Energy Forecasting for the Global Energy Forecasting Competition 2014

Semester Project Report

Fabian Brix
MSc Candidate
School of Computer & Communication Sciences
Swiss Federal Institute of Technology Lausanne
fabian.brix@epfl.ch

***Abstract*—The abstract goes here.**

I. Introduction

A. GEFCom 2014

The Global Energy Forecasting Competition (GEFCom 2014) is the second edition of a competition first held on Kaggle in 2012 that attracted hundreds of participants contributing many novel ideas to the energy forecasting field. The second edition lasted from 08/15/2014 to 12/15/2014 and was sponsored by several IEEE bodies and the International Journal of Forecasting. It included four competition tracks: electric load, electricity price, wind power and solar power forecasting. The respective data was published on the community platform of crowdanalytix.com on a weekly basis during the competition. The format of the competition was that of rolling forecasting requiring contestants to submit the next period of interest to forecast a day before the next set of data was published. To appear on the final leaderboard the contestants had to submit 99 quantiles for each step throughout the forecast horizon. In the course of this semester project we focused solely on the electric load forecasting track. This track had a weekly forecast horizon of one month and consisted of three trial periods and twelve competitive periods. In order to feature on the final leaderboard nine competitive submissions were required. Due to our lack of familiarity with the subject, participating in the competition was not prioritized after initial attempts. Instead the focus was on comparison of the performance of different methods on the whole dataset via cross validation.

B. Short Introduction to Energy Load Forecasting

Before we review related work we are going to introduce the area of Energy Load Forecasting. Load forecasting, as it is commonly referred to, is usually concerned with the prediction of hourly, daily, weekly, and annual values of the system demand and peak demand of an electric utility [1]. Such forecasts are sometimes categorized as short-term (up to 1 week), medium-term (1 week - 1 year) and long-term (¿ 1 year) forecasts, depending on the time horizon. In the load forecasting track of GEFCom 2014 we are concerned with forecasting the daily electricity demand/load of a utility for a whole month on a rolling basis given the data of the previous years. The task is therefore on the threshold between short-term and medium-term load forecasting.

Load Forecasting is necessary for the planning of energy system and their effective operation and maintenance. Forecasting accuracy therefore has a major impact on electric utilities and their regulators. In case of overestimation of future load/energy demand utility providers will operate too many units possibly driving energy demand and in case of long-term forecasts investment in the construction of new infrastructure can be wasted. Underestimation leads ot unmet demand and systems that are vulnerable to crashes. The output of such forecasts can either be point forecasts or estimates of the probability distribution of values of future demand as required during GEFCom 2014. Electricity demands follow a nonlinear, volatile pattern subject to several exogenous variables such weather conditions, randomness in human behavior leading to randomness in demand and economic conditions and demographic changes. In GEFCom 2014 the exogenous variables are limited to weather conditions in the form of recorded temperature at several sites and calendar effects such as the effects of weekends and holidays on the electricity demand. In this report we analyse the predictions produced by algorithms that are capable of capturing the nonlinear dependencies between these exogeneous variables and the load such as General Additive Models, Random Forests, Multilayer Perceptrons. We do not strive to build any algorithms ourselves, but employ the various implementations in R of the mentioned algorithms.

II. Review of Related Work

Review previous work from Gefcom 2012 and beyond including papers with related approaches including but not necessarily limited to those that can be found under the following link: http://blog.drhongtao.com/2014/08/recommended-papers-for-gefcom2014-contestants.html. Distinguish between approaches to temperature and load prediction.

Probabilistic Electric Load Forecasting: A Tutorial Review contains literature review for longterm probabilistic load forecasting

Tao Hang, Jason Wilson: Long Term Probabilistic Load Forecasting and Normalization With Hourly Information but papers is on Long Term Electric Load Forecasting: 1-X years [2]

Rob Hyndman, Shu Fan Density Forecasting for Longterm Electricity Demand Source for calendar effects [3]

Shu Fan, Rob Hyndman Short-term load forecasting based on a semi-parametric additive model [1]

III. Dataset and Evaluation Metrics

A. Dataset

The dataset provided by GEFCom 2014 includes hourly historical load and weather data of a utility in an undisclosed area on the east coast of the United States of America. The 25 weather stations in the dataset provide historical temperature for their respective zones. However, the load data consists only of the system level load in Mega Watts (MW) and not all of the zonal level load series. Therefore, forecasts in the context of Smart Grid Technology are not required. The temperature data made available consists of 25 series of temperature data in Fahrenheit from the 25 different weather statios dating from 01/01/2001 to 12/01/2011. The load data of the utility is recorded starting from the 01/01/2005 at 1am.

