

Original Research Paper

Forecasting Air Passenger Demand: A Comparison of LSTM and SARIMA

¹Quang Hung Do, ²Shih-Kuei Lo, ²Jeng-Fung Chen, ¹Chi-Luan Le and ¹Luong Hoang Anh

¹Faculty of Information Technology, University of Transport Technology, Vietnam

²Department of Industrial Engineering and Systems Management, Feng Chia University, Taiwan

Article history

Received: 02-05-2020

Revised: 03-07-2020

Accepted: 25-07-2020

Corresponding Author:

Quang Hung Do

Faculty of Information
Technology, University of
Transport Technology,
Vietnam

Email: hungdq@utt.edu.vn

Abstract: All airports need to have an accurate prediction of the number of passengers for their efficient management. An accurate prediction of the number of air passengers is crucial task since it provides information for planning decisions in the airport infrastructure to stabilize the service and maximize the profit. This study proposes a novel air passenger demand forecasting model based on Deep Neural Network (DNN), specifically, Long Short Term Memory (LSTM) algorithm. The developed models are applied on the data from Incheon International Airport to show its effectiveness and practicability. The Seasonal Auto-Regressive Integrated Moving Average (SARIMA) method is also applied to the research problem. The performance criteria including MAPE, MSE, RMSE and MAD are used to evaluate the forecasting accuracy. The experimental results show that both SARIMA and LSTM approaches provide accurate and reliable forecasting and have greater predictive capability; however, the LSTM model shows a superior forecasting performance.

Keywords: Forecasting, Air Passenger Demand, Deep Learning, LSTM, SARIMA

Introduction

Predicting airborne demand is the key research for airborne management and planning. It targets estimates the actual demand of specific point in advance according to needs of service provider. The core is to minimize the difference between estimation and measurement to stabilize the service and maximize the profit.

Due to the rapid growth of the aviation industry according to increased airborne demands, the world has entered the era of a one-day life zone. This leads more passengers to use the airport, making the airport as an important facility for international exchange beyond the means of transportation. Especially, the growth ratio of aviation transportation in Asia is high globally and moreover, the air freight volume is expected to be extended. According to International Air Transport Association (IATA), the number of global passengers would reach 8.2 billion in 2037 Fig. 1.

In this context, the airport tries to predict the number of passengers for efficient management because it plays a key role in overall planning. Currently, most airport focus on long-term management. However, for more efficient operation

and better quality of service, it is required to pay attention to mid-term and short-term schedule which include more fluctuation and variables. In this study, we proposed the prediction model for mid-term and short-term management.

Forecasting the number of air passengers is crucial because of the followings:

- (i) The airport is a facility for not only residences but also foreigners. Since it is the first sight of visitors in that country, the airport would make the first impression of the country. Therefore, planning and executing the efficient operating strategy based on accurate prediction will improve the national image
- (ii) The variation in airborne demand can be managed accurately by mid-term and short-term prediction of the number of passengers. The utilization of capacity, manpower requirement and financial planning of the operating capital project can be devised in more detail. In addition, it is possible to evaluate and improve airport services by making more efficient decisions on infrastructure

development and to reduce airline risks by objectively assessing the demand side of the air transportation business

- (iii) Mid-term and short-term forecasting provides important information for monthly operational management decisions, including aircraft scheduling decisions, maintenance plans, advertising and sales campaigns and the opening of new business locations and enables a relatively immediate response

In this study, we used Incheon International Airport as an example to evaluate the performance of prediction. Incheon International Airport is located in Incheon, South Korea. It is the largest international airport in Korea with an area of 22.39 million m^2 . It is an important facility which handles a number of tasks such as air transportation, passenger departure and exit, quarantine, maritime, bank, insurance, postal and sales service.

We conducted the prediction in two phases: The mid-term and the short-term which target monthly and weekly data. In the case of the mid-term prediction, we used the monthly passenger data provided by Incheon International Airport, which is collected during 192 months from January 2013 to December 2018. The number of days and holidays for each month are also considered. For the short-term prediction, we converted the daily passenger data of Incheon International Airport as weekly data which is collected during the recent 5 years, January 2013 - December 2018. Among them, we used the data of 260 weeks between 5 January 2014 and 29 December 2018. The number of weekends for each week was also considered.

The prediction model was developed using Long Short Term Memory (LSTM), one of the emerging artificial intelligent networks for series data analysis. For mid-term prediction, we trained the model with data of 10 years from 2003 to 2012 and evaluated the model using the data of 2013-2018. In short-term prediction, we used 60% and 40% of data for training and test, respectively. The Root Mean Square Error (RMSE) metrics was used to assess accuracy.

Lastly, we adopted a statistical analysis to compare the performance of LSTM and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) and verified the superiority of the proposed model.

The organization of the paper is as follows. Section 2 describes demand forecasting methods and related works about airline passengers. Section 3 describes the method to design the LSTM model and the ARIMA model for empirical studies. Section 4 analyzes the results of empirical studies and explains the statistical techniques used in the verification

process. Finally, section 5 summarizes the study, shows implications of this paper and presents direction of the future work.

Literature Review

Predictive Methods

The decision-maker needs to forecast the future in order to make a plan considering sales, demand and stock of the future. Several approaches including qualitative and quantitative method have been proposed to address this issue. The qualitative prediction is used for a new era without previously without historical data, or in the situation where mathematical modeling is not possible. Several popular qualitative methods are Delphi method, decision method, case analysis method and group discussion. The strong point of qualitative prediction is that it reflects a various situation of future as well as incomplete but intuitive and expert knowledge. However, the subjective view of the researcher may impact the result and the external environment with political will may distort the future demand. Moreover, the qualitative method can provide inconsistent results, therefore, it does not get full agreement from all sides.

The quantitative method predicts future demand using statistical analysis with previous market data. They are categorized into two major approaches: Casual model and time series analysis. The former includes econometric model and spatial balance model while the latter include moving average, trend analysis, Exponential smoothing, spectral analysis, adaptive filtering and ARIMA model.

Regression analysis, the most widely used method, analyzes the causality among variables. This approach assumes that causality between the independent variable and dependent variable. Based on this assumption, the main purpose of it is to find the independent variable which expresses the dependent variable logically, develop the model which represents their relationship more clearly and finally predict the demand.

However, time series analysis assumes that the future is a function of the past. It predicts the demand using the pattern which is derived by collecting and analyzing historical data. In the aviation industry, multiple regression analysis requires a number of the independent variable because various factors affect the future demand. However, collecting a sufficient amount of data with an acceptable level of reliability is difficult. In addition, it would produce contradictory regression with a relatively high value of r^2 even though there is no strong relation between independent and dependent variables. Hence, we adopted a time series analysis in this study.

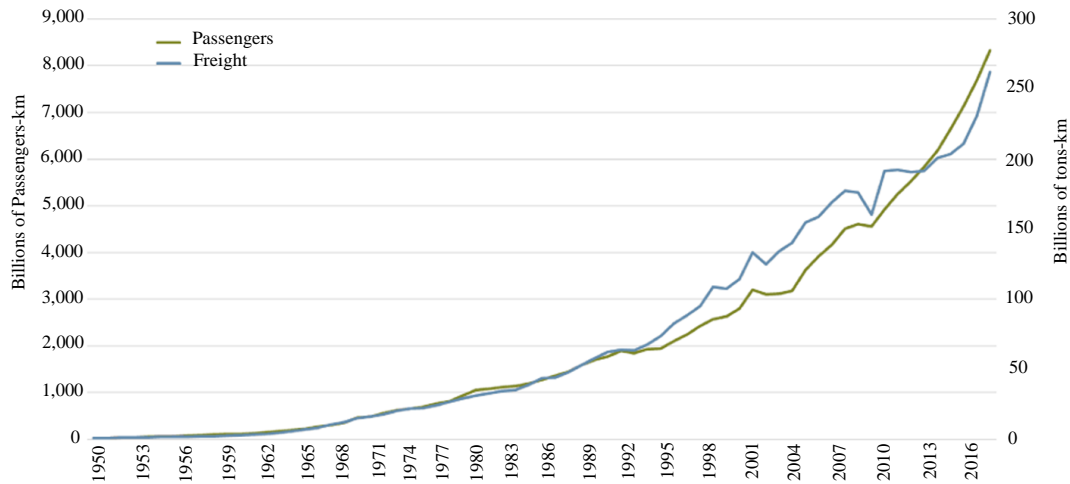


Fig. 1: The trend of increasing number of global airport passengers

Regardless of the prediction method, the estimation result and ground truth are always different. The accuracy of prediction is determined by the magnitude of the error. The less error is, the more accurate the model is, indicating that it is a good model. The metric of accuracy includes Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD).

The study used MSE as an evaluation index of LSTM and trained the model to minimize the MSE. In the evaluation, RMSE, MAPE and MAD indexes were also used.

Time Serial Analysis

The time series data is collected sequentially for a period of time, hence there is a relation among data collected at an adjacent time. Typically, birth rate, death rate, GDP, consumer price index, the closing price of the stock, precipitation, humidity and temperature are time series data. The goal of time series analysis is understanding the structure of historical data, developing the model which represents the structure efficiently and finally predicting the future using the model.

To this end, one should check the trend, any apparent sharp changes in behavior and outlier to acquire stationary by eliminating them. The stationary indicates that the consistency regardless of elapsed time. In other words, the stationary data always shows the same statistical characteristics such as average, variance, skewness and kurtosis. Also, there is no periodic pattern.

The stationary meets two conditions:

$$F_{t_1, t_2, \dots, t_n}(x_1, \dots, x_n) = F_{t_1+h, t_2+h, \dots, t_n+h}(x_1, \dots, x_n) \quad (1)$$

$$Cov(x_t, x_s) = Cov(x_{t+h}, x_{s+h}) \quad (2)$$

Equation (1) indicates that the CDF does not change as time goes by. The CDF after time h (Lag) is identical with one before time h .

Equation (2) means that the covariance after time h is the same with the previous one.

The time series analysis is divided into time domain analysis and frequency domain analysis. The time domain analysis assumes that the dependency or relateness among adjunct data regresses with previous data. Based on this assumption, it decomposes the time series data into deterministic factors such as trend and statistical factors such as noise. AR, MA, ARMA, ARIMA, SARIMA are typical approaches. The frequency domain analysis converts the normal time series data with a stable average and variance as a linear compound of strong periodic functions and then, it investigates the time series periodicity for individual periodic function. Representatively, the spectrum analysis belongs to this category.

In the time series analysis field, exponential smoothing which gives the biggest weight for the most recent data from observation point and reduces the weight gradually according to elapsed time is followed by the ARIMA proposed by Box-Jenkin which minimizes the error of measured data with noise and predicts the future recursively. Along with the growth of computer science, Artificial Neural Network (ANN) and fuzzy logic based on artificial neural network algorithm are proposed.

