Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Sequence to sequence analysis with long short term memory for tourist arrivals prediction

To cite this article: A A Rizal *et al* 2019 *J. Phys.: Conf. Ser.* **1211** 012024

Sequence to sequence analysis with long short term memory for tourist arrivals prediction

A A Rizal, S Soraya and M Tajuddin

Informatics, STMIK Bumi Gora, Cilinaya Mataram Nusa Tenggara Barat

Email: ashril.rizal@stmikbumigora.ac.id

Abstract. Prediction is one of the very important elements in decision-making. In general, the effectiveness of a decision depends on several factors. These factors cannot be observed when the decision is taken. Decisions should be taken based on existing and past data. Predictions can be done with two approaches. A first approach is a time-series approach. Time series model does not show the tendency of past data available. The second approach is the approach of showing a cause-effects method. This approach explained the occurrence of a situation (an explanatory method) by specific causes. The initial problem is how to make predictions model. In the beginning, to make predictions used the forecasting methods such as Autoregressive Integrated Moving Average. But, this method has limitations on waiver possibility of non-linear relationship, stationary and homokedastitas residual. At this time, forecasting methods of data with time-series have evolved with Neural Network approach. This research examines the prediction of tourist visits with the Long Short Term Memory (RNN LSTM) Recurrent Neural Network approach. The results of the research carried out by constructing a prediction model for tourist visits with the LSTM RNN using three models. The three LSTM models that are carried out namely LSTM regression, LSTM with sliding window and LSTM with time steps. There are no model that provides optimal results in terms of training and testing at once. The best results in the training process for predicting tourist visits were obtained using a regression model with RMSE 6529.42. Meanwhile, for the testing process, the best RMSE value is 13512.34 with the Sliding Window model.

1. Introduction

Natural resources and large-scale manufacturing industries do not exist on Lombok Island. Therefore, tourism is a mainstay sector in regional development. The contribution of the tourism sector is increasing every year. Foreign exchange in tourism areas has implications for people's income. The tourism sector contributes job opportunities in Lombok. These are the implication of tourist expenditure and investment in tourism. The positive impact of tourist spending on the economy is distributed to various sectors, not only hotels and restaurants. The distribution is also absorbed into the agricultural sector, industrial and craft sectors, the transportation and communication sector, the service sector and others [1]. Local governments generally prepare local tours on big events only. Even though, tourist visits are not only crowded in big events. As well as tourism actors, for example from the hotel, tourism or accommodation service providers and others. If tourist visits can be predict, tourism actors can prepare optimally.

Prediction is one of the most important elements in decision maker. The effectiveness of a decision depends on several factors that we cannot see when the decision is taken. The effectiveness of a decision is also based on current and past time data. Prediction carried out with two approaches. The first approach called the time-series model. Time series models do not show trends from past data. The second approach is an approach that shows a cause effects method. This approach explains the occurrence of a situation (explanatory method) by certain causes [2]

The problem is how to make predictions. At first, predictions were made by forecasting methods such as the Autoregressive Integrated Moving Average Model (ARIMA). This method has limitations on the possibility of non-linear relationships. ARIMA is also limited to stationary data. Now the data forecasting method with time-series has developed with the Neural Network approach. The ARIMA model and the Neural Network have differences. The ARIMA method better used to predict linear time series data. Neural networks are better used for linear and non-linear data.

There are two types of network models in the neural network. Network models in the neural network are feed forward and recurrent. Theoretically, both models approach non-linear functions. The characteristics of the feed forward network usually use an activation function that is fed forward. The forward feed carried out from input to output through the hidden layer. An example is Multilayer Perceptron. Mathematically feed forward applies static input-output mapping. The characteristics of Recurrent Neural Network (RNN) have at least one cyclic pathway from synaptic connections. All activities in neural networks carried out repeatedly and implement dynamic systems. RNN has a difference with feed forward networks. RNN is better at handling cases with inputs that have space and time structure. Recurrent Neural Networks Long Short Term Memory (LSTM) is specifically designed for sequential data that shows patterns over a period of time.

This research tries to examine time series data from tourist visits. Data studies conducted to predict tourist visits. The predictions are made with Recurrent Neural Network Long Short Term Memory approach.

