

Time Series Forecasting Models: A Comprehensive Review

Devyani Rawat, Vijay Singh, Shiv Ashish Dhondiyal, Sumeshwar Singh

Abstract: This comprehensive review provides an extensive overview of the existing Time Series Forecasting technique. This survey is not restricted to any single time series analysis; it provides forecasting of time series in different areas like marketing prediction, weather forecasting, technology prediction, financial forecasting etc. In this paper, we have analyzed forecasting in some areas namely, load forecasting, wind speed forecasting, prediction of energy consumption and short-term traffic flow prediction. Various models are available for prediction among them Autoregressive Integrated Moving Average model (ARIMA) is seen as a universal mechanism, these discussed forecasting areas utilizes different models that are combined with ARIMA. Hybrid models are the combination of classical models and modern methods, like ARIMA (classical method) combines with Artificial Neural Network (ANN) as well as with Support Vector Machine (SVM) (modern models). Hybrid model's performance is depending on the variety of data that are taken for forecasting.

Keywords: ARIMA, ANN, SVM, Time series, ARMA.

I. INTRODUCTION

The forecasting of time series is an essential field of predicting the future data by observing the relationship among past data. The use of time series analysis extracts useful data information along with other important data features. This approach is beneficial when a very small knowledge is available and the data is available in the form of non-stationary data for an example the data of weather, retails, electricity consumption, stock price and many more. A number of researchers have contributed in this field to design a time series forecasting model in different area [1].

The conducted survey in this paper covers the existing work that used different techniques such as machine learning to forecast the data in different fields. The major aim of this paper is to analyze the opportunities or the best approach for future work. The basic steps that are required to design a time series forecasting model are shown in Figure 1

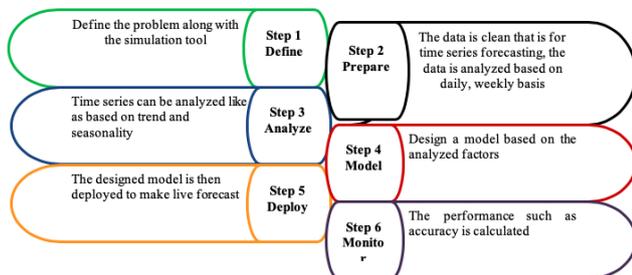


Figure 1: Basic Steps for Time Series Forecast Model

Revised Manuscript Received on September 25, 2019.

Devyani Rawat, Department of computer science and Engineering, Graphic Era Deemed to be University, Dehradun, India.

E-mail: devyani.rawat10@gmail.com

Vijay Singh, Department of computer science and Engineering, Graphic Era Deemed to be University, Dehradun, India.

Shiv Ashish Dhondiyal, Department of computer science and Engineering, Graphic Era Deemed to be University, Dehradun, India.

Sumeshwar Singh, Department of computer science and Engineering, Graphic Era Hill University, Dehradun, India.

II. TECHNIQUE USED

This section deals with the most commonly used approaches to forecasting time series. The ARIMA model is among the most frequently used techniques. This technique gain popularity due to its statistical characteristics along with the Box-Jenkins mechanism used for the designing of forecasting model [2]. The ARIMA model has been used to develop a number of time series modeling models such as exchange rate data forecasting, energy consumption forecasting and carbon price forecasting [3].

Besides, ARIMA models may implement different exponential smoothing models. Although the model is capable of representing several different time series, namely Autoregressive (AR), Moving averages (MA), and combined AR and MA (ARMA) series, their main limitation is the linear form of the previously thought-out model. In other words, a linear correlation structure is assumed between the time series values and therefore, the ARIMA model cannot catch any nonlinear patterns. To approximate the linear models into complex real-world problem is sometimes not satisfactory [4].

Jenkins developed the ARIMA model by integrating two time series approaches for instance; AR and MA and ARIMA. Mathematically can be defined as

$$z_t = K + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \dots + \theta_m z_{t-m} + G_n - \beta_1 e_{t-1} - \beta_2 e_{t-2} - \dots - \beta_q e_{t-q}$$

Where,

$\theta_1, \theta_2, \theta_3, \dots, \theta_m$ are the model parameters.

$K \rightarrow$ Constant.

$G_n \rightarrow$ The white Gaussian noise.

The values obtained from equation (1) represent the ARIMA predicted values. If the value of $n=0$, then ARIMA model becomes an AR- model with order of m and if the value of $m=0$ is an MA model or an order n [5].

