International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

# Short Term Load Forecasting by Using Data Mining Techniques

**Prof. Maya Shelke[1], Prashant Dattatraya Thakare[2]**

[1, 2] Symbiosis International University, Symbiosis Institute of Technology, Pune, India

**Abstract:** ***In this paper reports short term load forecasting is conducted by Holt-winter model and K-NN algorithm for classification. The different algorithms and techniques like Classification, Clustering, Regression, Decision trees, Nearest neighbors methods etc., are used for knowledge discovery from database. In a deregulated market it is much need for a generating company to know about the market load demand for generating near to accurate power. The paper focuses mainly those techniques that are most commonly used for short term load forecasting. For classification purpose the month wise electric load consumption dataset is gathered from Symbiosis Institute of technology, Pune campus of the year 2012 and 2013.Results have been classified by measuring Holt-winters model parameter and results has been conducted that adding parameter like temperature and event summarized to highlight the accuracy.***

**Keywords:** Holt-Winters Model, General load forecasting technique, k-NN, short term load forecasting.

## 1. Introduction

Forecast is an prediction of future events. Today the most used thing in the world is energy. Load forecasting is an important component for power system energy management system. Energy forecasting categorized as short term, medium term and long term. Short term load forecasting problems predict events only few time periods such as days, weeks and months into future. Medium term forecast predict event from one to two years and long term forecast predict through many years.

Now days, high activity of industry as IT leading and high dependence to electric power.Now need to say that accurate prediction of electric power demand has a necessary role for these requirements. With the recent move towards deregulation in the electric industry Short Term Load forecasting become important and constituent to market. Up to date many studies have been done on load forecasting. Generally they can be classified into two categories. Statistical methods(parametric) machine learning techniques(non-parametric). Statistical methods include regression model, time series, naive bayes, k-nearest neighbors and quadratic discriminant. On other hand machine learning technique include neural network, support vector machine, genetic algorithm, decision tree induction.The trend in current research tends to combine these techniques to create hybrid method i.e. concept evolution.

### 1.1 Types of Load forecasting

Electricity load forecasting has always been an important issue in power industry. Load forecasting mainly classified into four categories.

- Long term forecasting
- Medium term load forecasting
- Short term load forecasting

Load forecasting is an integral part of electric power system operations. Long lead time forecasts of up to 20 years ahead are needed for construction of new generating capacity as well as the determination of prices and regulatory policy. Medium term forecast predicts event from few 1 to 5 years ahead are needed for power scheduling, power sharing arrangements and setting of electricity cost so demand can be met with fixed capacity. Short term forecasts predict few hours to a few weeks ahead are needed for operations forecasting

**Table 1:** Overview of Forecasting

| Sr. No. | Nature of Forecast | Time | Application |
|---|---|---|---|
| 1. | Very short term | A few seconds to few minute or hour | Generation, distribution schedules, contingency analysis for system security |
| 2. | Short term | Half an hour to a few weeks | Operational planning and unit commitment, maintenance scheduling Planning for seasonal peak-summer, Winter. |
| 3. | Mid term | Few days to few months. | Planning for seasonal peak-summer, |
| 4. | Long term | Few months to few years. | Planning generation growth |

## 2. Forecasting Techniques

The paper mainly focuses on short term electric load forecasting techniques. Furthermore also discuss forecasting applications. This section present details of load forecasting techniques.

### 2.1 Short term load forecasting using k-Nearest Neighbors (NN) algorithm

Short term load forecasting using k-NN algorithm conducted by SARIMA and Holt-winters. After load classification using 1-year training set and 1-year test set forecast was performed through two models. By measuring their mean absolute percentage error, Holt winters models was shown to have better performance in short term load forecasting model. Advantage of NN is algorithm estimate target concept locally and differently for each new instance to be classified. NN algorithm learns dataset quickly and it is robust to noisy training data. Disadvantage of NN algorithm is it requires large storage requirements because it has store all data. NN algorithm is slow during instance classification

because all training instance have to be visited. Accuracy of NN algorithm degrades with increase of noise in training attributes.

### 2.2 Short-term power load forecasting based on SVM

Short term electric load forecast is also carried out by machine learning method is support vector machine (SVM). From recent years support vector machine become commonly used in short term electric load forecasting. SVM perform binary classification and regression estimate task. They are increasingly popular as classification and learning because of two factors unlike other classification technique SVM minimize expected error rather than minimize classification errors. SVM shows that the forecasting result based on support vector machine is data also increase with irrelevant better than those other methods. Besides the advantages of SVM from a practical point of view perhaps the most seriousproblems predict events only few time periods such as days, weeks and months into future.Problem with SVMs is the high algorithmic complexity and extensive memory requirements in large scale tasks. Although SVMs have good generalization performance, they can be poorly slow in test phase.

