UDC 621.382.2 (045)
DOI:10.18372/1990-5548.72.16941

[1]L. V. Sibruk,
[2]I. V. Zakutynskyi

RECURRENT NEURAL NETWORKS FOR TIME SERIES FORECASTING. CHOOSING THE BEST ARCHITECTURE FOR PASSENGER TRAFFIC DATA

Faculty of Air Navigation, Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine
E-mails: [1]sibruk@nau.edu.ua ORCID 0000-0002-4119-7984,
[2]ihor.zakutynskyi@nau.edu.ua ORCID 0000-0003-2905-3205

Abstract—*Accurately predicting the urban traffic passenger flow is of great importance for transportation resource scheduling, planning, public safety, and risk assessment. Traditional statistical approaches for forecasting time series are not effective in practice. They often require either strict or weak data stationarity, which is almost impossible to obtain with real data. An alternative method is time series forecasting using neural networks. By their nature, neural networks are non-linear and learn based on input and output data. With this approach, increasing the efficiency of the network is reduced to increasing the amount of data of the initial sample. Today, the class of recurrent neural networks is mainly used for forecasting time series. Another important stage is the choice of neural network architecture. In this article the use of long short term memory and gated recurrent units architecture is considered and also is compared their performance for passenger flow forecasting.*

Index Terms—Neural networks; recurrent neural networks; LST Marchitecture; GRU architecture; time series; passenger flow.

I. INTRODUCTION

Recurrent neural networks (RNN) are a class of artificial neural network which connections between nodes form a time-oriented graph. This creates an internal state of the network that allows it to exhibit dynamic behavior in time. Unlike feed forward neural networks, RNN's can use their internal memory to process arbitrary sequences of inputs.

In article a comparison of the efficiency for the two most popular architectures of recurrent neural networks long short term memory (LSTM) and gated recurrent units, controlled recurrent neurons (GRU) is considered.

The purpose of the work is the development of the neural network based on LSTM and GRU architectures, using Python and Keras library with the comparison of obtained results, and the choice of the best architecture for passengers flow forecasting.

II. PROBLEM STATEMENT

The neural network is intendedto predict passengers of international airlines, based on the dataset "Airline passengers January 1949 to December 1960" [1], which is publicly available. The data cover the period of 12 years with 144 observations (Table I).

As a result, the developed neural network should predict the number of passengers of international airlines for years that are not in this sample.

TABLE I. AIRLINE PASSENGERS FROM 1949 TO 1960

Year/Month	Number of Passengers
1960/05	472
1960/06	535
1960/07	622
1960/08	606
1960/09	509
1960/10	461
1960/11	390

III. THEORETICAL BASIS

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.

Recurrent neural networks is an artificial neural network which powerful to handle sequential data. As shown in Fig. 1 [2], RNN has the hidden layer which are sort of intermediate snapshots of the original input data.

Figure 2 shows the unfolding in time the data, which involved in forward computation. Also presented in Fig. 2, the output is produced from input through neural network [6]. The loops transfer the data to the next step. Via the loops, each independent data becomes dependent on each other.