In this study, both SARIMA and LSTM were utilized to compare and analyze the actual time series data. These approaches are known as a strong solution in previous research and emerging technique based on artificial neural network algorithm, respectively.

Prediction of Air Passenger

The study on airline passenger prediction has been conducted a long time before. A neural network was used to develop an airline passenger prediction model (Nam and Schaefer, 1995) and a regression model was designed to predict the airline passenger demand in Saudi Arabia (Abed *et al.*, 2001). The study used air passenger data from Saudi Arabia collected during about 22 years from 1971 to 1992. They adopted 16 independent variables, including per capita income, oil GDP, non-oil GDP and population. The final regression model was constructed by re-selecting seven independent variables with high correlation through analysis of correlation among 16 independent variables.

The Holt-Winters method was used to estimate the number of passengers in the UK from 2005 to 2030 (Barboza and Kimura, 2017). The Holt-Winters method is one of the smoothing methods and proved that although it cannot be used when the variance of seasonal components or errors is not independent, the adequate data conversion makes the method applicable. Monthly airline passenger data for 56 years from 1949 to 2004 were used.

A model was used for predicting the number of air passengers between airports in Berlin, the capital of German and those of 28 European countries using the gravity model (Grosche *et al.*, 2007). The authors used data from 9,091,082 passengers from January to August 2004 and 1,228 travel route data among 138 cities in Berlin and 28 European countries. The independent variables are the distance between two cities, the population, the average flight time of the passenger aircraft, Buying Power Index, Catchment and GDP.

A model was developed to predict the number of passengers at Lisbon airport using exponential smoothing (Samagaio and Wolters, 2010). With data from 1995 to 2007, it predicted the number of passengers from 2008 to 2020.

A forecasting model of Nigerian air passenger demand using regression analysis was presented (Aderamo, 2010). The model used Nigerian airline passenger data collected for about 32 years from 1975 to 2006. Independent variables are agricultural production, minerals production, manufacturing production, energy consumption, consumer price index, metrics related to electricity consumption, inflation rate, government expenditure and GDP.

A model based on system dynamics frameworks was developed to forecast air passenger demand and to assess several policy scenarios related with runway and passenger terminal capacity expansion to meet the future demand. It was found that airfare impact, level of service impact, GDP, population, number of flights per day and dwell time play a key role in determining air passenger volume, runway utilization and total additional area

needed for passenger terminal capacity expansion (Suryani *et al.*, 2010).

The number of passengers at Hong Kong International Airport from 2011 to 2015 was estimated through the ARIMA model. The used monthly airline passenger data of Hong Kong International Airport monitored during about 10 years from 2001 to 2010. The model produced an average error of about 3% (Tsui *et al.*, 2014).

The number of passengers visiting New Zealand was forecasted through a regression analysis model. The GDP and New Zealand dollar exchange rates of the United States, Australia, China, the United Kingdom, Korea and Japan, which account for the majority of New Zealand passengers, were used as independent variables (Duval and Schiff, 2011).

The post-mortem methods were utilized for evaluating air transportation forecasts focusing on time trend analysis and econometrics. The methods were proved their effectiveness by taking Rhodes Airport as an example (Profillidis, 2012).

A prediction model that combines the Holt-Winters model and the Integrated Mixture of Local Expert Model (IMLEM) models were developed to forecast air passengers (Scarpel, 2013). The model was trained using the airline passenger data at the Sao Paulo airport for 21 years from 1990 to 2010. The average error rate of IMLEM for 2011 and 2012 was 2.82%.

The number of passengers at Hong Kong International Airport was predicted by the use of SARIMA and ARIMAX methods (Tsui *et al.*, 2014). Data from January 1993 to November 2010 and December 2010 to August 2011 were used for training and assessing post-mortem prediction performance, respectively. Both forecasting models showed good performance and they predicted the growth rate of China, Taiwan and Africa as a negative value.

The Least Squares Support Vector Regression (LSSVR) model was applied to construct a prediction model (Xie *et al.*, 2014). LSSVR is an extended form of the ARIMA model. It generates a model by using several time series variables at once. Using monthly aviation passenger data of Hong Kong Airways International Airport for about four years from 1999 to 2013, a prediction model combining seasonal decomposition and LSSVR was made. The average error rate of the predicted model was lower than that of the ARIMA model with the same airline passenger data.

The number of passengers at Incheon International Airport was estimated by making a regression model using Internet search words (Kim, 2016). Data from June 2010 to August 2014 was used and the optimal prediction model used the number of search of 51 keywords before 8 months as independent variables. The model was evaluated through K-fold cross-validation.

The Egypt's international and domestic air passenger demand was predicted through back-propagation neural network and genetic algorithm (El-Din *et al.*, 2017). Data from 1970 to 2013 were used. Of the total 528 data, 372 data were trained and 156 were used as test data. The independent variables were population, PCI, GDP, GNP, economic growth rate and exchange rate.

A model was developed for predicting boarding time using the LSTM. The data was generated by direct measurement on site and simulation (Schultz and Reitmann, 2019). The statistical analysis of actual boarding and expected boarding progress proved that the LSTM model is promising for predicting boarding time. Table 1 provides a brief summary of related studies.

SARIMA

The ARIMA model is an evolved model of the ARMA model (Makridakis and Hibon, 1997). The ARMA model is a combined model of the autoregressive model $AR(p)$ and the moving average model $MA(q)$. The ARMA model determines the order of the AR and MA model for prediction. It can be applied only to data with

normality. When a time-series graph does not show a constant pattern and the Auto-Correlation Function (ACF) decreases gradually, ARMA is used for modeling. Time series data with abnormalities are normalized by transformation or differential and modeled by ARMA. This is called the ARIMA model.

Here, the $AR(p)$ model is the average regression model with the theory that the current time series data Y_t can be explained by the previous data from Y_{t-p} to Y_{t-1} . It is based on an assumption that the current time series data is dependent on the previous series data. If there is a little dependency between them, the current data is independent of the past and becomes a white noise time series data. The stronger dependency exists on the past, the more dependent on the past current data is, reaching to a random walk. The $AR(p)$ model analyzes the characteristics of the target time series data by analyzing the autocorrelation with the past. It means that the data before time p affects the present data. In Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) graphs, ACF decreases rapidly and PACF has a cut point at a certain point.

Table 1: Research on airport passenger prediction

Year	Research	Author	Method	Airport
1995	Forecasting international airline passenger traffic using neural networks	Nam and Schaefer (1995)	Neural networks	
2001	An econometric analysis of international air travel demand in Saudi Arabia	Abed <i>et al.</i> (2001)	Regression	Saudi Arabia
2007	Holt-winters forecasting: An alternative formulation applied to UK air passenger data	Bermúdez <i>et al.</i> (2007)	Holt-Winters	UK
2007	Gravity models for airline passenger volume estimation	Grosche <i>et al.</i> (2007)	Gravity model	Germany and 28 European countries
2010	Comparative analysis of government forecasts for Lisbon airport	Samagaio and Wolters (2010)	Exponential Smoothing	Lisbon
2010	Demand for Air Transport in Nigeria	Aderamo (2010)	Regression	Nigeria
2010	Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework		System dynamics framework (Simulation)	Taiwan Taoyuan International Airport (TTIA)
2011	Forecasting air passenger traffic; the case of Hong Kong International Airport	Tsui <i>et al.</i> (2014)	ARIMA	Hong Kong
2011	Effect of air services availability on international visitors to New Zealand	Duval and Schiff (2011)	Regression	New Zealand
2012	An ex-post assessment of a passenger demand forecast of an airport	Profillidis (2012)	Statistical methods, Econometric and Gravity methods	Rhodes Airport in Greece
2013	Forecasting air passengers at Sao Paulo International airport using a mixture of local experts model	Scarpel (2013)	Mixture of local expert model	São Paulo International Airport
2014	Forecasting of Hong Kong airport's passenger throughput	Tsui <i>et al.</i> (2014)	Seasonal ARIMA ARIMAX model	Hong Kong International Airport (HKIA)
2014	Short-term forecasting of air passenger by using hybrid seasonal decomposition and Least Squares Support Vector Regression approaches (LSSVR)	Xie <i>et al.</i> (2014)	Least Squares Support Vector Regression	Hong Kong International
2016	Forecasting short-term air passenger demand using big data from search engine queries	Kim (2016)	Big data (Engine query)	Incheon International Airport
2017	Airline Passenger Forecasting in EGYPT (Domestic and International)	El-Din <i>et al.</i> (2017)	Back-propagation neural network and Genetic algorithm	EGYPT
2019	Machine learning approach to predict aircraft boarding	Schultz and Reitmann (2019)	Long Short-Term Memory model	

ACF means the correlation coefficient of values separated by k periods. That is, the ACF indicates the order of correlation according to the time difference. Unlike the ACF, the Partial Auto-Correlation Function (PACF) is a pure correlation coefficient between two variables, which is calculated after eliminating the effect of all the different values between the observed values. If the PACF has a cut-off point at $k = 2$, it becomes an AR(1) model.

A general regression model of AR(p) is as follows:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t \quad (3)$$

$$Y_t = \mu + \alpha_1 (Y_{t-1} - \mu) + \alpha_2 (Y_{t-2} - \mu) + \dots + \alpha_p (Y_{t-p} - \mu) + \varepsilon_t \quad (4)$$

where, α_p is autoregressive coefficient, p autoregressive order, μ mean and ε_t the white noise with mean 0 and variance σ^2 . For example, in the case of AR(1), t is defined as the value at time $t-1$, which can be expressed as $Y_t = \alpha_1 Y_{t-1} + \varepsilon_t$.

The MA model is a moving average procedure and the current time series data is composed of a weighted average of past residuals. Since the residual term is white noise, the current data is described as the mean value of past white noise. Since the white noise has high normality and high average regression characteristic, the MA model based on the sum of them also has an average regression characteristic. ACF has a break point and PACF shows a rapid decrease.

Unlike the autoregressive model AR(p), the moving average model MA(q) is a model of weighted linear combination with white noise ε_t . The current time series data Y_t can be expressed by the continuous error terms $\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \dots, \varepsilon_{t-q}$. The general form of MA(q) model is as follows:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (5)$$

where, ε_t represents the white noise with average 0 and variance σ^2 , θ_q the moving average coefficient, q the order of moving average. Hence, MA(1) is expressed as $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$.

