A Survey on Forecasting of Time Series Data

G.Mahalakshmi
PG Scholar
CSE Department
Thiagarajar College of Engineering,
Madurai, India
maha0293g@gmail.com

Dr.S.Sridevi
Assistant Professor
CSE Department
Thiagarajar College of Engineering,
Madurai, India
sridevi@tce.edu

Dr.S.Rajaram
Associate professor
ECE Department
Thiagarajar College of Engineering,
Madurai, India
rajaram_siva@tce.edu

***Abstract*—Time series analysis and forecasting future values has been a major research focus since years ago. Time series analysis and forecasting in time series data finds it significance in many applications such as business, stock market and exchange, weather, electricity demand, cost and usage of products such as fuels, electricity, etc. and in any kind of place that has specific seasonal or trendy changes with time. The forecasting of time series data provides the organization with useful information that is necessary for making important decisions. In this paper, a detailed survey of the various techniques applied for forecasting different types of time series dataset is provided. This survey covers the overall forecasting models, the algorithms used within the model and other optimization techniques used for better performance and accuracy. The various performance evaluation parameters used for evaluating the forecasting models are also discussed in this paper. This study gives the reader an idea about the various researches that take place within forecasting using the time series data.**

Index Terms*—*Forecasting, Time series, Prediction, Temporal data mining

I. INTRODUCTION

The process of extracting useful and valid information from a large amount of data is termed as Data mining. The data can be of any format such as text, audio, video, etc. These large amounts of data that are available in business and market related areas are called as Big Data [1]. By extracting the useful information and patterns using data mining, these patterns can be analyzed for making future predictions. The prediction made helps the organization to make useful decisions for the growth of the company or organization. Apart from extracting data and making predictions, the other criteria such as data integrity and security should also be taken into consideration. But in this survey, prediction concept and methodologies is mainly focused.

The data are available in different formats and structures. Temporal data is a type of data that varies over the change in time. These data are represented using time stamps. The process of extracting or discovering patterns of data from temporal databases is temporal data mining. The major process involved in temporal data mining is the analysis of temporal data and finding useful patterns in temporal data. The data mining functionalities involved in temporal data mining are:

- Temporal classification: It is the process of finding a model that distinguishes data classes whose class label is unknown in the temporal database.
- Temporal prediction: It is the process of finding unknown future values from historical data.
- Temporal association: It searches for the temporal relationship between attributes.
- Temporal clustering: It is the process of grouping similar data objects from temporal database.
- Temporal regression: It attempts to find a function which models the data with the least error.
- Temporal summarization: It provides a more compact representation of the data set including visualization.
- Temporal outlier detection: Data objects that do not adhere with the generic behavior of other data within the database.

The temporal data mining is applied in many real time organizations such as marketing, banking, healthcare, business, stock market and weather forecasting or prediction.

Time series analysis and prediction [2] are part of the temporal data mining. Collection of large number of data values within a uniform time interval is termed as Time series data. The time can be represented as year, month, week, day, etc. The time series is analyzed to predict the changes that happens within the given data and to predict the changes that will happen in the future. That is, the characteristics of time series are analyzed to foresee the future data. The time series is represented by 5-tuple such as starting time, the pattern, the period value, confidence and the ending time.

In recent years, time series prediction is done in many applications that deals with numerical data. The prediction can be done on the basis of three different time spaces as:

- Short-Term period
- Mid-Term period
- Long-Term period

The Short-Term forecasting or prediction focus on a time frame or period of less than three months, whereas the Mid-Term period focuses on a time frame of three months to one year and the Long-Term considers a time period more than a year. Based on the time period and the type of data, many techniques have been used for prediction.

978-1-4673-8437-7/16/$31.00 ©2016 IEEE

In Earlier days, many machine learning algorithms and model has been applied for predicting the temporal data. Some of them employ hybrid methods by combining more than one model or using any optimization algorithm along with the prediction models. This paper provides the detail survey on the various forecasting models that have been used for prediction. The time series data such as electrical data, enrollment data, stock market data, seasonal & trend data, etc. are considered, they are classified and explained accordingly. The different uses of time series analysis, prediction in various applications have been discussed by classifying the methods used in different categories.

The second section deals with the various applications of time series forecasting. The third section briefs about traditional forecasting techniques. The fourth section deals with stochastic forecasting techniques. The fifth section deals with soft computing based forecasting techniques. The sixth section deals with fuzzy based forecasting techniques. The seventh section deals with evaluation parameters and the eighth section deals with the results and discussion.

II. APPLICATIONS OF TIME SERIES FORECASTING

The process of forecasting is applied in a wide range of areas where the forecasted values or conditions are helpful in many ways to make useful decisions. To demonstrate the usefulness of forecasting methods, this section discusses about the four major applications of forecasting in the world of business and personal investments.

A. Electricity Market Clearing Price

The price that subsists after clearing the shortages and surplus in electricity market is termed as Electricity Market Cleaning Price [3] (MCP). Electricity MCP forecasting predicts the future price of electricity referring to the various values such as the demand, the cost of fuel, temperature and other electricity related factors such as the volts, consumption, etc. The Electricity MCP is the result of overall bidding price for electricity in the market. After the electricity MCP is determined, the offering prices of all the suppliers with price less than or equal to that of the electricity MCP will be selected to provide electricity supply for that time period. The rate will always be less than that of the MCP. In this way the market balance is maintained and there is no possibility of market manipulation or fraud.