A. ANN in time series modeling

ANN is an effective and competitive alternative to ARIMA models for predicting the link between time series and distinctive features. This approach helps to make the designed model more consistent as well greatly accurate. A single hidden layer ANN with single output is used in this research [6].

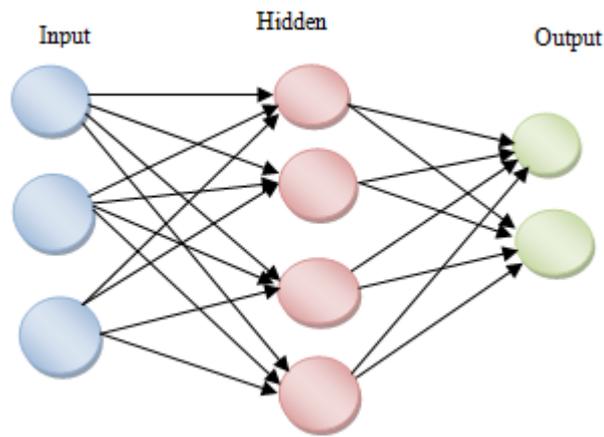


Figure 2: Layered structure of ANN

The output of ANN can be defined as

$$Z_t = \phi_0 + \sum_{j=1}^n \phi_{0j} h(\varphi_{0j} + \sum_{i=1}^m \varphi_{ij} Z_{t-i}) + \epsilon_t$$

Where,

$\phi_j \rightarrow (j=0,1,2,3,4,\dots,n)$

$\varphi_{ij} \rightarrow (i=0,1,2,3,4,\dots,m)$, both are the weight,

ϕ_0 and $\phi_{0j} \rightarrow$ Bias value

$\epsilon_t \rightarrow$ White Noise

$h \rightarrow$ hidden layer activation function of hidden layer [7]

Cadenas, E. & Rivera, W. [4, 2010], developed a Hybrid model ARIMA and ANN to forecast wind speed and compare this model with both models separately on the basis of mean square error (MSE), mean error (ME) and mean absolute error (MAE). The hybrid model produces significantly lower MSE value than ARIMA and ANN models by which clearly shows that proposed model predicted with high accuracy. The produced MSE value corresponds to ARIMA, ANN and proposed model in metre per second is correspondingly 0.6457, 0.7687 and 0.0089.

Kumar, S. V. & Vanajakshi, L. [5, 2015], proposed seasonal-ARIMA (S-ARIMA) model to forecast short-term traffic flow on less data, the data may be historical or, real-time short-term traffic flow forecasting on two peak periods such as morning and evening. The computed parameters i.e. mean absolute percentage error (MAPE) among observed and predicted morning and evening peak have been 4.37 and 3.83 per hour correspondingly. The obtained outcome showed that the proposed model perform better in predicting traffic flow in short-term, when real-time data has considered.

Khashei, M. & Bijari, M. [6, 2010], developed a unique ANN hybrid approach using ARIMA to achieve a higher performance scheme compared to ANN. This proposed model has provided more effective results than Zhang's hybrid model and both ARIMA as well as ANN models has been utilized independently over three different time periods, i.e. one month, six months and 12 months with both error measurements. The improvement in terms of MAE in this work over the Zhang hybrid model, ANN, along with ARIMA is 17.42%, 12.68%, and 20.98 % corresponds to the proposed research and 8.944 % is obtained.

B. SVM

SVM works based on the concept of decision planes, which defines the boundaries of the available problem. In Figure 3

categorized data is separated by the decision plane. This figure composed of two data form groups identified by colors green and red. The data type is separated by the hyperplane. From the figure 3, the red and green color nodes are located in the separating plane's right and left side. If any new data enters into the system and it is towards the green side then it comes under green data category otherwise come under red color category.