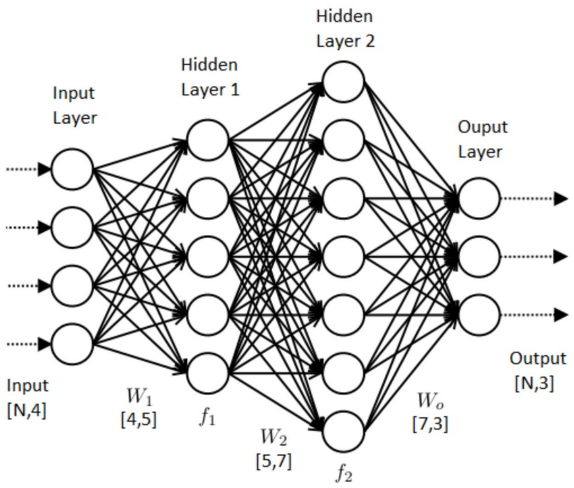

Fig. 1. RNN with hidden layers

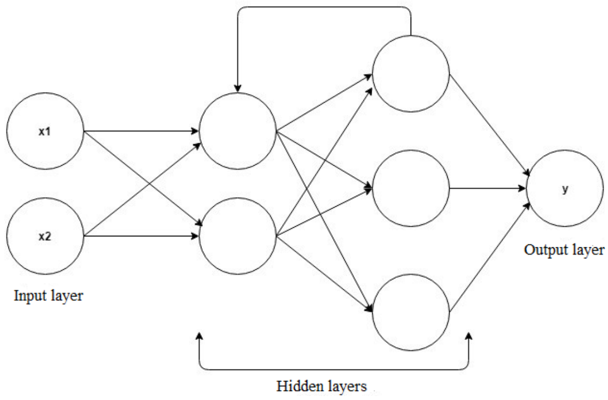

Fig. 2. Recurrent connections

Recurrent neural networks can be seen as multiple copies of the same network.

In classical architectures of recurrent neural networks, the information at each learning step is mixed with the information of the previous steps, thus erasing it during several iterations. The main problem of classical RNN architecture is short-term memory. If a sequence is long enough, they'll have a hard time carrying information from earlier time steps to later ones. So in recurrent neural networks layers, that get a small gradient update stops learning. Those are usually the earlier layers. So as these layers don't learn, RNN's can forget (1) what it seen in longer sequences, thus having a short-term memory [5]

$$W = W_t - kG, \tag{1}$$

where W is new weight value; W_t is current weight; k is learning rate; G is gradient value; t is learning step.

A. Long Short Term Memory

The architect's LSTM is designed in such a way that the data of a particular step is stored for both short and long time intervals, which avoids the problem of long-term dependence. A special feature is that the LSTM module does not use an activation function. Instead of neurons, LSTM networks have blocks of memory connected through layers.

A block contains gateways that control the state and output of the block. The block operates based on an input sequence, and each gate in the block uses sigmoidal activation blocks to control whether it fires or not, making the state change and addition of information passing through the block conditional.

In this case, LSTM neurons are used as nodes of a recurrent neural network.

In the first step, the LSTM determines what information can be removed from the current state. This logic performed by the sigmoid layer – Forget gate layer, Fig. 3 [3]. This layer checks hidden input data h_{t-1}, input data vector x_t and return 0 or 1 using relation (2).

$$f_t = \sigma\left(x_t U^f + h_{t-1} W^f\right). \tag{2}$$

where f_t is forget gate layer; σ is sigmoid function; U^f, W^f is weight vectors for forget gate.

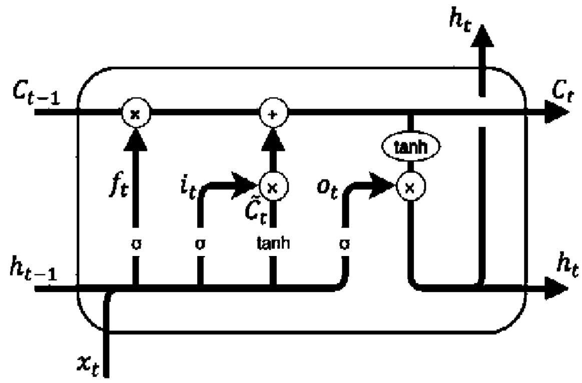

Fig. 3. LSTM cell

In Figure 3 C_t is hidden input vector; a_t, i_t are bias vectors.

At the next stage, it is decided which data from the current step will be stored. This process consists of two parts:

1) The sigmoid layer (input gate layer) determines which of the current data needs to be updated

$$i_t = \sigma\left(x_t U^i + h_{t-1} W^i\right). \tag{3}$$

2) The layer based on the hyperbolic tangent function builds a vector of new values, that will be added to the current ones

$$\overline{C}_t = \tanh\left(x_t U^g + h_{t-1} W^g\right), \tag{4}$$

where W^g is array of weights; $\overline{C}_t$ is vector of new candidate values.