2. Related work

This study refers to several references from previous studies. Makridakis [3] predicts time series data on Arima hybrid models with a Neural Network to predict time series data. The results show that the Arima model is good for predicting time series data that contains linear components. The results also show that ANN used to predict time-series data with non-linear components. The hybrid model of Arima-ANN has more accurate prediction accuracy than traditional Arima models. Meanwhile, Rumagit [4] tested the system on the implementation of the Arima method and Artificial Neural Network (ANN) and combined Arima and ANN. Data that is not stationary will be stationary using ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function).

The results of the study state that the largest calculation of MSE (Mean Square Error) and MAPE (Mean Absolute Percentage Error) occurs when using ANN. The values of MSE and MAPE when using the merge of ARIMA and ANN show mixed results from each input. In the following year, [5] conducted a study using the General Regression Neural Network (GRNN) method to predict the IHSG. The results showed that the MSE for the sample and outsourced IHSG predictions were 0.0136 and 0.0135. GRNN produces a smaller predictive MSE value than Arima. The strength of GRNN is faster in the calculation process and does not require data assumptions.

2.1 Neural network prediction

Other predictions applied the RNN [6] to forecast stock prices using Back Propagation Through Time (BPTT). The results of the trials conducted on forecasting stock prices using RNN-BPTT produce different error values. Rahmawati [7] applies the Elman Recurrent Neural Network (ERNN) and Principal Component Analysis (PCA) methods to predict electricity consumption. PCA used to determine the dominant factors that affect electricity consumption. The results of training with the network using ERNN have different parameters. The forecasting sample test results for the 5-year forecast period obtained an average value of MAPE for total consumption forecasting of 1 by 0.33%, total consumption of 2 by 0.64%, household 1.21%, industry 2.62%, business 3.25%, social 0.77% and public 0.49%.

Prediction uses RNN [8] by focusing research on online predictions. The task done is far more difficult than grammatical inference with neural networks offline. This work analysis uses discrete-time RNN and the RNN's ability to predict the next symbol in sequence. Barbounis [9] predict wind speed by using spatial information from a remote measurement station. The method used is the Local

Recurrent Neural Network. To improve the accuracy of predictions, online learning algorithms used based on Recursive Prediction Error (RPE).The RPE scheme was developed with weights being updated simultaneously. Simulation results show that this model shows good predictive results. The results are better than other types of networks.

2.2 Tourist arrivals predictions

Research on predictions of foreign tourist visits in the Yogyakarta City Museum by applying ARIMA, Genetic Algorithms (AG) and Neural Network (NN) is done by Setyaningsih [10]. The data used are time-series data that is not stationary so that predictions using ARIMA can not directly carried out, therefore differencing processes must done. Meanwhile, the chromosome formation in the AG model in the ARIMA model intended to get the best model parameters. Different from ARIMA, forecasting testing on the NN model does not need to do pattern recognition. In the following year, forecasted the number of tourist visits in the area at KUSUMA AGROWISATA BATU, Malang [11]. The method used in this study is the Box Jenkins method. This method uses the SARIMA model approach. SARIMA is a development of the ARIMA model. The first step is to look at stationary data. The next step is to identify the model from the ACF and PACF calculations. From the calculation of ACF and PACF a temporary ARIMA model can formed. After that, a model of parameter estimation carried out. The last step is diagnostic testing of residual results and normality.

The other tourist arrivals predictions use the Reccurent Neural Network with the Extended Kalman Filter. RNN-EKF provides better predictive results when compared to the Arima and ANN approaches. This research tries to predict foreign tourist arrivals on the island using the LSTM RNN approach.

3. Prediction methods

3.1 Recurrent neural network

Generally, there are two main types of neural networks, namely feed forward and recurrent. RNN has a difference with feed forward networks. But theoretically, both methods follow a non-linear function. The characteristic of feed forward networks usually uses an activation function that feed forward from input to output through the hidden layer. For example, like multilayer perceptron. Mathematically, feed forward applies static input-output mapping. Meanwhile, the characteristics of Recurrent Neural Network have at least one cyclic path with synaptic connections. All biological neural networks apply dynamic systems [12]. Cernansky [13] describing the basic part of a RNN neurons that are connected by synaptic links (connections) where the synaptic power is encoded by weight. In general, the layer on the RNN consists of input layers, hidden layers, context layers, and output layers. The input layer represented by u_i, the hidden layer with z_h, and the output layer with y_o. In addition, there is also a layer of context. Context layer (c_c) is a cycle process in which the value of neurons in the hidden layer when t will be used as input in neurons at the context layer at $t+1$. Jaeger [14] formally defined several types of RNN models with discrete models. Discrete-time mathematically as a map of iterations with time $t = 1, 2, 3, \ldots$.