B. Electricity Load Forecasting

Electricity load forecasting is a crucial and important parameter to plan and operate various electrical items. The magnitude of load and the geographical location of electricity over different time periods are predicted. Gross et al. [4] stated that load forecasting is associated with predictions made in hourly, daily, weekly and monthly periods. Later Srinivasan et al. proposed three different time periods for load forecasting: (1) Duration up to 1 day the Short-Term Load Forecasting (STLF), (2) Duration from 1 day till 1 year the Mid-Term Load Forecasting (MTLF) and (3) Duration from 1 year till 10 years the Long-Term Load Forecasting (LTLF).

The demand pattern of the load time series is complex due to the changes in electricity market. Hence it is difficult to find and implement a specific appropriate forecasting model. E.Almeshaiei et al. [5] stated a pragmatic method that can be used as a guide for building an appropriate Electric Power Load Forecasting (EPLF) models based on the needs. The methodology here uses the decomposition and segmentation techniques to split the load time series data. Many statistical analysis techniques have been employed to study the various load parameters and the accuracy of prediction.

C. Trend and Seasonal Time Series Forecasting

In most of the business and economic related data, the trends and seasonal time series data are available. It is important to accurately predict trends and seasonal time series for making important business decisions. Many techniques are available for predicting or forecasting the trend and seasonal time series data. The most used and basic method involves removing the seasonal changes within the data using certain seasonal adjustment methods. Other traditional methods include Box-Jerkins models and Autoregressive Integrated Moving Average (ARIMA) model [6]. ARIMA technique is applied only after converting the input time series to a stationary time series. But the underlying process used in the ARIMA model is of linear type and so it fails to capture the other nonlinear patters in the time series.

D. Stock Selection and Portfolio Construction

Nowadays, the area of the stock market and stock selection is crucial in investments and decision making. The main task here is the future prediction of the stock market data and the construction of portfolios. Using the historical stock market data, many soft computing approaches have been used for predicting financial time series data to avoid risk of reduction, make expected return of the investment and to construct portfolios successfully. Many approaches such as the Artificial Neural Networks (ANN), Evolutionary Algorithms (EA), Support Vector Machines (SVM) and Fuzzy Logic Systems (FLS) have been used for this purpose of stock selection and portfolio construction using the financial time series data. TC.Chu et al. and Zargham et al. designed a fuzzy multiple decision analysis model [7] and a fuzzy rule-based system, respectively to construct portfolio based on the selected stocks. Although the fuzzy based methods are strong in computation and accuracy, they lack in the learning process.

III. TRADITIONAL FORECASTING TECHNIQUES

In earlier days, the prediction and forecasting of future trend values were done using traditional forecasting techniques that make use of mathematical formula and techniques. These basic techniques can be improved further using various advancements in different tools and automation methods. The various traditional forecasting techniques are discussed below:

A. Regression Method

The regression method is an extensively used and easy to implement. This method is mostly implemented in a model that has various types of relationships between the data to be predicted and the other factors that it depends on. An assumption is made here in such a way that the data to be predicted is divided into standard base data and the linear data that depends on the factors that are influencing the data. The regression method depends on the representation of the various conditions that are used for predicting the future data and a method should be implemented to calculate these factors using the existing historical data.

By considering an electrical load forecasting problem, the load depends on the following factors such as the customer characteristics, the various days and the conditions of weather. To forecast the load in yearly manner the daily load change, annual change and the seasonal changes should be considered. A transformation based method is proposed that makes use of the transformation function for predicting the load. The transformation function contains reflection and translation techniques within. The radical change function converts previous year's existing load data into future predictions by implementing different functions. In general the mathematical model of the regression method is modeled as given below:

$$L(t) = L_n(t) + \sum a_i x_i(t) + e(t) \qquad (1)$$

Where,
$L_n(t)$ – normal load at time t
a_i – estimated varying coefficient
$x_i(t)$ – independent factors that affect load
$e(t)$ – noise component
n – number of data

Various types of transformation functions have been used for different types of time series prediction.

B. Multiple Regression Method

Multiple regression based model is also suitable for predicting the electricity load and demands using the existing data. They use the weighted least-squares estimation for predicting the load. The regression coefficients are computed by making use of weighted least-square estimations using the available historical data. G.A.N.Mbamalu et al. [8] proposed a regression model that uses this type of approach for load estimation or prediction by applying the following equation:

$$Y_t = v_t a_t + \varepsilon_t \qquad (2)$$

Where,
t – sampling time`
Y_t – total load at time t
v_t – vector of variables that affect the load
a_t – vector of regression coefficient
ε_t – model error

I.Moghram et al. studied and compared this regression model with the existing techniques. Later E.H.Barakat et al. [9] used this regression model for fitting values in the given dataset to check for the seasonal changes. Haida el al. proposed a regression based forecasting model that forecasts the daily peak value of the load using a transformation method. This method predicts the nominal load and the residual load using the regression model. Later Haida et al. proposed the extended version of this method by adding trend processing techniques within the regression model to reduce the errors during the transitional seasons of the load.