3) The previous value of C_{t-1} is replaced by the new state (5)

$$C_t = \sigma\left(f_t C_{t-1} + i_t \bar{C}_t\right). \quad (5)$$

At the last stage, the original data layer is determined using current state and some filters.

$$o_t = \sigma\left(x_t U^o + h_{t-1} W^o\right), \quad (6)$$

Firstly a sigmoidal layer o_t [0, 1] is used (6) for determination output data from the current state. After that the original values are normalized to the range [–1, 1]

$$h_t = \tan h\left(C_t\right) \cdot o_t. \quad (7)$$

B. Gated Recurrent Units

The Gated Recurrent Unit mechanism in recurrent neural networks was introduced in 2014. GRUis similar to long short memory with a forget gate, but have less parameters exit gate does not exist.

The main difference between the GRU architecture and the LSTM is that the filtering units (forgetting) and input filters are combined into one filter – the Update gate filter, Fig. 4 [4]. Also current state combines with hidden. As a result, this architecture significantly simplifies the model implementation.

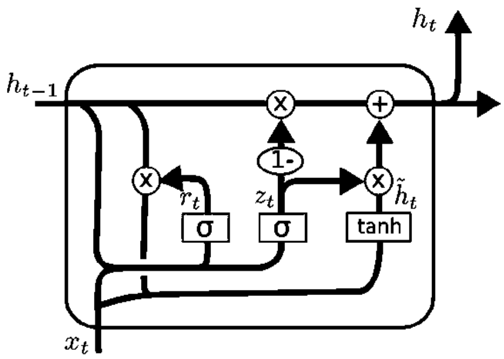

Fig. 4. GRU cell

Gated recurrent units does not possess any internal memory, they don't have an output gate that is present in LSTM.

IV. MODEL IMPLEMENTATION

The model is designed forthe international airline passengers prediction. For given input year and month it is necessary predict the number of international airline passengers in units of 1.000.

The data ranges from January 1949 to December 1960 or 12 years, with 144 observations (Fig. 5).There is a certain periodicity that corresponds to the holiday period, and the vacation period.

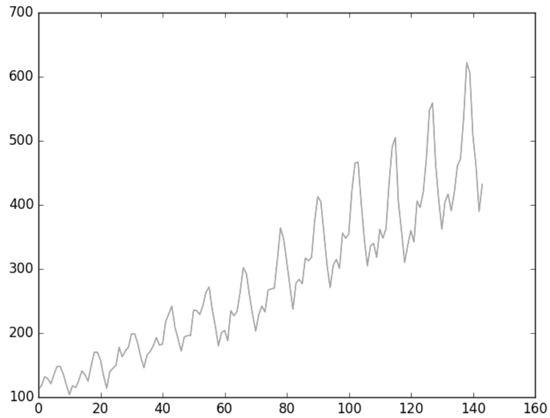

Fig. 5. Input dataset visualization

Long Short Term Memory and GRU networks are quite sensitive to the scale of the input data. Especially when sigmoid or tan *h* activation functions are used. So, the data must be normalized to the range [0, 1]. In Python for this is used the *MinMaxScaler* preprocessing class from the scikitlearn library.

Also, we need to divide the input data into training and testing samples. Input data is divided as follows – 67% initial sample (train), 33% test (test). For the network training the new dataset was created, where *X* is the number of passengers at a certain time(t) and *Y* is the number of passengers at the next time (t + 1).

An LSTM network expects that input data (*X*) is provided by some array structure in the form of [samples, time steps, features].

A neural network consists of: a visible layer with 1 input, 2 hidden layers with 64 blocks of LSTM neurons, and an output layer that makes a prediction of a single value. For LSTM blocks the default activation ReLu (8) function $f(u)$ is used, Fig. 6.

$$f\left(u\right) = \max\left(0, u\right). \quad (8)$$

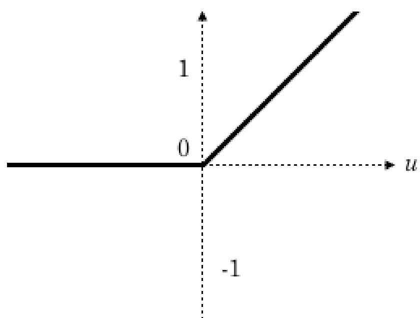

Fig. 6. RELU activation function

The network is trained in different epochs, hidden layers count with package size = 1. Applied Adam algorithm for model optimization is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments.