However, estimating the general time series data with only AR(p) or MA(q) is difficult. In that case, Auto-Regressive Moving Average (ARMA) which has characteristics of both models is used. ARMA model is a combined model of AR and MA model which assumes that current time series data is determined by the function of past time series data and past residuals. Since both the AR and the MA model have an average regression characteristic, the ARMA model also has it. Since the AR, MA and ARMA models have an average regression characteristic, they are suitable for time series

analysis which has normality always for all parameter values. The ARMA model calculates the approximation relatively more accurately and quickly with smaller parameters compared with existing AR or MA models. Since ARMA is a mixed model of AR and MA model, ARMA(1,0) and ARMA(0,1) is equal to AR(1) and MA(1), respectively. The general formula of the ARMA model is as follows:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (6)$$

Most of time-series data usually does not have normality, which shows increasing trends or increasing variance over time. The unstable time series means that the mean and variance of the time series change along with elapsed time, indicating the predicted value becomes invalid. Such time series data cannot be analyzed by AR, MA and ARMA models. Therefore, the data should be converted to time series with normality. To this end, log transformation, difference and seasonal differences are performed for converting according to the characteristics of the data. When the time series is normalized, it is analyzed using the ARIMA model.

The regression model using trigonometric functions and indicator functions or Winters' seasonal exponential smoothing is used to analyze time series models with seasonal pattern, but these methods can only be used when the seasonal time series data are independent of each other. However, the ARIMA model is most suitable because time series data are generally correlated with each other.

Certain data may have a time series average regression characteristic after difference even though the data itself does not have normality or average regression characteristics. The ARIMA model is an ARMA model applied to the differential time series. ARIMA model with difference value 0 is equal to the ARMA model. The differencing process is as follows:

$$\nabla Y_t = (1 - B)Y_t = Y_t - Y_{t-1} \quad (7)$$

$$\nabla^2 Y_t = (1 - B)^2 Y_t = (1 - 2B + B^2)Y_t = Y_t - 2Y_{t-1} + Y_{t-2} \quad (8)$$

where, B indicates the backshift operator which means $B^j Y_t = Y_{t-j}$. The difference is the process of subtracting the previous data from the original data until the time series data have normality. ARIMA has three orders p, d, q , expressed as ARIMA(p, d, q) where p is the number of autoregressive terms, d the number of nonseasonal differences needed for stationarity and q the number of lagged forecast errors in the prediction equation.

When the time series data show seasonal trends, seasonal ARIMA is generally used. The seasonal ARIMA is expressed as $SARIMA(p, d, q) (P, D, Q)_s$ by integrating the order of the nonseasonal time series model (p, d, q) and the order of seasonal time series model. The SARIMA model overcomes the limitations of the ARIMA model which cannot consider the seasonal or periodic characteristics of time series data. Here, P, D, Q indicate the number of seasonal autoregressive terms, the number of seasonal differences and the number of seasonal moving-average terms, respectively. The general formula of $SARIMA(p, d, q) (P, D, Q)_s$ is as follows:

$$\Phi_p(B^s)\phi_p(B)(1-B)^d(1-B^s)^D Z_t = \Theta_q(B^s)\theta_q(B)\epsilon_t \quad (9)$$

where, $\Phi_p(B^s)$ is the seasonal Autoregressive (AR) coefficient, $\phi_p(B)$ the nonseasonal Autoregressive (AR) coefficient, $(1-B)^d$ the number of nonseasonal difference d , $(1-B^s)^D$ the number of seasonal difference D , Z_t the observation of seasonal time series without normality, $\Theta_q(B^s)$ seasonal moving average coefficient (MA), $\theta_q(B)$ is the nonseasonal moving average coefficient (MA) and ϵ_t the error term or white noise. If the order of seasonal time series model is zero, it is the same with ARIMA.

Deep Neural Network

An artificial Neural Network (ANN) is a mathematical model to simulate the network of biological neurons that make up a human brain so that the computer will be able to learn things and make decisions in a humanlike manner. A deep Neural Network (DNN) is an ANN with more than the three layers. With more hidden layers, DNNs have the ability to capture highly abstracted feature from training dataset. Fig. 2 shows a deep neural network with three hidden layers. In comparison with conventional shallow learning architectures, DNN has capability to model deep complex non-linear relationship by using distributed and hierarchical feature representation. Various deep learning architectures such as Convolution Neural Network (CNN), Recurrent Neural Network (RNN) have been applied to the domain of computer vision, speech recognition and natural language processing.

In a traditional ANN, it is assumed that all inputs (and outputs) are independent of each other. Recurrent Neural Networks (RNNs) perform the same task for every element of a sequence, with the output being depended on the previous computations. RNNs are networks with inner loops in them, allowing information to persist (Schmidhuber, 2015).

RNN is an artificial neural network which solved the problem of traditional neural network. It is powerful to

handle sequential data. As shown in Fig. 3, RNN has a loop at the hidden layer which helps the iteration of data.

Figure 4 shows the unfolding in time of the computation involved in its forward computation. Also presented in Fig. 4, the output h_t is produced from input x_t through neural network A. The loop transfers the data to the next step. Via the loop, each independent data becomes dependent on each other. RNN can be seen as multiple copies of the same network.

Figure 5 indicates that RNN excels in short-term memory. The output h_3 contains the information of input X_0, X_1 . However, Fig. 6 shows that RNN is not good at long-term memory. The output h_{t+1} cannot consider the information of input X_0, X_1 . RNN processes the next data by memorizing the recent data but it loses the information of previous data as time elapsed. This problem is called the problem of long term dependencies. As the distance between output and input increase, RNN cannot learn the information of input data.

LSTM

LSTM is a specific version of RNN. LSTM outperforms other RNN-based models (Hochreiter and Schmidhuber, 1997). It is useful because both the long term dependency problem and gradient vanishing problem which occurs during backpropagation are solved. LSTM sums the weights instead of multiplication to solve the vanishing gradient problem. Also, the model continuously transfers the information of historical data to solve long term dependency problem. The structure of LSTM is given in Fig. 7.

LSTM has 4 network layers for each module. It calculates the hidden layer using memory cell, instead of neural. The yellow box represents the trained network layer (hidden layer). The green circle indicates arithmetic operation such as vectored sum. The arrow is the flow of vector, which transfers the entire single vector from the output of a node to the input of another node.

LSTM is able to add or remove the information to cell state via the gate. It carefully controls this procedure in the gate. As shown in Fig. 8, LSTM updates the information selectively. The gate is responsible to add or remove information selectively and LSTM controls the gate to discard of memory the previous information. In addition, the gate adds or eliminates new information. The gate is composed of multiplication for each factor and the sigmoid network layer. The output of sigmoid layer is between 0 and 1, which indicates the number of factors to be passed. The gate discards or eliminates the information for output 0 whereas memorizes or adds the information for output 1. Figure 9 represents the LSTM network cell at time step t .