C. Exponential Smoothing

Exponential smoothing is a classical method used in many forecasting based applications and especially for forecasting the load. In this approach initially the data to be predicted is modeled based on the previously existing data and then this model is used to predict the future data. I.Moghram et al.[10] proposed an exponential smoothing model using a function as given below:

$$y(t) = \beta(t)^T f(t) + \varepsilon(t) \qquad (3)$$

Where,
$f(t)$ – fitting function
$\beta(t)$ – coefficient vector
$\varepsilon(t)$ – white noise

One of the seasonal time series prediction methods that use exponential smoothing is the winter's method. A total of three smoothing parameters such as trend, seasonality and stationary are used for seasonal time series data. E.H.Barakat et al. proved that it is difficult to analyze and predict by directly applying the winter's methods. Data sets usually contain sparse values. To overcome the sparsity problem El-Keib et al. proposed a combined strategy where the exponential smoothing method was combined with other methods such as the autoregressive models and spectrum analysis. An assumption is made that states that a time series is stationed locally with slow varying mean when using averaging and smoothing methods. To avoid this, a varying average is used to evaluate the current mean value. Compensating the degree of smoothing may lead to strike some kind of ideal balance between the performances, but the complexity of the model increases as it relies on lots of computational factors.

IV. STOCHASTIC FORECASTING TECHNIQUE

The unique patterns of the demand and load of the electricity forecasting and other applications are difficult to forecast using normal time series techniques. But the existing basic time series methods are the standard and popular approaches that are available. Using these time series methods first a model is built and then the future data are predicted or forecasted using these models. These types of models are stochastic methods and some of them are explained below:

Xing Yan et al. compared the Support Vector Machine and Least Squares Support Vector Machine for forecasting MCP based on mid-term electricity [3]. The advantage of using SVM over the other forecasting method such as the Artificial Neural Networks (ANN) is that the problem like data over fitting, the large out of the sample types of data and the local minimum problems can be avoided in SVM. The SVM learning method is much simpler and easier to model than the other complex models used in ANN and Bayesian Networks. The authors provided the comparison between SVM and

LSSVM forecasting methods by predicting hourly electricity MCP for mid-term period. The dataset is taken from the existing PJM interconnected electric market. The data values from 1/1/2009 to 31/12/2009, excluding June month data are considered as training data. The remaining data from 1/6/2009 to 30/6/2009 is taken as testing data. During experimental implementation of these two methods, a comparison is made between them. During the training process SVM uses a quadratic formula, whereas LSSVM executes a linear equation. The next variation is that, the SVM selects only those data with static co-efficient as support vectors and the LSSVM selects training data as vectors.

Xing Yan et al. proposed another combined forecasting model for MCP forecasting based on Mid-Term using Support Vector Machine (SVM) and the AutoRegressive Moving Average with External input (ARMAX) [11]. The same PJM interconnected electricity market data was used for the experimentation. The proposed SVM-ARMAX hybrid model is compared with the existing models such as the single SVM, single LSSVM, single ARMAX and a hybrid LSSVM-ARMAX [12]. The comparison shows that the SVM-ARMAX hybrid method is more accurate than other methods listed above. This is because the SVM model can obtain a better forecasting accuracy by accumulating a linear module. SVM also has the capability to handle the outlier data values during the training stage.

Later Xing Yan et al. proposed different SVM models for electricity MCP forecasting. The author used multiple SVM and multiple LSSVM [13] models instead of using single SVM. The data pre-processing task is done first based on the various price zones by designing a data classification and forecasting model. The price zones are categorized as Low, Medium, High and Peak. The multiple SVM and LSSVM improve the forecasting accuracy of the peak prices and the overall performance is high compared to that of the single SVM and LSSVM. The PJM interconnected electricity market data is used for the analysis. The multiple LSSVM model consist of many layers of LSSVM to get improved results. This model consists of a LSSVM based classification stage followed by a LSSVM multi layered prediction layer. A similar model was also proposed with a normal multiple SVM layered model instead of the LSSVM. The multi layered LSSVM provides better results compared to the single approaches of SVM and LSSVM.

The LSSVM models are the basic type of models in regression. The efficiency of these models depends on the choice of the kernel factors and the hyper factors that are defined within the functions of these models. G.Rubio et al. proposed a heuristic approach for the parameter approximation in LSSVM models [14] to be used for anticipating the future time series data. The Non-parametric based Noise Estimation (NNE) is used to make the guess of the hyper parameter γ and the training-data based guess. It is used to estimate the parameter of the Gaussian kernelσ. These guessed parameters are then cross validated using the l-fold cross validation error [15] and optimization of parameters are done for the LSSVM model. The Levenberg-Marquardt (LM) or the Conjugate Gradient (CG) scheme can also be used with the different optimization approaches.

The characteristics of the stock market data are that, they are noisy and have multiple dimensionalities. Thus, most of the above mentioned methods are proven to be inconsistent and unpredictable in terms of prediction performance. This is due to the use of nonlinear process within the models for time series forecasting. To overcome these issues Vapnik developed a financial time series prediction model [16] using a well-known novel machine learning algorithm, the Support Vector Machines (SVM) that uses linear solutions for solving these kinds of problems. The SVM researches have shown that these models overcome the issues in the existing nonlinear based method such as the ANN based prediction, Discriminate Analysis, Case based Reasoning and Back-propagation NN.