3.2 Long short term memory

LSTM is part of the RNN. When training is done with back propagation through time, LSTM can reduce errors. LSTM allows the system to run repeatedly with very many steps (iterations), so it can open a more complex cause and effect relationship. The general LSTM scheme is shown in Figure 1. Figure 1 explains that LSTM starts with the dataset input process. From the input layer, the network will perform a feed forward process. The result will given an activation process and then stored in the block memory. From a memory block, a feed forward process performed to get the output layer value. This process is repeated until the error condition is met. In the next iteration, the neurons in the memory block will be reused as input values on the network.

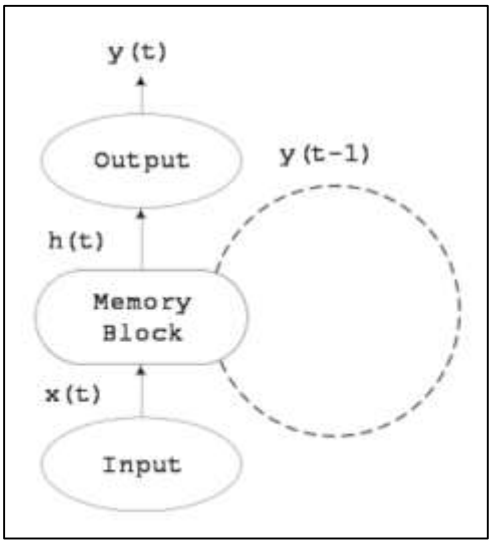

Figure 1. Long Short Term Memory (LSTM), starts with the dataset input process, From the input layer, the network will perform a feedforward process. From a memory block, a feedforward process is performed to get the output layer value.

LSTM contains information outside the normal flow of the recurrent network in the gate cell. The information generated can be stored, written or read from cells, such as data in computer memory. Cell gives a decision about the weight that must be stored. The cell also controls the permission to read, write and delete data through open or closed gates. This gate is analogous, implemented by element-wise multiplication by sigmoid at intervals [0-1]. Analog has the advantage over digital. Analog signals can be differentiated, making them suitable for backpropagation. The gate acts based on received signals and is similar to neurons on ANN. Gate blocks or transmits information based on the strength that goes into it and will be filtered into the weight of the gate itself. The weight is equal to the weights in the input and hidden layers which are adjusted through the learning process on the recurrent network. That is, cell learning when allowing data to enter, exit or be deleted through an iterative process. To produce guesses, errors in backpropagation are carried out by gradient differentiation.

3.3 Model Long Short Term Memory

Time series data analysis is done after the data is collected. Time series data analysis is done to predict tourist visits. The prediction model used is LSTM RNN. The following are the steps in the prediction process using the RNN LSTM method.

3.3.1. Input Dataset

The dataset consists of training data and testing data. The percentage of training data is 70% of the total dataset. Each neuron in the input layer represents an input vector that involves training data. Training data is stored in a comma-separated values file.

3.3.2. Normalization

Preprocessing in this research includes normalization of training data. Before making a prediction process, the input data will be normalized. Data normalization is done to adjust the network output according to the activation function used. These data are normalized into intervals [-1,1]. The interval is the value limit for the hyperbolic tangent activation function. Normalization follows equation (1) below.

$$n_i = \frac{2(x_i - x_{min})}{x_{max} - x_{min}} - 1 \quad (1)$$

where:
n_i = Normalized data
x_i = Data-i
x_{min} = Data with minimum value
x_{max} = Data with maximum value

3.3.3. Sliding Windows Design
The sliding windows design is done to see the sequence of data used. The sliding windows design will affect the architecture that will be used in the system. The design model will determine the input and output models of the LSTM RNN used.