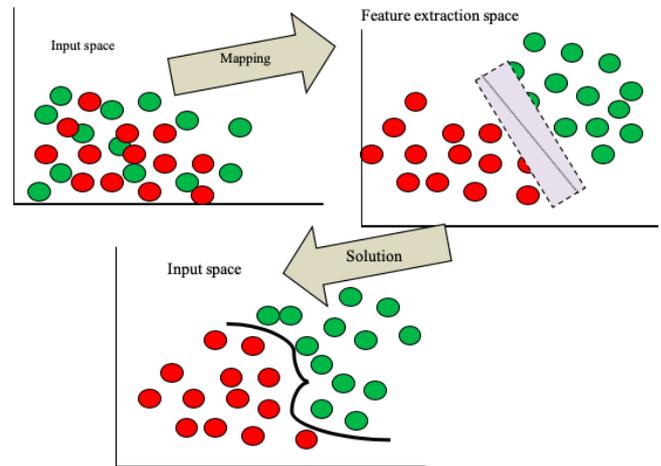


Figure 3: SVM work flow

Wang et al. [7, 2017], presented a hybrid GARCH to analyze time series data with probably conditional Heteroskedasticity characteristics and to forecast volatility of PM_{2.5} concentrations. The parameter MAE computed for the model ARIMA, SVM and GARCH among the actual and expected value is 6.07, 5.84 and 5.53. Because of the SVM and ARIMA based hybrid model, MAE values decrease by 8.98 % and 5.23 respectively.

Karthika et al. [8, 2017], deployed hybrid model with ARIMA-SVM, based on past results, ARIMA is utilized for forecasting the load by using past information. SVM is used to improve the accuracy of the sensitive load components named as weekly temperature and time delay. The percentage error (PE) is used with the deviation approach to find and correct outliers. It must be noted that the hybrid approach minimizes MAPE's value to 4.15 % and has improved its convergence speed.

Xuemei et al. [9, 2010], presented methods of forecasting named as ANN, SVM, ARIMA to forecast the building cooling load. A hybrid model with Seasonal-ARIMA (SARIMA) along with SVM has been utilized here and compares the performance of this hybrid model with SARIMA and SVM based on relative mean errors (RME), root mean squared errors (RMSE) and mean absolute relative error (MARE). The hybrid model provides the greater overall average MAPE obtained is 5.6% for air conditioning load forecasting.

Wang et al. [11, 2012], demonstrated ARIMA and SVR in order to predict financial time series, the text data used here for prediction. The performance is measure on the basis of MAE, MAPE and RMSE. Performance of this present model is contrast with two taken models individually and hybrid with SVM.



The obtained MAPE for ARIMA (Model I), ARIMA-SVM (Model II) and proposed work is 26.45%, 30.79% and 23.06% by which it has to be concluded that the error value is reduced for proposed model by which accuracy is improved.

Authors Name	Proposed work	Techniques used	Parameters	Obtained value (%)
Khashei, M., & Bijari, M. [6, 2010]	Time series Forecasting	ARIMA and ANN	MAE	8.99
Wang et al. [7, 2017]	concentrations forecasting	ARIMA and SVM	MAE	5.53
Karthika et al. [8, 2017]	short term load forecasting	ARIMA and SVM	MAPE	4.15
Xuemei et al. [9, 2010]	building cooling prediction	ARIMA and SVM	RME, RMSE, MARE and MAPE	5.6
Wang et al. [11, 2012]	financial time series forecasting	ARIMA and SVR	MAE, MAPE and RMSE	23.06

In given table 1 discussed the work done previously for time series forecasting on different areas like prediction of load forecasting, prediction of cooling in building, stock price prediction and so on. The performance to predict the future values in given time frame such as sometimes years has been specified from upcoming year's duration, Ediger, V. Ş., & Akar, S. (2007). By which authors have to predict the time series, the main aim of represent this previous work in tabular form is to analyze the performance of different techniques

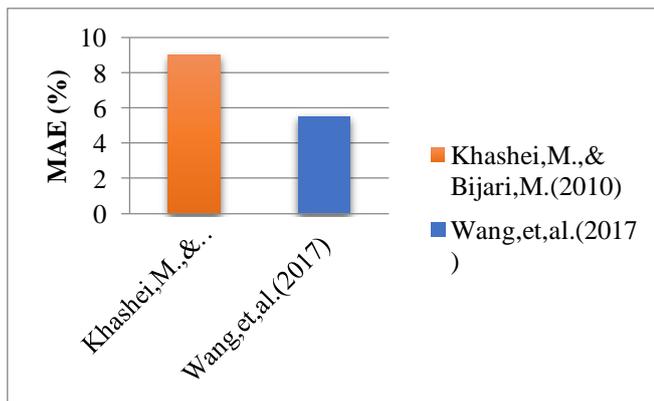


Figure 4: MAE (%)

