Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 8, August 2016, pg.104 – 109

A SURVEY ON ARIMA FORECASTING USING TIME SERIES MODEL

Z. Asha Farhath[1], B. Arputhamary [2], Dr. L. Arockiam[3]

[1] Research Scholar, Department of Computer Science, Bishop Heber College, Tiruchirappalli, India
[2]Assistant Professor, Department of Computer Applications, Bishop Heber College, Tiruchirappalli, India
[3]Associate Professor, Department of Computer Science, St. Joseph's College, Tiruchirappalli, India
[1] asha.zakir@yahoo.com, [2] arputhambaskaran@rediffmail.com, [3] larockiam@gmail.com

Abstract: Time series modeling and forecasting has a fundamental importance to numerous realistic domain names. Therefore a whole lot of lively research works goes on in this subject at some point of several years. Many vital models were proposed in literature for enhancing the accuracy and efficiency of time collection modeling and forecasting. The intention of this paper is to expect the forecast the use of arima version gift a concise description of some famous time series forecasting models used in exercise, with their salient functions. Whilst fitting a model to a dataset, special care is taken to pick the maximum parsimonious one to evaluate forecast accuracy as well as to evaluate among distinct fashions suited to a time series models. ARIMA stands for [Autoregressive Integrated Moving Average Models]. ARIMA is actually a forecasting technique that projects the future values of a series based entirely on its own inertia. ARIMA is a superior to exponential smoothing techniques when the data is reasonably long and the correlation between past observations is stable. The following is a literature survey study on the various prediction models proposed, modeled and suggested by various authors earlier.

Keywords: Linear Forecasting, Prediction Analysis, Time Series, Auto Regression, ARIMA

I. INTRODUCTION

Time series analysis is a main part in statistics, which analyzes data set to study the characteristics of the data and it predicting future values of the time series based on the characteristics. A time series contain records of a single variable is termed as univariate. But if records of more than one variable are considered so it is termed as multivariate. A time series can be continuous or discrete. In a continuous time series observations are measured at every instance of time, whereas a discrete time series contains observations measured at discrete points of time. For example temperature readings, flow of a river, concentration of a chemical process etc. It can be recorded as a continuous time series. On the other hand population of a particular city, production of a company, exchange rates

between two different currencies may represent a discrete time series. Usually in a discrete time series the consecutive observations are record at evenly spaced time interval such as hourly, daily, weekly, monthly or yearly time separations. The main aim of time series model is to collect and carefully study the past observations of a time series to develop an appropriate model which describes the natural structure of the series. This model is then used to generate future values for the series to make forecasts. Time series forecasting can be termed as the act of predicting the future by understanding the past. Due to the indispensable importance of time series forecasting in numerous practical fields such as economics, finance, science and engineering, etc. The proper care should be taken to fit an adequate model to the underlying time series. It is a successful time series forecasting depends on an appropriate model fitting. A lot of efforts have been done by the researchers over many years for the development of efficient models to improve the forecasting accuracy. Time series prediction has a critical role within the region of choice-making. Analyses are made from past observations to form a mathematically the instant series, to try to get upcoming values, with a suitable margin of blunders by way of the choice maker. The time series prediction models are implemented, while a touch understanding approximately the method of collection technology. Examples of typical methods of forecasting are container and Jenkins and Artificial Neural Network (ANN), the primary is capable of version information which have linear and desk bound characteristics, the second one is carried out for complex data relation to conquer boundaries of the standard forecasting models and to growth the accuracy of the predictions, hybrid fashions were used. These styles of forecasters are composed of diverse fashions that can specific linear and nonlinear sample, and grow to be capable of modeling any kind of information. Whereas the actual data sometimes has each characteristics, and this makes using the standard methods unsatisfactory final results.

II. RELATED WORKS

Kalid Yunus et al [1] presents this paper, a modified auto regressive integrated moving average (ARIMA) modeling technique which could capture time correlation and possibility distribution of determined wind-pace time-collection records is offered. The technique introduces frequency decomposition (splitting the wind-speed information into high frequency (HF) low-frequency (LF) components), shifting, and limiting further to differencing and energy transformation that are used within the trendy ARIMA modeling system.

