



A LITERATURE SURVEY OF TIME SERIES FORECASTING APPROACHES

Smitkumar Arvindbhai Patel, 2023

Abstract: This literature review offers a comprehensive analysis of time series forecasting techniques. It explores traditional methods such as autoregressive integrated moving average (ARIMA) and exponential smoothing, focusing on their strengths and limitations. The review discusses challenges in forecasting, including model selection and uncertainty estimation. It also highlights emerging trends in the field, such as the integration of external factors and ensemble methods. This review provides valuable insights for researchers and practitioners seeking a deeper understanding of time series forecasting and its potential applications.

I. INTRODUCTION TO TIME SERIES FORECASTING

The narrative method used in this review of literature will help readers understand the topic of the study. The collection of such observations is known as time series data. A time series is a group of observations, each of which was recorded at a certain time. Data are analyzed to derive statistical information, data features, and output predictions. Time series analysis and related methods have made prediction easier since time series data may have a tendency to follow a pattern, which makes it challenging for the Machine Learning model to forecast accurately. Former researchers' techniques and the outcomes they attained through them will be outlined, including several time series techniques and a comparison of them.

II. WHAT IS TIME SERIES ANALYSIS

Time series analysis involves fitting a suitable model to a given time series by estimating its parameters based on the available data. This process, termed as time series analysis, aims to understand the nature of the series and is valuable for future forecasting and simulation (Adhikari & Agrawal, 2013).

Time series forecasting, a subset of time series analysis, involves developing mathematical models based on past observations to predict future events. This approach is particularly useful when there is limited knowledge about the statistical patterns in the data or when an explanatory model is lacking. Accurate forecasting is crucial as it influences strategic decisions and precautionary measures. Researchers, such as Ratnadip Adhikari and R. K. Agrawal, have made significant efforts in developing and improving time series forecasting models over the past several decades (Adhikari & Agrawal, 2013).

The trend, seasonality, and variability of the time series data are the major three things that are discovered through visualization and analysis.

Trend: A time series' trend is a long-term pattern or direction that shows an underlying, gradual shift over time. Forecasting and comprehending the overall behavior of the series requires a comprehension of and modeling for the trend component (Cryer & Chan, 2008).

Seasonality: Patterns within a time series that repeat at regular intervals are referred to as seasonality; these patterns are frequently connected to cyclical or recurring events. For reliable forecasting and the detection of cyclic patterns, seasonality must be detected and taken into account (Cryer & Chan, 2008).

Heteroscedasticity: Indicators of heteroscedasticity include varied degrees of volatility or dispersion in the data across time. It implies that the series' variability does not remain constant throughout time. For accurate statistical inference and model performance, it's critical to address heteroscedasticity (Cryer & Chan, 2008).

Few approaches perform poorly in forecasting when the data is seasonal, and even fewer perform poorly when the data has trends. As a result, trends, seasonality, and heteroscedasticity must be addressed while determining the optimal statistical method for predicting.

III. VARIOUS METHODS USED IN TIME SERIES FORECASTING AND COMPARING FEW METHODS

There are few Classical methods for time series forecasting.

1. Autoregression (AR)
2. Moving Average (MA)
3. Autoregressive Moving Average (ARMA)
4. Autoregressive Integrated Moving average (ARIMA)
5. Seasonal Autoregressive Integrated Moving Average (SARIMA)
6. Vector Autoregression (VAR)
7. Vector Autoregression Moving Average (VARMA)
8. Simple Exponential Smoothing (SES)
9. Holt's Winter Exponential Smoothing (HWES)

To anticipate future output, all approaches consider trend, seasonality, or heteroscedasticity. Time series data needs to be deconstructed based on the results of data analysis. Data must be classified as trend or seasonality based on the outcomes of the investigation.

A. Stationarity :

Time series analysis relies heavily on the concept of stationarity. A stationary series has statistical characteristics that include mean, variance, covariance, and standard deviation that do not change over time. In a nutshell, stationary patterns in Time Series refers to the absence of a Pattern or Periodic component(Manuca, R. and Savit, R., 1996).

(A. Witt, J Kurths and A. Pikovsky) suggest a method for verifying time series stationarity through the combination of an experiment of time independence using an one-dimensional probability density with any of the spectrum densities. They apply their test approach to a variety of numerically produced time series, ranging from basic linear stochastic procedures to high-dimensional transient chaos, as well as observational information from geophysics and physiology.