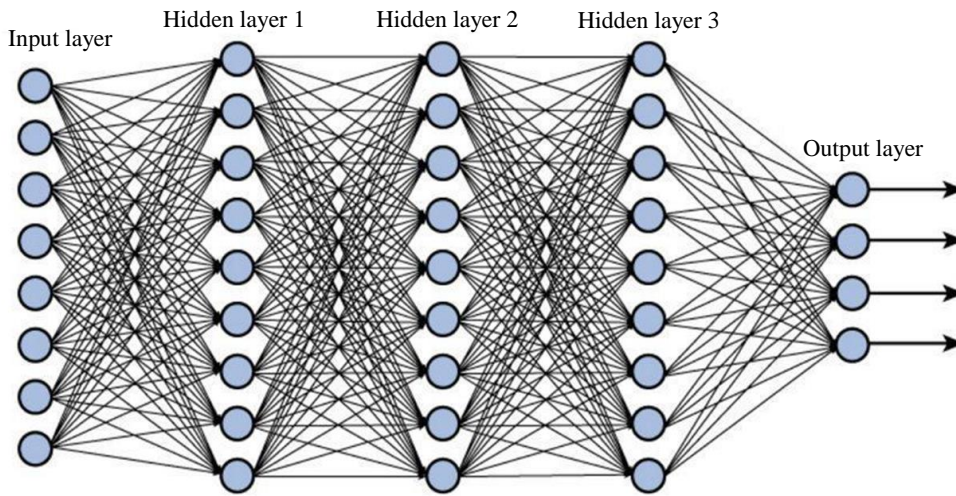


Fig. 2: A deep neural network with multiple layers

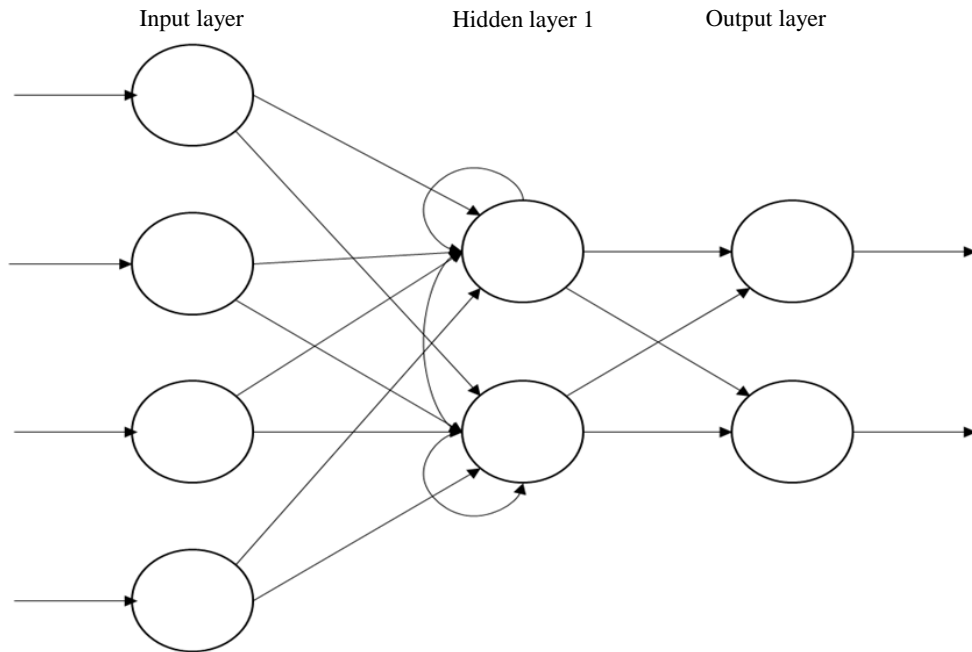


Fig. 3: The structure of RNN 1

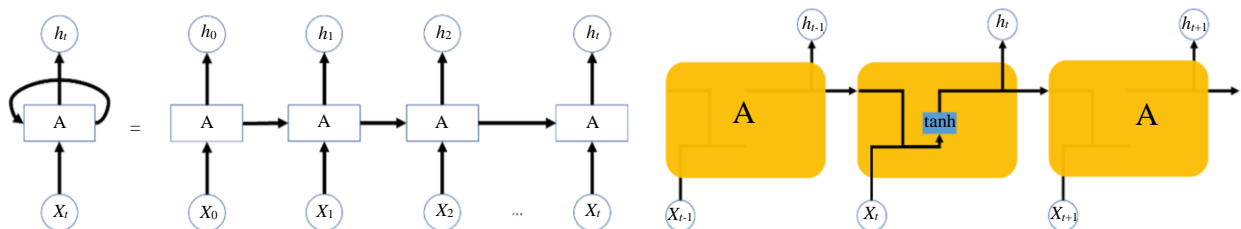


Fig. 4: The structure of RNN 2

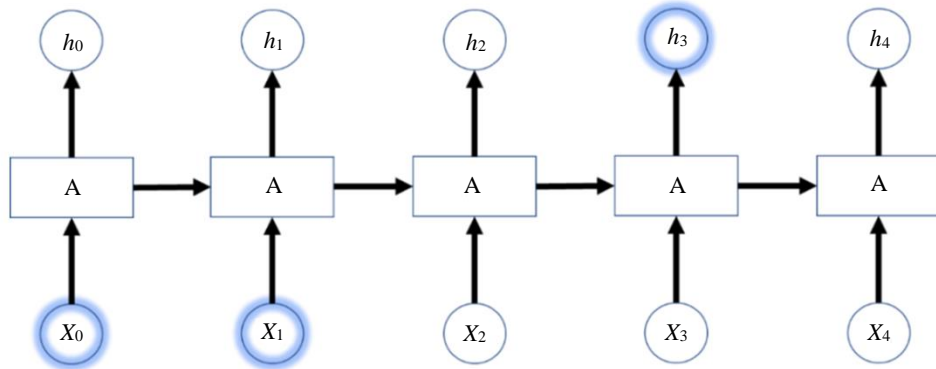


Fig. 5: Short-term memory of RNN

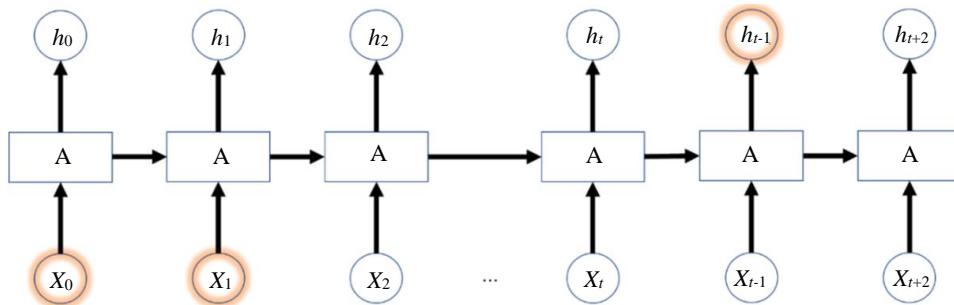


Fig. 6: Long-term memory of RNN

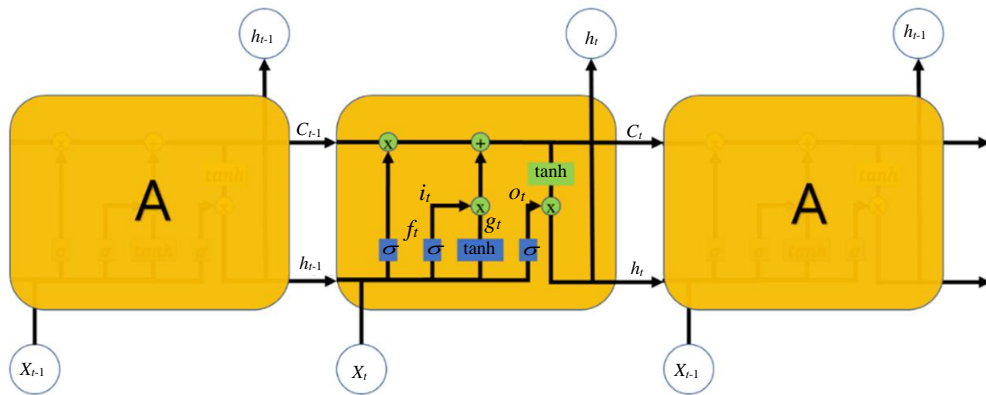


Fig. 7: The structure of LSTM

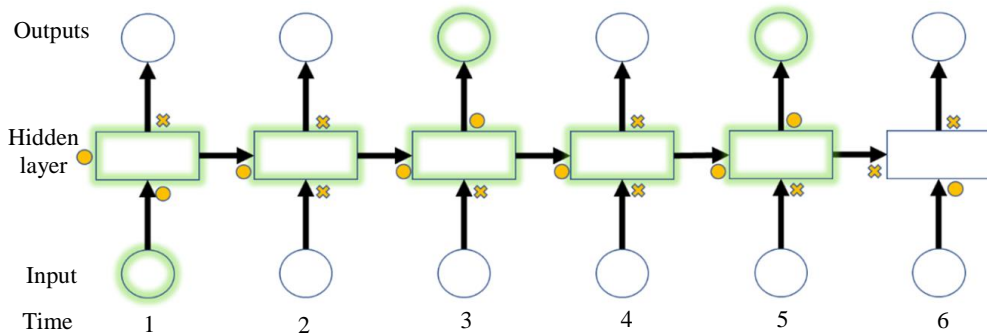


Fig. 8: LSTM neural network consisting of the unrolled LSTM cells

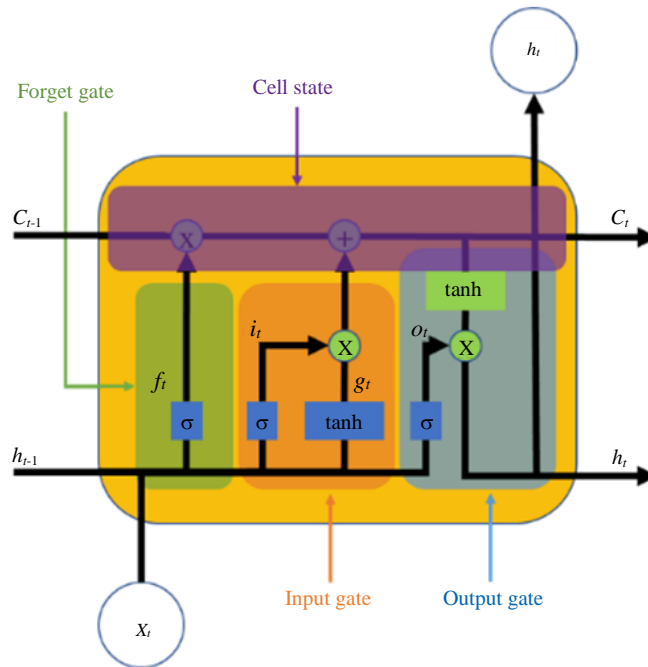


Fig. 9: LSTM network cell at time step t

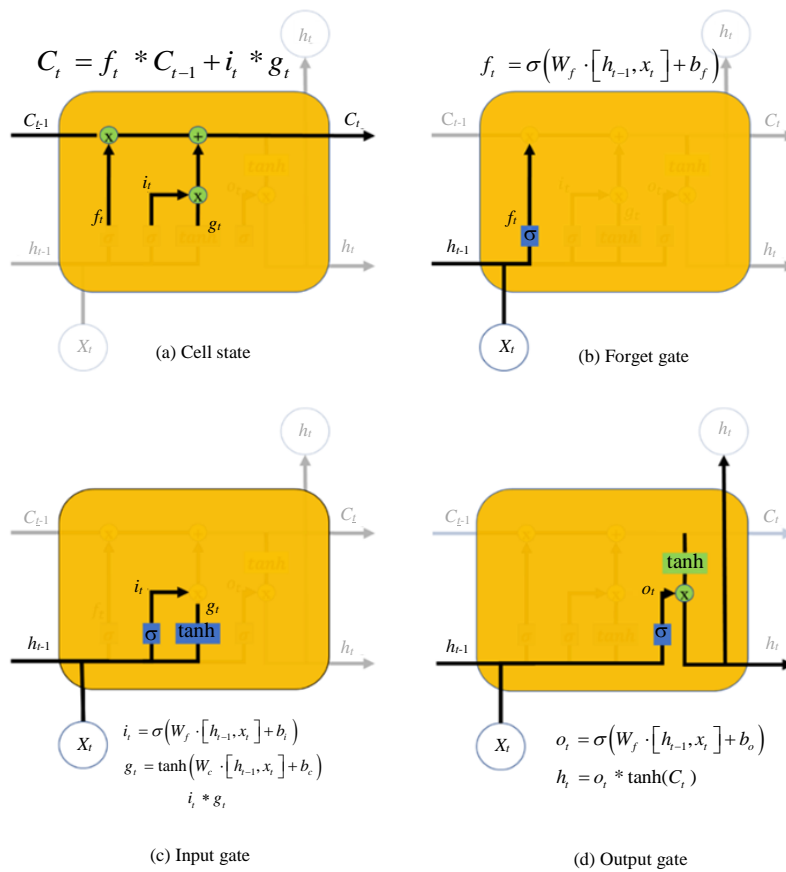


Fig. 10: The internal operation process of LSTM

Figure 10 depicts the operation process of LSTM's memory cell.

Forget gate determines which information of previous cell state to be eliminated. It is composed of sigmoid functions and behaves depending on the output of forget gate. Output 1 and 0 indicates retention and elimination, respectively.

Input gate determines whether to store the new data to cell state or not. In the input gate, the value to be updated is determined by sigmoid function and the vector to be added to cell state is generated by tanh function.

Cell state updates the previous cell state to a new state.

Output gate decides the final output. It outputs the filtered value based on cell state.

Methodology

The Proposed Framework

The overall research process is shown in Figure 11. First, we collect data about the number of airport passengers and then, the pre-process is performed for analysis. In that process, NaN data and abnormal data are removed and normalized is performed after extracting the necessary data and converting them to time series data. The pre-processed data is separated as training and test data. LSTM is trained with the training data set. Through the validation step with test data, the optimal model is produced. Using the optimal model, we predict the number of passengers at Incheon International Airport.

Model Development

We implemented the LSTM model using Tensor Flow which is an open source library developed and opened to the public by Google. For training LSTM, input variables, output variables, activation function and hyper parameters should be determined. Especially, since hyper parameter and activation function affects the

performance significantly, the careful tuning for these values is critical.

