

Accurate Prediction of Data Traffic in Wireless Networks based on Deep Artificial Intelligence

Christopher Olabisi Olutobi Adeogun*

contact@doctoroluwatobiadeogun.com.ng

Adefemi Alabi**

Abstract: Accurate prediction of data traffic in telecom network is a challenging task for a better network management. It advances dynamic resource allocation and power management. This study employs deep neural networks including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) techniques to one-hour-ahead forecast the volume of expected traffic and compares this approach to other methods including Adaptive Neuro-Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN) and Group Method of Data Handling (GMDH). The deep neural network implementation in this study analyses, evaluates and generates predictions based on the data of telecommunications activity every one hour, continuously in one year, released by Viettel Telecom in Vietnam. The performance indexes, including RMSE, MAPE, MAE, R and Theil's U are used to make comparison of the developed models. The obtained results show that both LSTM and GRU model outperformed the ANFIS, ANN and GMDH models. The research findings are expected to provide an assistance and forecasting tool for telecom network operators. The experimental results also indicate that the proposed model is efficient and suitable for real-world network traffic prediction.

Keywords: Telecom Networks, Data Traffic Prediction, LSTM, GRU, ANFIS, ANN, GMDH

Introduction

With the development of mobile devices with Internet connectivity and network applications, the demand for telecom data has increased dramatically. Network operators are facing the problem of how to enhance Quality of Service (QoS) and end-user experience. Accurate traffic prediction plays an important role in the network management, network monitoring, routing optimization and other network activities. With an accuracy prediction of network traffic, the network congestion can be prevented and the utilization rate of the network can be maximized (Jiang *et al.*, 2015; Mahmassani, 2001). Analysis of historical traffic data to make accurate predictions is therefore a crucial task. A suitable prediction technique for particular application should be selected on the basis of the characteristics of the time series.

A time series is a sequence of numerical data points in successive order. The time series prediction problem is the prediction of future activity based on the historical values, associated patterns (i.e., trend analysis, cyclical fluctuation and issues of seasonality) and the current

value of the time series. Several existing time series prediction techniques can be considered for traffic prediction and these techniques can be grouped into two kinds of methods including conventional and intelligence computational techniques. Conventional techniques consider the base station traffic flow as discrete time sequences. Several well-known traditional methods are Auto Regressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SA), Markov Modulated Poisson Process models (MMPP) and Kalman filtering. A multiplicative seasonal ARIMA/GARCH model was proposed in describing and forecasting the mobile communication network traffic (Tran *et al.*, 2015). The data traffic was collected from EVN Telecom mobile communication network. The results showed that shows a good estimation when dealing with volatility clustering in the data series. Gao *et al.* (2017) utilized multiplicative seasonal ARIMA and Holt-Winters models to model traffic prediction. The two methods were used to analyze the trend of mobile network traffic per hour, build and validate models and then predict mobile network traffic within a given period of time. It

was indicated that Holt-Winters model is better than multiplicative seasonal ARIMA model. The authors also concluded that a prediction model in reality is decided by the different characteristics of the specific data and a specific model cannot always be the optimal model for all. Although the conventional forecasting models have been successfully applied to prediction, these models have somewhat several limitations in their ability to make predictions in certain situations (Hernandez Benet *et al.*, 2016; Park and Lee, 1995). In recent years, intelligence computational techniques including machine learning techniques, wavelet transform and fuzzy methods have been proposed to form analysis method for prediction. These methods can improve the quality of predictions and deal with high-frequency non-linear data (Lu and Wang, 2006). Gowrishankar and Satyanarayana (2009) predicted network traffic using neural network and statistical methods. The performances of both methods were compared on different time scales. The obtained results indicated that neural network predictors better than statistical models. Tian *et al.* (2017) analyzed the network traffic based on the hybrid neural network model. The proposed model can predict the network traffic more accurately. Abdulkarim and Lawal (2017) proposed a new prediction model based on a cooperative neural network strategy in order to improve the prediction accuracy of mobile data in UMTS-based mobile data networks. The proposed method produced superior results in comparison with the results obtained on the same problems from the traditional method. Kaushik *et al.* (2018) presented a Deep Neural Network (DNN) model for evaluating and predicting traffic activity levels based on the telecommunication activity of calls, SMS and Internet. The model produced predictions for the traffic activity levels along with their probabilities which match the expected traffic levels with an efficiency of 98.6-99.8%.