3.3.4. Neural Network Process
In the RNN LSTM architecture, there are three types of weights. The weights in the RNN areweights from the input layer to the hidden layer, weights from the hidden layer to the output layerand weights from the context layer to the hidden layer. Besides the weight, the RNNLSTMprocess also initializes learning parameters. The learning parameters in the LSTM RNN in this research are the learning rate, epoch and activation functions. The training process will be stopped when the error value meets the target or the maximum iteration that has been set is met.

3.3.5. Denormalization
Before calculating the accuracy of the predicted results, the denormalization process is performed on the network output. Denormalization is done to get real value from the results of the prediction given. Meanwhile, accuracy is calculated by looking at the percent accuracy of the predicted results. Denormalization process is calculated by equation (2).

$$dn_i = \left((x_i + 1).\,(x_{max} - x_{min})\right) + \frac{2(x_{min})}{2} \quad (2)$$

where:
dn_i = Denormalization Data
n_i = Data - i
x_{min} = Data with a minimum value
x_{max} = Data with a maximum value

3.3.6. Output
The output of the system is the result of prediction and accuracy of prediction results. The results of prediction accuracy are calculated by analyzing the percentage of accuracy.

4. Result and Discussion
There are three types of gates in one unit:
1. Forget Gate : Conditionally, it will decide whether information must be forwarded to the block or not
2. Input Gate : Conditionally, determine the value of the input to be used to update state memory
3. Output Gate : Conditionally, determine the output based on input and blocking memory.

4.1. LSTM Regression Prediction
LSTM can be modeled into a linear regression model. Prediction by regression will analyze the number of arrivals in month *t* to predict arrivals in the month *t+1*. Prediction results with LSTM regression are shown in Figure 2. The blue line in figure 2 shows a graph of the actual data. The green line shows a prediction graph in the training process. And the red line shows the prediction results in the testing process.

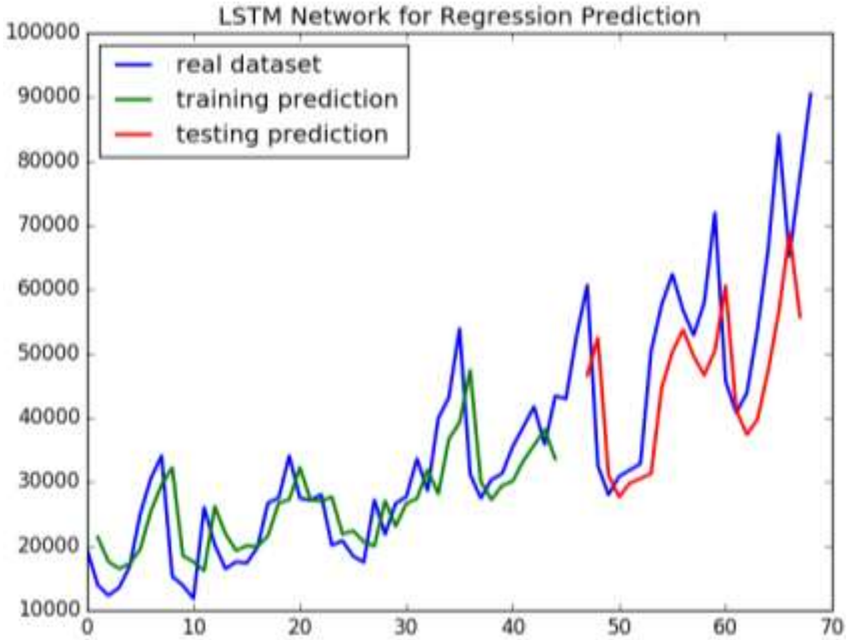

Figure 2. Graph of prediction result with LSTM regression. The blue line shows a graph of actual data. Meanwhile, the green line shows a prediction graph in the training process, and the red line shows the prediction results in the testing process.

4.2. LSTM With Sliding Windows

A prediction model with more than one-time variable to predict the next step is the sliding window model. For example, the value at t and value at $t+1$ is used to predict the value at time $t+2$. The model can be developed using the current time t and previous times $t-1$ as input variables to predict $t+1$. When made into a regression model, the input variable is $t-1$ and t and the output variable is $t+1$. The prediction results given are shown in figure 3. The blue line in figure 3 shows a graph of the actual data. The green line shows a prediction graph in the training process. And the red line shows the prediction results in the testing process. Errors given by the sliding windows method slightly increased from the regression method. This is because window size and network architecture affect predictive results.