The SVM based financial time series models are not only used for forecasting in financial applications, but they are also employed to anticipate the future of stock market index or a specific stock. A coarse-grained classification procedure based SVM was proposed by Fan et al. to classify each of the stocks into winning and losing groups by analyzing the time series patterns. In later research, the author used SVM for regression of stock results that serve as a duplicate or predicted values of the actual stock results that imply their quantity and ranking. The performance of SVM and Support Vector Regression (SVR) depends on the various input features of the stock data that is taken for processing within the models. By combining the SVM and SVR regression models with other parameter Optimization techniques such as Genetic Algorithm (GA) the forecasting can be made much better.

H.Frohlich et al. was the first one to provide a model on this using SVM with GA for financial time series forecasting [17]. A much similar method was proposed later by CL.Huang et al [18] that involve concurrent optimization process and they proved that the classification efficiency in SVM is better using some University of California Irvine (UCI) financial data sets. Later Huang proposed a similar type of hybrid model for stock selection or analysis using the Genetic Algorithm (GA) and Support Vector Regression (SVR). In this approach, various features of the data are extracted depending on the learning algorithm used in the SVR that uses the wrapper approach. The SVR is used to generate the forecasts of the actual stock results based on ranking. The highly rated stocks are finally taken for construction of portfolio. The parallelization of parameter optimization and modeling is made using GA.

V. SOFT COMPUTING BASED FORECASTING TECHNIQUES

These days another most popular learning method called Neural Networks (NN) or the Artificial Neural Networks (ANN) is applied for analysis of trends and seasonal time series data. The NN based models are more preferable for certain types of time series prediction and it is also valuable for predicting trends and seasonal time series. Many investigations and researches have been done with NN for predicting trend and seasonal time series. GP.Zhang et al. provided an investigation [19] on how to model the time series prediction of trend and seasonal data with various types of NNs. They used Neural Networks to model quarterly prediction of time series for trend and seasonal patterns. C.Hamzacebi proposed a new type of NN model [20] for improving the performance of seasonal time series prediction. An examination of hybrid NN techniques used for time series prediction was given by TT.Temizel et al. [21] to train the NN model; many random search algorithms and gradient search algorithms are used depending on the type of time series data.

Apart from these existing methods, the various coefficients models have been used in recent days to explore and analyze the dynamic or time series data. The Radial Bias Function network based Autoregressive model (RBF-AR) is one of the widely used coefficients modeling. The RBF-AR is a linear model that is derived from a set of RBF networks. It is used for approximating the coefficients. The quasi-linear auto regression model also belongs to the same category as that of the RBF-AR model. M.Gan et al. proposed and implemented the RBF-AR model to forecast the trend and seasonal time series [6]. The various criteria to be considered for building the RBF-AR model is the type of input variables, the selection of a suitable state vector representation and the approximation of all the parameters such as the number of nodes, weight values and the center values.

The proposed method by M.Gan et al. uses a hybrid Genetic Optimization and a Gradient based Optimization for modeling the RBF-AR model [22]. This model is tested against a total of five time series data obtained from the US Retail Sales from US census Bureau. All the collected data are available from Jan 1992 till Dec 2001. The forecasts are made for these entire five trends and seasonal time series and the behavior is compared with other existing approaches such as Support Vector Regression (SVR) [18], Artificial Neural Network (ANN) based models [19-21], ARIMA [6] and Time Delayed Neural Network (TDNN) models.

Earlier, M.Gan et al. proposed and experimented a similar RBF NN based architecture for prediction on nonlinear time series data. The input parameters and the nodes of the RBF network are evolved and optimized using the Genetic Algorithm (GA). For each of the chromosomes with certain inputs and nodes the respective parameters are optimized using a gradient-based parameter augmentation function. The parameter optimization method employed in this method was the Structured Nonlinear Parameter Optimization Method (SNPOM) [23]. The performance of the proposed hybrid RBF network with GA is experimented by many nonlinear time series data such as Box-Jenkins time series, Wolf's sunspot data and Mackey-Glass time series. The performance is evaluated using the various evaluation parameters.

TS.Quah et al. [24] proposed and experimented an ANN based stock selection system that can be used to select stocks that are top ranked in terms of performance. N.Chapados et al. [25] also used a learning based NN to predict the behavior of the various assets in the stock market to provide decision making in stock assets. But the use of NN worked in some applications, but in other cases, they face the over fitting problems. K.Kim et al. designed a Genetic Algorithm (GA) based method [26] to discretize the stock features and to find out the connection weights in the ANN model to predict the stock market index. Many other researches and modifications have been done in the GA based stock models that are used to rank the stocks from high to low based on the pre-defined fitness functions.

VI. Fuzzy Based Forecasting Techniques

In real time, many time series data have more than one value of observations or attributes. One such example of such date is the stock exchange data that have different attributes like low, medium and high in a single day. For such type of data it is not possible to use conventional time series analysis. By modeling these types of time series using a fuzzy set [27] will be more accurate. The fuzzy approach for time series depends on the fuzzy logic systems or fuzzy sets proposed by Zadeh. The three main phases in the fuzzy time series analysis or prediction are the fuzzification stage, defining the fuzzy rules & relations and the defuzzification stage.