Currently, Recurrent Neural Network (RNN), a deep learning model, has been considered one of the most effective models in processing sequential data. Long Short-Term Memory (LSTM) is the most successful RNN architectures. LSTM has a unit of computation, a memory cell that replaces traditional artificial neurons in the hidden layer of the neural network. Thus, LSTMs can grasp the structure of data dynamically over time with high prediction capacity. Bian *et al.* (2018) proposed an ensemble method based on LSTM method for time series prediction with the aim of increasing the prediction accuracy of Internet traffic time series. The results showed that the results show that the proposed method was more accurate. In comparison with different techniques, the result showed that the actual values and predicted values by LSTM model were in good agreement and accurately revealed the future developing

trend of telecom traffic data. Gated Recurrent Unit (GRU) Network, a variant of LSTM, was proposed to make the implementation of RNN simpler. The previous studies demonstrated the superiority of both LSTM and GRU models (Chung *et al.*, 2015; Fu *et al.*, 2017). Therefore, no concrete conclusion was drawn on which kind of DNNs is the most efficient model for predicting and analyzing data traffic in telecom networks. The objective of this study is to investigate the effectiveness of LSTM and GRU deep neural networks in time series forecasting of telecom traffic. The main contributions of this study are the followings: (1) Evaluating two forecasting deep neural network models based on LSTM and GRU by comparing and analyzing through several performance indexes; (2) developing the model using real data from Viettel Telecom, Vietnam; and (3) proposing a suitable prediction model of telecom traffic for Vietnam.

Deep Neural Network

An ANN is a mathematical model to simulate the network of biological neurons that mimic a human brain so that the computer will be able to learn things and make decisions in a humanlike manner. Recently, deep learning or deep neural network (DNN) gained a great interest from practitioners and academia (Le *et al.*, 2019). 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. 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. RNN is an artificial neural network which solved the problem of traditional neural network. It is powerful to handle sequential data. RNNs are networks with inner loops at the hidden layers, allowing information to persist (Schmidhuber, 2015). In a traditional ANN, it is assumed that all inputs (and outputs) are independent of each other. Whereas, RNNs perform the same task for every element of a sequence, with the output being depended on the previous computations. LSTM, a specific version of RNN, outperforms other RNN-based models (Hochreiter and Schmidhuber, 1997). It is efficient 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.

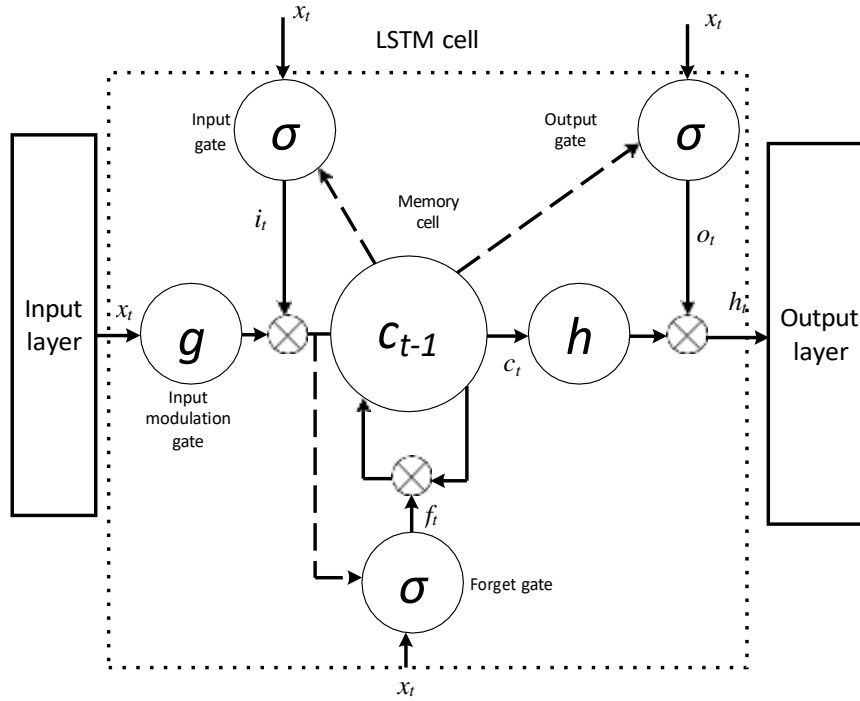


Fig. 1: The structure of LSTM cell

A typical structure of LSTM cell is shown in Fig. 1. The cell consists of four gates: Input gate, input modulation gate, forget gate and output gate. Cell state is modified by the forget gate placed below the cell state and also adjust by the input modulation gate.

Suppose that the input time series is $X = (x_1, x_2, \dots, x_n)$, hidden state of cells is $H = (h_1, h_2, \dots, h_n)$ and output time series is $Y = (y_1, y_2, \dots, y_n)$:

$$h_t = H(W_{hx} x_t + W_{hh} h_{t-1} + W_{hc} c_{t-1} + b_h) \quad (1)$$

$$p_t = W_{hy} y_{t-1} + b_y \quad (2)$$

where, W is the weight matrix, b denotes the bias vector. The hidden state of memory cells is calculated as follows:

$$i_t = \sigma(W_{ix} x_t + W_{ih} h_{t-1} + W_{ic} c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{fx} x_t + W_{fh} h_{t-1} + W_{fc} c_{t-1} + b_f) \quad (4)$$

$$c_t = f_t * c_{t-1} + i_t * g(W_{cx} x_t + W_{ch} h_{t-1} + W_{cc} c_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(W_{ox} x_t + W_{oh} h_{t-1} + W_{oc} c_{t-1} + b_o) \quad (6)$$