In figure 4 depicts the performance measurement of time series forecasting using ARIMA and ANN in Khashei, M., & Bijari, M. (2010) based on mean absolute error. The lower error value tends to be enhanced performance, as per the above figure the graph plotted of MAE respect to author's name. The representation of parameter provided along x-axis and corresponds to y-axis plotted author's name. From plotted graph the performance of wang et al. (2017) is better due to less MAE value; here the hybrid technique ARIMA with SVM has been utilized

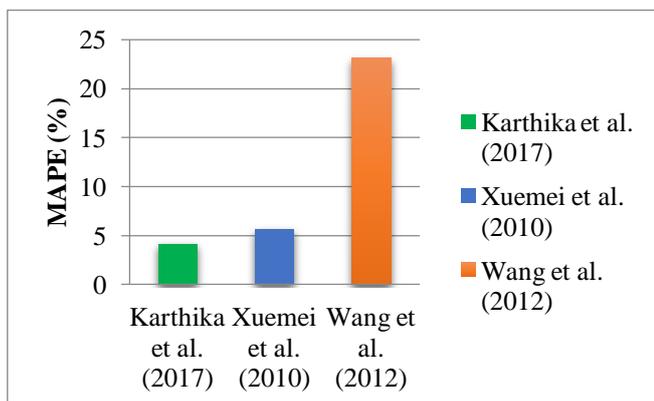


Figure 5: MAPE (%)

In figure 5, the previous work has been evaluated on the basis of mean absolute percent error (MAPE) Karthika et al. (2017) and Xuemei et al. (2010) used hybrid technique ARIMA along with SVM, Wang et al. (2012) have done own work by utilizing ARIMA and SVR (Support vector regression) and obtained enhanced result as per the graph plotted. In which distinct work has been represented through different colors such as green, blue and red line in the form of column graph. The name of authors with published year and obtained MAPE value in percentage has been plotted corresponds to x-axis and y-axis respectively.

III. CONCLUSION

Time series analysis consists of methods to examine the specific data to derive meaningful stats and other features of the data. The forecasting of time is the use of a model to predict future values on the basis of previously observed values. In this review work, different time series forecasting models has been discussed and it has also analyzed that sometimes it's not good to compare the models to analyze the performance. Because it always depends on the types of data i.e. the patterns and features of data, the hybrid model ARIMA-ANN or, ARIMA with SVM provides better accuracy. The increase of accuracy in hybrid models is due to the combination of linear and nonlinear models. The parameter for time series forecast is taken as MAPE, RMSE, MARE etc are used in the previous work.

REFERENCES

- Zhang, G. P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175, 2003.
- Nichiforov, C., Stamatescu, I., Făgărășan, I., & Stamatescu, G. Energy consumption forecasting using ARIMA and neural network models. 5th ISEEE (pp. 1-4). IEEE, October 2017.
- Zhu, B., & Chevallier, J., Carbon price forecasting with a hybrid Arima and least squares support vector machines methodology. In *Pricing and Forecasting Carbon Markets* (pp. 87-107). Springer, Cham., 2017.
- Cadenas, E., & Rivera, W., Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model. *Renewable Energy*, 35(12), 2732-2738, 2010.
- Kumar, S. V., & Vanajakshi, L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review*, 7(3), 21, 2015.
- Khashei, M., & Bijari, M., An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*, 37(1), 479-489, 2010.
- Wang, P., Zhang, H., Qin, Z., & Zhang, G. A novel hybrid-Garch model based on ARIMA and SVM for PM2. 5 concentrations forecasting. *Atmospheric Pollution Research*, 8(5), 850-860, 2017.
- Karthika, S., Margaret, V., & Balaraman, K., Hybrid short term load forecasting using ARIMA-SVM. *i-PACT* (pp. 1-7). IEEE, April 2017.
- Xuemei, L., Lixing, D., Yuyuan, D., & Lanlan, L., Hybrid support vector machine and ARIMA model in building cooling prediction. *International Symposium on Computer, Communication, Control and Automation (3CA)* (Vol. 1, pp. 533-536). IEEE, May 2010.
- Bedi, J., & Toshniwal, D., Deep learning framework to forecast electricity demand. *Applied energy*, 238, 1312-1326, 2019.
- Wang, B., Huang, H., & Wang, X., A novel text mining approach to financial time series forecasting. *Neurocomputing*, vol. 83, pp. 136-145, 2012.
- Ediger, V. Ş., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy policy*, 35(3), 1701-1708.