus from the below iterations of the different linear models which includes multivariate linear regressions and general additive models the predictions are a lot more tighter than the ones that were shown in the proposal, also using the MAPE as a metric we see that

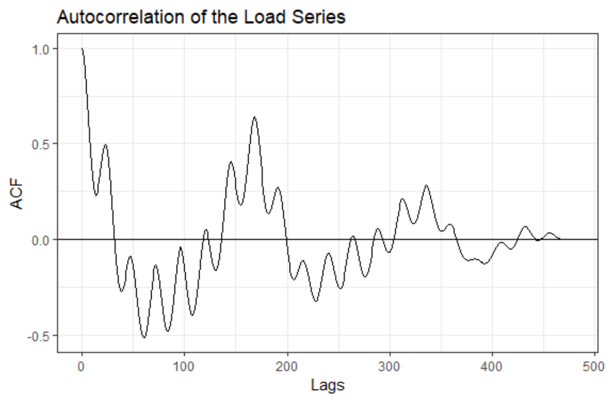

Figure 2. Autocorrelation of the load

model g5 performs better on the test data rather than the train data, whereas model g4 performs marginally better on the train data.

Model		R-Square	MAPE
MLR	l1	0.9746	16.3521
	l2	0.9678	9.8751
	l3	0.9668	9.8751
	l4	0.9744	13.8519
	l5	0.9710	18.4913
	l6	0.9747	16.3794
	l7	0.9784	22.1864
	l8	0.9779	23.1454
	l9	0.9735	16.4806
GAM	g1	0.5802	23.0785
	g2	0.8858	4.4722
	g3	0.6674	35.8305
	g4	0.9794	2.7210
	g5	0.9749	2.6543
	g6	0.5514	24.3911

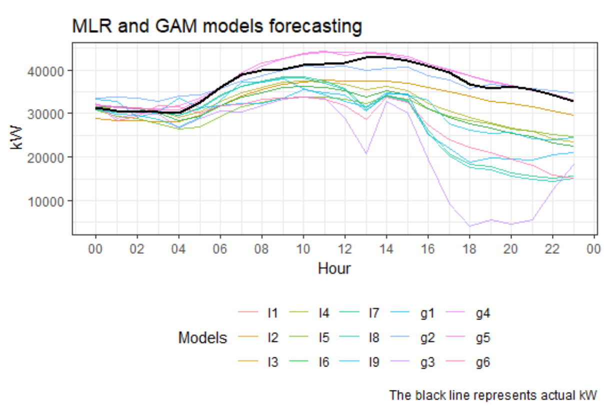

Figure 3. Model Performance

It is surprising to see that an ensembled model to perform such poorly on the test set, which hints that the model parameters may not have been set properly or the the decomposition of the series might not have happened properly into its components

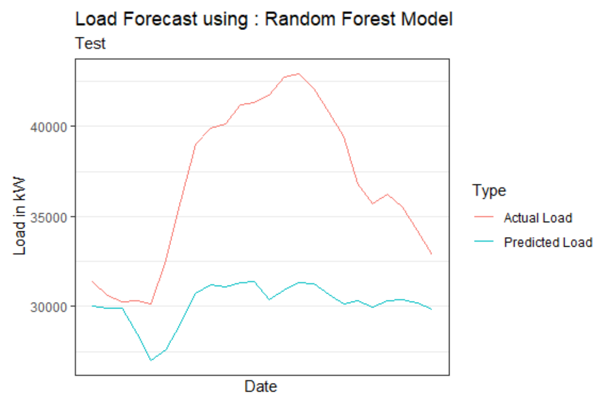

Figure 4. Model Performance