First, the input variables and output variables should be set before training. It is important for the user to determine the appropriate value considering the characteristics of data because input variables may contain output variables and the output changes according to the structure. In Tensor Flow, Basic RNN Cell and BasicLSTMCell are cell function for developing RNN. When the network is generated by a stacked layer of multiple cells, the network is called a deep neural network.

The hyper parameter is a variable that is must be tuned directly by the user for model learning in deep learning. The examples of hyper parameter are the sequence length, the number of input variables, the number of output variables, the number of neural network layers, the number of learning iterations and the learning rate. Sequence length determines the length of input data to be entered to model at a time. The sequence length decides how many previous inputs will affect the output. The neural network layer produces good results as it is stacked more, but when stacked too much, it may be slowed and generate error depending on the situation. Hence, the optimal value should be searched. Also, extremely small learning rate decreases the learning speed while too large value will make the model impossible to find the optimal value and terminated. Therefore, it is important to hold the appropriate value and the iteration number of learning.

The RMSE and MSE are commonly used as the loss functions in time series prediction. RMSE is the square root of the MSE. These values show the difference between the actual observation and the predicted value. The smaller the loss function is, the better the prediction accuracy is. Optimizer function is used to minimize the error value. Adam (Adaptive Moment Estimation) is generally used as an optimization function. It is faster and easier to use than the typical slope descent method.

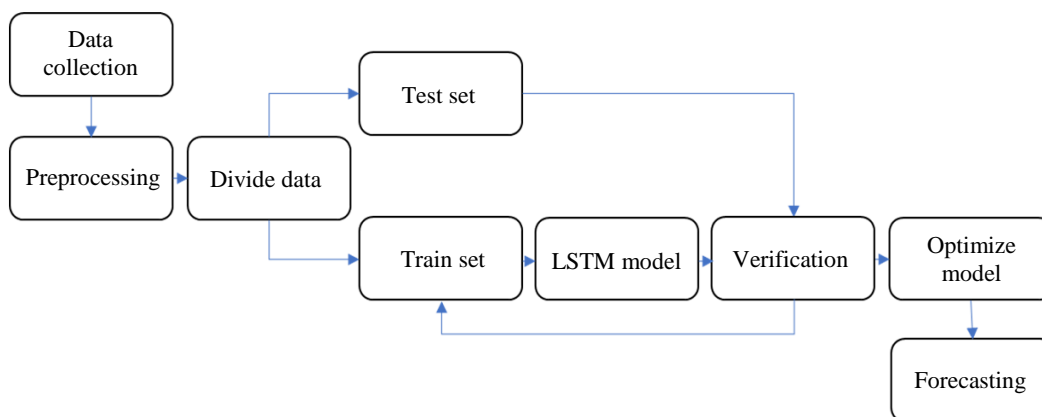


Fig. 11: The development of LSTM-based model

The RNN exploits the information in the previous step to interpret the information in the next step. For example, the number of passengers at an earlier time helps to understand the number of passengers at the next time. LSTM has the advantage of learning not only short-term time dependence but also long-term time dependency. In this study, we use BasicLSTMCell to estimate the number of passengers over a long period of time.

Overfitting often occurs during model training. Generally, this is due to complex data with a large number of parameters. Over fitted models can filter out very small errors, producing very good results in a learning data set even when applied to new data according to the situation, but in most cases, they show poor performance. Therefore, generalization should be done as much as possible in learning phase. Like hyper parameters tuning problem, the overfitting problem is also one of the most important challenges. One of the solutions is a dropout. This makes only a part of the whole neural network is used during the learning and each neuron is dropped out stochastically. Generally, the dropout rate is specified as 30% in the training phase and in the test phase the whole data is without dropout. If too many layers are stacked or are overfitting occurs, the ratio of drop out should increase. Otherwise, when underfitting is seen, the ratio should be lowered. Once dropped out, the data is not lost, but it is activated again when the learning is repeated. In other words, since some neurons are removed at each learning step, it prevents the specific neurons with the characters from being fixed. This balances the weights and finally makes the model avoid overfitting. Because dropout omits some neurons for learning, the learning phase takes longer. But it is worth the time and effort to get a good model. Figure 12 shows the dropout neural network model.

SARIMA-Based Model Design

In this study, we design SARIMA as presented in Fig. 13.

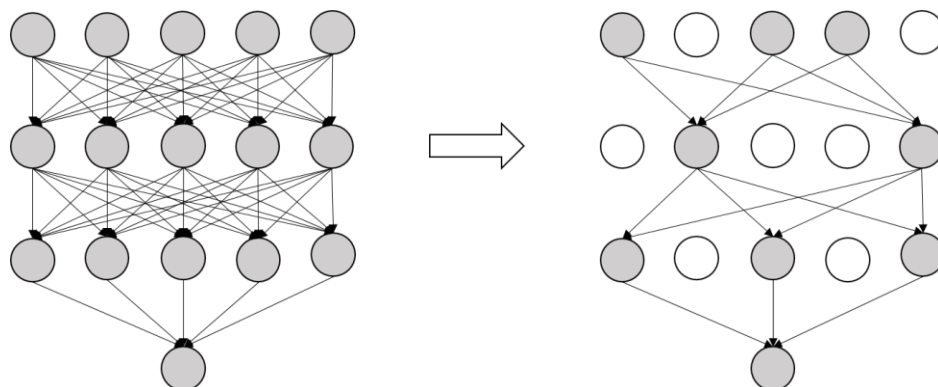


Fig. 12: Dropout neural network model: (i) A standard neural network, with no dropout and (ii) Neural network with dropout applied

In the identification step, we figure out the characteristics of data by visualizing the data with the graph. We determine whether the data has a seasonal pattern of normality. When the data does not have normality, the difference is performed until the data have normality.

In this study, the SARIMA technique is used because the time series data has seasonal characteristics. During this process, the parameter ‘d’ and ‘D’ of SARIMA (p, d, q) are determined. For other parameters, the order of non-seasonal time series model ‘p’ and ‘q’ and the order of seasonal time series model ‘P’ and ‘Q’ are calculated by using Autocorrelation Function (ACF) and PACF which is found in correlogram of time series data.

In the estimation step, we determine the most appropriate value among the values of ‘p’, ‘d’, ‘q’, ‘P’, ‘D’ and ‘Q’ calculated in the identification step and then the AIC value of the model made of the determined values is checked. The smaller the AIC value, the better. Representative estimation methods for estimating the parameters of each term include Least Squares Estimation Method, Nonlinear Estimation Method, Maximum Likelihood Estimation Method and Method of Moment Estimation.

In the verification step, the estimated model is verified statistically. Through the verification, it is judged whether the model is statistically significant from the overall point of view and whether it is appropriate as a prediction model. In this phase, the entire model is evaluated mainly by overfitting diagnosis and residual analysis. The ACF graph is plotted against the residuals and the model is valid when the residuals show the characteristics of white noise. If the estimated model seems invalid in the verification process, we repeat the identification step to estimate the model again. Otherwise, it is selected as the optimal model and used as a prediction model. Also, after the forecast period has elapsed, the accuracy of the prediction model can be confirmed by comparing the predicted values with the measured values and the results can be reflected in future predictions.

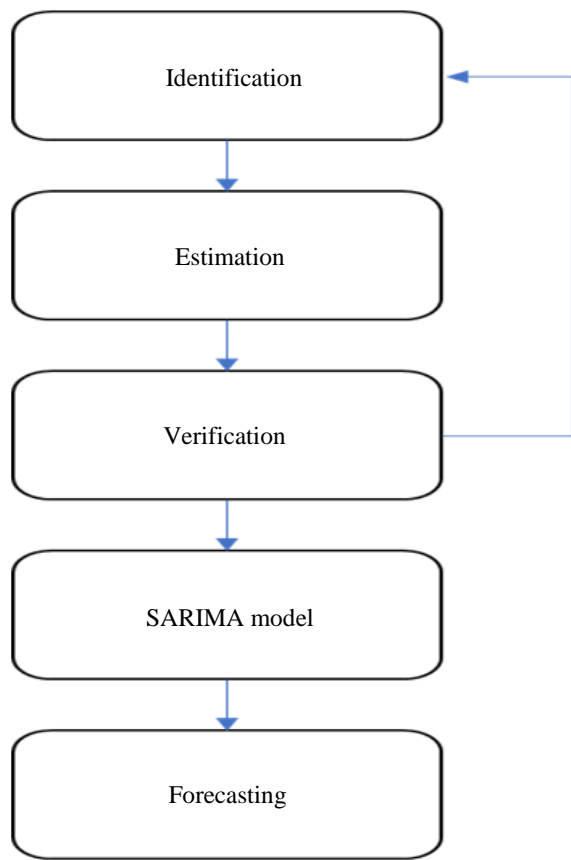


Fig. 13: The overall design procedure of SARIMA

A Case Application

LSTM Forecasting

Data Collection

The data of the number of passengers of Incheon International Airport is available at its website. The website also provides statistical aviation data including delay, cancellation and flight per day of the week, time, region and airline. Monthly data shows the number of flights, the number of passengers and the number of cargoes by country while daily data shows the number of passengers in Japan, China, Northeast Asia, Southeast Asia, the Americas, Oceania, Europe, the Middle East and other regions.

Data Preprocessing

The collected data is public raw data without any processing. Before analysis, the data should be pre-processed into the appropriate form. After finding outliers and NaN values, we removed them. The outlier data and the NaN data observed when the data is not normally transmitted due to a computational error of the

system or an error in the equipment. If we train a prediction model with the data containing outliers and NaN data, the model would show poor performance. Therefore, the outlier data is replaced to the mean value of the factor and the NaN data is treated as 0 to avoid training the model inaccurately.

After removing the abnormal data, we extracted the necessary data and transformed its form. Passenger data were categorized by Japan, China, Northeast Asia, Southeast Asia, the Americas, Europe, Oceania, the Middle East and others and then it is converted into time series data aligned with month and week units.

Before analyzing time series data, we normalized the data using Min-Max scaling. Since the number of holidays, the number of days and the number of passengers have different scales each other, they are normalized individually. Figures 14-16 present the steps of data pre-processing and the obtained data after pre-processing.

Short-Term Forecasting

The hyper parameters of LSTM were tuned as shown in Table 2.

The sequence length indicates the number of inputs. According to Table 2, the prediction is made with 12 previous data. The number of hidden layers is 200. The forget bias was 1 to reduce the amount of forgetting data at forget gate.

We set the number of stacked layers, the ratio of dropout and an epoch which indicates the number of iteration for training as 1, 10% and 240, respectively. The learning rate was 0.01.

Figure 17 shows the RMSE against epoch where the yellow line and the blue line indicates the training and test data, respectively. When the training was iterated 240 times, the minimum RMSE of training data was 0.050954554, whereas 0.06993915 for test data.