where, $*$ denotes the scalar product of two vectors, g and h are the extend of the standard sigmoid function with

the range in $[-2, 2]$ and $[-1, 1]$, respectively. σ Is the standard sigmoid function as follows:

$$\sigma(x) = \frac{1}{1 + e^x} \quad (7)$$

The objective function is the square loss function given by the following equation:

$$e = \sum_{i=1}^n (y_i - p_i)^2 \quad (8)$$

where, y_i is the actual output and p denotes the forecasting (predicted) value. In the training phase, a modified stochastic gradient descent with adaptive learning rate, Adam optimizer, is utilized to minimize the training error and avoid the local minimum points. In order to reduce the overfitting, dropout method is used to regularize a deep neural network (Srivastava *et al.*, 2014).

The GRU, proposed by (Cho *et al.*, 2014), is similar to a Long Short-Term Memory (LSTM) with forget gate but has fewer parameters than LSTM, as it is without an output gate. GRU has two gates including reset r and update gate z . Update gate determines how much previous state memory to keep, whereas reset gate determines how to combine new input values with old cell state. Since GRU has less parameters, it trains faster and is little more efficient than LSTM. The hidden state output at time t is calculated using of the hidden state at time $t-1$ and the input time series value at time t , as follows:

$$h_t = f(h_{t-1}, x_t) \quad (9)$$

The Research Framework

This section provides the steps in developing the prediction model. Figure 2 presents the research framework. After data collection step, data is preprocessed through feature scaling. Data is scaled according to the following equation:

$$v' = \frac{v - \min(x)}{\max(x) - \min(x)} \quad (10)$$

where, x is the list of data values. The collected dataset is then divided into two groups - training and testing datasets. The training and testing dataset are used for the construction and validation of the models, respectively. In this study, 70 and 30% of the dataset is for training and testing, respectively. The model development is done by the use of training dataset. In the model development, the number of layer, the optimizer, learning rate and loss function need to be identified. This stage is to design the structure of the model and adjusting the parameters. Calculated data is then scaled back to original scale to calculate error. Through the validation step with testing dataset, the optimal model is

obtained. Using the optimal model, the time series forecasting is conducted.

The Model Development

This section presents the detail of major steps in model development. It also provides the data pattern used in the study. The model parameters and evaluation criteria are also presented.

Data Description

The data used in model development is telecommunication activity dataset released by Viettel Telecom - the largest telecommunications service provider in Vietnam - from October, 2017 to October, 2018. The dataset consists of call in and out activity, SMS in and out activity and Internet traffic activity records and logs. If the user is involved in a telecommunication activity, an eNodeB will handle the interaction and communication via the access network and delivers the signal information (control plane) and data (data plane) to the core network. All this information and communication are recorded in the form of Call Detail Records (CDR). The Internet or data activity occurs and is recorded whenever the user connects or disconnects to the internet network. A CDR is generated and aggregated every 60 min for all activities