Erol Egrioglu et al. proposed a novel hybrid method by combining the Fuzzy C-Means (FCM) with Artificial Neural network (ANN) for forecasting fuzzy time series [28]. The FCM clustering algorithm is applied to the fuzzy time series as the fuzzification stage to calculate the various cluster centers and the corresponding membership values for each of the data in the time series. Then the relationship establishment is done using the concepts of feed forward NN and the fuzzy rules are defined. Based on these relationships the fuzzy forecasts are obtained from the NN. These are the defuzzified forecasts. University of Alabama's enrollment data is taken as the initial dataset. This novel method is implemented for the yearly recorded time series data from the year 1971 to 1992.

QiSen Cai et al. presented a new hybrid forecasting model for stock data using fuzzy time series and Genetic Algorithm (GA) [29]. This hybrid stock forecasting model Fuzzy Time Series Genetic Algorithm (FTSGA) develops the efficiency of the forecasting using the GA optimization techniques such as the selection, crossover, fitness calculation, mutation, etc. This model is experimented using a large dataset named TAIEX that contains closing prices from the year 1990 till 1999.The data from January to October, is used as the estimation data and the remaining data from November to December is used for forecasting the stock values. The performance evaluation shows that the FTSGA hybrid method is better in performance and forecasting accuracy compared to that of other existing methods.

E.Bas et al. proposed a modified Genetic Algorithm (GA) approach to infer the future values of fuzzy time series [30]. Genetic Algorithm optimization is mostly used along with the fuzzy time series approach for their good performance link in the above FTSGA model. But the randomization of the GA mutation, leads to negative results. Thus a modified GA is proposed to calculate the optimal interval in the mutation phase that will control these negative effects. The Modified genetic Algorithm (MGA) based fuzzy time series method is applied to three datasets such as the enrollment data from University of Alabama, TAIFEX dataset and the data from Belgium vehicle road accidents. The performance examination shows that MGA approach is superior compared to certain existing forecasting models.

Fuzzy based time series prediction is more preferable in cases of non-linear and non-stationary data. M.Shah [31] provided a study on fuzzy time series predictions and proposed a fuzzy based time series forecasting and trend prediction using the IF-THEN fuzzy logic rules to seize the trend that is available in the time series data. An enhanced fuzzy time series forecasting algorithm is implemented. The trend is identified as different classes such as increasing, decreasing, steady etc. The proposed method is experimented in three datasets such as University of Alabama's enrollment data, Sales data of a Propylene manufacturing firm and Gross domestic capital of India.

Many approaches have been applied in the fuzzification phase of the fuzzy time series. One such method is the use of the Fuzzy C-Means (FCM) clustering technique. Apart from this, other clustering techniques like hierarchical clustering and Gustafson-Kessel fuzzy clustering are also used. For the defuzzification phase mostly the Neural Networks such as the Feed Forward NN and Centroid based defuzzification is used. In time invariant fuzzy time series, the forecasting results depend on the second phase that determines the relationship of the fuzzy logic. In existing method, they use fuzzy logic relationship tables for this purpose, but in some scenarios, there are more chances for information loss. To avoid these flaws, the Particle Swarm Optimization (PSO) algorithm is applied for identifying the fuzzy relationships by using membership values obtained from the FCM.

CH.Aladag et al. [32] suggested a hybrid invariant fuzzy time series forecasting technique that uses FCM clustering and PSO optimization algorithms to predict the fuzzy time series. As discussed before the FCM clustering technique is used to identify the cluster centers and membership values for all the time series data in the dataset. In the next phase, the PSO algorithm is implemented to compute the membership values using the fuzzy relationship matrix. Using the relationship matrix, defuzzified forecast values are determined for the testing and training data. The PSO algorithm calculates the defuzzified forecasts for the training data and using the best obtained position of the particle to calculate forecasts for testing dataset. The designed method is applied to three datasets such as University of Alabama's enrollment data, Index 100 dataset in stocks and Istanbul's bond exchange market dataset and finally the TAIFEX data. Erol Egrioglu also proposed a similar PSO based invariant time series analysis that uses the FCM clustering. The author provided a comparison between IMKB data, TAIFEX data and the TAIEX stock exchange datasets.

VII. Evaluation Parameters

The evaluation parameters are the performance evaluation metrics [32] that are used to observe the overall efficiency of the prediction of time series data. Any prediction technique is compared with other techniques using these performance evaluation metrics or parameters. Some of the evaluation parameters are discussed below:

For forecasting or analysis of time series data, the most used performance evaluation factors are given below: For a total of N given historical data in the dataset, p_x and the corresponding forecasted or predicted value $\hat{p}_x$, such that $x = 1 \; to \; N$, the performance evaluation factors are given as,

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\sum_{x=1}^{N}|p_x - \hat{p}_x| \tag{4}$$

- Normalized Mean Absolute Error (NMAE)

$$NMAE = \frac{MAE}{p_{max} - p_{min}} \tag{5}$$

- Mean Absolute Percentage Error (MAPE)

$$\mathrm{MAPE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{x}=1}^{\mathrm{N}}\left|\frac{\mathrm{p_x} - \hat{\mathrm{p}}_\mathrm{x}}{\mathrm{p_x}}\right| * 100 \tag{6}$$

- Mean Square Root Error (MSRE)

$$MSRE = \frac{1}{N}\sum_{x=1}^{N}\sqrt{(p_x - \hat{p}_x)^2} \tag{7}$$

- Mean Square Error (MSE)

$$MSE = \frac{1}{N}\sum_{x=1}^{N}(p_x - \hat{p}_x)^2 \tag{8}$$

- Root Mean Square Error (RMSE).

$$RMSE = \sqrt{\frac{1}{N}\sum_{x=1}^{N}(p_x - \hat{p}_x)^2} \tag{9}$$

Where p_{max} and p_{min} are maximum and minimum values of the obtained forecasts. Observing the performance of the proposed forecasting technique will help the researcher to further improve the model by adding new techniques or algorithms within. Also by comparing with other models the gaps between the different models and their specific characteristics can be identified.