### 2.3 Short-term load forecasting with artificial neural network [ANN].

Short-term load forecasting has become an essential part of human's life. In the past few decades, many forecasting models have been presented. By using neural network model, which divides the electric load into two parts: the load scaled curve and the day maximal load and minimal load. The day maximal load and divides the electric load into two parts: The development of this model consist of three phases in which initially start from historical data, every day is classified according to its load profile by means of self-organizing feature maps the second consists of building and training the neural networks for each class and the third is an online operation phase in which prediction is carried out by previously trained recurrent neural network. In ANN structure number of input parameters, number of hidden neurons are system dependent mainly they determined by size of training set and number of input variables.

### 2.4 Short term load forecasting by using regression method

Regression method is statistical method which also used for electric load forecasting. Regression method used to understand which independent variables are related to dependent variable and find the forms of these relations. In regression method forecasting also done through linear regression, multiple regression, auto regression model. Forecast by using linear regression is simple statistical process and it is easy to carry out. But linear regression only looks linear relationship between dependent and independent variable. Other hand multiple regression estimates of unknown parameters obtained from linear least square regression are optimal. It uses data very efficiently. In multiple regression good result is obtained from small dataset. The drawback of this model is its dependency on the accuracy of previously recorded load (this will be a common drawback for all STLF

## 3. Forecasting Method

What makes good for forecast? It should be timely as accurate as possible, it should be reliable, it should be presented in writing and the method should be easy to use and understand in most cases. Till now, many forecasting methods have been utilized which are classified into two basic types: qualitative and quantitative methods. In qualitative forecastingtechniques) and temperature data which will greatly affect the forecast yield. This model is very sensitive to the fluctuation of temperature. methods generally use the opinions of experts to predict future load subjectively. Such types of methods are useful when historical data are not available. These methods are: subjective curve fitting and technological comparisons. On other hand quantitative methods are: regression methods, decomposition methods, smoothing, holt winters and the ARIMA.

### 3.1 Holt-Winters Model

A time series is simply defined as a collection of observations of well-defined data items obtained through repeated measurements over long time period. For example, measuring the value of electricity consumption each month of the year would comprise a time series. This is because electricity consumption units are well defined and it will constantly measured at equal spaced intervals. Irrelevant data and data collected irregularly or only once are not time series. In time series it is observed that time series can be dividing into two components: the trend that is long term direction, the seasonal that is change by season.

Holt (1957) and winters (1960) extended Holt's method to capture seasonality. This model was first proposed in the early 1960s. It uses a process which is known as exponential smoothing. All data values in time series contribute to the calculation of the prediction model.

The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations — one for the level ℓt, one for trend bt, and one for the seasonal component denoted by st, with smoothing parameters α. We use m to denote the period of seasons in a year i.e. for quarterly data season is m=4, and for monthly data season is m=12.

#### 3.1.1 Season Factor

In simple way description of seasonal effect is a systematic and calendar related effect. Some examples include electricity consumption is larger which occurs around Christmas period or diwali period or in holidays as compare to the ordinary period. Seasonal effect arise influences such as:

1) In Natural Conditions: weather fluctuations that are representative of the season
2) In Business and Administrative procedures: start and end of the school term
3) In Social and Cultural behavior: Christmas and diwali holidays.

### 3.1.2 Trend Effect

The trend effect simply defined as the 'long term' movement in a time series without calendar related and irregular constants effects. The following figure 1 shows time series without any trend and figure 2 shows a series in which there is an obvious upward trend over time:

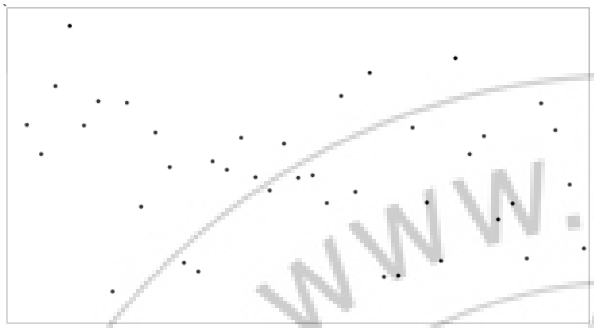

**Figure 1:** Time series without trend

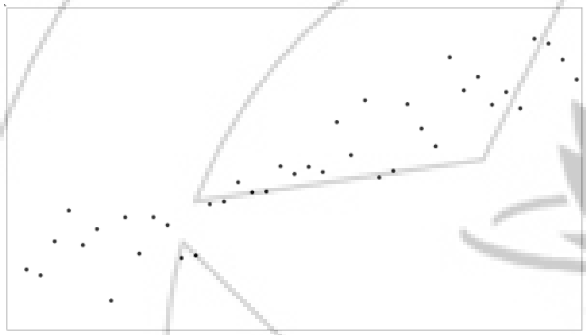

**Figure 2:** Time series with trend

### 3.1.3 Types of Holt-Winters Model

There are two types of Holt-winters model named as Addictive model and other one is multiplicative model.

### 3.2 Addictive Models

Addictive model is type of Holt Winters model in which addictive model considers that for monthly data an addictive model assumes that the difference between the January and July values is approximately the same each year.

### 3.3 Multiplicative Model

Multiplicative model is another type of Holt Winters model. Suppose season factor and trend factor are consider in model at that time multiplicative model is best suitable for forecasting.For example in electricity consumption the load on different holidays like charismas or any season such as summer behave differently like for example, in summer vacations so may have structurally different loads than ordinary days. This happening particularly true during the summer time. Generally Holidays are more difficult to forecast than non- holidays because of their relative infrequent occurrence. Load variation with time reflects the arrangement of people's daily life: i.e. working time, leisure time and sleeping time. Basically there are some rules of load variation with time that observed. At the weekend or holiday load curve is lower than the weekday curve, due to the decrease of working load. At this situation multiplicative model is best suited.

In the addictive model the components are added and in the multiplicative model they are multiplied. Using *T* for trend, *S* for season and *L* for level, we can represent these models as follows:

*Addictive model:* F (t) =Trend+Season+Level (1)
*Multiplicative model:* F (t) =Trend*Season*Level (2)

### 3.3.1 Actual implementation of Holt-Winter model:

To perform short term load forecasting need electric load consumption data. We used two year historical electric load data. For classification purpose the month wise electric load consumption dataset is gathered from Symbiosis Institute of technology, pune of the year 2012 to 2014. By many researchers observation it is clear that multiplicative model can better performance than addictive model. So we can implement holt-winter by using multiplicative model.
is used for forecast calculation:

**Table 2:** Unit Consumption for year 2012.

| Area | Months | Units | Total |
|---|---|---|---|
| Canteen | Jan | 1290 | 18650 |
| | Feb | 1485 | |
| | March | 1300 | |
| | April | 1900 | |
| | May | 1500 | |
| | June | 1525 | |
| | July | 1450 | |
| | August | 1450 | |
| | Sep | 1600 | |
| | Oct | 2050 | |
| | Nov | 1600 | |
| | Dec | 1500 | |

The general form of multiplicative model for Forecasting is calculated as follows:

$$Ft = (lf + (tf * m)) * sf$$

Where,
Ft=forecasting
lf=level factor of particular period
tf=trend factor of particular period
sf=seasonal factor of particular period
m=period of forecasting

Initializing the process firs calculate the season factor, level factor and trend factor.
Effect of January
$S_f$ =(Jan Actual Value/Total) * 12
$L_f$=(Jan Actual Value/January Season Factor )
$T_f$=(Jan Actual Value*(User Defined trend))

Calculate Season factor, Level factor and trend factor of canteen campus example:

### 3.3.2 Season Factor

=(Jan Actual Value/Total) * 12
= (1290/18650)*12
=0.8300

### 3.3.3 Level Factor

= (Jan Actual Value/January Season Factor )
=(1290/0.8300)
=1554.167

### 3.3.4 Trend Factor

=(Jan Actual Value*(User Defined trend))
=(1290*(-0.05))
= -64.5

[Here we use user defined value set to [-0.05]]

Forecast value Ft =(lf+(tf*m))*sf
=(1554.167+((-64.5)*1))*0.8300
=1236.42

## 4. Classification by using K-NN Algorithm

To perform short term load forecasting need electric load consumption data. We used two year historical electric load data. The data to classify monthly electric load consumption in units. For classification purpose the month wise electric load consumption dataset is gathered from Symbiosis Institute of technology, pune campus of the year 2012 and 2013. K-Nearest Neighbours [NN] have been used in statistical estimation and pattern recognition already in the beginning of 1970's. The K-nearest neighbours algorithm is a method for classifying objects based on closest training examples in the feature space.