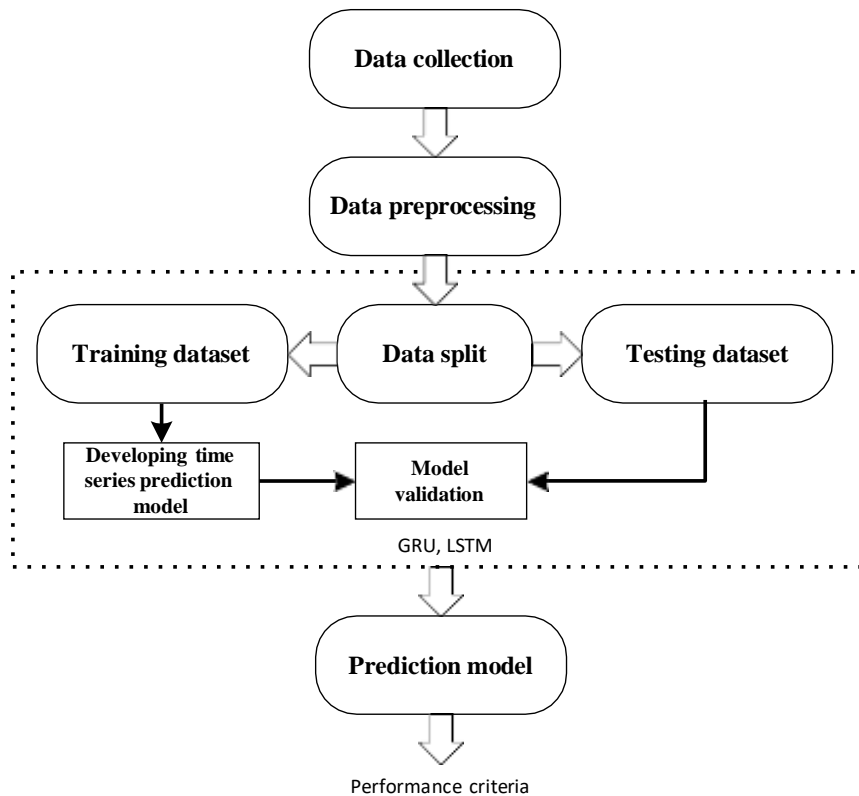


Fig. 2: The research framework

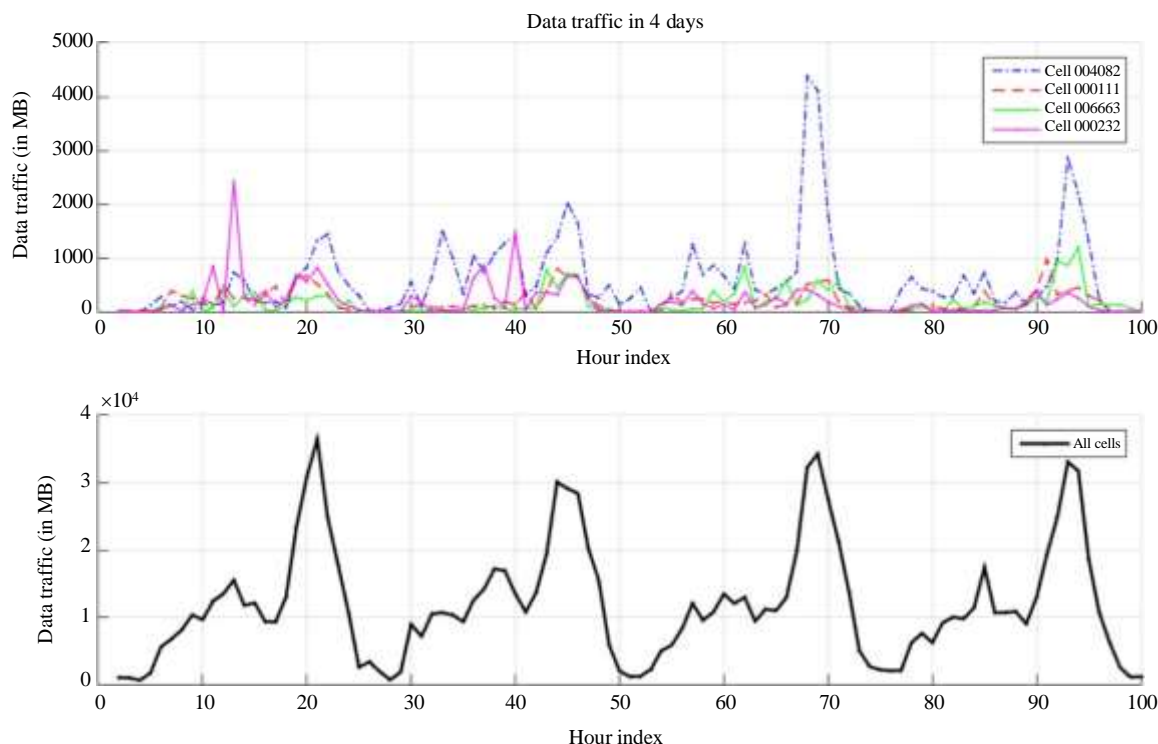


Fig. 3: Daily data traffic pattern

Basically, telecomm network traffic shows cyclical characteristics. There are shorter and longer cyclical effects on the load data, e.g., daily, weekly. Also monthly and even seasonal trends can be observed. In Fig. 3, we show busy-hour load (daily peak load) curve of Viettel Telecom network over the period of four days. Figure 3 also presents the data traffic pattern of four individual cells. The total traffic data represents the aggregated hourly measurement from 57 different cell sites of Viettel UMTS-based cellular operator. The nature of traffic pattern will vary. The load is lower at weekend than on working day. It is noticed that that on working days, social activities are at a higher level than on Saturdays and Sundays and therefore the load is also higher. So, in load forecasting, days are often divided into several day types, each of which has their own characteristic load patterns. It is clear that Saturdays and Sundays have different load curves than other days. The pattern is also different for special events, such as Lunar new year, long weekends and holidays, especially New Year's Eve, Valentine's Day. When a user uses a telecom service, the mobile device will be served by a nearby cell. The total data capacity of all users served by a cell within an hour is called the traffic of that cell within one hour. For example, cell 000232 is serving 50 subscribers; each subscriber in one hour uses an average of 10 Mb. So the traffic of this cell in one hour = $50 \times 10 = 500$ Mb.

Experimental Structure

All the models were coded in Python (Anaconda Python 3). To develop LSTM model, the number of layer was set equal to 5. The sequential structure with a linear stack of layers was applied to the model development. The main parameters were set as follows: The dimensionality of the output space is 50; the rectified linear unit was used as the activation function. The cost function was Mean Squared Error (*MSE*).