The network quality is checked with a loss function *cost* (mean squared error), which determines the difference between the original (predicted values) and the previously known values

$$cost = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2. \tag{9}$$

The basis of neural network learning is the minimization of *cost* value.

For both networks (LSTM and GRU) the similar training process (Fig. 7) was applied.

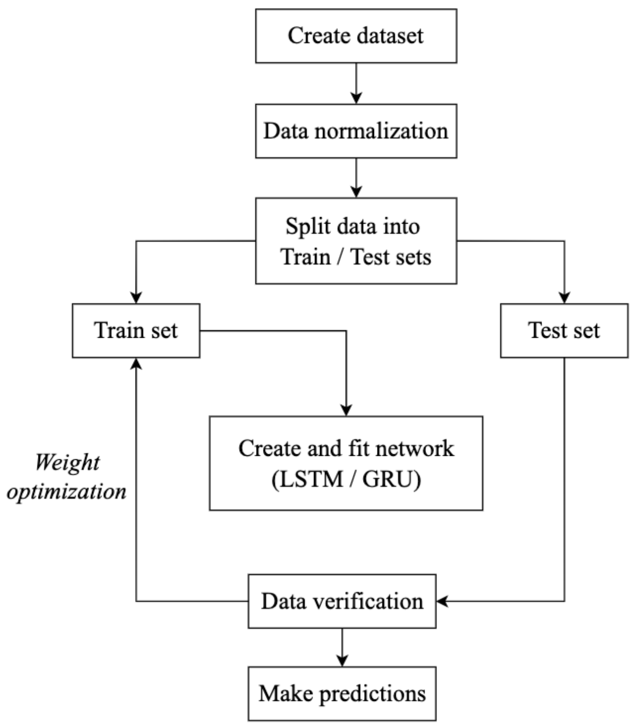

Fig. 7. Learning Process structure

The data on model accuracy according to epochs are in Table II.

According to the above data the best result (86.83% accuracy) got with LSTM architecture and 50 "learning" epochs.

Also can be noticed the decline model performance point (LSTM > 50k Epochs). This is the moment of saturation, or network retraining moment.

In the result graphs (Figs 8 – 11) the best results for both networks (LSTM – 100 Epochs, GRU – 10 Epochs) are given.

TABLE II. LSTM AND GRU MODELS ACCURACY

Model	Accuracy	Error rate
LSTM (10 Epochs)	71.18%	28.82%
GRU (10 Epochs)	67.11%	32.89%
LSTM (30 Epochs)	73.64%	26.36%
GRU (30 Epochs)	61.02%	38.98%
LSTM (50 Epochs)	86.83%	13.17%
GRU (50 Epochs)	67.12%	32.88%
LSTM (100 Epochs)	86.1%	13.9%
GRU (100 Epochs)	67.18%	32.82%
LSTM (1000 Epochs)	71.18%	28.82%
GRU (1000 Epochs)	67.11%	32.89%
LSTM (10k Epochs)	81.39%	18.61%
GRU (10k Epochs)	67.05%	32.95%
LSTM (50k Epochs)	76.82%	23.18%
GRU (50k Epochs)	66.22%	33.78%
LSTM (100k Epochs)	53.68%	46.32%
GRU (100k Epochs)	66.16%	33.84%

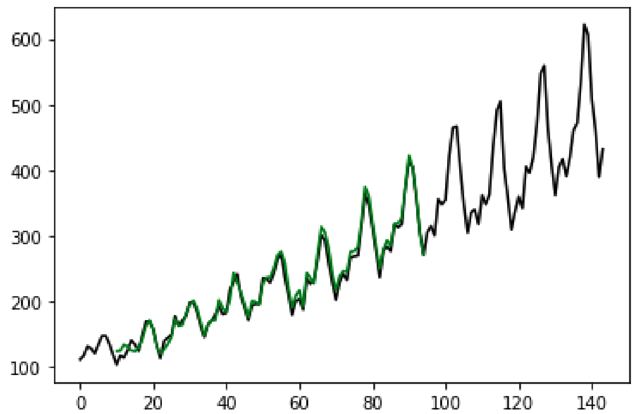

Fig. 8. LSTM network accuracy with train data

In these graphs *X* axis is the month number, *Y* axis is the passengers count. The main line shows the input data, the auxiliary one – the result of the model forecast.