### 4.1 KNN algorithm

K-NN algorithm is method for classification and regression that predicts object values based on k-closest training example. K-Nearest Neighbours (KNN) classification divide data into a learning set and a training set. For each row of the learning set, the K nearest formula which is known as Euclidean distance apply on training set objects and the classification is determined by majority vote with on training examples.

There are many reasons to choose K-Nearest Neighbours [NN] algorithm for classification purpose. Following description shows the strength and weakness of K-NN.

A distance measure is needed to determine the "closeness" of instances. Classify an instance by finding its nearest neighbours and picking the most popular class among the neighbours. To determine the class of a new example A:

1) Calculate the distance between A and all examples in the training set.
2) Select K-nearest examples to E in the training set
3) Assign A to the most common class among its K-nearest neighbours
4) A new example is assigned to the most common class among the (K) examples that are most similar to it.

In K-NN algorithm closeness is in terms of the Euclidean distance between two examples. The Euclidean distance between X=(x1, x2, x3,…xn) and Y =(y1,y2, y3,…yn) is defined as:

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

## 5. Result

The result was performed on two year historical electric data of Symbiosis Institute of technology, pune campus collected for training and test data to carry out forecasting load demand. Holt-winters model were supplied with monthly load unit consumption which were classified by using K-NN algorithm. Figures shows depictions of Holt-winters model accuracy with adding parameters event and temperature. Figure 5(a),5(b),5(c) and 5(d) are depictions of canteen,hostel,café_70 and SSBS area in which black line shows actual load consumption, blue line shows existing method load consumption and pink line shows proposed method with adding extra parameter temperature and event load consumption

