USE OF RMSE VALUE AS THE BENCHMARK FOR COMPARING THE PERFORMANCE OF THE MODEL BUILT USING THE MACHINE LEARNING TOOL

Vivek Singh[1], Dr.Arun Jain[2], Dinesh Sharma[3]

[1]*M.Tech, Dept. of Computer Science and Engineering, SIET, Lucknow(India)*
[2]*Professor, Dept. of Computer Science and Engineering, SIET, Lucknow(India)*
[3]*Asst. Professor, Dept. of Electrical and Electronics Engineering, SIET, Lucknow(India)*

ABSTRACT

Accurate Load Forecasting of electricity demand is vital for any power utility to reduce losses and to increase efficiency. This is vital for implementation of Smart Grid across the country as Load Forecasting helps to provide reliable and quality power supply to the consumer. In this paper we shall compare some of the existing methods that are used for load forecasting and compare the performance of each of those methods using the RMSE value to measure the performance. The method with the least RMSE value performs better having reliable level of accuracy. We have considered the holidays and temperature of the area information also in our work as it greatly influences the load consumption pattern among the consumers.

Key Words: -Data Mining, Load Forecasting, Regression, Smart Grid, RMSE

I. INTRODUCTION

The Smart Grid **[1]** refers to the digitized electrical grid which uses state-of-the-art technology for connecting the utilities with the customers for delivering reliable and adequate and quality power supply. The Smart Grid also involves integration of renewable energy sources to meet the demand of various regions by supporting micro-grids which act as mini electricity generation centres. The figure 1 shows some of the features of Smart Grid. Load Forecasting descriptions of each category. Load forecasting techniques are classified into nine categories. In subsequent sections, one section is devoted to each category, where a brief description is given of the technique and a literature review offers a representative selection of principal publications in the given category. Arranged in roughly chronological order, the nine categories of load forecasting techniques to be discussed are: - multiple regression; exponential smoothing; Iterative reweighted least-squares; adaptive load forecasting; stochastic time series; ARMAX models based on genetic algorithms;. Fuzzy logic; Load Forecasting**[2]**is an important and crucial process in the planning and operations of the power utilities. The power utilities have to understand the various factors such as customer types, geographic location among other factors for generation and distribution of the electricity without incurring losses. It is well known that the electricity cannot be stored in vast quantities and as and when the electricity is generated it should be consumed.

Hence accurate predictions or forecasting of demand is necessary for efficient running of the power utilities. Load Forecasting has been broadly categorized into three types as Short Term, Medium Term and Long Term forecasting. The Short and Medium term forecasting is for durations ranging from one day to one week respectively whereas the long term forecasting usually ranges for more than a year. Many researchers have studied the problem of electric load forecasting but have not been able to come to an agreement on a single method of accurate forecasting that can be used for all the regions of the world. This is because the customer consumption, temperature among the various other factors that affect the forecasting varies from one region to another.

Figure 1 Shows an Overview of Smart Grid Architecture

II. LITERATURE REVIEW

The electric load forecasting method has many methods, however for the present study we shall consider the following machine learning methods:-

a. Regression [3]: Regression is a statistical method where one value is said to be dependent on another value. The dependant value is usually known as the dependant variable while the other variable is known as the independent variable. In this model an equation is built which can be used to predict the future values as and when required. The equation clearly differentiates the dependant from the independent variables.

$$Y = a + bX \quad \text{--------------------------------} \quad (1)$$

From Eq. (1), we can see that 'Y' is the dependant variable while 'X' is the independent variable which along with the constants 'a' and 'b' which are defined as the intercept value (intercept is the value of y when the value of x is zero) and the slope respectively.

b. Exponential smoothing [4]: Exponential smoothing is one of the commonly used methods for load forecasting. Thiswas the first approach to build a model based onprevious data, then to use the builtmodel to make load forecasting. This is also one of the prediction models used for analysis purposes. The formula can be shown as equation (2)

$$Yn = Yn-1 + (Xn-1 - Yn-1 \quad \text{---------------------} \quad (2)$$

Where Y n is the forecasted value, Yn−1 is previous forecast, αis known as the smoothing constant ($\alpha \geq 0$ but $\alpha \leq 1$),Xt−1 is value of actual demand of the preceding period.