VIII. Results and Discussion

For the results and discussion, the fuzzy logic based time series prediction methods are considered since they have been applied in the research of forecasting and prediction in most of the cases. E.Egrioglu [28] proposed hybrid fuzzy time series forecasting model and compared his work along with other existing approaches to prove that the hybrid method is more efficient compared to that of the other fuzzy based times series prediction models. The first comparison is made by using University of Alabama's enrollment data from year 1971 to 1992; error rate is calculated and compared using MSE parameter. This comparison is shown in Table I and Figure I.

TABLE I. COMPARISON OF UNIVERSITY OF ALABAMA DATASET

METHOD NO (M.NO)	METHOD	AUTHORS	MSE
1.	First order time series forecasting model using large fuzzy sets	Q.Song and BS.Chissom [33]	412449
2.	Fuzzy logic relationship are established using tables instead of matrix	C.Chen [34]	407507
3.	Transition matrices were used that are based on the Markov chain to establish fuzzy relationships	J.Sullivan and Woodall [35]	386005
4.	Multi-attribute fuzzy time series approach	C.Cheng et al. [37]	228918
5.	Adaptive expectation model	C. Cheng et al. [36]	192086
6.	Establishment of Fuzzy relation using Fuzzy time series forecasting approach	R.C. Tsaur et al. [38]	134923
7.	First order forecasting model that calculates forecasts using max-min composition	SR.Singh [39]	133700
8.	Higher order forecasting using neural networks	CH. Aladag et al. [40]	78073
9.	Forecasting using optimization of the interval lengths in fuzzy time series.	E.Egrioglu et al[41]	66661
10.	Forecasting using Gustafson–Kessel fuzzy clustering	E.Egrioglu et al[42]	60140
11.	Hybrid invariant fuzzy time series forecasting technique with FCM and PSO	CH.Aladag et al. [32]	46422
12.	Hybrid fuzzy time series forecasting technique based on FCM and NN	E.Egrioglu et al[28]	32849

Another comparison is made by using the TAIEX (Taiwan Stock Exchange) data set for the observations recorded between 01.01.2004 and 31.12.2004. The results are compared based on the evaluation parameter RMSE. The comparison is shown in Table II and Figure II

TABLE II. COMPARISON OF TAIEX DATASET

METHOD NO (M.NO)	METHOD	AUTHORS	RMSE
1.	First order time series forecasting model using large fuzzy sets	Q.Song and BS.Chissom [33]	77.86
2.	Fuzzy logic relationship are established using tables instead of matrix	SM.Chen [34]	77.18
3.	Higher order forecasting using neural networks	CH.Aladag et al. [40]	69.80
4.	Ratio based length of intervals	Huarng and Yu [44]	63.57
5.	A novel multivariate heuristic method	Huarng et al.[45]	72.35
6.	A bi-variate fuzzy time series approach	Yu and Huarng [46]	67
7.	Higher order combined invariant fuzzy time series forecasting technique using FCM and PSO	E.Egrioglu [43]	51.14

In this comparison also the hybrid invariant time series prediction model has less error rate and is said to be more efficient compared to that of the other methods.