Other models were also developed to the research problem. In GMDH model development, the maximum number of layers was set to 5; the maximum number of neurons in a layer was equal to 50. When developing GMDH model, one parameter called "Selection pressure" is a proper threshold value to determine the number of neurons in each layer. After calculating the coefficients for all the neurons, those which produce the poorest performance according to the selection criterion (*MSE*) will be eliminated from the layer. The selection pressure is from 0 (no pressure) to 1 (the maximum pressure of selection). In this study, the selection pressure is set to 0. For ANN model, the Multilayer Peceptrons (MLP) was utilized in this study. The used MLP is feed forward fully connected network, with two hidden layers, along with Levenberg-Marquadt training algorithm. Two most commonly nonlinear and linear transfer functions sigmoid and tangent, were used in first

and second hidden layers, respectively. After a number of trial, the architecture of the MLP network for prediction was determined. For ANFIS model development, the grid partition method was used for Fuzzy Inference System (FIS) generation. In our study, two membership functions were chosen for each input in the model. The Gaussian membership function was used in ANFIS model. The membership function output is linear. Use lagged versions of the variable, five input variables are given [(x-5), (x-4), (x-3), (x-2), (x-1)].

Model Evaluation

To evaluate the performance of the prediction model, several performance indexes were used. These criteria are applied to the developed model to know how well it works. The criteria were used to compare predicted values and actual values. They are as follows:

Root Mean Squared Error (RMSE)

This index estimates the residual between the actual value and desired value. A model has better performance if it has a smaller *RMSE*. An *RMSE* equal to zero represents a perfect fit:

$$RMSE = \sqrt{\frac{1}{m} \sum_{k=1}^m (t_k - y_k)^2}, \quad (11)$$

where, t_k is the actual (desired) value, y_k is the predicted value produced by the model and m is the total number of samples.

Mean Absolute Percentage Error (MAPE)

This index indicates an average of the absolute percentage errors; a model has better performance if it has a smaller *MAPE*:

$$MAPE = \frac{1}{m} \sum_{k=1}^m \left| \frac{t_k - y_k}{t_k} \right| \quad (12)$$

Mean Absolute Error (MAE)

This index indicates how close predicted values are to the actual values. a model with a lower *MAE* means it has better performance:

$$MAE = \frac{1}{m} \sum_{k=1}^m |t_k - y_k| \quad (13)$$

Correlation Coefficient (R)

This criterion reveals the strength of relationships between actual values and predicted values. The correlation coefficient has a range from 0 to 1 and a model with a higher *R* means it has better performance:

$$R = \frac{\sum_{k=1}^m (t_k - \bar{t})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^m (t_k - \bar{t})^2 \cdot \sum_{k=1}^m (y_k - \bar{y})^2}}, \quad (14)$$

where, $\bar{t} = \frac{1}{m} \sum_{k=1}^m t_k$ and $\bar{y} = \frac{1}{m} \sum_{k=1}^m y_k$ are the average values of t_k and y_k , respectively.

Theil's U-Statistic

This index is an accuracy measure that emphasizes the importance of large errors as well as providing a relative basis for comparison with naïve forecasting methods. Theil's equation is shown as below:

$$U = \frac{\sqrt{\sum_{k=1}^m (t_k - y_k)^2}}{\sum_{k=1}^m t_k^2 + \sum_{k=1}^m y_k^2} \quad (15)$$

The *U* value is bound between 0 and 1, with values closer to 0 indicating greater forecasting accuracy.

Results and Discussion

The performance statistics for each model were calculated and are presented in Table 1. The performance criteria *RMSE*, *MAPE*, *MAE*, *R* and *Theil's U* were respectively calculated as 13830.3706, 0.3839, 10309.7391, 0.8863 and 0.1323 for GRU model; 13971.4670, 0.3745, 10337.0889, 0.8866 and 0.1350 for LSTM model. Theoretically, a forecasting model is regarded as good when *RMSE*, *MAPE* and *MAE* are small, *R* is close to 1 and *Theil's U* is close to 0. In Table 1, the value marked in bold and italic indicates the best performance. It can be seen that GRU and LSTM models archived the best performance according to the five criteria in all predictions. The performance criteria indicate that the assessed result is highly correlated and precise.

The simulation results by GRU and LSTM models are also presented in Fig. 4 and 6. Due to the limited space available, only the predicted results in the last two weeks are presented. It can be clearly observed that the errors obtained from the testing data in Fig. 5a and 6a; the error plot in Fig. 5b and 6b, along with the histogram of error (Fig. 5c and 6c).

Figure 7 presents the scatter diagrams by the LSTM and GRU models in all the testing set. The scatter plot illustrates the degree of correlation between forecasting values and actual values. In the figure, the 1:1 line was drawn as a reference. In a scatter diagram, the 1:1 line represents that the two sets of data are identical. The more the two data sets agree, the more the points tend to concentrate in the vicinity of the 1:1 line. It may be observed that most predicted values are close to the actual values in Fig. 7 and this indicates a good

agreement between the predicted values obtained by the GRU and LSTM models and the actual values.