c. Simple Moving Average (SMA)[5]: SMA is the most basic method of the moving averages. A simple moving average (SMA) calculated by finding the average value of an attribute over a setnumber of periods.A 5-day simple moving average is the five day sum of closing prices divided by five. A moving average is an average that is calculated on the go. Old data is discarded as new data becomes ready and available for analysis. The formula can be shown as equation (3)

M (t) = A (t) +A (t-1) +A (t-2)….A (t-N+1) / N ------ (3)

Here M is the Moving Average value and A (t) up to A (t-N+1) gives the values of the previous time period values.

d. Decomposition Method [6]: Decomposition method is a very generic term that is used for finding the solutionsfor various problems. Here the idea is to decompose the problem into many sub problems and solve it. It is an effective method for the analytical solution for a variety of systems without linearized behaviour. The formula can be shown as equation (4)

$Xa = UTaSaCaQa$ ------------------------------------- (4)

Here in the above equation (4) the value of Q denotes the random error.

III. METHODOLOGY

The historical dataset from 2006-2009 is analysed using the Machine Learning Tool Weka 3.6.11[9] software for each of the methods and performance measure is recorded to be identified using the software. The root mean square error (RMSE) value is used as the benchmark for comparing the performance of the model built using the machine learning tool. For efficient and accurate values, we perform 10-fold cross validation for the analysis. The dataset is first prepared in the excel sheet form and then converted into the CSV variant and then fed into the explorer tab of the Weka software for analysis. For the purpose of evaluating the performance for the given dataset, the value of RMSE is taken into consideration. The Root Mean Square Error (**RMSE**) (also called the root mean square deviation, RMSD) is a frequently used measure of the difference between values predicted by a model and the values actually observed from the environment that is being modelled. These individual differences are also called residuals, and the RMSE serves to aggregate them into a single measure of predictive power.

IV. ANALYSIS AND COMPARISON

The whole dataset was first made into comma separated (CSV) file format and fed into the machine learning tool Weka. After this process we selected the variables that were required for analysis using the select feature. This gives out a graphical output showing each variable with its data like number of distinct values and so on. This leads us further to select a method for comparison where we choose and apply a selected method with 10% cross-validation. After the analysis we tabulated the results by having RMSE values mapped with the data mining technique against which it was obtained. This is plotted against each other to visually compare the performance of each method. This is one of the metric used for performance analysis in load forecasting problems to compare the accuracy of the considered model along with MAPE (Mean Absolute Percentage Error) The formula RMSE is the Square root of Mean Square Error and is given below.

Table 1 shows the Performance Record

Performance Record	RMSE			
Name of the Classifier	**15,000 Instances**	**30,000 Instances**	**45,000 Instances**	**60,000 Instances**
Linear Regression	200.423	1136.119	1446.311	1593.222
LeastMedSq	189.564	1198.114	1226.318	13492.295
MultiLayer Perceptron	178.216	1058.173	1020.815	1037.859
Decision Stump	202.564	1058.173	1226.815	1242.859

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{1,i} - X_{2,i})^2}{n}} \qquad \text{-----------} \qquad (5)$$

The smaller the Root Mean Square Error indicates a better performance for the given dataset. The below table shows the performance of each of the method and the method giving out the least RMSE values has the higher accurate performance compared to rest of the methods.

We have taken different number of instance values for analysis.

The same was then plotted using a line graph using the graphical tool. The X-axis is the RMSE values and the Y-axis is the number of instance values of the historical dataset used for the analysis. As the number of instances values increases we can see that there is gradual change in the RMSE values also and hence we can see that the multi-layer perception can be used has the lower RMSE value indicating better performance than the rest of the methods used for the analysis. The mean square error value is calculated using the formula as given in the equation (5) shown above and hence is better performance indicator for the judging the results. The load on the system is considered main here for forecasting and this can be used for the electricity price forecasting also. This can also be used to better understand and prevent losses for a power utility and increase efficiency in its operation

