

Tennessee State University

Digital Scholarship @ Tennessee State University

Computer Science Faculty Research

Department of Computer Science

4-19-2019

Forecasting the Number of Monthly Active Facebook and Twitter Worldwide Users Using ARMA Model

Qasem Abu Al-Haija

Tennessee State University

Qian Mao

Tennessee State University

Kamal Al Nasr

University of Texas at San Antonio

Follow this and additional works at: <https://digitalscholarship.tnstate.edu/computerscience>



Part of the [Computer Sciences Commons](#), and the [Data Science Commons](#)

Recommended Citation

Abu Al-Haija, Q., Mao, Q. & Al Nasr, K. (2019). Forecasting the Number of Monthly Active Facebook and Twitter Worldwide Users Using ARMA Model. *Journal of Computer Science*, 15(4), 499-510.
<https://doi.org/10.3844/jcssp.2019.499.510>

This Article is brought to you for free and open access by the Department of Computer Science at Digital Scholarship @ Tennessee State University. It has been accepted for inclusion in Computer Science Faculty Research by an authorized administrator of Digital Scholarship @ Tennessee State University. For more information, please contact XGE@Tnstate.edu.

Forecasting the Number of Monthly Active Facebook and Twitter Worldwide Users Using ARMA Model

¹Qasem Abu Al-Haija, ²Qian Mao and ³Kamal Al Nasr

¹Department of Computer Information and Systems Engineering, Tennessee State University, Nashville, USA

²Department of Electrical and Computer Engineering, Tennessee State University, Nashville, USA

³Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

Article history

Received: 30-01-2019

Revised: 22-03-2019

Accepted: 18-04-2019

Corresponding Author:

Qasem Abu Al-Haija
Department of Computer
Information and Systems
Engineering, Tennessee State
University, Nashville, USA
Email: qabualha@my.tnstate.edu

Abstract: In this study, an Auto-Regressive Moving Average (ARMA) Model with optimal order has been developed to estimate and forecast the short term future numbers of the monthly active Facebook and Twitter worldwide users. In order to pickup the optimal estimation order, we analyzed the model order vs. the corresponding model error in terms of final prediction error. The simulation results showed that the optimal model order to estimate the given Facebook and Twitter time series are $ARMA[5, 5]$ and $ARMA[3, 3]$, respectively, since they correspond to the minimum acceptable prediction error values. Besides, the optimal models recorded a high-level of estimation accuracy with fit percents of 98.8% and 96.5% for Facebook and Twitter time series, respectively. Eventually, the developed framework can be used accurately to estimate the spectrum for any linear time series.

Keywords: Time Series, ARMA Model, Signal Estimation, Signal Prediction, Final Prediction Error, Facebook, Twitter

Introduction

Time in general is a crucial factor to organize and coordinate many real life phenomena and can be used to ensure the success of systems and business. With the time, various metrics are measured over regular time intervals, making a time series such as sunspot activity, weather data, stock prices, industry forecasts and many others.

A times series can be defined as an ordered sequence of values of a variable at equally spaced time intervals (NIST/SEMATEC, 2018). Time intervals can be of any scale including seconds, minutes, hours, months and years. For example, the earthquake time series (USGS, 2018) counts the average number of quakes every year as illustrated in Fig. 1 which shows the yearly quake numbers span from 1919 to 2018 (100 years).

Its a common practice to use time series to understand the underlying forces and structure that produced the observed data (NIST/SEMATEC, 2018) as well as to fit a model and proceed to forecasting, monitoring or even feedback and feed-forward control. Therefore, time series analysis comprises several methods or services to process the time eries such as: data compression, phenomena explanation such seasonal factors (temperature, humidity, pollution, pressure), signal processing such as signal description, signal

classification, signal transformation and signal prediction (i.e., using a specific model to predict future values of the time series).

Time series phenomena can be classified as either discrete or continuous, deterministic or stochastic, linear or nonlinear (Imdadullah, 2014). Also, time series signals can be modeled using different modeling techniques. Recently, the modern theory of digital signal processing (Proakis and Manolakis, 2007) offers several linear processing models such as the parametric models (Proakis and Manolakis, 2007) including: Auto-Regressive models $AR(p)$, Moving Average models $MA(q)$ and Auto-Regressive Moving Average models: $ARMA(p,q)$. These models can be used to regenerate (describe) the time series past values and then, accordingly, forecast (predict) the future values.

The idea of signal prediction or time series forecasting is illustrated in Fig. 2 which basically applies the past values using some prediction model (e.g., ARMA Model) to regenerate the signal and then predict the short-term future values.

In this study, we employ the Auto-Regressive Moving Average (ARMA) Model to regenerate and analyze the time eries given by active users on Facebook and Twitter social networks. The time series is representing the number of monthly active users

worldwide. Facebook time series is for the last 10 years (from the end of 2008 to 2018) while Twitter time series is for 9 years (from 2010 to 2018). This can be accomplished by using the optimal modeling order that minimizes the estimation error and maximize the fit percent. It should be noted that while the concept of ARMA modeling has been used to model and predict the time series of several systems and phenomena, to the best of our knowledge, this is the first contribution at hand to forecast the number of monthly active Facebook and Twitter worldwide users using ARMA Model. Specifically, the main contributions of this paper can be summarized as follows:

- We develop an ARMA model for the collected Facebook and Twitter time series that can maintain

optimal degree of ARMA modeling with minimum modeling error to optimize the signal estimation to forecast the system for a given time period.

- We employ the optimal derived ARMA model to re-generate the time series of the measured data and predict the short-term future values of possible global numbers of monthly active users for both aforementioned time signals to the end of year 2019 (i.e., the 12 months of the year are divided into 4 quarters of estimation)
- We provide simulation plots for the original collected signal along with the forecasted signal with analysis to gain insight into the developed model and the solution technique