Figure 18 depicts the prediction result of the test data. The red and blue line means the ground truth and estimation result of LSTM, respectively. The graph indicates that the prediction traces the real value accurately. This model predicted that the number of passengers in next week (2018.12.30~2019.01.05) will be 1,369,368.

Mid-Term Forecasting

In the mid-term forecasting, most of the parameters were the same with daily prediction except for epoch. Since as epoch increases the accuracy of the model becomes high, we set the epoch as 630. The other parameters are set as shown in Table 3.

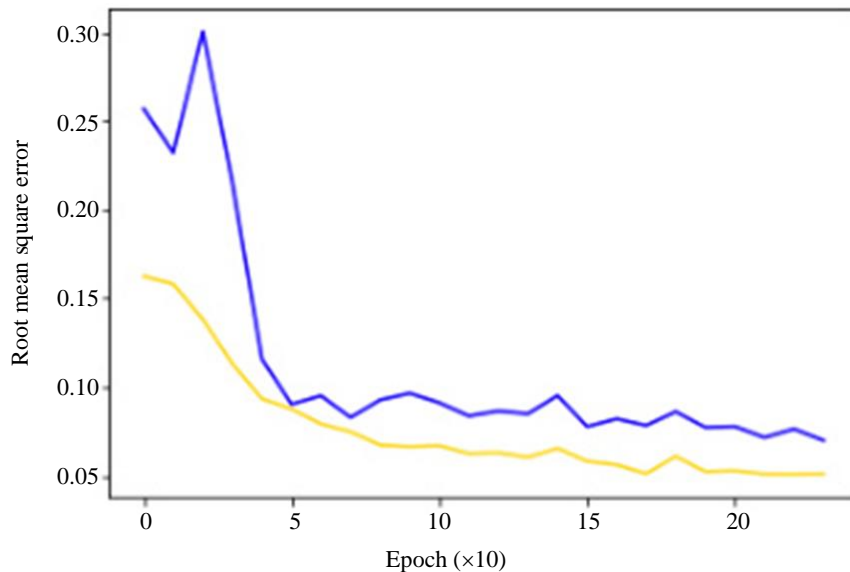


Fig. 17: RMSE of Short-Term LSTM Forecasting Model

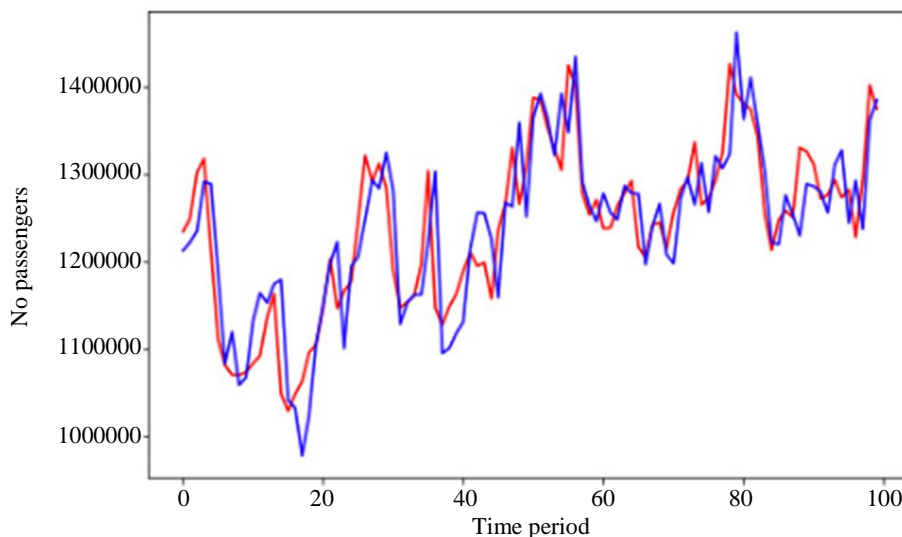


Fig. 18: Short-Term LSTM Forecasting

In Fig. 19, the yellow line is RMSE of training data against epoch whereas blue line is RMSE of test data. The RMSE of training data and test data were 0.017396053 and 0.047278523, respectively.

The red line in Fig. 20 means the ground truth while the blue line is estimated value. The graph shows that the estimation result follows the ground truth precisely. Using this model, the number of passengers in next month was predicted as 5,968,707.

SARIMA Forecasting

Short-Term Forecasting

We analyzed the trend or seasonal characteristics by plotting the weekly data and decomposing it Fig. 21. Fig.

22 shows the difference of weekly data (left), the weekly data after eliminating seasonal factor. Also, as shown in Fig. 24, the data is normalized by the difference.

There is no cut point in ACF, PACF. The ARIMA coefficient is checked by using Auto. Arima functionality provided by R Fig. 23. Auto Arima is an open package that automatically finds the order of SARIMA. We checked the residuals to validate that these coefficients are not abnormal. The analysis result shows that time-normalized residuals do not show any particular trend, autocorrelation functions have mostly zero values and Ljung-box tests have p-values higher than 0.05. Therefore, the null hypothesis "H0: autocorrelations are independent of each other" cannot be rejected. We set the ARIMA coefficient and predicted the number of passengers in the next 100 weeks Fig. 24.

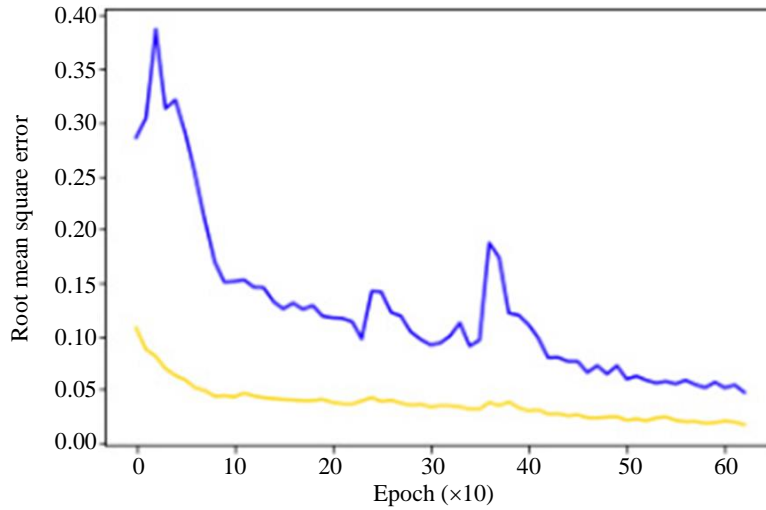


Fig. 19: RMSE of mid-term LSTM forecasting model

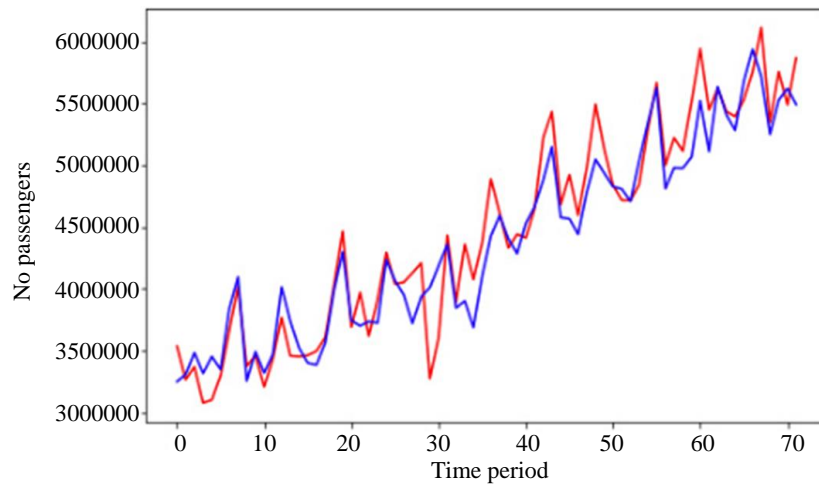
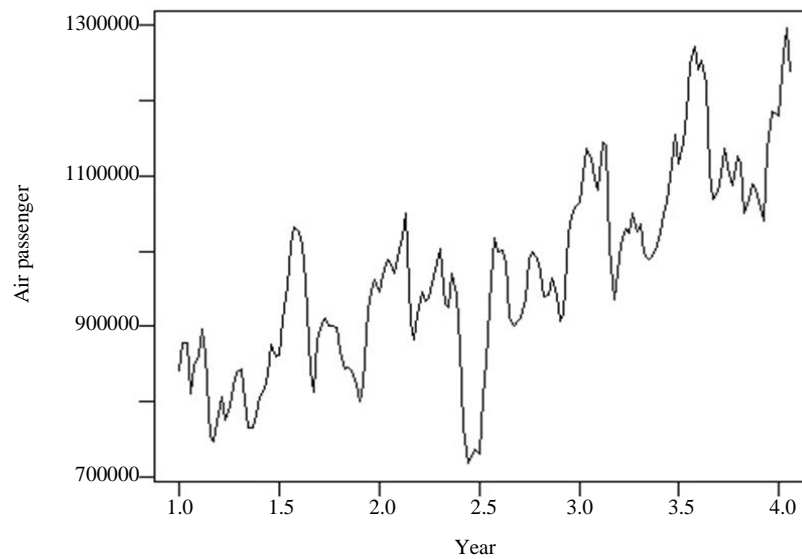


Fig. 20: Mid-Term LSTM forecasting



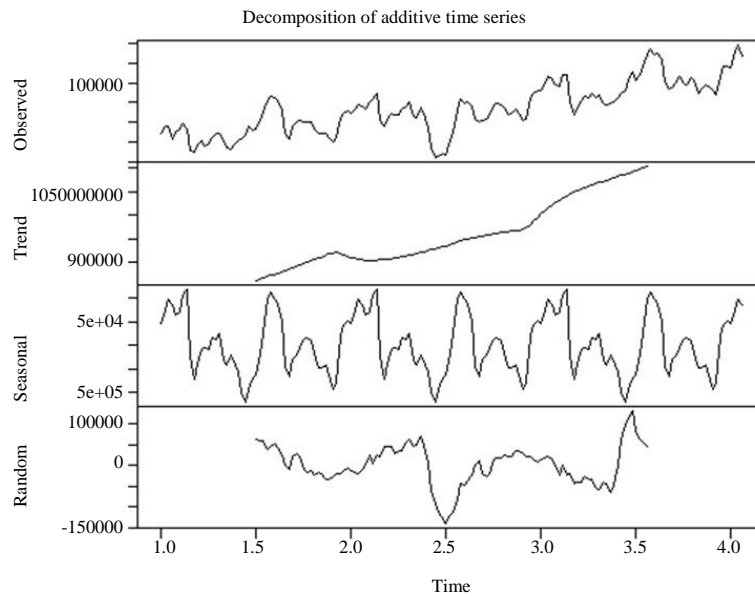


Fig. 21: Weekly data (left) and decomposed data (right)

