EPJ Web of Conferences **318**, 02010 (2025) https://doi.org/10.1051/epjconf/202531802010
AAPM-III 2025

Selecting the neural network architecture in the problem of geomagnetic SME index modeling

Yurii Polozov*

Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS, Mirnaya st., 7, Paratunka, Kamchatskiy krai, 684034, Russian Federation

Abstract. The paper solves the problem of selecting the recurrent neural network architecture to construct geomagnetic SME index architecture. The description is given for the architectures providing the processing of long-term dependencies in data. Neural network performance quality was estimated on the basis of errors. The Dropout layer effect on the modeling results is shown separately. The constructed neural network models can be used to forecast the SME index variations.

1 Introduction

Natural time series carry important information on the processes in geospheres. The investigation and analysis of such data depend on the complexity of investigated signals, developed models and theoretical means. At the present time, neural networks (NNs) are the effective and available tool for data analysis and construction of process models. NNs have a large number of architectures and can be used for solving the problems of classification, approximation, forecast, etc. [1, 2]. Multi-year development of NNs theoretical apparatus made it possible to apply them in a large number of applied investigations when computing instruments of satisfactory performance became available [2].

One of the architectures often used in NNs is the recurrent architecture [1]. Recurrent NN (RNN) allow one to maintain hidden state during time series processing. In a general case, the hidden state (depending on an actual RNN architecture) undergoes the impact of the time series variations, which have been earlier supplied to the NN input. This property of the RNNs provides modeling of the dependencies inside a time series. The RNNs can be applied in a variety of ways. For example, a technology for detecting step-like anomalies in head motion parameters for a functional MRI was shown in the paper [3]. The authors suggested using a neural network model to detect such anomalies and the procedure for network training. The authors in [4] apply a set of several RNNs for household use. The paper [4] describes interaction of sensors with a head unit in the form of a smart phone where the sensor data are processed by separate RNNs. Trained RNNs classify user actions that provides filtration of the data transmitted to a smart phone.

In this paper, we investigate different RNN architectures (LSTM, BiLSTM, GRU, Projected) in connection with the Dropout layer and without it in the problem of model construction for a natural time series of the SME geomagnetic index [5]. This index provides

* Corresponding author: up_agent@mail.ru

an estimate of the energy penetrated into the magnetosphere during magnetic storms and is an important parameter during magnetic variation analysis. Based on the earlier paper [6], the wavelet transform [7] is used together with neural networks. Wavelets make it possible to represent time series in the form of frequency-time components that is used when creating training vectors for neural network models.

The paper presents the estimate of the quality of geomagnetic SME index neural network models for different RNN architectures. The impact of different layers and NN architectures on data approximation error is shown. The obtained results allowed us to detect the RNN architectures providing construction of the SME model with the least error.

2 Methodology

2.1 Data preparation for constructing the neural network models

As it was illustrated in the papers [5-6], the SME index depends on the Sun energy penetrating into the Earth magnetic field. Physical processes reflect in the relations between the vertical component Bz variations of the interplanetary magnetic field (IMF) and auroral current parameters, in particular, the SME index [5-6, 8]. The recorded data of the IMF Bz and SME have nonstationary structure, contain oscillations of different frequency and amplitude. Based on the papers [6], data preliminary processing included the representation of the IMF Bz and SME index in the form of wavelet decomposition components. Initial data were transformed into the sets of wavelet coefficients as [7]

$$\left(W_\Psi s_{y,x}\right) := |x|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} s(t)\Psi\left(\frac{t-y}{x}\right) dt,\ s \in L^2(R),\ x,\ y \in R,\ x \neq 0, \quad (1)$$

where Ψ is the basic wavelet, x is the scale, y is the time. The wavelet coefficients, obtained on the basis of (1), were grouped relatively 0 as

$$W_\Psi s_{y,x} = \begin{cases} (W_\Psi s_{y,x})^+ \ , if\ W_\Psi s_{y,x} > 0 \\ (W_\Psi s_{y,x})^- \ , if\ W_\Psi s_{y,x} \leq 0 \end{cases}. \quad (2)$$

The intensity, obtained on the basis of the wavelet coefficients, at different ranges of scales D was used at the NNs input [7]

$$T_y^{\pm} = \sum_D (W_\Psi s_{y,x})^{\pm}. \quad (3)$$

Initial vectors and those, obtained on the basis of (1)-(3), in the quantity of 14 lines were formed into two-dimensional matrices, each from which contained information on magnetic data state over 72 hours.

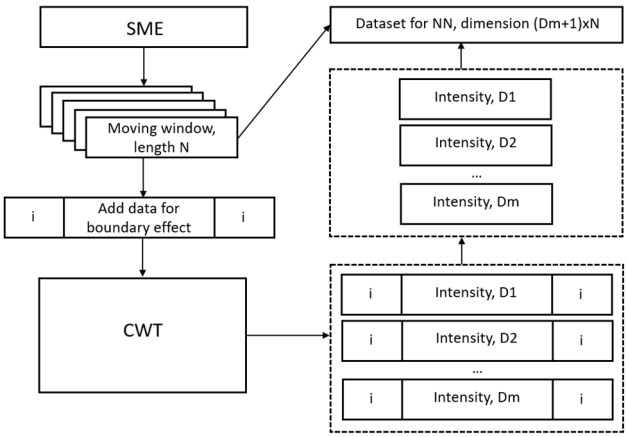

Fig. 1. Formation of input data for neural network.