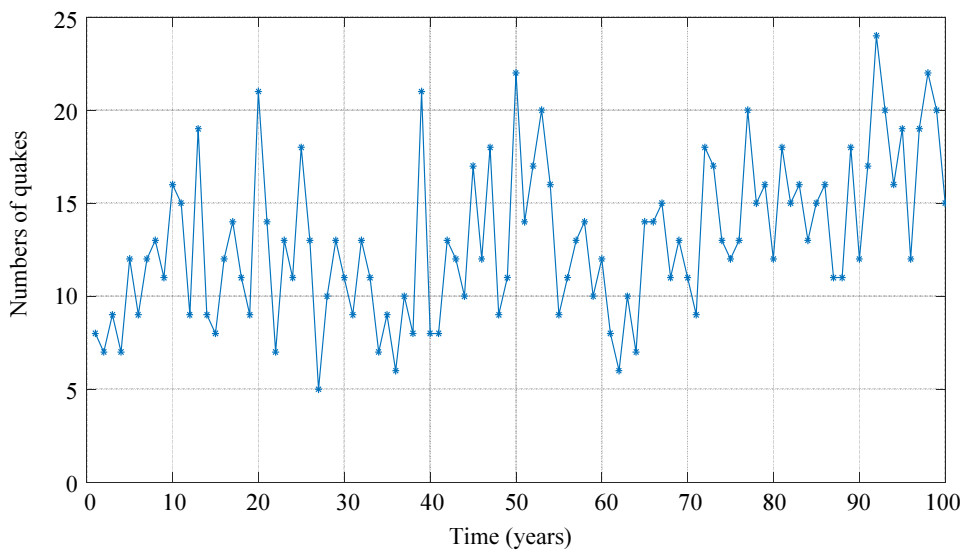


Fig. 1: Time series for yearly averaged quake numbers 1919-2018

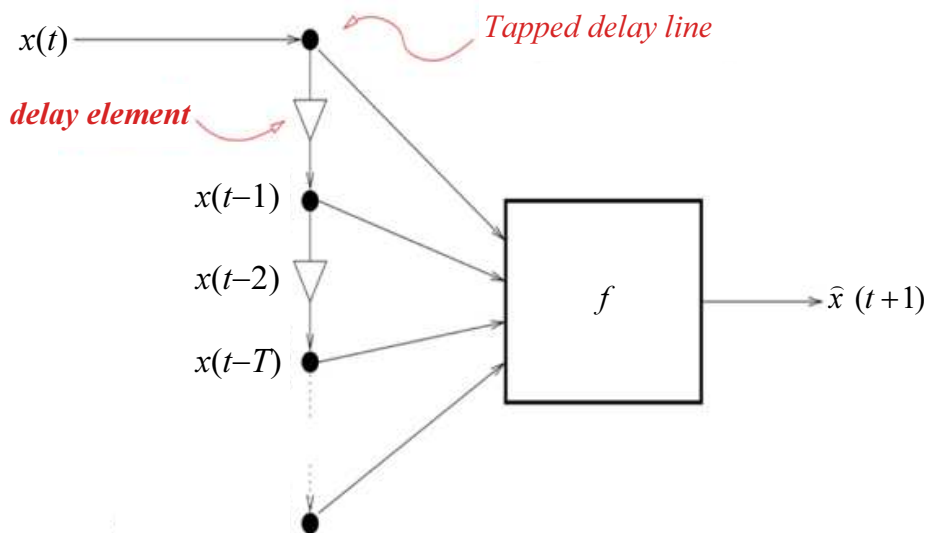


Fig. 2: Signal forecasting principle: Using the past to predict the future

The rest of this paper is organized as follows. Section II reviews the related research in the area. Section III describes the system modeling using ARMA technique with its parametric computation and signal processing. Section IV provides the problem formulation and proposed solution approach for derivation of proper model order and minimization of system prediction error. Section V presents and discusses the simulation results and comparisons by considering several scenarios. Finally, Section VI concludes the paper.

Related Work

Recently, ARMA model has been widely used for prediction purposes of various time series applications such as the works conducted in (Liu *et al.*, 2015; Ratnam *et al.*, 2015; Liu and Shao, 2016; Gankevich and Degtyarev, 2018; Karimpour *et al.*, 2017) For instance, Liu *et al.* (2015) employed the ARMA model to analyze the impact of climate change (air temperature and precipitation) on the stream-flow that originated from mountain glaciers Urumqi River basin area, China. To generate the proper model, they collected their time series of the monthly air temperature and precipitation over a period of 48 years. As a result of their ARMA model analysis, they concluded that the amount of runoff is increasing every 10 years by $1 m^3/s$ due to the climate change especially for those related to the precipitation which contributed more than the air temperature to the stream-flow recording more effective regression coefficient of 0.163.

Another noticeable work is the ARMA model developed in (Ratnam *et al.*, 2015) to forecast the short term future values of the Vertical Total Electron Content (VTEC) considering the dual frequency GPS (Global Positioning System) receiver at KL University, India. Since TEC is the basic representation of ionosphere, its considered as an accountable factor that contribute to the range error of GPS system. However, the results of their forecasting model for VTEC time series proved that ARMA model would be useful to set up an early warning system of ionospheric disturbances.

Moreover, Liu and Shao (2018) developed an ARMA model for the time series of India's weekly tea auction price for the years 2013 to 2014. They primarily used the model to predict the tea price of the last week in 2014 and for the first two weeks in 2015. Hence, because of the efficiency of the developed ARMA model, the authors were able to recommend that China as well needs to establish a sensible tea auction market as it is considered one of the biggest exporter of tea production.

Eventually, ARMA Model guided them to set an early warning mechanism for the tea prices to guide the production of tea cultivation and sales activities.

Furthermore, Gankevich and Degtyarev (2018) proposed an ARMA Model to generate simulated sea waves of arbitrary amplitudes to analyze the impact of

external excitation on a ship hull. They compared their proposed model with model based on linear wave theory. Their simulation results showed that the sea waves simulation-based ARMA model is superior since it provided higher performance and accuracy for both shallow and deep water cases. The simulation results were verified against the ones of real sea surface for benchmarking purposes. Finally, Karimpour *et al.* (2017) proposed an online traffic flow time series to enhance the response time against the traffic congestion. To do so, they developed an ARMA model using the continuous real time data of the traffic flow for a certain intersection to predict the upcoming traffic condition. Accordingly, their experimental results demonstrated that the model was able to predict the traffic flow with 88.74% for 15 minutes ahead.

In this study, we are employing ARMA Model to regenerate, analyze and forecast the short term upcoming spectrum for the number of monthly active Facebook users worldwide as of 2008 to 2018 (in millions) and the number of monthly active Twitter users worldwide from 2010 to 2018 (in millions). To enhance the signal processing for ARMA Model, an error modeling technique has been implemented by plotting the model order vs. order error to pickup the optimal order number to be used for the ARMA modeling. However, all signal processing mechanisms such as signal generation, estimation and error modeling techniques were developed via MATLAB. Thus, ARMA Modeling, estimation, prediction, results and analysis are reported in this paper.

System Modeling Via ARMA (p,q)

Auto-regressive Moving Average (ARMA) model is a parametric method for spectrum estimation (Proakis and Manolakis, 2007). It is used to provide a linear framework to the approximation of the signal dynamics over time and to predict (forecast) the short-term future behavior based on past behavior. The prediction is conducted based on the past behavior of the signal by employing linear regression techniques on the current time series data against one or more past values in the same series. The realization structure of ARMA model is depicted in Fig. 3 where: $y(n)$ is the target model signal to be described and predicted by ARMA Model, $e(n)$ is a prediction noise (error) and $A(q^{-1})$ and $C(q^{-1})$ are polynomial coefficients for the ARMA Model. According to Fig. 3, ARMA Model is expressed as:

$$\begin{aligned} y(n) &= (1 - A(q^{-1}))y(n) + C(q^{-1})e(n) \\ \Rightarrow y(n) &= y(n)A(q^{-1}) + C(q^{-1})e(n) \\ 0 &= -A(q^{-1})y(n) + C(q^{-1})e(n) \\ \Rightarrow A(q^{-1})y(n) &= C(q^{-1})e(n) \end{aligned}$$

where, $A(q^{-1})$ and $C(q^{-1})$ are used as a time shift for $y(n)$ and $e(n)$ respectively, Thus:

⇒ First order ARMA [1,1] Model is derived as:

$$y(n) = a_1 y(n-1) + c_1 e(n-1) + e(n)$$

⇒ Second order ARMA[2,2] Model is derived as:

$$y(n) = -a_1 y(n-1) + c_1 e(n-1) + e(n)$$

Thus, the generalized ARMA [p,q] model of any order ARMA[p,q] can be derived as follows:

$$y(n) = -\sum_{i=1}^p a_i y(n-i) + \sum_{i=1}^q c_i e(n-i) + e(n) \text{ or}$$

$$y(n) = \hat{y}(n) + e(n) \text{ and}$$

$$e(n) = -\sum_{i=1}^p a_i y(n-i) - \sum_{i=1}^q c_i e(n-i) + e(n) \text{ or}$$

$$e(n) = -\sum_{i=1}^p a_i y(n-i) - \sum_{i=1}^q c_i e(n-i), c_0 = 1$$

where, $y(n)$ is the measured data, $\hat{y}(n)$ is the predicted data and $e(n)$ is the prediction error (noise variance). For optimal estimation, we solve the minimization problem of cost factor $J(n)$ as follows:

$$\min_{ac} J(n) = \min_{ac} E \{ e(n)^2 \} \tag{1}$$

As a result, the solution of this minimization criterion will result in a model to compute the ARMA parameters as follows:

$$-R_{yy}(j) = -\sum_{i=1}^p a_i R_{yy}(i-j) - \sum_{i=1}^q c_i R_{ey}(i-j) \text{ For } j = 1, 2, 3, \dots, p$$

$$R_{ye}(j) = -\sum_{i=1}^p a_i R_{ye}(i-j) + \sum_{i=1}^q c_i R_{ee}(i-j) \text{ For } j = 1, 2, 3, \dots, p$$

Given that:

$$R_{yy}(k) = \sum_{n=0}^{N-1} y(n)y(n-k)$$

$$R_{ye}(k) = \sum_{n=0}^{N-1} y(n)e(n-k)$$

Where:

$$R_{ye} = R_{ey}$$

R_{yy} = Auto-correlation Function (ACF) of signal y

R_{ye} = Cross-correlation Function (CCF) of signals y and e

R_{ee} = Prediction error noise of unit variance (white noise)

Recall the 2nd order equation for ARMA Model ($p = q = 2$). Here, we need to find the model parameters a_1, a_2, c_1, c_2 and then 2nd ARMA Model can be developed as:

$$y(n) = -a_1 y(n-1) - a_2 y(n-2) + c_1 e(n-1) + c_2 e(n-2) + e(n)$$

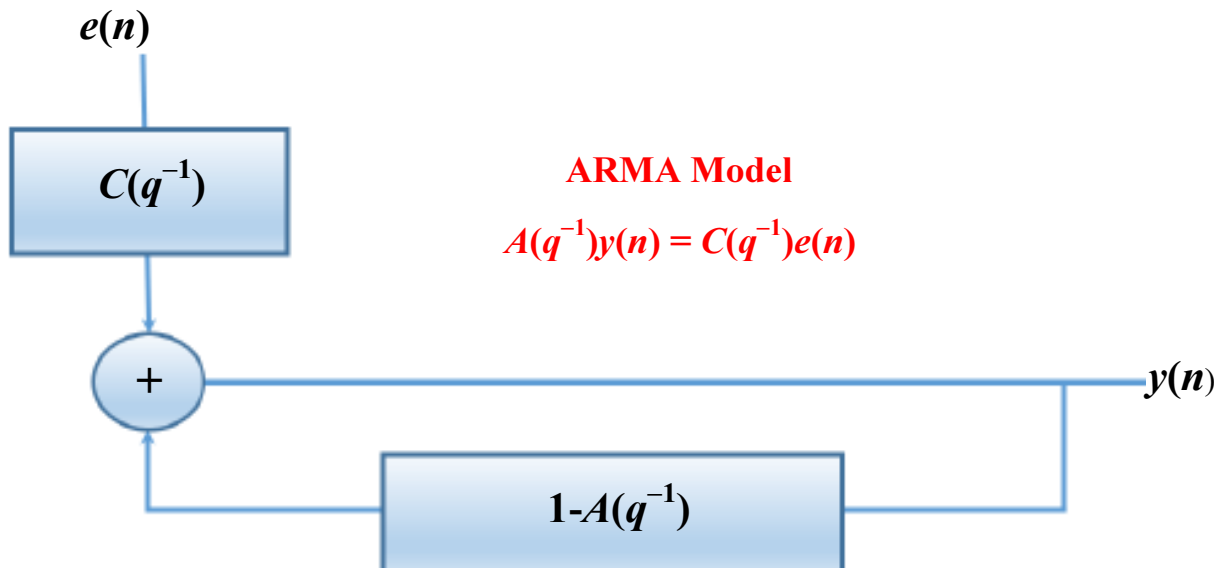


Fig. 3: General organization of ARMA model

To find 2^{nd} parameters, we need to solve the following system of equations:

$$\begin{bmatrix} -R_{yy}(1) \\ -R_{yy}(2) \\ R_{yy}(1) \\ R_{yy}(1) \end{bmatrix} = \begin{bmatrix} R_{yy}(0) & R_{yy}(1) & -R_{ey}(0) & -R_{ey}(1) \\ R_{yy}(-1) & R_{yy}(0) & -R_{ey}(-1) & R_{ey}(0) \\ -R_{ey}(0) & -R_{ey}(1) & -R_{ee}(0) & R_{ee}(1) \\ -R_{ey}(-1) & -R_{ey}(0) & R_{ee}(-1) & R_{ee}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \end{bmatrix}$$

In addition to ARMA modeling development, we can find the transfer function $H_{ARMA}(Z)$ and power spectrum density $SARMA(Z)$ of ARMA model as follows:

$$H_{AR}(Z) = \frac{C(Z)}{A(Z)} = \frac{1 + c_1 Z^{-1} + c_2 Z^{-2} + \dots + c_{Nc} Z^{-Nc}}{1 + a_1 Z^{-1} + a_2 Z^{-2} + \dots + a_{Na} Z^{-Na}}$$

$$S_{AR}(Z) = [H_{AR}(Z)]^2 R_{ee} \Delta T = H(Z)_{AR} H(Z)_{AR}^* R_{ee} \Delta T$$

where, ΔT is the sampling interval, $Z = e^{j\omega}$ and R_{ee} is error/noise variance (Proakis and Manolakis, 2007). However, the power spectral density can be calculated in various ways such as:

- Power Spectrum using Periodogram method (Welch, 1967): Periodogram is an estimation of the spectral density of a signal. It computes the power spectra for the entire input signal as follows:

$$Periodogram = \frac{|abs(F(signal))|^2}{N} \quad (2)$$

where, $F(signal)$ is the Fast Fourier Transform (FFT) of the signal and N is the normalization factor.