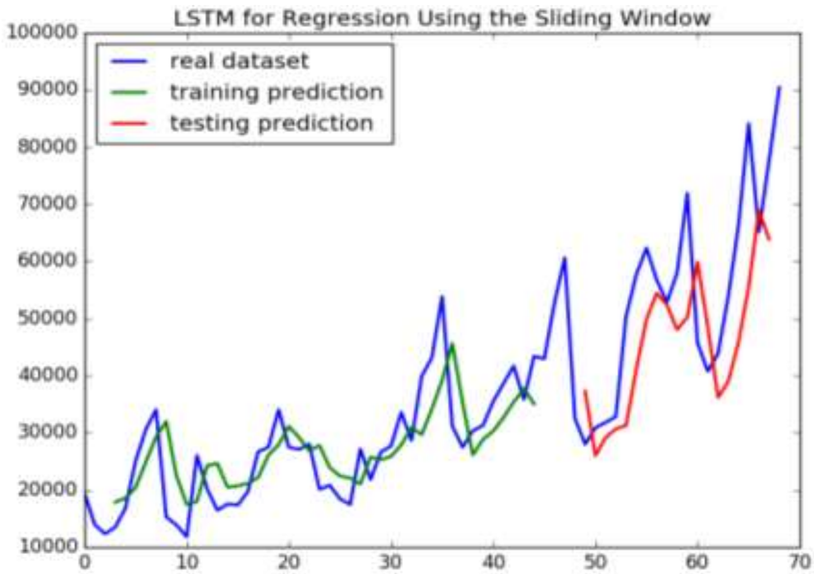

Figure 3. Graph of prediction result using sliding windows. Errors given by sliding windows method slightly increased from regression method.

4.3 LSTM with time steps

This method applies rolling projection or it called a walk-forward validation model. Each time step of the test dataset will run at the same time. The model is used to make estimates one time ahead. Actually, it same as conventional prediction. Observation of tourist visits will be available every month and will be used for the following month's estimates.

Observations are made on the previous three steps (*t-1*, *t-2*, *t-3*) and used as inputs to predict observations at the current time (*t*). Predictions using multi-time step follow the concept of a sequence to sequence with a longer sequence. The sequence used in this study is implemented as follows:

```
January      February    March     →    April
February     March       April     →    May
March        April       May       →    June
April        May         June      →    July
...          ...         ...      ...   ...
Month t-2    Month t-1   Month t   →    Month t+1
```

The prediction results given are shown in figure 4. The blue line in figure 4 shows a graph of the actual data. The green line shows a prediction graph in the training process. And the red line shows the prediction results in the testing process. Before the training process, a dataset transformation is carried out to get the scale according to activation requirements. In this research, rescale data was performed at a value between -1 and 1 to fulfill the activation function of the tangent hyperbolic tangent of the LSTM model. This transformation will be given an inversion process after the prediction process to return the value to the original scale.