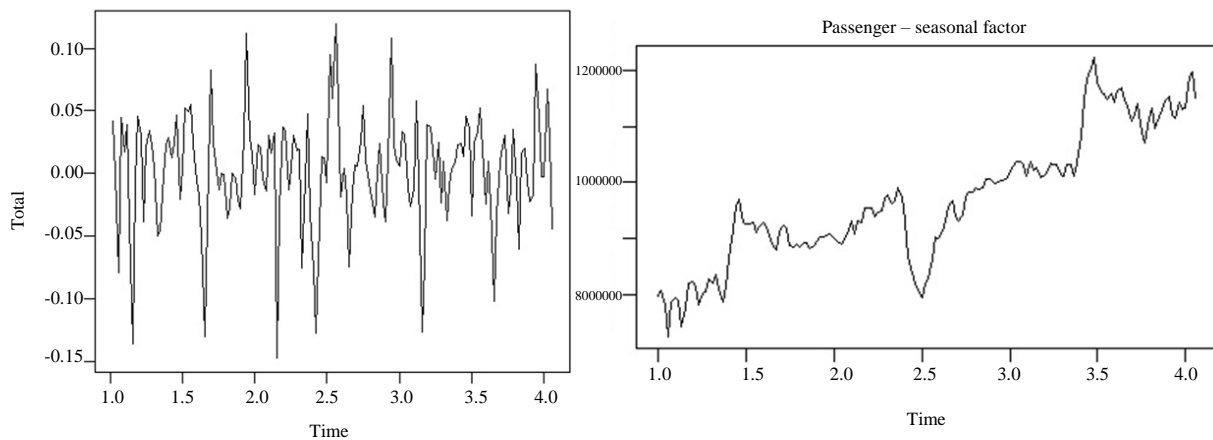


Fig. 22: The difference of weekly data (left), the weekly data after eliminating seasonal factor

The mid-term forecasting model is generated in the same way as the short-term prediction. We analyzed the trend or seasonal characteristics by visualizing monthly data as shown in Fig. 25.

The data is normalized through difference Fig. 26.

No cut point is found in ACF, PACF. The ARIMA coefficient is checked by using Auto Arima provided by R Fig. 27.

These coefficients are checked to determine whether there is any abnormality. Through the analysis, we observed that time-standardized residuals do not show any particular trend, most autocorrelation functions output zero values and Ljung-box test produces high p-

values. Therefore, the null hypothesis "H0: Autocorrelations are independent of each other" cannot be rejected. We input the ARIMA coefficients and made a forecast for the next 72 months Fig. 28.

Comparison Analysis

We compared the performance of LSTM with SARIMA using a statistical method such as RMSE, MAPE and MAD.

The Root Mean Square Error (RMSE) is a metric suitable for expressing accuracy when the target of the model is the estimated value or the difference between predicted and actual observations. Each difference value

is called a residual and the mean square deviation is used to integrate the residuals into a single metric. The equation of RMSE is as follows:

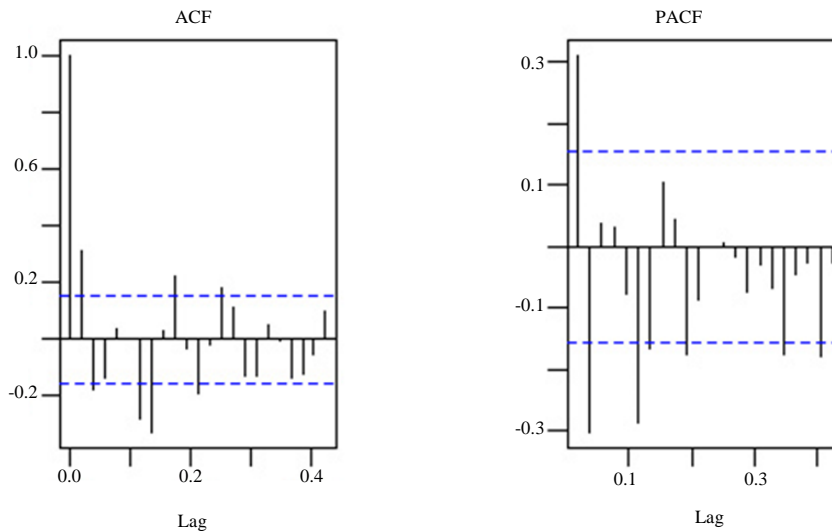
$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (10)$$

Mean Absolute Percentage Error (MAPE) represents the error ratio of prediction compared with the ground truth. Using this metric, the residuals between models can be compared. RMSE compares the magnitude of the error whereas MAPE compares the model in the ratio of error. It is also possible to compare the magnitude differences of the average error. The general formula is as follows:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (11)$$

Mean Absolute Deviation (MAD) is one of the scatter maps which indicates the average distance among the mean and individual observations. It is the absolute value of extracting the average of the total from each measurement, representing the arithmetic mean of the deviations. It is used to solve the problems caused by extremely small or large outliers. The equation is as follows:

$$MAD = \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (12)$$



ARIMA(0,1,1)(0,1,0)[52]

Coefficients:

 ma1
 0.2938
 s.e. 0.0794

sigma^2 estimated as 1.23e+09: log likelihood=-1271.14
 AIC=2546.28 AICC=2546.39 BIC=2551.62

Fig. 23: The ACF and PACF of weekly data (left) and its ARIMA coefficient (right)

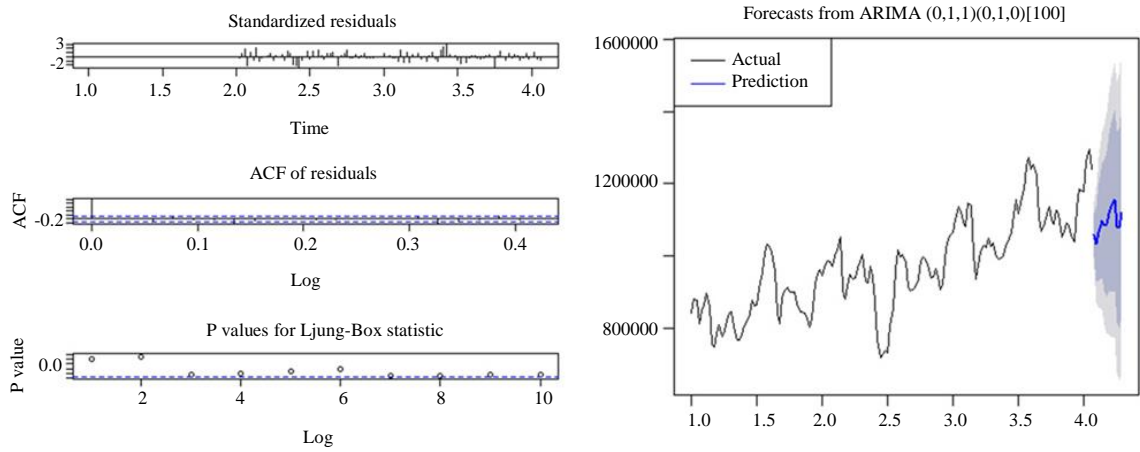


Fig. 24: The residual analysis for weekly data (left) and the prediction result of SARIMA (right)

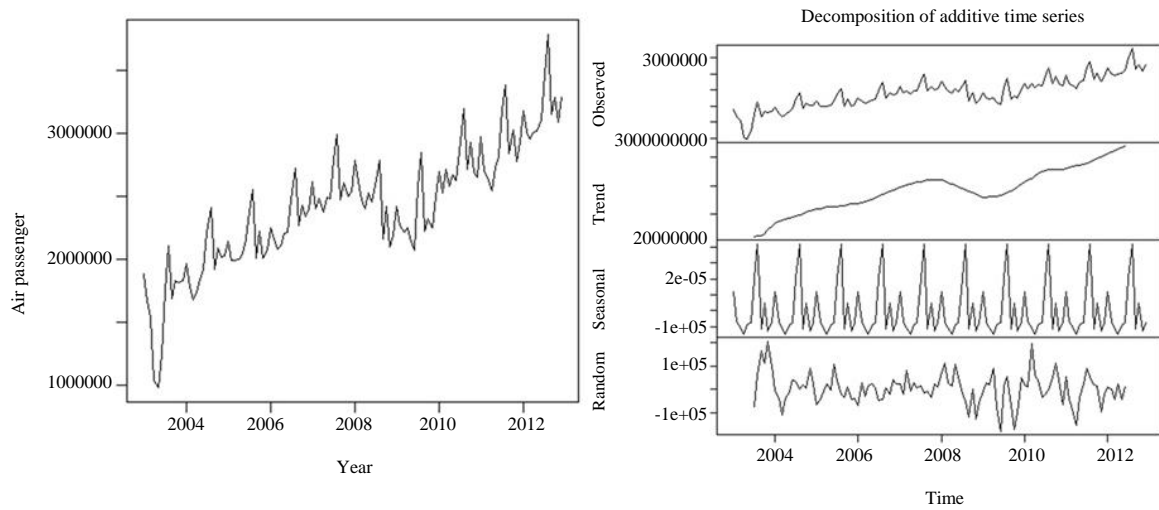


Fig. 25: Monthly data (left) and decomposed data (right)

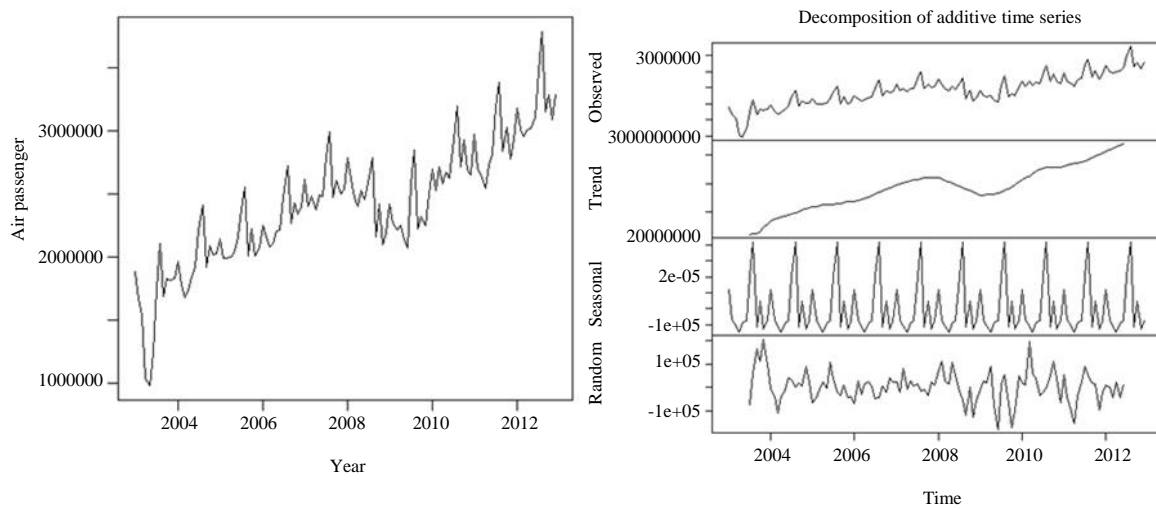


Fig. 26: The difference of monthly data (left) and the data after removing seasonal factor (right)

Table 4 comparison result of the short-term and mid-term prediction. In both forecastings, the RMSE, MAPE and MAD of LSTM are lower than SARIMA, indicating LSTM outperforms than SARIMA.