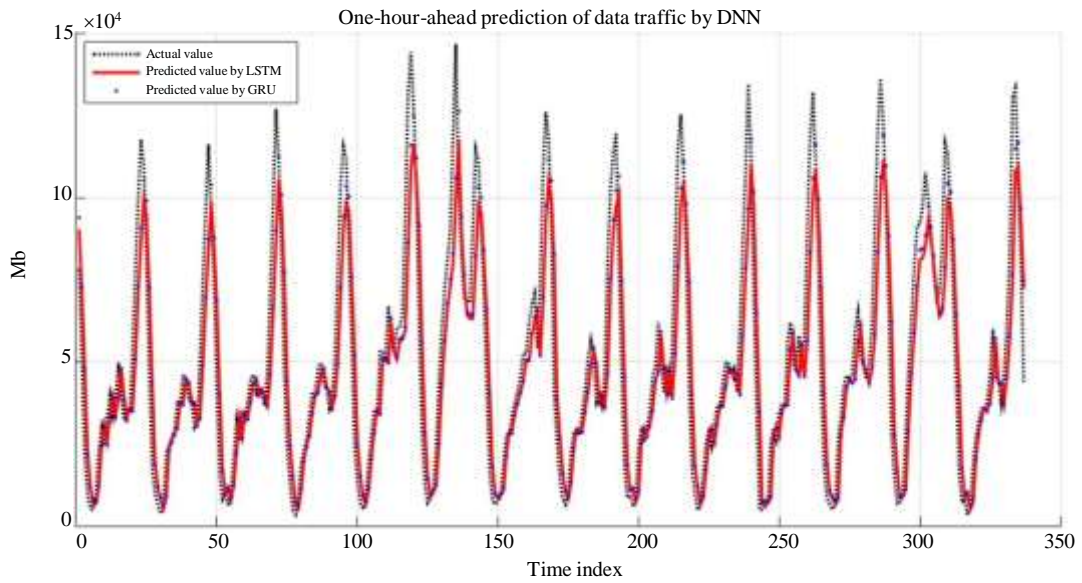


Fig. 4: Comparison between actual and predicted values by LSTM and GRU model in the last two weeks

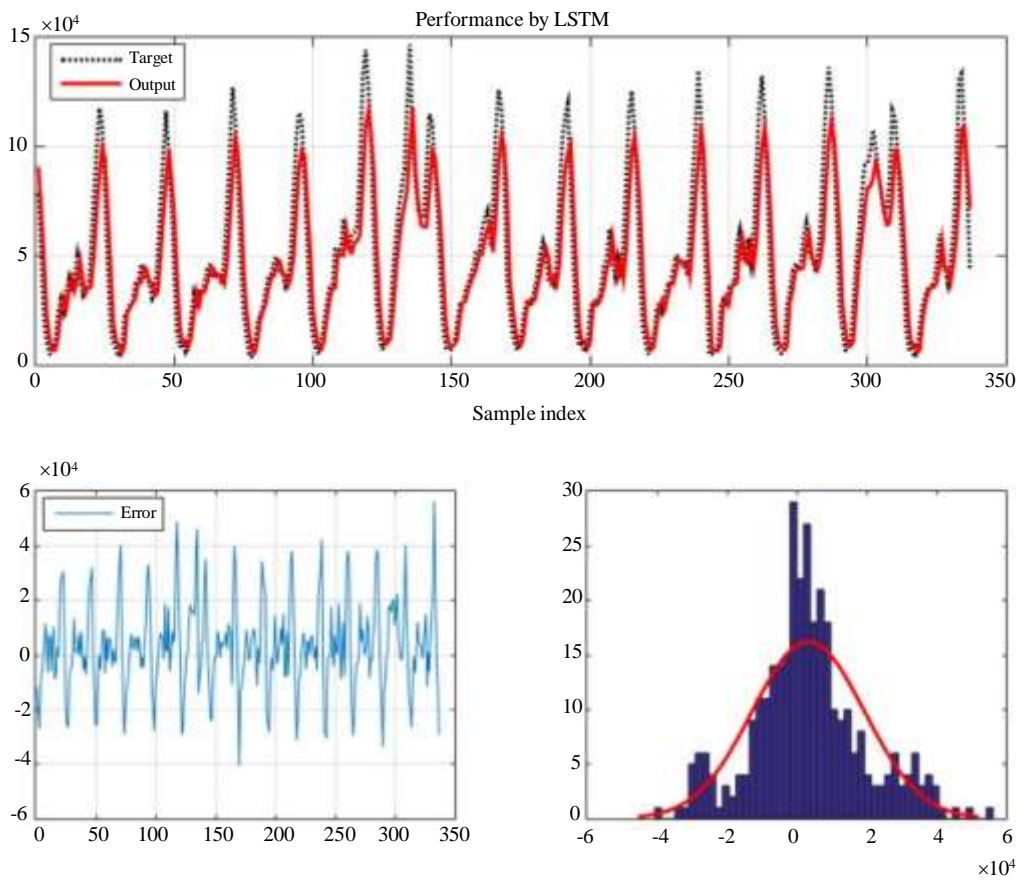


Fig. 5: Predicting performance by LSTM model: (a) actual (target) value vs predicted (output) value, (b) error value and (c) standard deviation