A. Vaccaro et al [2] the paper proposes hybrid architecture for electricity price forecasting. The proposed architecture combines the reward of the easy-to-use and comparatively simple to- tune Auto regressive Integrated Moving Average (ARIMA) model and the approximation power of local learning techniques. The architecture is robust and more accurate than the individual forecasting methodologies on which it is based, since it combines a reliable built-in linear model (ARIMA) with an adaptive dynamic corrector (Lazy Learning algorithm). The corrector model is sequentially updated, in order to adjust the whole architecture to varying market conditions. Detailed simulation studies show the effectiveness of the proposed hybrid learning methods for forecasting the volatile Hourly Ontario Energy Prices (HOEPs) of the Ontario, Canada, and electricity market.

Guoqiang Liu et al [3] in the software program reliability boom section, the character of the failure records in an experience, decided by means of the software testing method. A hybrid version is proposed for medium and lengthy-term software program failure time forecasting on this paper. The hybrid version consists of two techniques, Singular Spectrum Analysis (SSA) and ARIMA on this version, the time series of software failure time are firstly decomposed into numerous sub-series corresponding to some tendentious and oscillation (periodic or quasi-periodic) components and noise by using the usage of SSA after which every sub-series is expected, respectively, through the best ARIMA version, and lastly a correction system is conducted for the sum of the prediction results to ensure the residual to be a natural random collection.

Takaomi HIRATA et al [4] Time series records examine and prediction may be very essential to the observe of nonlinear phenomenon. Studies of time series prediction have protected records when you consider that last century, linear models along with ARIMA, and nonlinear models consisting of multi-layer perception (MLP) are well-known. Because the nation-of-art approach, a deep belief net (DBN) using a couple of Restricted Boltzmann machines (RBMs) changed into proposed these days. In this have a look at advise a novel prediction method which composes not simplest a form of DBN with RBM and MLP however additionally ARIMA. Prediction experiments for the time collection of the actual records and chaotic time collection were performed, and effects confirmed the effectiveness of the proposed method.

Ling Wang et al [5] presents Based totally on the evaluation of the measured put on records and the wear characteristics in terms the wheels of Guangzhou Metro Line1, the accumulative put on prediction approach of metro wheels based totally on the ARIMA (p, d, q) model is proposed in this paper. According to the time series modeling method of the ARIMA (p, d, q) model, the stationarity evaluation and transformation of the metro wheel

wear records are described at first. Then, with the application of the AIC criterion and the most likelihood Estimation method, the model order is determined and the version parameters are derived for the ARIMA (p, d, q) model in the end, the flange thickness and the diameter of the metro wheels will be anticipated by means of this advanced ARIMA(p, d, q) model. The effects show that the proposed prediction method is straightforward and effective for the quick-term prediction of the metro wheel.

Theresa Hoang Diem Ngo et al [6] a time series is a put of ethics of a exacting variable that happen over a age of time in a positive pattern. The most ordinary patterns are rising or falling trend, cycle, seasonality, and uneven fluctuations. To replica a time sequence event as an occupation of its history values, analysts recognize the outline with the supposition that the pattern motivation keep on in the future. Applying the Box-Jenkins style, this paper emphasizes how to recognize a fitting time series replica by identical behaviors of the taster autocorrelation purpose (ACF) and fractional autocorrelation purpose (PACF) to the academic autocorrelation functions. In addition to model recognition, the paper examines the meaning of the stricture estimate, checks the diagnostics, and validates the forecasts.

Eliete Nascimento Pereira et al [7] present to pick up time series forecasts the wavelet decompose has been applied. The mixture of forecasting method as the Autoregressive incorporated Moving Average (ARIMA) and false Neural Networks have been use to realize a superior merit time series forecasting than. This paper proposed a hybrid model collected of wavelet decompose, ARIMA and neural network Multilayer Perception. These models are mutual linearly then yielding the time sequence forecasting. The sequence deliberate is the Wolf's sunspots and the British pound/US dollar swap in excess of rate data. The contrast of the projected model in this paper with writing indicates an efficient way to advance forecasting.