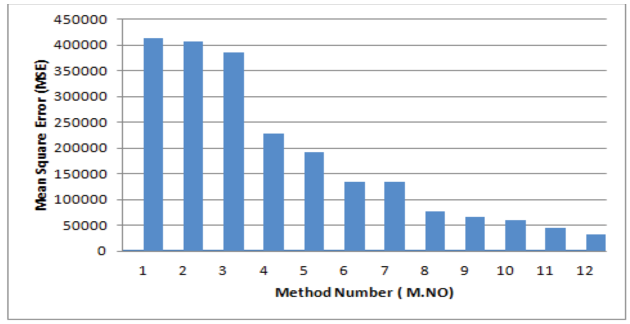

Figure I. Comparison of MSE for University of Alabama dataset

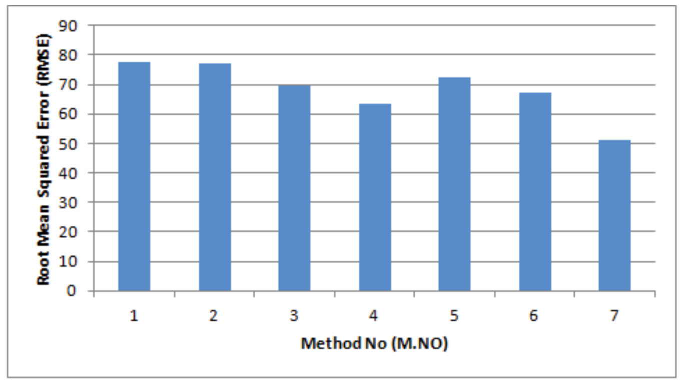

Figure II. Comparison of RMSE for TAIEX dataset

IX. CONCLUSION

A time series data keeps on changing as the time goes on and these changes can be recorded in different time periods such as hourly, daily, weekly, monthly and yearly. By analyzing these changes, the future of the time series can be predicted or forecasted using various forecasting models that implements machine learning algorithms and optimization approaches. This survey paper provides a detail study about the various types of time series data that are available, how they can be analyzed and predicted using the various forecasting models. By comparing and analyzing the forecasting model for the two different datasets such as University of Alabama and TAIEX based on the performance evaluation parameter it is proven that the hybrid forecasting model yields good results compared to other model. Apart from the methods provided in the above survey, many other novel techniques are also available for forecasting unique time series contents. The research area keeps increasing after the introduction of the Big Data and Big Data analytics. So the future work focuses on effectively implementing the forecasting techniques in Big Data analytics.

REFERENCES

[1] B.Thakur, M.Mann, “Data Mining for Big Data: A Review”, International Journal of Advanced Research in computer Science and Software Engineering, Vol.4, Issue 5,pp. 469-473, May 2014

[2] P.Esling, C.Agon, “Time Series Data Mining”, ACM Journal of Computing Surveys, Vol. 45, Issue 1, November 2012.