- Power Spectrum using Auto-correlation method (Proakis and Manolakis, 2007): Auto-correlation function measures the similarity of a signal with a delayed version of itself as given previously and it defines signal energy: $E = R_{yy}(0)$. The power spectrum can be found by employing the auto-correlation and then the Fast Fourier transform (FFT).

Problem Formulation and Methodology

In order to develop a successful estimation model, its has been recommended by (Hanke and Wichern, 2008) to have a minimum of $2xS$ of data points where S is the seasonal period such as for monthly data $S = 12$ and thus for 50 data points would be $50 = 12 = 4$ years of data. However, it also depends on the regularity of the data. If the seasonal pattern is quite regular then 3 years is OK for such a case. In this paper, the seasonal period is given quarterly (i.e., $S = 4$) with almost 9 to 10 years of data (i.e., 36 to 40 data points) of regular seasonal patterns and smooth tendency figures. Therefore, we were able to

derive the proper ARMA model (i.e., parameters) to re-generate the time series signal for the number of monthly users involved in the social networking of Facebook and Twitter. The developed model has been implemented using MATLAB computing platform. However, the work development phases can be described as follows.

Collecting and Preparing Time Series for Modeling

This phase is about creating a time series signal such as the data generated by the computer speaker or finding/downloading any published time series data. However, we have collected two data sets from STATISTA portal (Statista, 2018) which are the number of monthly active Facebook users worldwide as of 2008-2018 (in millions) and the number of monthly active Twitter users worldwide from 2010 to 2018 (in millions). Then, the collected data set files are converted into a spreadsheet with two columns (time and value) to be effectively imported into the MATLAB for ARMA modeling.

Visualizing The Time Series Signal

This phase is about plotting the original measured data before estimating it with ARMA modeling to see how it behaves throughout the duration of time and how the values are distributed over the plane. Figure 4 and 5 show the plots of this phase which illustrates the original measured data sets for both target time series. The plot for Facebook time series (Fig. 4) illustrates the data sets collected for 10 years discretized in 40 quarters starting from the 3rd of year 2008 up to the 3rd quarter of year 2018, whereas the plot for Twitter time series (Fig. 5) illustrates the data sets collected for 10 years discretized in 35 quarters starting from the 1st quarter of year 2010 up to the 3rd of year 2018 (Q^1 of 2010 to Q^3 of 2018). According to the plots, it is expected that the estimation model of Facebook time series is to be more precise as its data set includes more data items (i.e., 40 vs 35) and they tend to have more linearity over the data set items measured for twitter time series.

Providing an Arbitrary Model Order Method

This phase is about configuring the program to be of general type so that user can enter the choice of model order number. Accordingly, the implemented simulation will respond with the analysis corresponding to the signals for the user-selected model order. Analyzing the estimation errors for model orders This phase is about generating and plotting the ARMA model estimation errors vs. model orders in order to pickup the optimal model order number that minimize the error and the design cost as well. The prediction error can be calculated by different methods such as by simply calculating (Norm(e)/Norm(y)) for each model order. However, we have used the well known efficient Akaike's Prediction Errors (Niedzwiecki and Ciolek, 2017).

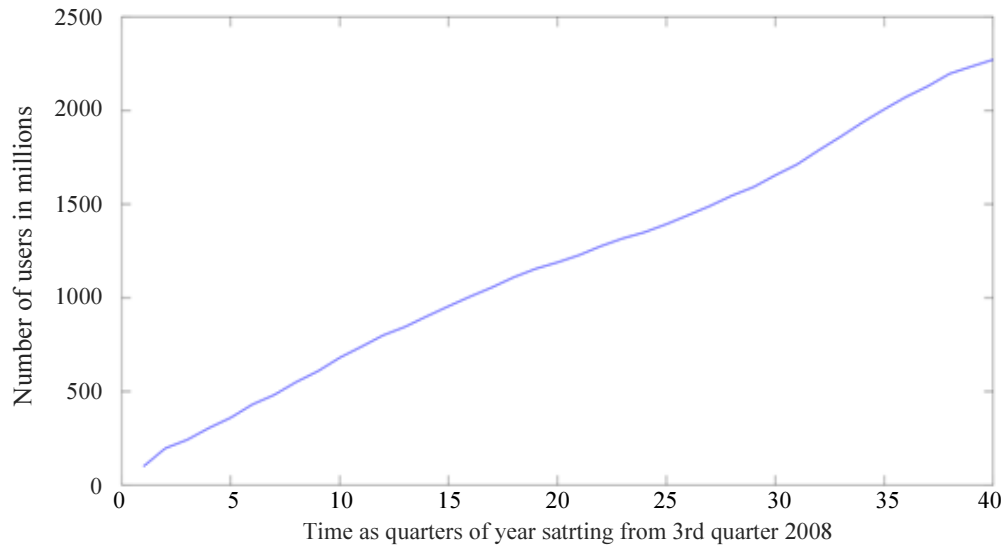


Fig. 4: Facebook: Number of monthly active users

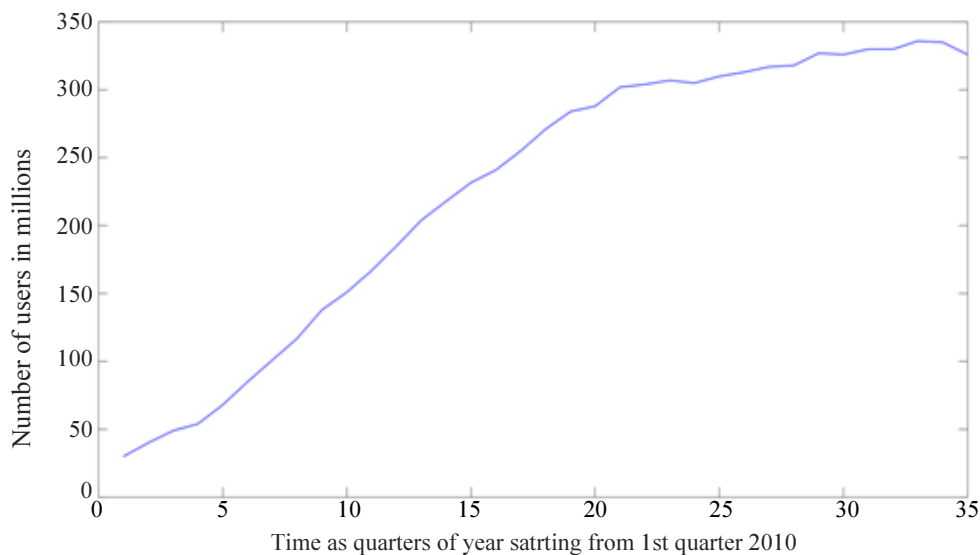


Fig. 5: Twitter: Number of monthly active users worldwide 2010-2018

Figure 6 shows the error modeling plot for Facebook and Twitter time series in terms of model order vs. Final prediction error. As can be clearly seen from Fig. 6, the optimal ARMA model orders to estimate the Facebook time series and the Twitter time series are $ARMA [5, 5]$ and $ARMA [3, 3]$ respectively. This can be easily depicted from the figure since they correspond to the minimum acceptable prediction error values.