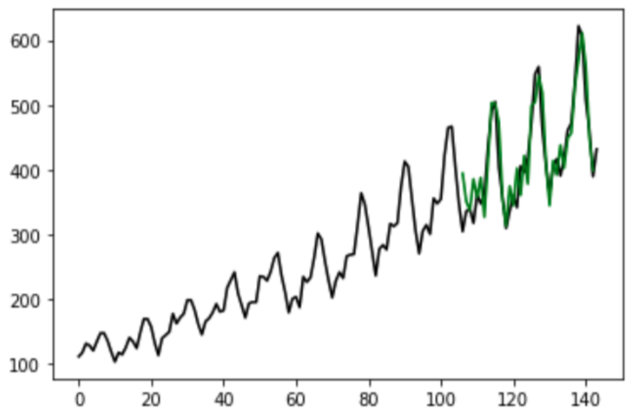

Fig. 9. LSTM network accuracy with test data

Figures 8 and 10 represent network accuracy with train data. It this case both networks show similar good results. But real network performance can be seen at Figs 9 and 10 (network prediction based on the test data).

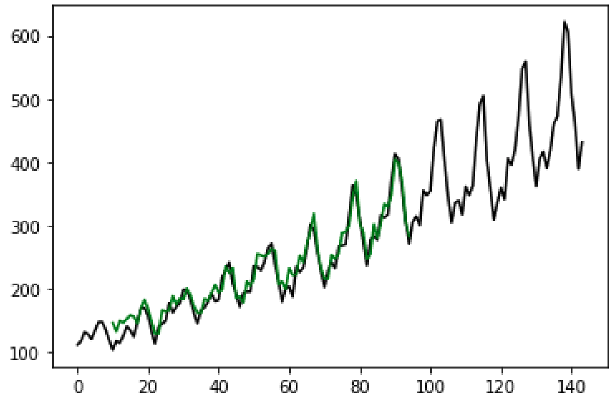

Fig. 10. GRU network accuracy with train data

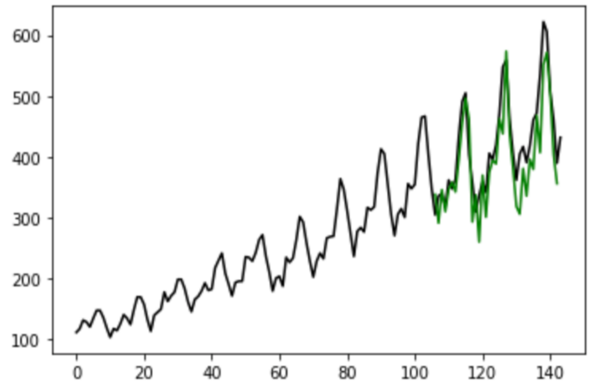

Fig. 11. GRU network accuracy with train data

The results can be considered satisfactory.

V. CONCLUSIONS

The results show that RNN's are well suited for forecasting time series data (like passengers flow), but have some problems with shot-memory. Therefore for airline passenger prediction the architect's LSTM and GRU were considered.

For this dataset the LSTM network showed better results (86.83% accuracy) in comparison with GRU (67.12% accuracy). According to the small size of the input data set, both results can be considered as good. An increase in forecasting accuracy is associated with greater volume of the input data.

So, neural networks based on LSTM architecture can be used for nonlinear, long-term data prediction with high accuracy. Recurrent neural networks based on GRU architecture is less accurate than the LSTM but has the advantage of being relatively simple with relatively good performance and low computing time.