[3] Xing Yan, Chowdhury; “A Comparison Between SVM and LSSVM in Mid-term Electricity Market Clearing Price Forecasting”, 26th Annual IEEE Canadian Conference of Electrical and Computer Engineering, May 2013.
[4] Gross, G; Galiana, F.D, “Short Term Load Forecasting”, Proceedings of the IEEE International Conference on Load Forecasting,Vol.75, pp.1558-1573, December 1987.
[5] E.Almeshaiei, H.Soltan; “A Methodology for Electric Power Load Forecasting”, ELSEVIER Alexandria Engineering Journal, Vol. 50, Issue 2, pp. 137-144, June 2011.
[6] M. Gan, Y.Cheng, K.Liu, GL.Zhang; “Seasonal and trend time series forecasting based on a quasi-linear autoregressive model”, Applied Soft Computing, Vol. 24, pp. 13-18, November 2014.
[7] TC.Chu, CT.Tsao, YR. Shiue,” Application of fuzzy multiple attribute decision making on company analysis for stock selection”,Proceedings of Soft Computing on Intelligent Systems and Information Processing, pp. 509–514, December 1996.
[8] G.A.N.Mbamalu and M.E.El-Hawary,”Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation,” IEEE Transaction on Power System, Vol.8, Issue 1,pp.343-348, February 1993.
[9] E.H.Barakat,M.A Qayyum,M.N.Hamed and S.A. Al-Rashed,” Short term peak demand forecasting in fast developing utility with inherent dynamic load characteristics,”IEEE Transactions on Power Vol. 5, Issue 3,pp.813-824, August 1990.
[10] I.Moghram, and S.Rahman, “Analysis and evaluation of five short-term load forecasting techniques.”IEEE Transaction on Power System,Vol. 4, Issue 4, pp.1484-1491,November 1989.
[11] Xing Yan, Nurul Chowdhury; “Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and auto-regressive moving average with external input”,Journal of Electrical Power and Energy Systems, Vol. 63, pp. 64-70,December 2014.
[12] Xing Yan, Nurul A Chowdhury; “Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach”, Electrical Power and Energy Systems, Volume 53, pp. 20-26, December 2013.
[13] Xing Yan, Nurul A Chowdhury; “Mid-term Electricity Market Clearing Price Forecasting Using Multiple Least Squares Support Vector Machines”, The Institution of Engineering and Technology, Generation, Transmission & Distribution, Vol.8, Issue 9, pp.1572-1582, September 2014.
[14] G.Rubio, H.Pomares, I.Rojas, LJ. Herrera; “A heuristic method for parameter selection in LS-SVM: Application to time series prediction”, International Journal of Forecasting, Vol. 27, Issue 4, pp. 725-739, 2011.
[15] S.An, W.Liu ,S.Venkatesh; “Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression”, Pattern Recognition, Vol.40, Issue 8, pp. 2154-2162, August 2007.
[16] V.N.Vapnik, “Statistical Learning Theory”, Springer-Verlag, New York,pp:203-223, 1998.
[17] H.Fröhlich, O.Chapelle, B.Schölkopf,” Feature selection for support vector machines by means of genetic algorithms”, Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, pp. 142–148, November 2003.
[18] CL.Huang, CJ.Wang, “A GA-based feature selection and parameters optimization for support vector machines”, Expert Systems with Applications, Vol. 31, Issue 2, pp.231–240, August 2006.
[19] GP. Zhang, M.Qi; “Neural network forecasting for seasonal and trend time series”, European Journal of Operational Research, Vol. 160, Issue 2, pp. 501-514, January 2005.
[20] C.Hamzacebi,“Improving artificial neural networks’ performance in seasonal time series forecasting”, Journal of Information Science, Vol. 178, Issue 23, pp. 4550-4559, December 2008.
[21] TT.Temizel, Matthew C. Casey M; “A comparative study of autoregressive neural network hybrids”, Journal of Neural Networks, Vol. 18, Issue 5-6, pp. 781-789, July 2005.
[22] M.Gan, H.Peng, XP.Dong; “A hybrid algorithm to optimize RBF network architecture and parameters for nonlinear time series prediction”, Applied Mathematical Modelling, Vol. 36, Issue 7, pp. 2911-2919, July 2012.
[23] H.Peng, T.Ozaki , VH.Ozaki V, Y.Toyoda; “A parameter optimization method for radial bias function type models”, IEEE Transactions on Neural Networks, Vol. 14, Issue 2,pp. 432-438, March 2003.
[24] TS. Quah, B.Srinivasan, “Improving returns on stock investment through neural network selection”, Expert Systems with Applications , Vol. 17, Issue 4,pp.295–301,November 1999.
[25] N.Chapados,Y.Bengio,“Costfunctions and model combination for VaR-based asset allocation using neural networks”, IEEE Transactions on Neural Networks Vol. 12, Issue 4, pp.890–906, July 2001.
[26] K.Kim, I.Han, “Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stockprice index”, Expert Systems with Applications”, Vol.19, Issue 2, pp.125–132, August 2000.
[27] Q.Song, RP..Leland, BS. Chissom, “A new fuzzy time-series model of fuzzy number observations”. Fuzzy Sets and Systems, Vol. 73, Issue 3,pp. 341–348, August 1995.
[28] E.Egrioglu, CH.Aladag, U.Yolcu, “Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks”, Expert Systems with Applications, Vol. 40, Issue 3,pp. 854-857, February 2013.
[29] QiSen Cai, Defu Zhang, Bo Wu, Stephen CH Leung; “A novel stock forecasting model based on fuzzy time series and genetic algorithm”, Procedia Computer Science, Vol. 18, pp. 1155-1162, 2013.
[30] E.Bas, VR. Uslu, U.Yolcu, E.Egrioglu; “A modified genetic algorithm for forecasting fuzzy time series”, Application Intelligence, Vol. 41, Issue 2, pp. 453-463, September 2014.
[31] M.Shah; “Fuzzy based trend mapping and forecasting for time series data”, Expert Systems with Applications, Vol. 39, Issue 7, pp. 6351-6358, June 2012.
[32] CH. Aladag, U. Yolcu, E.Egrioglu, Ali Z Dalar; “A new time invariant fuzzy time series forecasting method based on particle swarm optimization”, Applied Soft Computing, Vol. 12, Issue 10, pp. 3291-3299, October 2012.
[33] Q.Song, BS.Chissom, “Forecasting enrollments with fuzzy time series”, Fuzzy Sets and Systems Vol.54, pp. 1–10, 1993.
[34] SM. Chen, “Forecasting enrollments based on fuzzy time-series”, Fuzzy Sets and Systems, Vol. 81, pp. 311–319, 1996.
[35] J. Sullivan, Willam H. Woodall,” A comparison of fuzzy forecasting and Markov modeling”, Fuzzy Sets and Systems Vol.64, Issue 3, pp. 279–293, 1994.
[36] C.Cheng, T.L. Chen, H.J. Teoh, C.H. Chiang, Fuzzy time series based on adaptive expectation model for TAIEX forecasting, Expert Systems with Applications 34, pp. 1126–1132, 2008.
[37] C.Cheng, G.W. Cheng, J.W. Wang, Multi-attribute fuzzy time series method based on fuzzy clustering, Expert Systems with Applications Vol.34, pp. 1235–1242, 2008.
[38] R.C. Tsaur, J.C. Yang, H.F. Wang, Fuzzy relation analysis in fuzzy time series model, Computers and Mathematics with Application Vol.49, pp. 539–548, 2005.
[39] S.R. Singh, A simple method of forecasting based on fuzzy time series, Applied Mathematics and Computation Vol, 186, pp. 330–339, 2007.
[40] CH.Aladag, Basaran, E.Egrioglu,U.Yolcu, VR.Uslu; “ Forecasting in high order fuzzy time series by using neural networks to define fuzzy relations”, Expert Systems with Applications, Vol. 36, pp. 4228-4231, 2009.
[41] Erol Egrioglu, Cagdas Hakan Aladag, Basaran, Vedide Rezan Uslu, Ufuk Yolcu, “A new approach based on the optimization of the length of intervals in fuzzy time series”, Journal of intelligent and fuzzy systems, Vol. 22, pp. 15-19, 2011.
[42] E.Egrioglu, CH.Aladag, U.Yolcu, VR.Uslu, “Fuzzy time series forecasting method based on Gustafson-Kessel fuzzy clustering” Expert Systems with Applications, Vol.38, pp. 10355-10357, 2011.
[43] E. Egrioglu, “PSO based higher order time invariant fuzzy time series method:Application to stock exchange data”,Economic Modelling,Vol.38, pp.633-639,2014.
[44] Huarng, K., Yu, T.H.-K., “Ratio-based lengths of intervals to improve fuzzy time series forecasting”. IEEE Transaction System. Man Cybern. B Cybern. Vol.36, pp.328–340,2006
[45] Huarng, K., Yu, T.H.K., Hsu, Y.W., “A multivariate heuristic model for fuzzy time series forecasting”, IEEE Transaction System. Man Cybern. B Vol. 37, pp.836–846, 2007.
[46] Yu, T.H.K., Huarng, K.H., “A bivariate fuzzy time series model to forecast the TAIEX”,Expert System Application. Vol.34, pp.2945–2952, 2008.