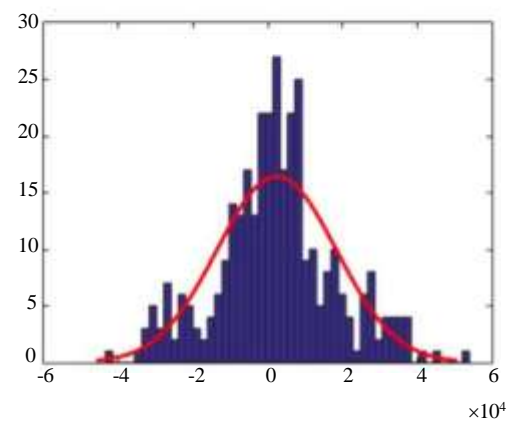
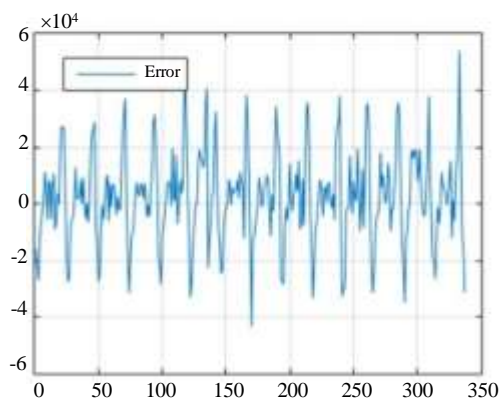
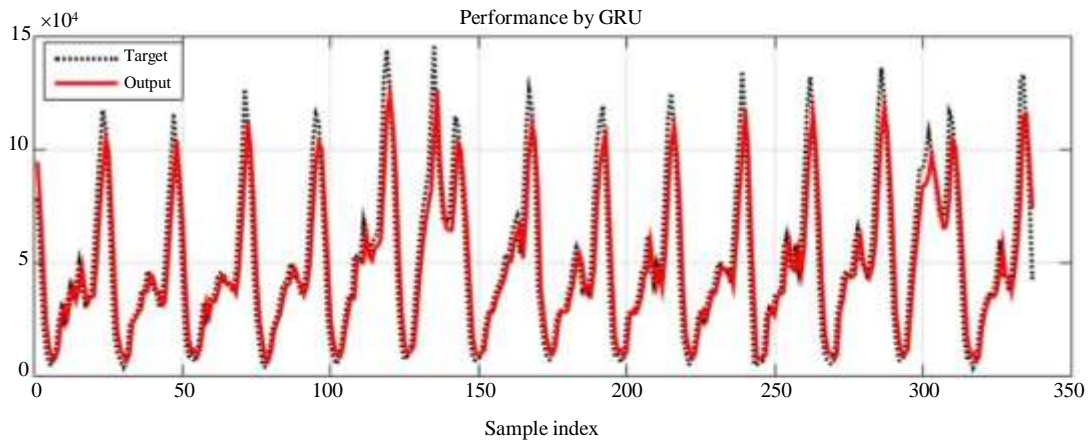


Fig. 6: Predicting performance by GRU model: (a) actual (target) value vs predicted (output) value, (b) error value and (c) standard deviation

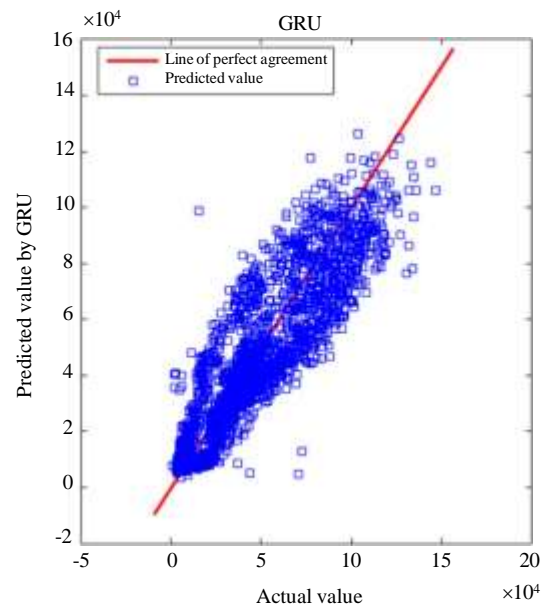
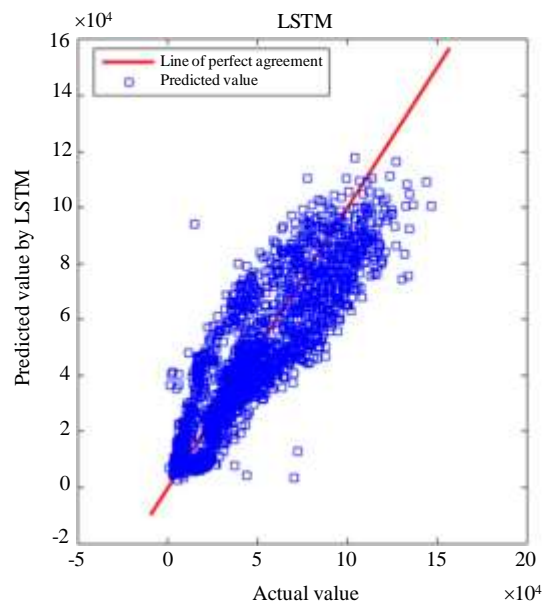