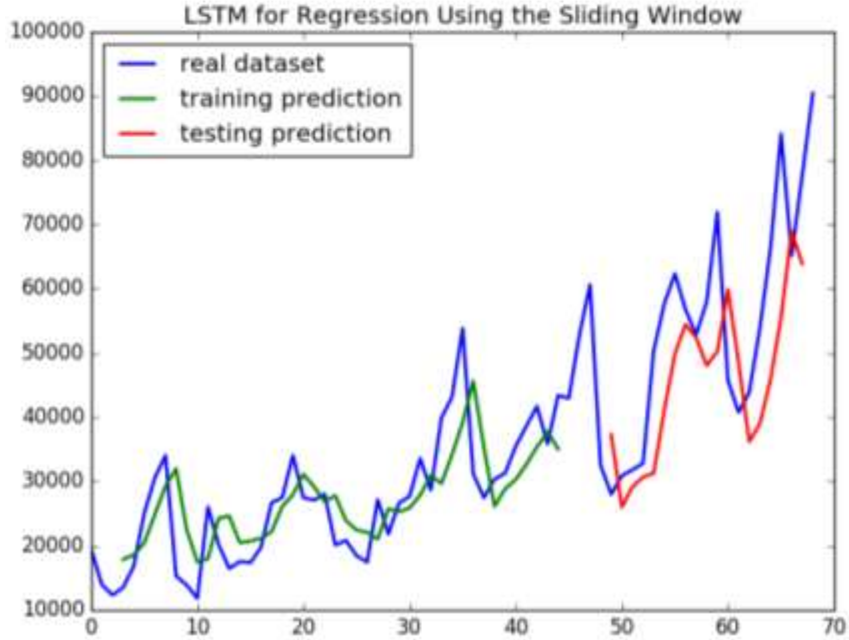

Figure 4. Graph of prediction result withtime steps. Tourist visits will be available every month and will be used for the following month’s estimates.

5. Conclusion

Based on the results of research for predicting tourist visits with Recurrent Neural Network Long Short Term Memory (LSTM RNN), there is no model that provides optimal training and testing results. The best results in the training process for tourist visit prediction were obtained using a regression model with RMSE 6529.42. Meanwhile, for the testing process, the best RMSE value was obtained at 13512.34 with the sliding window model. The RMSE results with a time step of 6888.37 in the training process and 14684.33 in the testing process.

Acknowledgement

We express our highest gratitude and appreciation to all members of the research team andDRPM Ristekdikti that gave the Beginner Lecturer Research grant so this research can be completed.

IOP Conf. Series: Journal of Physics: Conf. Series **1211** (2019) 012024 doi:10.1088/1742-6596/1211/1/012024

References

[1]Wijaya I N 2011 The Effect of Foreign Tourist Amounts, Length of Stay, and US Dollar Exchange Rate on the Receipt of Gross Regional Domestic Product of the Tourism Industry in Badung (Universitas Udayana).
[2]Rizal A A and Hartati S 2016 Recurrent neural network with Extended Kalman Filter for prediction of the number of tourist arrival in Lombok *2016 International Conference on Informatics and Computing (ICIC)* pp 180–185.
[3]Munarsih E 2011 Application of the Hybrid Arima-Neural Network Model for Time Series Forecasting (Universitas Gadjah Mada).
[4]Rumagit A 2013 Prediction of Electricity Usage Using Artificial Neural and ARIMA Networks in the Sullutenggo Region *IndoCEISS* **7** pp 139–148.
[5]Adnyani L P W and Subanar 2015 General Regression Neural Network on Forecasting Dollar Exchange Rates and Composite Stock Price Index *Factor Exacta* **8** pp 137–144.
[6]Susanti L A D, Fariza A and Setiawardhana 2011 Stock Price Forecasting Using Recurrent Neural Network With Backpropagation Through Time Algorithm *PENS-ITS* pp 1–8.
[7]Rahmawati T 2013 Application of the Elman Recurrent Network Method and Principal Component Analysis for Forecasting Electricity Consumption *Teknomatika* **5** pp 1–11.
[8]Pérez-Ortiz J A, Calera-Rubio J and Forcada M L 2001 Online text prediction with recurrent neural networks *Neural Process. Lett.***14** pp 127–140.
[9]Barbounis T G and Theocharis J B 2007 Locally recurrent neural networks for wind speed prediction using spatial correlation*Inf. Sci. (Ny).***177** pp 5775–97.
[10]Setyaningsih F A 2015 Comparison of Genetic Algorithms and Arima Statistical Methods for Time Series Data Prediction (Case Study: Museum Visits in Yogyakarta City) in *Prosiding Semirata 2015 Bidang Teknologi Informasi dan Multi Disiplin* pp 69–82.
[11]Lestari N et al 2012 Forecasting Tourism Visits with the SARIMA Model Approach (Case Study: Kusuma Agrowisata) **1**.
[12]Jaeger H 2013 ESNTutorialRev **2** 1–46.
[13]Cernansky M and Lubica B 2013 Recurrent Neural Networks trained by RTRL and Kalman Filter Algorithms*Neural Netw. World* **3** pp 223–234.
[14]Jaeger H 2005 A tutorial on training recurrent neural networks , covering BPPT , RTRL , EKF and the " echo state network " approach*ReVision* **2002** pp 1–46.