Developing ARMA Model Using Optimal Order

This phase is about finding the ARMA model coefficients for optimal order and predicted model output. This also used to compare the measured data (y) with the estimated data by plotting them on the top of each other. This phase is illustrated in the next section.

Forecasting the Short Term Future Values

This phase is about using optimal order ARMA for predicting and visualizing the short term future period (i.e., generate the next 1-5 time slots). This phase is illustrated in the next section.

[Optional] Analyzing the Power Spectral Density

This phase is about finding the power spectrum of the signal y using ARMA method and compare the power spectrum using any method to compute the power spectrum such as periodogram and auto-correlation (AFC) based methods to analyze the signal energy distribution over the frequency components of the signal. However since this phase is optional, we have provided some plots for power spectrum using periodogram and some other useful plots.

Simulation Results

ARMA model is an Infinite Impulse Response (IIR) filter that uses the feedback to generate the internal dynamics for the predicted signal. Therefore, its accuracy is heavily based on the signal estimation for past time series values. Thus, for optimal ARMA modeling, we have investigated the relationship between the different model orders against the final prediction errors as illustrated in Fig. 6 earlier and accordingly found that ARMA [5, 5] and ARMA [3, 3] are the optimal model orders to estimate and predict the time series values for the collected time series numbers of Facebook and Twitter social networks respectively. In this section, we provide all the simulation results in which we generated them by implementing the proposed models using MATLAB. However, all results including figures and the comparison table will be discussed the figures in the next section, Discussion section.

Discussion

Indeed, the aforementioned results revealed that the developed forecasting models were superior due to the high levels of confidence achieved from applying the optimal model orders for both time series.

Figure 7 shows the amount of model fit percent as a result of applying the derived optimal orders for both time series. As noticed from the figure, both time series have been accurately estimated with slight higher figures related to ARMA model of Facebook time series with 98.9% and 96.5% for ARMA model of Twitter time series. This is an expected behavior since the collected Facebook time series has more linearity than that for the Twitter values. However, both cases showed that the derived ARMA models are precise in the signal estimation and can be safely (i.e., trusty) used to predict (forecast) the short term future for both signals.