The complete data set acted both as training and validation set except for the first and last month of the data published for the reason that the data was released on a weekly basis. It consists of 15 spreadhseets in the format of Comma-Seperated Values (CSV) for the load forecasting track. The first spreadsheet contains data starting from 01/01/2001 up until midnight on the 10/01/2014 at 1am from when on the incremental spreadsheets released every week contain only one month of data.

The forecasts were required to be made starting from 10/01/2010 on a monthly rolling basis for 15 months. The nature of the data provided required the contestants to produce their own temperature forecasts for the month ahead in the dataset.

B. Evaluation Metrics

The Evaluation Metric employed to score the contestants' submissions of quantile forecasts is the tilted loss/error function also known as the pinball loss/error function. In the following paragraphs let y denote an observation and $\hat{y}$ denote a corresponding forecast while ξ is defined as the residual $y - \hat{y}$.

TODO: Include short explanation.

$$L_\tau(\xi) = \begin{cases} \tau\xi & \text{if } \xi \geq 0 \\ (\tau - 1)\xi & \text{if } \xi < 0 \end{cases} \quad \text{where } \xi = (y - \hat{y})$$

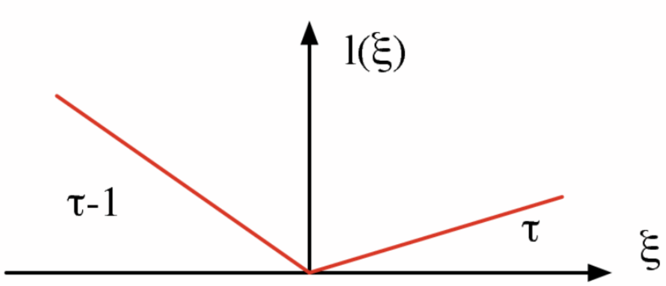

Fig. 1. The tilted loss function for 50th, $\tau = 0.5$, and 75th, $\tau = 0.75$, quantile [4].

TODO: Include the proof for why the "pinball" loss estimates the τ-quantile [4].

In our evaluation of forecasts, we further use some well-established point error metrics, for the simple reason that we generate our quantile predictions starting from a point prediction. The measures are the Mean Average Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

The MAE and RMSE measures are the two most common scale-dependent errors, meaning that the residuals ξ_i (for ith observations) are on the same scale as the data. Hence, MAE and RMSE are in units of Fahrenheit or Mega Watt for our data set.

$$\text{MAE} = \mathbb{E}[\xi_i]$$

$$\text{RMSE} = \sqrt{\mathbb{E}[\xi_i]}$$

As a percentage error the MAPE measure is scale-independent. In our case it is useful for giving an immediate sense of the relative magnitude? of the error.

$$\text{MAPE} = \mathbb{E}\left[\left|\frac{100\xi_i}{y_i}\right|\right]$$

The measure is undefined for $y_i = 0$. Fortunately, in our dataset all values are several integers larger than zero for both load and temperature series so that the measure is neither undefined or affected by extreme values.

C. Data Cleaning

An annoying feature of the dataset is that the timestamps are not saved in the international ISO 8601 standard, but as "MMddYYYY H:m" without leading zeros for both days and months. Fortunately, the dataset was provided continuously without gaps and therefore the problem could be easily solved by hard-coding the first and last datetimes and using these to generate the needed sequence of datetimes.n

D. Data Selection

The Cross Correlations of the temperature series of the 25 weather stations suggest that they can be explained to over 90% by the first series.

[include correlation plots here]

A temperature of 60 degrees Fahrenheit would therefore allow for an error of maximum 6 degrees of Fahrenheit or 3 degrees Celsius respectively. As we will see later in this report this error is negligable when taking into account the inaccuracy of/the error introduced in the temperature prediction.

IV. Feature Selection

Description of features obtained from the data. Name all features in this part or only the ones in the final selection?