Fig. 7: The scatter diagrams by the LSTM and GRU models in the testing set

Table 1: Performance statistics of different models

Model	RMSE	MAPE	MAE	R	Theil's U-statistic
ANFIS	29750.2178	0.4718	12595.3005	0.6503	0.2654
ANN	16838.9119	0.3271	10111.0545	0.8508	0.1682
GMDH	1920553.0813	1025.9252	6404251.2836	0.0271	0.9998
GRU	13830.3706	0.3839	10309.7391	0.8863	0.1323
LSTM	13971.4670	0.3745	10337.0889	0.8866	0.1350

Conclusion

The historical time series data collected by the telecom network operators has characteristics of classical nonlinearity and instability. In this study, we have analyzed and compared the ability of the GRU and LSTM models in analyzing and predicting data traffic in the network of Viettel Telecom, Vietnam. Several criteria namely *RMSE*, *MAPE*, *MAE* and *R* were used to evaluate the performance of the develop models. The results indicated that deep neural networks can be promising tools for analyzing time series data of telecom network. It was confirmed that GRU and LSTM models was the most robust and powerful method, with respect to all performance criteria, for research area. The study findings show the forecasting potential of the artificial intelligence models in developing data traffic model and are expected to provide an assistance and forecasting tool in telecommunication network planning and network optimization. For future research, the authors are combining the other parameters, such as the ratio between on- and off-net traffic, the percentage of off-net traffic sent to a particular peering point and the ratio between upstream and downstream traffic, so as to improve the accuracy of the predicting model. Also for future research, the next step is to explore more deep learning structures and combine the characteristics of network traffic data to explore the possibility of further improving the forecasting accuracy.

References

- Abdulkarim, S. A., & Lawal, I. A. (2017). A cooperative neural network approach for enhancing data traffic prediction. *Turkish Journal of Electrical Engineering & Computer Sciences*, 25(6), 4746–4756.
- Bian, G., Liu, J., & Lin, W. (2018). Internet Traffic Forecasting using Boosting LSTM Method. *DEStech Transactions on Computer Science and Engineering*.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2015). Gated feedback recurrent neural networks. *32nd International Conference on Machine Learning, ICML 2015*.
- Fu, R., Zhang, Z., & Li, L. (2017). Using LSTM and GRU neural network methods for traffic flow prediction. *Proceedings - 2016 31st Youth Academic Annual Conference of Chinese Association of Automation, YAC 2016*.
- Gao, Y., Zheng, L., Zhao, D., Wu, Y., & Wang, G. (2017). Short Term Prediction Models of Mobile Network Traffic Based on Time Series Analysis. *International Conference on Machine Learning and Intelligent Communications*, 205–211. Springer.

- Gowrishankar, S., & Satyanarayana, P. S. (2009). A Time Series Modeling and Prediction of Wireless Network Traffic. *International Journal of Interactive Mobile Technologies (IJIM)*.
- Hernandez Benet, C., Kassler, A., & Zola, E. (2016). Predicting expected TCP throughput using genetic algorithm. *Computer Networks*, 108, 307-322.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Jiang, D., Xu, Z., & Xu, H. (2015). A novel hybrid prediction algorithm to network traffic. *Annals of Telecommunications-Annales Des Télécommunications*, 70(9-10), 427-439.
- Kaushik, P., Singh, S., & Yadav, P. (2018). Traffic Prediction in Telecom Systems Using Deep Learning. 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 302-307. IEEE.
- Le, X. H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of Long Short-Term Memory (LSTM) neural network for flood forecasting. *Water (Switzerland)*.
- Lu, J. J., & Wang, Z. Q. (2006). Prediction of network traffic flow based on chaos characteristics. *Journal of Nanjing University of Aeronautics & Astronautics*, 38(2), 217-221.
- Mahmassani, H. (2001). *Dynamic Network Traffic Assignment and Simulation Methodology for Advanced System Management Applications. Networks and Spatial Economics*.
- Park, Y. K., & Lee, C. (1995). Applications of neural networks in high-speed communication networks. *IEEE Communications Magazine*, 33(10), 68-74.
- Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*.
- Tian, H., Zhou, X., & Liu, J. (2017). A Hybrid Network Traffic Prediction Model Based on Optimized Neural Network. 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), 284-287. IEEE.
- Tran, Q. T., Ma, Z., Li, H., Hao, L., & Trinh, Q. K. (2015). A multiplicative seasonal ARIMA/GARCH model in EVN traffic prediction. *International Journal of Communications, Network and System Sciences*, 8(04), 43.