REFERENCES

[1] Chirag Kothari, US Airline passengers from 1949 to 1960. *Kaggle.* 2018, April. [Online]. Available: https://www.kaggle.com/datasets/chirag19/air-passengers.

[2] Jayesh Bapu Ahire, The Artificial Neural Networks. Part 1, 2018, August. [Online]. Available: https://www.datasciencecentral.com/the-artificial-neural-networks-handbook-part-1.

[3] Savvas Varsamopoulos, and Koen Almudever, “Designing neural network based decoders for surface codes,” pp. 1–12, Nov. 2018.

[4] Christopher Olah, Understanding LSTM Networks. 2015, August. [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs.

[5] Michael Phi, “Illustrated Guide to LSTM and GRU,” *Medium Blog.* 2018, September. [Online]. Available: https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21.

[6] Fulvia Ceccarelli, and Marco Sciandrone, “Prediction of chronic damage in systemic lupus erythematosus by using machine-learning models,” *Plos One.* 12(3), 2017, March. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174200. https://doi.org/10.1371/journal.pone.0174200

[7] Kyungdoo Nam and Thomas Schaefer, Forecasting international airline passenger traffic using neural networks, pp. 239–252. 1995, September. [Online]. Available: https://www.researchgate.net/publication/233684273_Neural_network_forecasting_for_airlines_A_comparative_analysis.

[8] Diederik P. Kingma, and Jimmy Lei Ba, “Adam: A method for Stochastic Optimization,” *3rd International Conference for Learning Representations, San Diego.* 2015, June. [Online]. Available: https://arxiv.org/abs/1412.6980

[9] Bowen Du, and Hao Peng, “Deep Irregular Convolutional Residual LSTM for Urban Traffic Passenger Flows Prediction,” *IEEE Transactions on Intelligent Transportation Systems*, 2020, March, pp. 972–985. https://doi.org/10.1109/TITS.2019.2900481

[10] W. Min, and L. Wynter, “Real-time road traffic prediction with spatiotemporal correlations,” *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, August 2011. https://doi.org/10.1016/j.trc.2010.10.002

[11] A. Abadi, T. Rajabioun, and P. A. Ioannou, “Traffic flow prediction for road transportation networks with

limited traffic data," *IEEE Trans. Intell. Transp. Syst*., vol. 16, no. 2, Apr. 2015, pp. 653–662.

[12] Quang Hung Do, "Forecasting Air Passenger Demand: A Comparisonof LSTM and SARIMA," *Journal of Computer Science,* 16(7): 1063, Jul, 2020. https://doi.org/10.3844/jcssp.2020.1063.1084

[13] S. Haykin, *Neural networks. A comprehensive Foundation. Second edition*, in Prentice Hall*,* 2020, pp. 919–925.

Received January 28, 2022

Leonid Sibruk. ORCID 0000-0002-4119-7984. Doctor of Engineering Science. Professor.
Radio Electronic Devices and Systems Department, Faculty of Air-navigation, Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: Kyiv Institute Engineers of Civil Aviation, Kyiv, USSR, (1974).
Research area: Antennas, propagation, radio monitoring, communications.
Publications: more than 180 papers.
E-mail: sibruk@nau.edu.ua

Ihor Zakutynskyi. ORCID 0000-0003-2905-3205. Post-graduate student.
Radio Electronic Devices and Systems Department, Faculty of Air-navigation, Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (current time).
Research area: Neural networks, Software development, Automation systems.
E-mail: ihor.zakutynskyi@nau.edu.ua

Л. В. Сібрук, І. В. Закутинський. Рекурентні нейронні мережі для прогнозування часових рядів. Вибір оптимальної архітектури для прогнозування пасажиропотоку