The variation in mid-term prediction was 4 and 9% for LSTM and SARIMA, respectively.

In the case of ARIMA, the lower and upper bound of prediction were 80 and 95%, respectively. The upper bound of the mid-term SARIMA model was also considered. In the mid-term prediction, the variation of LSTM, SARIMA and SARIMA (Hi-95) was 5, 29 and 12%, respectively. Figure 29 shows the actual values and the prediction values obtained from different models.

Conclusion

In this study, we developed a short-term and mid-term prediction model based on LSTM. The mid-term model forecasts monthly while short-term focus on weekly prediction. The RMSE from validation data verified that the performance is outstanding. We have

tuned parameters such as sequence length, hidden layer count, stack layer count, epoch, dropout rate and learning rate to generate a model with the highest accuracy.

The accuracy of the LSTM model was evaluated by comparing with the SARIMA model which is widely used for analyzing the time series data. Both models showed good performance in short-term prediction. We guess that this is because the amount of short-term data is sufficient to build a prediction model. The performance difference between the two models was noticeable in the mid-term prediction rather than the short-term prediction.

The LSTM model can be a powerful predictor because it is able to learn nonlinear data, has long-term memory and is less affected by normality, but it takes a relatively long time and requires a lot of data for high accuracy. It also has the disadvantage of tuning many parameters. The SARIMA model is less accurate than the LSTM but has the advantage of being relatively simple, time-consuming, with relatively good performance even with a small amount of data.

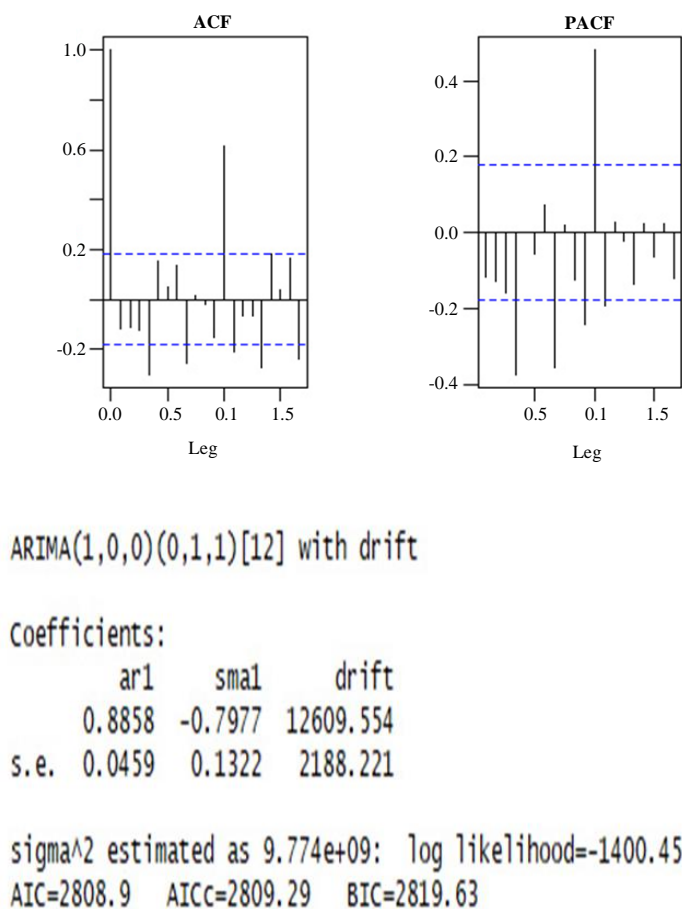


Fig. 27: ACF and PACF of monthly data (left) and its ARIMA coefficient (right)

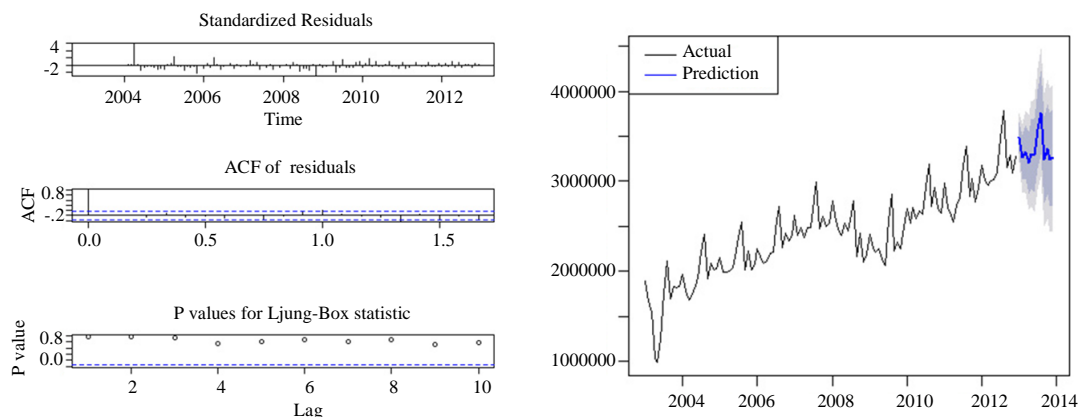


Fig. 28: Residual analysis for monthly data (left) and prediction result of SARIMA (right)

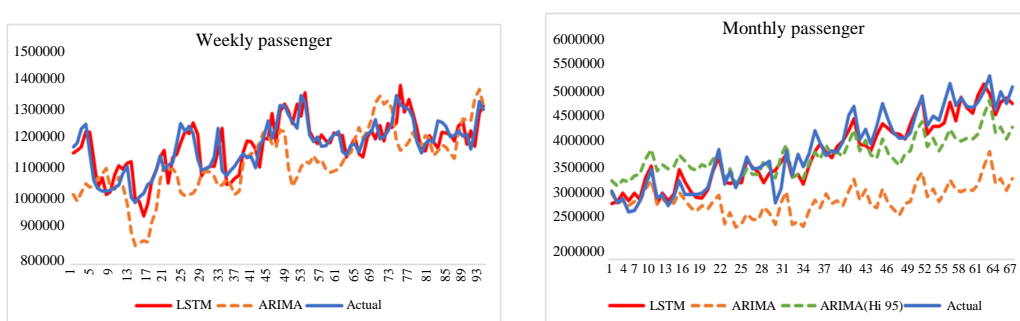


Fig. 29: The comparison of prediction result in short-term (left) and mid-term (right)

In this study, we presented the possibility of forecasting the number of passengers at the airport using LSTM. We expect that a more accurate model would be generated with more data and repeated tuning. We are also looking forward to making a synergy by combining with other forecasting techniques.

Acknowledgement

Authors would like to thank editors and anonymous reviewers for valuable comments.

Author's Contributions

Quang Hung Do conceived of the study. Jeng-Fung Chen conducted the literature review. All of the authors developed the research design and implemented the research. The final manuscript has been approved by all authors. All authors have read and approved the final manuscript.

Ethics

This paper is original and contains unpublished material. No ethical issues were involved and the authors have no conflict of interest to disclose.

References

- Abed, S. Y., Ba-Fail, A. O., & Jasimuddin, S. M. (2001). An econometric analysis of international air travel demand in Saudi Arabia. *Journal of air transport management*, 7(3), 143-148.
- Aderamo, A. J. (2010). Demand for air transport in Nigeria. *Journal of Economics*, 1(1), 23-31.
- Barboza, F., & Kimura, H. E. Altman. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-17.
- Bermúdez, J. D., Segura, J. V., & Vercher, E. (2007). Holt–Winters forecasting: an alternative formulation applied to UK air passenger data. *Journal of Applied Statistics*, 34(9), 1075-1090.
- Duval, D. T., & Schiff, A. (2011). Effect of air services availability on international visitors to New Zealand. *Journal of Air Transport Management*, 17(3), 175-180.
- El-Din, M. M., Farag, M. S., & Abouzeid, A. A. (2017). Airline passenger forecasting in Egypt (domestic and international). *International Journal of Computer Applications*, 165(6), 1-5.
- Grosche, T., Rothlauf, F., & Heinzl, A. (2007). Gravity models for airline passenger volume estimation. *Journal of Air Transport Management*, 13(4), 175-183.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

- Kim, S. (2016). Forecasting short-term air passenger demand using big data from search engine queries. *Automation in Construction*, 70, 98-108.
- Makridakis, S., & Hibon, M. (1997). ARMA models and the Box–Jenkins methodology. *Journal of Forecasting*, 16(3), 147-163.
- Nam, K., & Schaefer, T. (1995). Forecasting international airline passenger traffic using neural networks. *The Logistics and Transportation Review*, 31(3), 239-252.
- Profillidis, V. A. (2012). An ex-post assessment of a passenger demand forecast of an airport. *Journal of Air Transport Management*, 25, 47-49.
- Samagaio, A., & Wolters, M. (2010). Comparative analysis of government forecasts for the Lisbon Airport. *Journal of Air Transport Management*, 16(4), 213-217.
- Scarpel, R. A. (2013). Forecasting air passengers at São Paulo International Airport using a mixture of local experts model. *Journal of Air Transport Management*, 26, 35-39.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- Schultz, M., & Reitmann, S. (2019). Machine learning approach to predict aircraft boarding. *Transportation Research Part C: Emerging Technologies*, 98, 391-408.
- Suryani, E., Chou, S. Y., & Chen, C. H. (2010). Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework. *Expert Systems with Applications*, 37(3), 2324-2339.
- Tsui, W. H. K., Balli, H. O., Gilbey, A., & Gow, H. (2014). Forecasting of Hong Kong airport's passenger throughput. *Tourism Management*, 42, 62-76.
- Xie, G., Wang, S., & Lai, K. K. (2014). Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches. *Journal of Air Transport Management*, 37, 20-26.