W. Jacobs et al [8] presents This study aims to predict the standards of the time series of milk insist in a dairy industry by combining forecasting of Box-Jenkins and artificial neural network replica and contrast the results to the individual models, exemplifying the joint forecast for the production planning. Eight prediction combining technique were used and, behind the use of statistical methods, the consequences obtain by fitting the Box- Jenkins and artificial neural network template were contrast with the results obtained in the future combinations. The results showed that the mixture of seasonal Box-Jenkins and deseasonalize artificial neural network model by the inverse denote square method, provide a performance in the forecast for six months in front 66.5% higher than the individual models, where the combination of forecasts provided a root mean square error of 1.43 and mean absolute percentage error of 2.16. The forecast for 12 months ahead, the presentation of the mixture was 56.5% higher compare to entity models, with root mean square error of 2.86 and mean absolute percentage error of 3.70%. In both cases, the combination of predictions showed superior results.

Ratnadip Adhikari et al [9] presents development of time series forecasting correctness through combining numerous models is an important as well as a dynamic area of research. As a result, a variety of forecasts combination methods have been urban in literature. However, most of them are based on straightforward linear ensemble strategies and therefore ignore the possible relationships between two or more participate models. In this paper, propose a robust weighted nonlinear group technique which considers the entity forecasts from different models as well as the correlation among them while combining. The planned ensemble is constructing using three well-known forecasting models and is experienced for three real-world time series. A comparison is complete among the proposed scheme and three other widely used linear combination method, in terms of the obtain forecast errors. This contrast shows that our ensemble scheme provide significantly lesser forecast errors than each entity model as well as each of the four linear combination methods.

Yi-Shian Lee et al [10] which is a conventional numerical method, is employed in many fields to build models for forecasting time series. Although ARIMA can be adopting to obtain a highly accurate linear forecasting model, it cannot exactly forecast non linear time series. Artificial neural network (ANN) can be use to build more accurate forecasting model than ARIMA for nonlinear time series, but explanation the meaning of the hidden layers of ANN is hard and, moreover, it does not yield a mathematical equation. This revision proposes a mixture forecasting model for nonlinear time series by combine ARIMA with genetic programming (GP) to improve upon both the ANN and the ARIMA forecasting model. Finally, some real data sets are adopted to show the efficiency of the proposed forecasting model.

III. FORECASTING METHODS AND MODELS

The forecast procedure provides a quick and regular way to generate the forecasts for many time series in one step. The procedure had forecast hundreds of series at a time with the series organized into separate variables or across by groups. Forecast uses extrapolative forecasting methods where the forecasts for a series are function only of time and past values of the series, not of additional variables. Forecasting is the system of create predictions of

the future based totally on past and present information and study of trends. A common example is probably opinion of a few variable of hobby at some separate future date. Prediction is a comparable, however extra accepted time period. Utilization can vary between regions of utility for instance, in hydrology the conditions "forecasting" time to time kept for evaluation of values at positive exact future times, whilst the term "prediction" is used for enhanced well-known estimates, along with the amount of times floods will arise over a protracted duration.

A. ARTIFICIAL NEURAL NETWORK

Artificial neural networks (ANNs) approach has been recommended as an alternative technique to time series forecasting and it gain immense popularity in the last few years. The basic objective of ANNs is to build a model for mimicking the intelligence of human brain into machine. Similar to the work of a human brain, it will try to identify regularities and patterns in the input data learn from knowledge and then provide general results based on their known previous knowledge. Although the development of ANNs was mainly biologically motivated but afterwards they have been applied in many different areas, especially for forecasting and classification purposes. First, ANNs are data-driven and self-adaptive in nature. There is no need to specify a particular model form or to make any a priori statement about the statistical distribution of the data the desired model is adaptively formed and it based on the features presented from the data.

B. TIME SERIES FORECASTING USING SUPPORT VECTOR MACHINE

Initially SVMs were considered to solve pattern classification problems such as optimal character recognition, face identification and text classification, etc. But soon they found wide applications in other domains such as function approximation, regression estimation and time series prediction problems. SVM techniques are based on the Structural Risk Minimization rule. The objective of SVM is to find a decision rule with good generalization capability through selecting some particular subset of training data called support vectors. In this method, an best possible separating hyper plane is constructed and after nonlinearly mapping the input space into a higher dimensional feature space. Thus, the quality and complexity of SVM solution does not depends directly on the input space. Another important characteristic of SVM is that the training process is equivalent to solving a linearly inhibited quadratic programming problem. In contrary to other networks training, the SVM solution is always exclusive and globally optimal. However a major difficulty of SVM is that when the training size is large, it requires an enormous amount of computation which increases the time complexity of the solution.