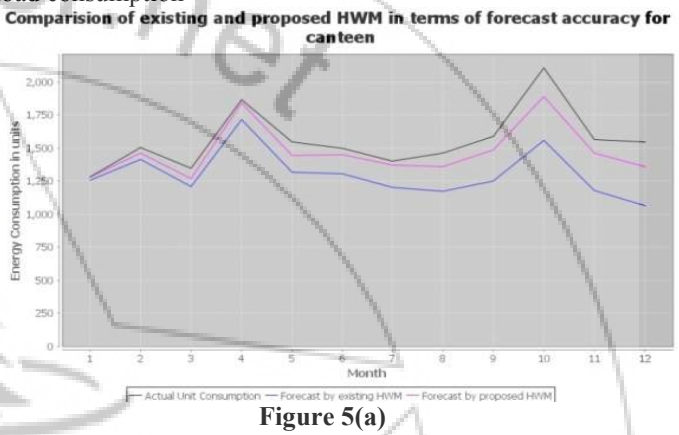


**Figure 5(a)**

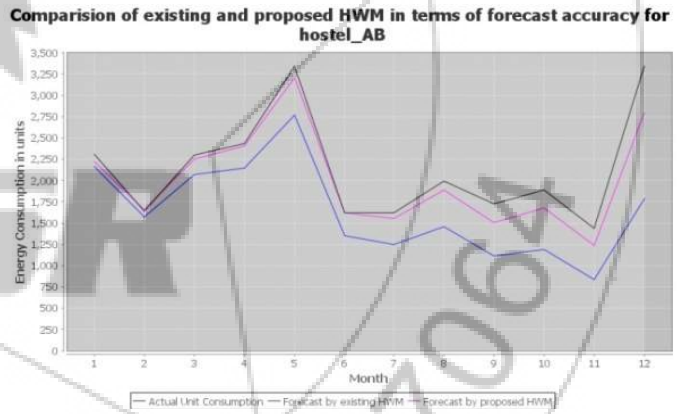


**Figure 5(b)**

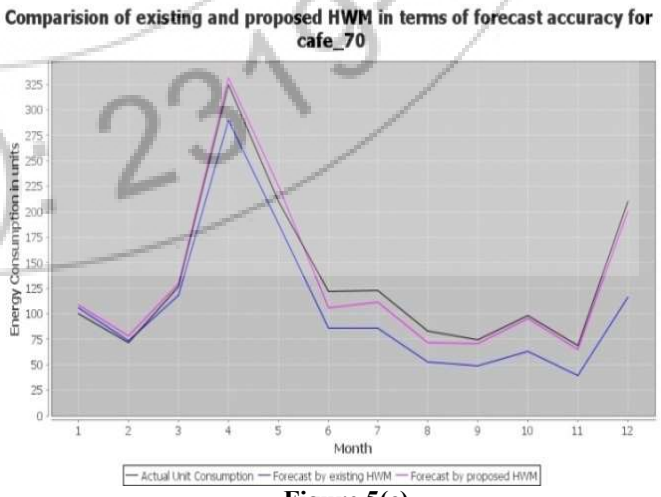


**Figure 5(c)**

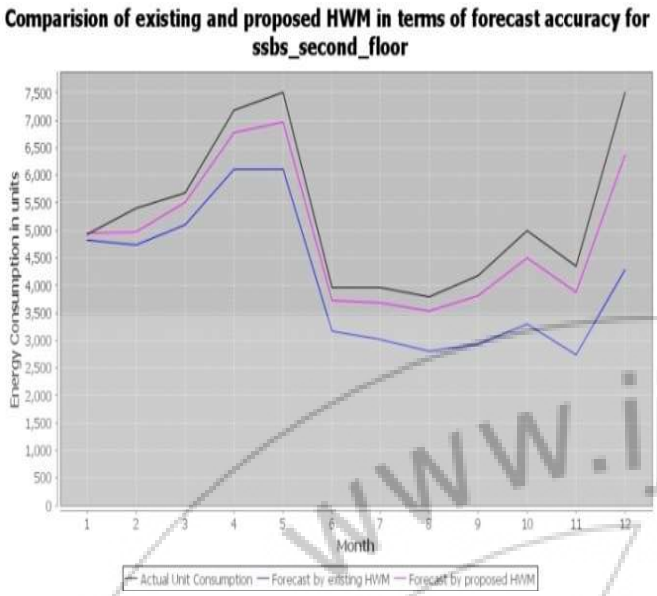


**Figure 5(d):** Comparing the accuracy of model shows that the proposed Holt-winter model which adding parameter is more accurate.

## 6. Conclusion

This paper suggested that short term load forecasting classification by K-NN algorithm and Holt-winter model. As preprocessing classification was successfully completed. By comparing with existing Holt-winters model and Proposed Holt-Winter model which considering extra parameter like event and temperature, however Proposed system performs better than existing.

## References

[1] D.P Kothari& I.J Nagrath, "Modelling and forecastingelectricityloads", Tata McGrawHill, 2003.
[2] H.Lee. Willis, "Spatial Electric Load Forecasting", Marcel Dekker, Inc, 2002.
[3] JuergenSchlabbach, Karl-Heinz Rofalski, "PowerSystemEngineering", Wiley&Verley Co,2008.
[4] Online: http://docs.oracle.com/cd/E16582_01/doc.91/e15111/und_forecast_levels_methods.htm [AccessedonSept; 2013]
[5] Online: https://www.otexts.org/fpp/7/5 [Accessedon Nov; 2013]
[6] G. Juberias, R. Yunta, J. Garcia, Moreno C. "Mendivil, A New ARIMA ModelforHourly Load Forecasting",Transmission and distributionConference, 1999IEEE Vol.1, 11-16 April 1999, pp:314-319 vol.1
[7] Dae-Yong Kim, Chan-Joo Lee, Yun-Won Jeong,Jong-Bae Park, Joong-Rin Shin, Development ofSystem Marginal Price Forecasting Method UsingSARIMA Model, Proceeding of conference KIEENov. 2005, pp.148-150
[8] HyeonjoongYoo, Russel L. Pimmel, "Short-TermLoad Forecasting using a Self-Supervised AdaptiveNeural Network", IEEE Transaction on PowerSystem, Vol.14, No.2, May 1999
[9] William W. S. Wei, Time Series Analysis(UnivariateAnd Multivariate Methods) 2nd Ed., Addison-Wesley,2005
[10] KohHee-Soek, Lee Chung-Sik, Kim Hyun-Deok,Lee Hee-Chul, "Short-term Load Forecasting usingNeural Network", Proceeding of conference KIEE1993, pp.29-31, Nov. 1993
[11] Xindong Wu, Vipin Kumar, J. Ross Quinlan, JoydeepGhosh, Qiang Yang, Hiroshi Motoda, Geoffrey J.McLachlan, Angus Ng, Bing Liu, Philip S. Yu, ZhiHuaZhou, Michael Steinbach, David J. Hand, DanSteinberg "Top 10 algorithms in data mining",Knowledge and Information Systems, vol.14, pp1~37,2007

## Author Profile

**Prashant Dattatraya Thakare** received B.E. in Computer Engineering from S.S.V.P.S. College of engineering, Dhule from North Maharashtra University, Jalgaon. (2012). Currently pursuing M.Tech [Computer Science & Engineering] from Symbiosis Institute of Technology, Pune. Area of interest is Data Mining and Modeling.