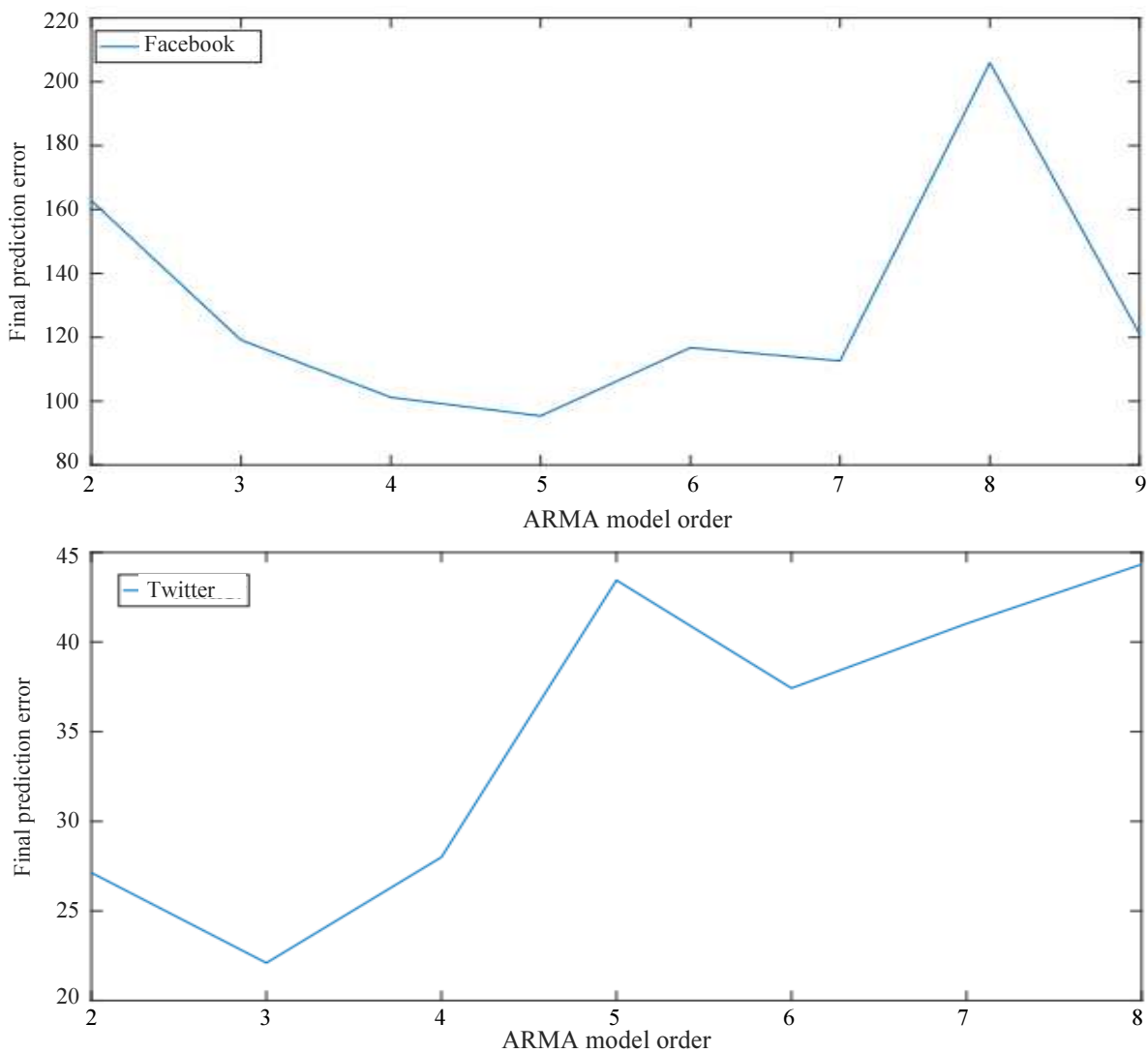


Fig. 6: Error modeling technique: Model order vs. final prediction error

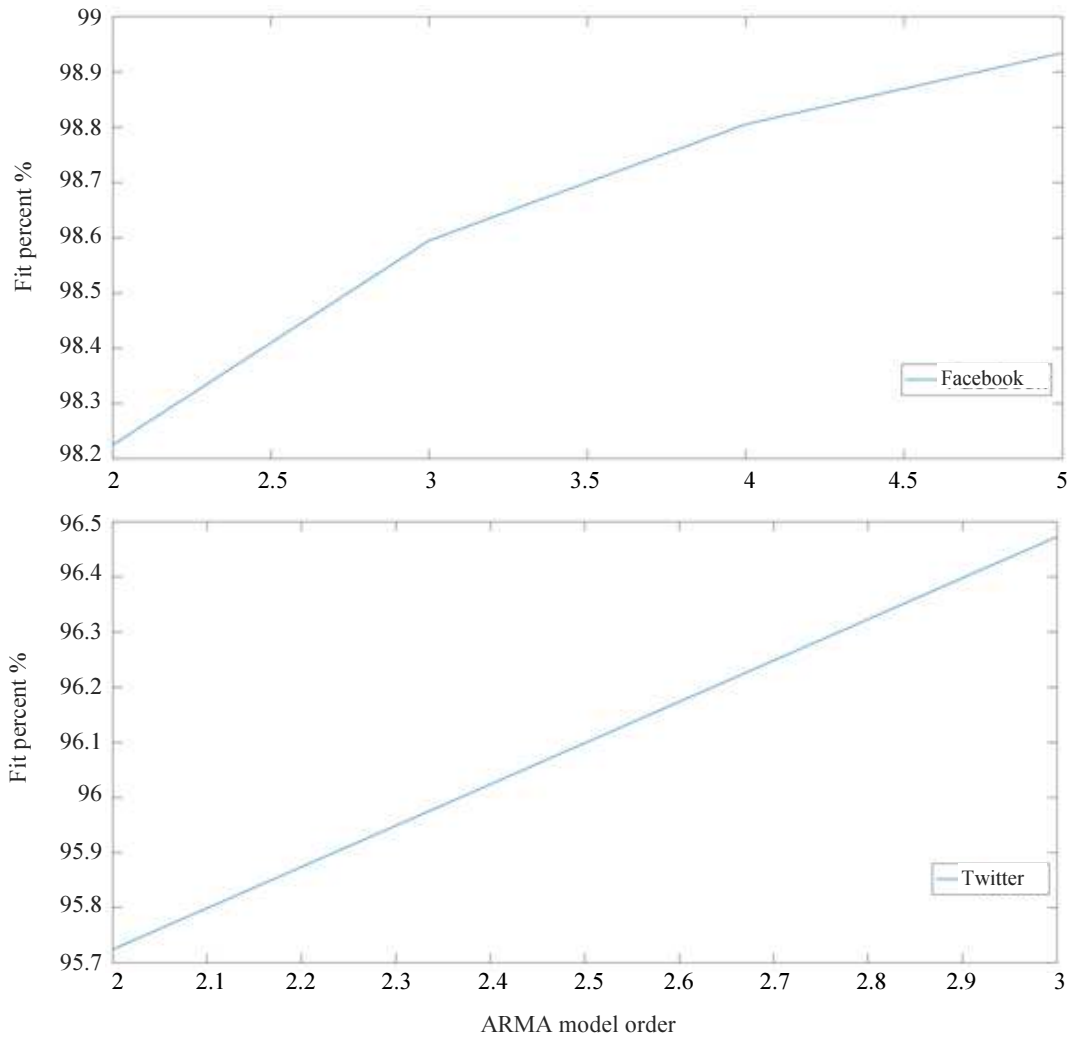
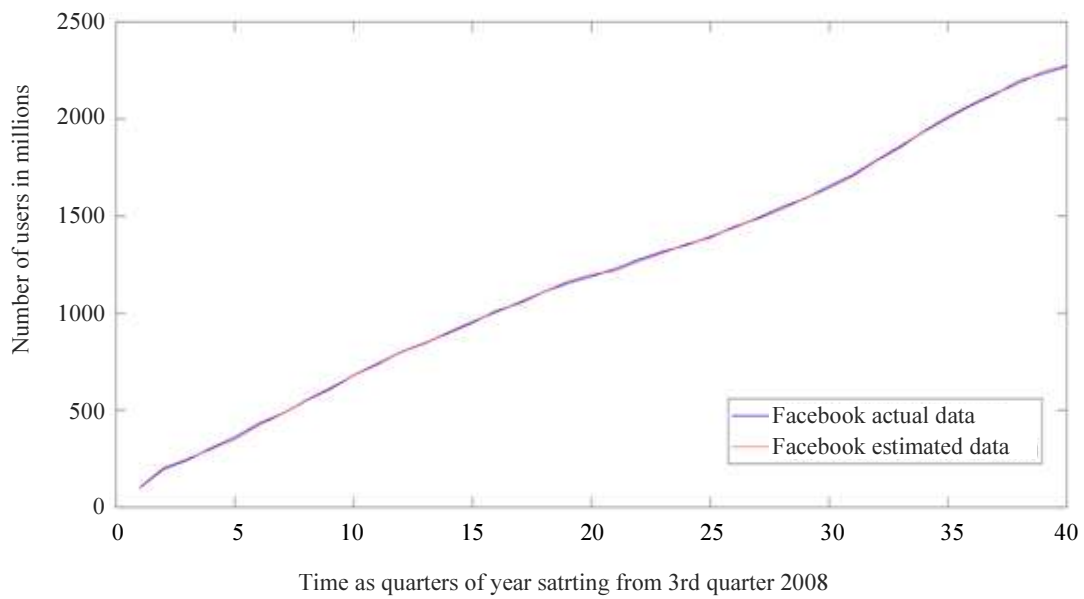