Manuca and Savit present a novel family of methods for testing, modeling, and describing non-stationary processes, whereas Witt, Kurths, and Pikovsky suggest a method for verifying stationary behavior through integrating an experiment for time independence using a one-dimensional probability density with a measurement of the spectral density.

B. Exponential Smoothing models :

Time-series data is based on the idea that an observation at one point in time is dependent on previous findings in time (Cryer, J.D., 1986). Previous observations are weighted as they contribute to future prediction. weighing is done using an equation called 'Theta' (Cryer, J.D., 1986). To discover the best possible figure for theta, we have to calculate the total number of the squared deviations between the actual and predicted values of the preceding observation. We can anticipate the next value using this procedure, but it does not help much to predicting multiple values because the prediction will be the same as the prior value.

A few concepts, such as stationarity and differencing, must be mastered in order to comprehend the approaches and assess distinct models. Both of these notions aid in the interpretation of the approaches' key principles.

C. Differencing :

The notion of differencing is used to render trending and seasonal data stationary. In the method of differencing, the current observation is subtracted from the preceding observation. This aids in keeping the mean constant. Dickey, D.A., and Pantula, S.G. (1987) authored a study on how to deal with non-stationary behavior in time series via computing the initial variations and applying a model to the distinct dataset unless the differentiated series also appears non-stationary.

D. Autoregressive models (AR) :

An autoregressive (AR) model is a sort of linear time series model in which the current value of a variable is represented as a linear mixture of its historical values. The standard equation describing an AR(p) model, where p is the model's order, is as follows:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

Where,

y_t is the current value of the variable,
 c is a constant,
 $\phi_1, \phi_2, \dots, \phi_p$ are the coefficients of the model, and
 ϵ_t is white noise.

The number of previous values included in the model is indicated by the model's order, p (Cryer, J.D., 1986).

AR model of order one is an algorithm that relies exclusively on one lag in the past (Shibata, R., 1976). Autoregressive models are sometimes known as long memory models since they must remember all lags up until the first start point and determine their value. If there was a shock incidence in the past that caused data fluctuations, it will have an influence on the present value, making the model very susceptible to shocks (Shibata, R., 1976).

E. Moving Average (MA) :

A moving average is an approach for smoothing out time series data by computing the average of a predetermined number of previous observations. Moving averages are classified into two types: simple moving average (SMA) and exponential moving average (EMA).

The formula for an n-period SMA is:

$$SMA_t = \frac{1}{n} \sum_{i=0}^{n-1} y_{t-i}$$

Where, y_t is the current value of the variable.

The formula for an n-period EMA is:

$$EMA_t = \alpha y_t + (1 - \alpha) EMA_{t-1}$$

Where, α is the smoothing factor, calculated as $\alpha = \frac{2}{n+1}$.

F. Autoregressive Moving Average (ARMA) model

The ARMA model is a hybrid of the AR and MA models. When the AR model of order one approaches the beginning point, the equation will have an infinite moving average (Choi, B., 2012). In the ARMA model, p and q must be specified, where p is the number of important terms in the ACF and q is the number of significant terms in the PACF.

To determine the optimal value for p and q there are two ways:

1. Plotting patterns in correlation
2. Automatic selection techniques

1. Plotting patterns in correlation

When plotting patterns in correlation, two functions are used:

- a) Auto correlation factor (ACF): The correlation between observations at the present time stamp and observations at the preceding time stamp (Hagan, M.T., and S.M. Behr, 1987).
- b) Partial auto correlation factor (PACF): the relationship between two separate time stamps, presuming that both data points are connected to another time stamp (Hagan, M.T., and Behr, S.M., 1987).

2. Automatic selection techniques

There are three generally used ways for selecting a time series model automatically:

- a) Minimum information criteria (MINIC): This generates different model combinations based on a grid search of AR and MA phrases. It then identifies the model with the lowest Bayesian information criterion (Stadnytska, T., Braun, S., and Werner, J., 2008).
- b) Squared canonical correlations (SCAN): This method examines the data's correlation matrix and compares it to its lags. It then examines the eigenvalues of the correlation matrix in order to determine a mixture of AR and MA that has SCAN as 0. It chooses the pair with the quickest convergence (Stadnytska, T., Braun, S., and Werner, J., 2008).
- c) The expanded sample auto correlation function (ESACF): AR and MA are known to be connected. It essentially filters out the AR phrases until just the MA portion remains. This procedure is continued until the fewest AR terms and the most MA terms remain (Stadnytska, T., Braun, S., and Werner, J., 2008).

It is entirely up to the person to select any of the strategies for determining the ideal value of p and q for improved model performance.