C. STATIONARITY AND DIFFERENCING MODELS

A motionless time series is one whose property does not depend on the time at which the sequence is observed. So time series with trend, or with seasonality, are not stationary and seasonality will have a result on the value of the time series at dissimilar times. On the other hand, a white noise sequence is stationary it does not substance when you observe it should look much the same at any stage of time. In general, a stationary time series will have no unsurprising patterns in the long-term. Time plots will show the series to be roughly level with constant variance.

D. AUTOREGRESSIVE MODELS

In a multiple regression model estimate the variable of attention using a linear combination of predictors. In an auto regression model guess the variable of notice using a linear combination of past ideals of the variable. The term auto regression specifies that it is a regression of the variable against itself. Autoregressive models are remarkably flexible at handling a broad range of different time series patterns. Moving standard models Rather than use past values of the forecast changeable in a regression, a moving average model uses history forecast errors in a regression model. A moving average model is second-hand for forecasting prospect values while moving average smoothing is used for approximating the trend-cycle of past values.

E. MOVING AVERAGE MODELS

Rather than use present values of the forecast variable in a regression, a moving average model use past forecast errors in a regression-like model.

$Yt = c+et+\theta 1et-1+\theta 2et-2+\cdots+\theta qet-q$, $yt = c+et+\theta 1et-1+\theta 2et-2+\cdots+\theta qet-q$

It refers to this as an MA **(**q q**)** replica. Of course, do not observe the values of etet, so it is not actually regression in the usual sense. Notice that each value of ytyt can be consideration of as a weighted affecting average of the past few forecast errors. But moving average models should not be mystified with moving average smoothing. A rousing average model is used for forecasting vision values while moving average smoothing is used for estimate the trend-cycle of past values.

F. Non-Seasonal ARIMA MODELS

ARIMA model is also called Box-Jenkins model that may possibly include autoregressive terms, moving average terms, and differencing operations. It is differencing with an auto regression and a moving average model. ARIMA is a short form for Autoregressive Integrated Moving Average model ("integration" in this background is the reverse of differencing). When a model only involved autoregressive it may be referred as an AR model. When a model only involves moving average terms, it may be referred to as an MA model.

IV. ARIMA MODEL

In common models for time series data can have many forms and stand for different stochastic processes. There are two commonly used linear time series models in literature, viz. Autoregressive (AR) and Moving Average (MA) models. Combining these two, the Autoregressive Moving Average (ARMA) and Auto regressive Integrated Moving Average (ARIMA) models have been proposed in literature. The Autoregressive Fractionally Integrated Moving Average (ARFIMA) model generalizes ARMA and ARIMA models. For seasonal time series forecasting, a variation of ARIMA, viz. the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used. ARIMA model and its different variations are based on the well-known Box-Jenkins principle and so these are also broadly known as the Box-Jenkins models. An ARIMA (p, q, d) model is a combination of AR (p), I (d) and MA (q) models and is suitable for univariate time series modeling. In an AR (p) model the future value of a variable is assumed to be a linear combination of p past observations and a random error mutually with a constant term. Mathematically the AR (p) model can be expressed in,

$$Y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \in_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \in_t$$

Here $y_t \ and \ \in_t$ are respectively the actual value and random error (or random shock) at time period t) ϕi (i = 1, 2,..., p) are model parameters and c is a constant. The integer constant p is known as the order of the model. Sometimes the constant term is not there for simplicity. Generally, the random shocks are assumed to follow the typical normal distribution. Thus conceptually a moving average model is a linear regression of the current observation of the time series aligned with the random shocks of one or more prior observations. Fitting an MA model to a time series is more difficult than fitting an AR model because in the previous one the random error terms are not fore-seeable. Once describing different time series models, the next issue to our concern is how to select an appropriate model that can produce accurate forecast based on explanation of historical pattern in the data and how to determine the optimal model instructions. Statisticians George Box and Gwilym Jenkins developed a realistic approach to build ARIMA model, which best fit to a given time series and also assure the parsimony principle. Their concept has fundamental importance on the area of time series analysis and forecasting. The Box-Jenkins methodology does not assume any particular pattern in the past data of the series to be forecasted. It uses a three step iterative approach of model identification, parameter estimation and diagnostic checking to verify the best parsimonious model from a general class of ARIMA models. This three-step process is repeated more than a few times until a satisfactory model is finally selected. Then this model can be used for future forecasting values of the time series.