Точне прогнозування пасажиропотоку міського транспорту має велике значення для планування транспортних ресурсів, громадської безпеки та оцінки ризиків. Традиційні статистичні підходи до прогнозування часових рядів не ефективні на практиці. Вони часто вимагають або суворої, або слабкої стаціонарності даних, яку майже неможливо отримати за реальними даними. Альтернативним методом є прогнозування часових рядів за допомогою нейронних мереж. За своєю природою нейронні мережі є нелінійними і навчаються на основі вхідних і вихідних даних. При такому підході підвищення ефективності мережі зводиться до збільшення обсягу даних вихідної вибірки. Сьогодні для прогнозування часових рядів в основному використовується клас рекурентних нейронних мереж. Ще одним важливим етапом є вибір архітектури нейронної мережі. У даній статті розглядається використання архітектури нейронних мереж LSTM і GRU, а також порівнюється ефективність їх використання для прогнозування пасажиропотоку.

Ключові слова: нейронні мережі; рекурентні нейронні мережі; архітектура LSTM; архітектура GRU; часові ряди; пасажиропотік.

Сібрук Леонід Вікторович. ORCID 0000-0002-4119-7984, SCOPUS ID 16064809600
Доктор технічних наук. Професор.
Кафедра електроніки, робототехніки і технологій моніторингу та Інтернету речей, Факультет аеронавігації, електроніки та телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Київський інститут інженерів цивільної авіації, Київ, СССР, (1974).
Напрям наукової діяльності: антени, поширення радіохвиль, радіомоніторинг, радіозв`язок.
Кількість публікацій: більше 180 наукових робіт.
E-mail: sibruk@nau.edu.ua

Закутинський Ігор Володимирович. ORCID 0000-0003-2905-3205. Аспірант.
Кафедра електроніки, робототехніки і технологій моніторингу та Інтернету речей, Факультет аеронавігації, електроніки та телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2019).
Напрям наукової діяльності: нейронні мережі, розробка програмного забезпечення, системи автоматизації.
E-mail: ihor.zakutynskyi@nau.edu.ua

Л. В. Сибрук, И. В. Закутинский. Рекуррентные нейронные сети для прогнозирования временных рядов. Выбор оптимальной архитектуры для прогнозирования пассажиропотока

Точное прогнозирование пассажиропотока городского транспорта имеет большое значение для планирования транспортных ресурсов, общественной безопасности и оценки рисков. Традиционные статистические подходы к прогнозированию временных рядов не эффективны на практике. Они часто требуют либо строгой, либо слабой стационарности данных, которую почти невозможно получить на реальных данных. Альтернативным методом является прогнозирование временных рядов с помощью нейронных сетей. По своей природе нейронные сети нелинейны и обучаются на основе входных и исходных данных. При таком подходе

повышение эффективности сети сводится к увеличению объема данных исходной выборки. Сегодня для прогнозирования временных рядов в основном используется класс рекуррентных нейронных сетей. Еще одним важным этапом является выбор архитектуры нейронной сети. В данной статье рассматривается использование архитектуры нейронных сетей LSTM и GRU, а также сравнивается эффективность их использования для прогнозирования пассажиропотока.
Ключевые слова: нейронные сети; рекуррентные нейронные сети; архитектура LSTM; архитектура GRU; временные ряды; пассажиропоток.

Сибрук Леонид Викторович. ORCID 0000-0002-4119-7984. SCOPUS ID 16064809600.
Доктор технических наук. Профессор.
Кафедра электроники, робототехники и технологий мониторинга и Интернета, Факультет аэронавигации, электроники и телекоммуникаций, Национальный авиационный университет, Киев, Украина.
Образование: Киевский институт инженеров гражданской авиации, Киев, СССР, (1974).
Направление научной деятельности: антенны, распространение радиоволн, радиомониторинг, радиосвязь.
Количество публикаций: более 180 научных работ.
E-mail: sibruk@nau.edu.ua

Закутинский Игорь Владимирович. ORCID 0000-0003-2905-3205. Аспирант.
Кафедра электроники, робототехники и технологий мониторинга и Интернета, Факультет аэронавигации, электроники и телекоммуникаций, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина, (2019).
Направление научной деятельности: нейронные сети, разработка программного обеспечения, системы автоматизации.
E-mail: ihor.zakutynskyi@nau.edu.ua