Fig. 7: Optimal order fitting percentages



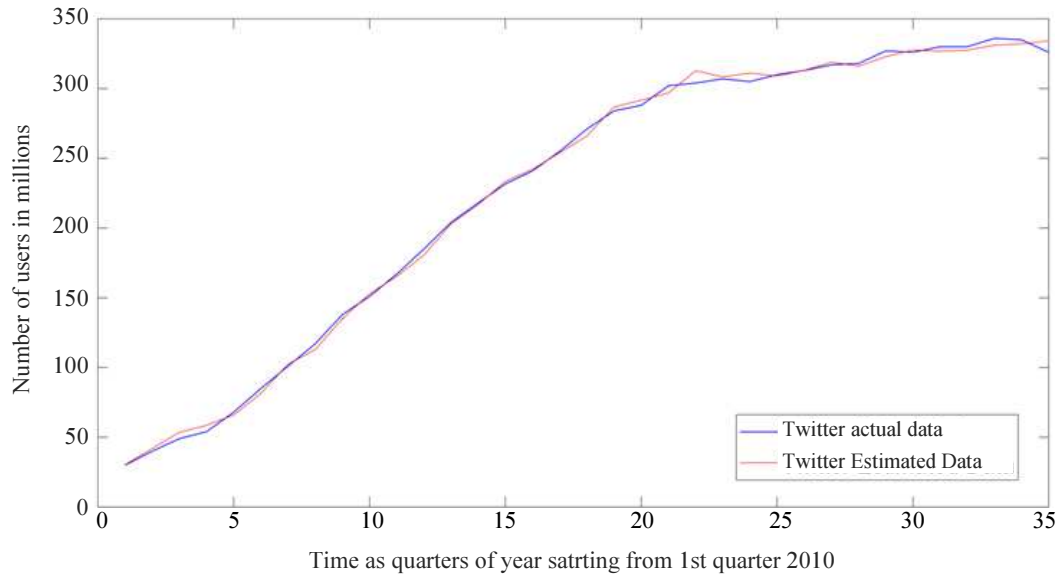


Fig. 8: Monthly active users for both time series: Actual vs. estimated

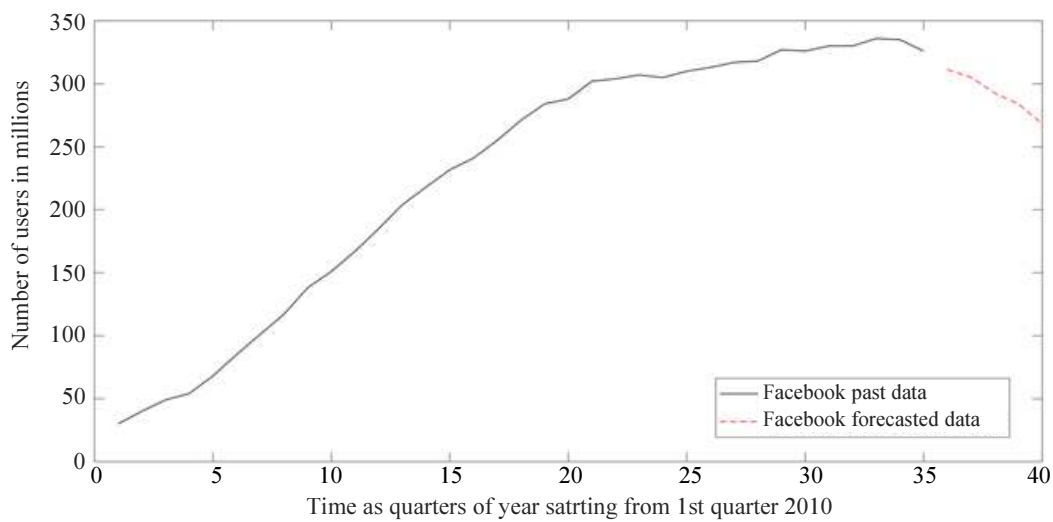
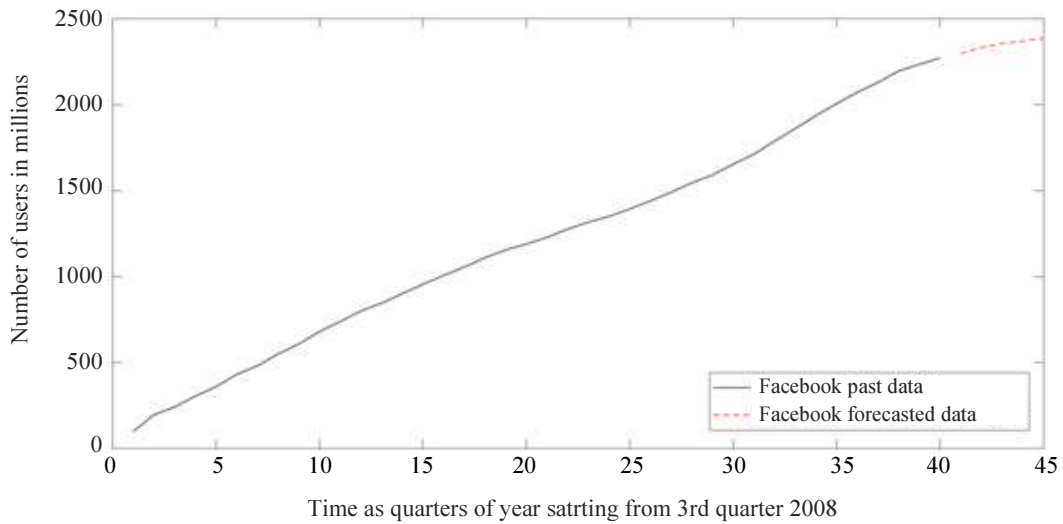


Fig. 9: Forecasting five quarters for both time series

Table 1: The comparison with others in the literature

Model Ref.	Prediction technique	Phenomena of-study	Series length	Accuracy level
(Liu and Shao, 2018)	ARMA(1, 1)	Weekly tea-auction prices	104 weeks	95%
(Karimi <i>et al.</i> , 2013)	ARMA (3, 3)	Sea-level darwin-harbor	100 hours	95.1%
(Valipour, 2016)	ARIMA-ARMA	Monthly rainfall amount	60-588 data	81-96%
(Zhang <i>et al.</i> , 2017)	ARMA(2, 2)-GARCH(1, 1)	Sea surface target detection	5000 data	95%
(Yan and Ouyang, 2019)	DDECM	Wind power prediction	35, 040 data	86.64-93.2%
(Jiang and Gong, 2014)	VECM model	Construction markets forecasting	36 quarters	90%
(Parmar and Bhardwaj, 2014)	Arima model	Water quality prediction	120 months	95%
(Rehman <i>et al.</i> , 2014)	RCGPANN	Foreign currency exchange rates	1000 days	98.8%
Proposed	ARMA (3, 3)	Monthly twitter users	35 quarters	96.5%
Proposed	ARMA (5, 5)	Monthly facebook users	40 quarters	98.8%

Figure 8 illustrates the plots of both signals: The measured signal (actual) and the ARMA's estimated signal for both of the collected time series (Facebook and Twitter). According to the figure, it is obviously observed that the ARMA's estimated signals are very precise and compatible for almost all the components of the measured signals for both time series. The reason of these highly accurate behaviors is due to the strong linearity tendency for the measured data of both time series which has a very high impact in the estimation process of linear ARMA modeling. However, the figures recorded for Facebook signal seem to be much more accurate as its measured values are highly coupled with the estimated values while a very minor/slight variations have been observed to appear in the plot of Twitter signals. Indeed, this seems to be a normal tendency as the time series for Facebook tends to be more linear (and slightly larger) than its counterpart for Twitter which allows more estimation accuracy using ARMA modeling.

In addition, both case studies (for Facebook and Twitter) showed that the derived optimal ARMA models are very precise with luxurious signal fitting percents in the signal estimation. Consequently, they can be reliably used to forecast the short term future of both signals. Therefore, Fig. 9 presents the forecasting of five future quarters for both time series (i.e., 4th of 2018, 1st of 2019, 2nd of 2019, 3rd of 2019 and 4th of 2019). However, due to the increasing tendency in the ARMA [5, 5] forecasting results of Facebook social networking and slowly decreasing tendency in the ARMA [3, 3] forecasting results of Twitter social networking, it seems that people are going to be more interesting being active social communicators on Facebook than on Twitter.

Finally, the comparisons with other prediction models might be inconsistent because of the use of different modeling techniques, prediction orders and different time series with different series lengths and linearity levels. However, the use of ARMA model was strongly present in the related researches which prove the efficiency of using ARMA for prediction applications. For example, the ARMA prediction model for the online traffic flow time series presented in (Karimpour *et al.*, 2017) predicts the traffic flow for 15 minutes ahead with a 88.74% as level of confidence

whereas the proposed prediction modeling predicts the aforementioned phenomena (i.e., the number of monthly active users of Facebook and Twitter) with a much higher level of confidence (i.e., more than 96%). This result is competitive with many other dedicated models. Besides, the most compatible way for comparison is to compare the level of confidence (i.e., accuracy percentage of prediction model) gained from applying the developed prediction models with the corresponding time-series. Therefore, Table I compares the proposed model results with eight other different prediction (i.e., forecasting) models reported in the literature, such as: ARIMA (AutoRegressive Integrated Moving Average) Model (Valipour, 2016; Parmar and Bhardwaj, 2014), Data-Driven Error Correction (DDEC) Model (Yan and Ouyang, 2019), Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) model (Valipour, 2016), VECM (Vector Error Correction) Model (Jiang and Gong, 2014), RCG (Recurrent Cartesian Genetic) Model (Rehman *et al.*, 2014) and Programming evolved Artificial Neural Network (PANN) model (Rehman *et al.*, 2014). The comparison in the table is carried out in terms of: prediction technique and model order, time series length and level of accuracy. From Table 1, it can be noticed that the proposed predictions are very accurate and superior. However, all the provided models were accurate especially those built using ARMA with a large number of observations.