Use this part to motivate selection of time of year (TOY) variable:

The effect of the lag on the Time Series Correlation Coefficient can be demonstrated using the Autocorrelation function acf() of the R stats package. Here we display five plots showing the Autocorrelation Function for different maximum lags: As can

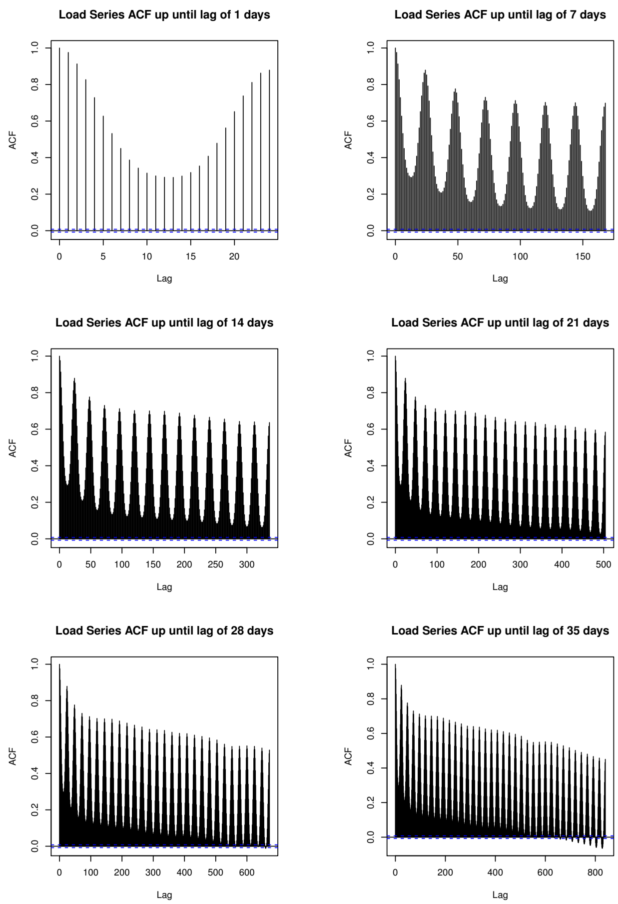

Fig. 2. Plots of Autocorrelation Function Estimates for hourly load data in MW for different maximum lags.

be seen the correlation diminishes exponentially up until a lag of 72h from any point in the time series, then stays more or less constant for up to 7-8 days, whereafter it diminishes near linearly (up to a lag of 35 days).

Autocorrelation plots for temperature and load side by side for one year (daily avg)

include seasonal decompositions?

A. Calendar Features

hour, TOY vs. month [1] [3]

V. Models

Description of the Models (LM), GAM (more extensive), NN, RF

VI. Analysis

What combination of features and models for temperature and load provide us with a good prediction accuracy with respect to Gefcom leaderboard?

A. Temperature Modeling

1) Data Processing: average temperature vs. principal component

2) Effect on Load Prediction: Effect of temperature on load prediction evaluated using different methods:
Mean over past years (yearly lag), LM, GAM, NN, RF vs. true temperature

3) Evaluate Results of weekly vs. monthly Temperature Prediction: plot MAPE & PINBALL scores for different methods over load training + CV period in a 1x2 plot of the form:
monthly scores weekly scores

B. Load Modeling

1) Evaluate the Influence of the Lag on Load Prediction: set the basis by plotting MAPE & PINBALL scores by week over w1, w2, w3, w4; one curve for every month in CV
do this for every method
as well as a comparison of the best performing configuration of every method among each other

2) Evaluate Performance for different Method Configurations: Use temp method that provides best score as shown in Temperature Modeling Section

Different GAM formulas:
plot MAPE & PINBALL scores for all GAM formulas over CV period in a 2x2 plot of the form:
monthly load with monthly temp monthly load with weekly temp
weekly load with monhtly temp weekly load with weekly temp

Different NN hidden units:
plot MAPE & PINBALL scores in 2x2 plot

Different RF ntrees:
plot MAPE & PINBALL scores in 2x2 plot

3) Compare Performance of different Methods: choose best scoring configuration for every method and plot the results in one 2x2 plot

VII. Conclusion

The conclusion goes here.

Acknowledgment

References

[1] S. Fan, S. Member, and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Transactions on Power Systems*, no. August, pp. 1–8, 2010.

[2] T. Hong, J. Wilson, J. Xie, and A. Member, "Long Term Probabilistic Load Forecasting and Normalization With Hourly Information," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 456–462, 2014.

[3] R. J. Hyndman and S. Fan, "Density Forecasting for Long-Term Peak Electricity Demand," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1142–1153, May 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5345698

[4] I. Takeuchi, Q. V. Le, T. D. Sears, and A. J. Smola, "Nonparametric Quantile Estimation," *Journal of Machine Learning Research Nonparamteric Quantile Estimation*, 2005.