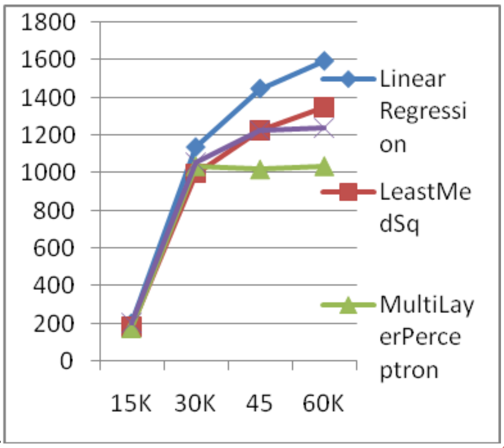

Figure 2 Shows the Graphical Analysis of Performance Comparison

V. CONCLUSION

The performance of each of the method was analysed and tabulated along with the RMSE values obtained against each of these and it can be seen that the Multi-Layer Perception which is a type of the Artificial Neural Network has better performance against other methods. So we can conclude that this method is most suitable for performing load forecasting in the smart grid environment as this method is also scalable considering the huge data influx from the smart meters in the smart grid. This method works for any number of instances which are meter reading data and hence can applied to big geographical areas also.

REFERENCE

[1] Towards Accurate Electricity Load Forecasting in Smart Grids by Zeyar Aung, Mohamed Toukhy et al. - 2012

[2] Forecasting Electricity Consumption: A Comparison of Models for New Zealand by Zaid Mohamed and Pat Bodger

[3] Bianco V, Manca O, Nardini S. Electricity consumption forecasting in Italy using linear regression models. Energy 2009; 34(9):1413e21.

[4] Benaoudaa D, Murtaghb F, Starckc JL, et al. Wavelet- based nonlinear Multi scale decomposition model for electricity load forecasting. Euro computing 2006;70 (1):139–54.

[5] Pappas SS, Ekonomou L, Karamousantas DC, et al. Electricity demand loads modelling using autoregressive moving average (ARMA) models. Energy 2008; 33(9):1353–60.

[6] Christiaanse W. Short-term load forecasting using general exponential smoothing. Power Apparatus Syst IEEE Trans 1971;PAS- 90(2):900–11.

[7] T. Cipra, “Robust exponential smoothing,” Int. J. Forecast., vol. 11, no. 1, pp. 57–69, 1992.

[8] Gelper, S., Fried, R., Croux, C., 2010. Robust forecasting with exponential and Holt winters smoothing. J. Forecast. 29, 285–300.

[9] (2013) The Weka Website [Online]. Available: http://cs.waikato.com/.

[10] [10] PrayadBoonkham, SomsakSurapatpichaiA New Method for Electric Consumption Forecasting in a Semiconductor Plant

[11] J. Shafer, R. Agrawal, and M. Mehta. SPRINT : A scalable parallel classifier for datamining. VLDB’2001

[12] L. Kaufman and P. J. Rousseeuw. Finding Groups in Data: an Introduction to Cluster Analysis. John Wiley & Sons, 2005

[13] G. M. Jenkins, "Practical experiences with modeling and forecast", Time Series, 1979.

[14] H. T. Yang and C. M. Huang, "A new short-term load forecasting approach using self-organizing fuzzy ARMAX models", IEEE Trans. Power Systems, vol.13, no.1, pp.217–225, 1998.

[15] H. T. Yang, C. M. Huang and C. L. Huang, "Identification of ARMAX model for short term load forecasting: An evolutionary programming approach", IEEE Trans. Power Systems, vol.11, no.1, pp.403–408, 1996.

[16] Z. Yu, "A temperature match based optimization method for daily load prediction considering DLC effect", IEEE Trans. Power Systems, vol.11, no.2, pp.728–733, 1996.

[17] W. Charytoniuk, M. S. Chen and P. Van Olinda, "Non parametric regression based short-term load forecasting", IEEE Trans. Power Systems, vol.13, no.3, pp.725–730,1998.

[18] Harvey and S. J. Koopman, "Forecasting hourly electricity demand using time-varying splines", J. American Stat. Assoc., vol.88, no.424, pp.1228– 1236,1993.

[19] J. W. Taylor and S. Majithia, "Using combined forecast swith changing weights for electricity demand profiling", J. Oper. Res. Soc., vol.51, no.1, pp.72–82, 2000.

[20] R. F. Engle, C. Mustafa and J. Rice, "Modeling peak electricity demand", J. Forecast., vol.11, pp.241–251,1992.