G. Autoregressive Integrated Moving Average (ARMA) :

To comprehend the ARIMA model, we must first comprehend the ARMA model, as this is only an extension of the ARMA model. Essentially, we need to keep data immobile in order to provide it to a machine learning model. It is accomplished by differentiating. ARIMA models are theoretically stated as $ARIMA(p,d,q)$, where p and q are the same as ARMA models, but d is the number of initial differences (Yu, G., and Zhang, C., 2004, May).

H. Seasonal Autoregressive Integrated Moving Average (SARMA) :

SARIMA models were introduced to handle seasonality in the data. Seasonality is different from stationarity; however, seasonality can be handled using stationarity up to some extent, but seasonal correlations cannot be eliminated completely. SARIMA models are mathematically written as $SARIMA(p,d,q)(P,D,Q)s$, where P = Number of seasonal AR terms, D = Number of seasonal differences, Q = Number of seasonal MA terms and s = Length of the season. Removing seasonality will help the model to perform better but getting rid of seasonality in data is a difficult task to do.

I. Comparing ARIMA method with SARIMA method :

SARIMA models beat ARIMA in long-term runoff forecasts in the United States, according to Valipour (2015). They were, however, vulnerable to parameter changes, resulting in poor performance. Wang, Li, and Lim (2019) found ARIMA and

SARIMA models less effective for Linear System Analysis, Spectra Analysis, and Digital Filtering. When picking a method, Wang and Li underlined the necessity of data analysis, with ARIMA being better suited for trend data and SARIMA for seasonality data.

IV. ADVANTAGES AND DISADVANTAGES OF TIME SERIES FORECASTING:

Advantages:

- Forecasting Trends and Patterns: One of time series analysis' most major advantages is its ability to foresee future trends and patterns. We can construct models that properly forecast future trends and patterns by studying past data sets.
- Identifying Hidden Dependencies: Time series analysis can assist in identifying hidden dependencies between variables.

Disadvantages:

- One of the fundamental disadvantages of time series analysis is its limited application area.
- Dealing with Outliers and Noise: Handling with outliers and noise in data can be difficult when utilizing time series analysis.

V. CONCLUSION

In this article, we examined time series forecasting, quickly defined the methodologies utilized, compared several approaches to see which one was superior, and highlighted the benefits and drawbacks of time series forecasting. When comparing AR with MA, ARIMA and SARIMA produced similar results based on the use case and data. As a result, it is reasonable to argue that each model is unique to its use case and the data used in model construction; methodologies should be chosen depending on data. Time series forecasting takes into account a variety of natural elements such as trends, seasons, and so on. It is expected that new algorithms that can manage both seasonality and trends will be developed in the next years. However, this might be difficult job and should be undertaken by researchers in this field.

REFERENCES

1. An Introductory study on Time series Modeling and Forecasting by Ratnadip Adhikari and R. K. Agarwal.
2. Hagan, M.T. and Behr, S.M., 1987. The time series approach to short term load forecasting. IEEE transactions on power systems, 2(3), pp.785-791
3. Choi, B., 2012. ARMA model identification. Springer Science & Business Media.
4. Cryer, J.D., 1986. Time series analysis (Vol. 286). Boston: Duxbury Press.
5. Dickey, D.A. and Pantula, S.G., 1987. Determining the order of differencing in autoregressive processes. Journal of Business & Economic Statistics, 5(4), pp.455-461.
6. Manuca, R. and Savit, R., 1996. Stationarity and nonstationarity in time series analysis. Physica D: Nonlinear Phenomena, 99(2-3), pp.134-161.
7. Witt, A., Kurths, J. and Pikovsky, A., 1998. Testing stationarity in time series. physical Review E, 58(2), p.1800
8. Shibata, R., 1976. Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika, 63(1), pp.117-126.
9. Stadnytska, T., Braun, S. and Werner, J., 2008. Comparison of automated procedures for ARMA model identification. Behavior research methods, 40(1), pp.250-262.

10. Yu, G. and Zhang, C., 2004, May. Switching ARIMA model-based forecasting for traffic flow. In 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 2, pp. ii-429). IEEE.
11. 14. Valipour, M., 2015. Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications*, 22(3), pp.592-598
12. 15. Wang, S., Li, C. and Lim, A., 2019. Why are the ARIMA and SARIMA not sufficient. arXiv preprint arXiv:1904.07632.
13. Time series analysis With Applications in R by Johnthan D. Cryer and kung-Sik Chan. Springer Second Edition.