Conclusion and Remarks

An Auto-Regressive Moving Average (ARMA) model for estimating the short term future of the number of monthly active Facebook users worldwide and the number of monthly active Twitter users worldwide (in millions) with optimal model order number using MATLAB simulation package has been developed and reported in this paper. The simulation results showed that ARMA model error decreases fast and then fluctuates while increasing the model order number. However, the optimum model order is the order in which the model error has the smallest value with acceptable design cost. For our case, the optimal model order to estimate the given Facebook and Twitter time series were ARMA [5, 5]

and ARMA [3, 3], respectively, where they recorded a high-level of estimation accuracy with fit percents of 98.8% and 96.5% for Facebook and Twitter time series, respectively. Finally, the larger the time series and the more linear measured data signal (y) is the more precise estimation and prediction with ARMA model. However, the results of ARMA Models point out that Facebook is expected to have more turnout than twitter in for the upcoming year (2019).

In the future, we will consider the use of our developed model in modeling and estimating some other time series with high degree of linearity, specifically, the future perdition of global cyber crimes using the data set time series of world wide cyber crimes for the last couple of years. Also, we will consider to incorporate the non-linear modeling techniques such as the artificial neural networks to model some nonlinear time series such as the global oil prices, the world wide smart-phones market and others. Finally, for further prediction accuracy, we may seek to incorporate several levels of prediction techniques.

Acknowledgement

Authors would like to thank the Department of Computer Science at University of Texas at San Antonio, for supporting the publication of this work.

Author's Contributions

Qasem Abu Al-Haija: Developed the prediction model, participated in all experiments, coordinated the data analysis and contributed to the writing of the manuscript.

Qian Mao: Participated in reviewing the system model, performed data analysis and contributed to the writing of the manuscript.

Kamal Al Nasr: Verified the research method, provide publication recommendations, and contributed to the writing of the manuscript as well as support the publication of this manuscript.

Ethics

This article is original and contains unpublished material. All authors have read and approved the manuscript and no ethical issues involved.

References

- Gankevich, I. and A. Degtyarev, 2018. Simulation of Standing and Propagating Sea Waves with Three-Dimensional ARMA Model. In: The Ocean in Motion, Velarde, M., R. Tarakanov and A. Marchenko (Eds.), Springer Oceanography, ISBN-10: 978-3-319-71933-7, pp: 249-278.
- Hanke, J. and D. Wichern, 2008. Business Forecasting. 1st Edn., Pearson.
- Imdadullah, D., 2014. Time series analysis. Basic Stat. Data Analysis.
- Jiang, H. and H. Gong, 2014. A panel vector error correction approach to forecasting demand in regional construction markets. J. Construct. Manag. Economics, 32: 1205-1221.
DOI: 10.1080/01446193.2014.977800
- Karimi, S., O. Kisi, J. Shiri and O. Makarynsky, 2013. Neuro-fuzzy and neural network techniques for forecasting sea level in Darwin Harbor, Australia. Comput. Geosciences, Elsevier, 52: 50-59.
DOI: 10.1016/j.cageo.2012.09.015
- Karimpour, M., A. Karimpour, K. Kompany and A. Karimpour, 2017. Online Traffic Prediction Using Time Series: A Case Study. In: Integral Methods in Science and Engineering, Constanda, C., R.M. Dalla, P. Lamberti and P. Musolino (Eds.), Springer, ISBN-10: 978-3-319-59387-6, pp: 147-156.
- Liu, H. and S. Shao, 2016. India's tea price analysis based on ARMA model. Modern Economy, 7: 118-123.
- Liu, Y., J. Wu, Y. Liu and B.X. Hu, 2015. Analyzing effects of climate change on streamflow in a glacier mountain catchment using an ARMA model. Quaternary Int. Elsevier, 358: 137-145.
DOI: 10.1016/j.quaint.2014.10.001
- Niedzwiecki, M. and M. Ciolek, 2017. Akaike's final prediction error criterion revisited. Proceedings of the 40th International Conference on Telecommunications and Signal Processing, Jul. 5-7, IEEE Xplore Press, Barcelona, Spain. DOI: 10.1109/TSP.2017.8075977
- NIST/SEMATEC, 2018. e-Handbook of Statistical Methods, Engineering Statistics. NIST/SEMATEC, <http://www.itl.nist.gov/div898/handbook>
- Parmar, K.S. and R. Bhardwaj, 2014. Water quality management using statistical analysis and time-series prediction model. Applied Water Sci., 4: 425-434.
DOI: 10.1007/s13201-014-0159-9
- Proakis, J.G. and D.K. Manolakis, 2007. Digital Signal Processing. 4th Edn., Pearson.
- Ratnam, D.V., M. Harshali and M.V. Gopala, 2015. Forecasting of ionospheric time delays using ARMA model under Geomagnetic storm conditions. Proceedings of the IEEE 2nd International Conference on Signal Processing and Integrated Networks, Feb. 19-20, IEEE Xplore Press, Noida, India. DOI: 10.1109/SPIN.2015.7095303
- Rehman, M., G.M. Khan and S.A. Mahmud, 2014. Foreign currency exchange rates prediction using CGP and recurrent neural network. IERI Procedia, Elsevier, 10: 239-244.
DOI: 10.1016/j.ieri.2014.09.083
- Statista, 2018. The portal for statistics. Statistics. <https://www.statista.com/>

- USGS, 2018. Earthquake Hazards, Earthquake data. USGS, <https://earthquake.usgs.gov/>
- Valipour, M., 2016. How much meteorological information is necessary to achieve reliable accuracy for rainfall estimations? *Agri. MDPI*, 4: 53-59. DOI: 10.3390/agriculture6040053
- Welch, P.D., 1967. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transact. Audio Electroacoustics*, 15: 70-73. DOI: 10.1109/TAU.1967.1161901
- Yan, J. and T. Ouyang, 2019. Advanced wind power prediction based on data-driven error correction. *Energy Convers. Manag.*, 180: 302-311. DOI: 10.1016/j.enconman.2018.10.108
- Zhang, K.Y., Y. Zhanga, Z. Denga, X. Zhang and H. Liua, 2017. Sea surface target detection based on complex ARMA-GARCH processes. *Digital Signal Process.*, Elsevier, 70: 1-13. DOI: 10.1016/j.dsp.2017.07.017