V. CONCULSION

ARIMA is a trendy method to analyze stationary univariate time series data. There are generally three main stages to build an ARIMA model, with model identification, model estimation and model checking, of which model classification is the most crucial stage in building ARIMA models. Thus the survey provides an insight into the various time series prediction and forecasting models with reference to ARIMA. Also a lot of real world

applications conducted by the various persons were studied and it has come to prove that ARIMA is a real world toll for time series prediction, forecasting and analysis with accuracy.

REFERENCES

[1] W. Jacobs, A. M. Souza and R. R. Zanini – "*Combination of Box-Jenkins and MLP/RNA Models for Forecasting Combining Forecasting of Box-Jenkins*", IEEE LATIN AMERICA TRANSACTIONS, 14(4), 2016.

[2] S. I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou and A. G. Bakirtzis, "*Comparison of SARIMAX, SARIMA, Modified SARIMA and ANN-based Models for short-term PV generation forecasting*", IEEE International Energy Conference (ENERGYCON), Leuven 1-6, 2016.

[3] Rodrigo N. Calheiros ; Dept. of Comput. & Inf. Syst., Univ. of Melbourne, Melbourne, VIC, Australia; Enayat Masoumi; Rajiv Ranjan; Rajkumar Buyya, "*Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications*", QoS IEEE Transactions on Cloud Computing, 3(4), 2015.

[4] Takaomi HIRATA, Takashi KUREMOTO, Masanao OBAYASHI, Shingo MABU, "*Time Series Prediction Using DBN and ARIMA*", International Conference on Computer Application Technologies, IEEE, 2015.

[5] G. Liu, D. Zhang and T. Zhang, "*Software Reliability Forecasting: Singular Spectrum Analysis and ARIMA Hybrid Model*", Theoretical Aspects of Software Engineering (TASE), 2015 International Symposium on *Nanjing*, 111-118, 2015.

[6] Sornpon Wichaidit, Surin Kittitornkun, "*Predicting SET50 Stock Prices Using CARIMA*", International Computer Science and Engineering Conference (ICSEC), 2015.

[7] A. Vaccaro, T. H. M. EL-Fouly, C. A. Cañizares and K. Bhattacharya, "*Local Learning-ARIMA Adaptive Hybrid Architecture for Hourly Electricity Price Forecasting*", PowerTech, 2015 IEEE Eindhoven, 1-6, 2015.

[8] Kalid Yunus, Torbjorn Thiringer, and Peiyuan Chen, "*ARIMA-Based Frequency-Decomposed Modeling of Wind Speed Time Series*", IEEE Transaction on Power Systems, 31(4), 2546-2556, 2016.

[9] S. Li, H. Wang, Y. Tian, Y. Shen and A. Aitouche, "*Wind Speed Forecasting Based on Fuzzy-Neural Network Combination Method*", The 27th Chinese Control and Decision Conference (2015 CCDC), *Qingdao*, 4811-4816, 2015.

[10] E. Guresen, G. Kayakutlu, T. U. Daim, "*Using Artificial Neural Network Models in Stock Market Prediction*", Expert Systems with Applications, 38, 10389-10397, 2011.

[11] T. M. Choi, Y. Yu, K. F. Au, "*A Hybrid SARIMA Wavelet Transform Method for Sales Forecasting*", Decision Support Systems, 51, 130-140, 2011.

[12] Y. S. Lee, L. I. Tong, "*Forecasting Time Series Using a Methodology Based on Autoregressive Integrated Moving Average and Genetic Programming*", Knowledge-Based Systems, 24, 66-72, 2011.

[13] J. H. Stock, M. Watson, "*Combination Forecasts of Output Growth in a Seven-Country Dataset*", Journal of Forecasting, 23, 405-430, 2004.