The target data for the NNs had the form of SME index vector values for one hour ahead that made it possible to construct neural network models for SME forecast. A general scheme of data formation is illustrated in Figure 1.

2.2 Applied architectures of NNs

2.2.1 LSTM

Development of RNNs resulted in the solutions intended for processing the data with long-term dependencies. The LSTM architecture provides processing of such data based on the introduction of the mechanisms for preservation of the information passing through the NN blocks [9]. Each block forms a computational unit of a defined structure (Figure 2).

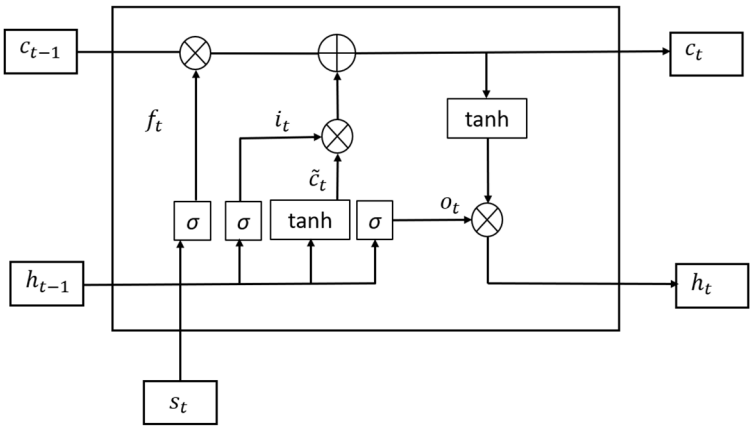

Fig. 2. LSTM structure block.

Estimates in LSTM block are carried out according to the following formulas:

$$f_t = \sigma\big(w_f[h_{t-1}, s_t] + b_f\big),$$
$$i_t = \sigma(w_i[h_{t-1}, s_t] + b_i),$$
$$\tilde{c}_t = \tanh(w_c[h_{t-1}, s_t] + b_c),$$
$$c_t = c_{t-1}f_t + i_t\tilde{c}_t,$$
$$o_t = \sigma(w_o[h_{t-1}, s_t] + b_o),$$
$$h_t = o_t\tanh(c_t),$$

where s_t are the input data; w_f, w_i, w_c, w_o are the weight matrices states; b_f, b_i, b_c, b_o are the biases; h_t is the LSTM hidden state; f_t, i_t, o_t are the values of the forget gate, input gate and output gate, respectively; $c_t, \tilde{c}_t$are the current and temporal states of a cell; $\sigma, \tanh$ are the activation functions.

2.2.2 BiLSTM

LSTM architecture development. It represents two connected LSTMs, one of which processes the input vectors forward, the other processes the vectors in a backward way. This architecture makes it possible to expand the information obtained from a sampling and associated with the context. BiLSTM may be effective during the processing of IMF Bz and SME data which have long-term dependencies. Figure 3 illustrates the BiLSTM network architecture.

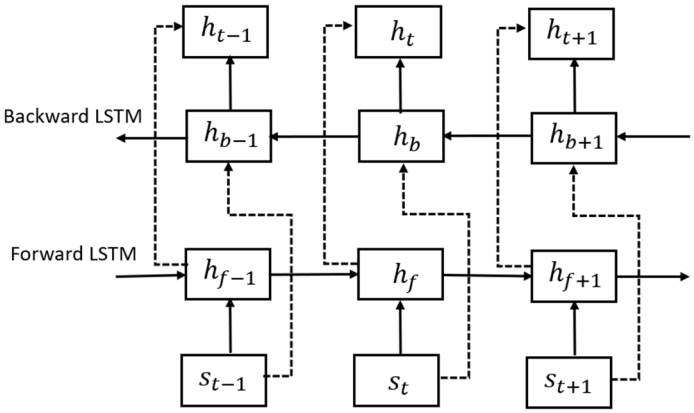

Fig. 3. BiLSTM network architecture.

2.2.3 GRU

Gated recurrent unit (GRU) is a variant of LSTM modification [10]. It provides the least number of estimates for comparable and better results depending on a problem solved. The GRU block scheme is illustrated in Figure 4. Block estimates are the following:

$$z_t = \sigma(w_z[h_{t-1}, s_t] + b_z),$$
$$r_t = \sigma(w_r[h_{t-1}, s_t] + b_r),$$
$$\tilde{h}_t = \tanh(w[r_t \cdot h_{t-1}, s_t] + b),$$

where w, w_z, w_r are the weight matrices states; b, b_z, b_r are the biases; $\cdot$ is the element by element multiplication.

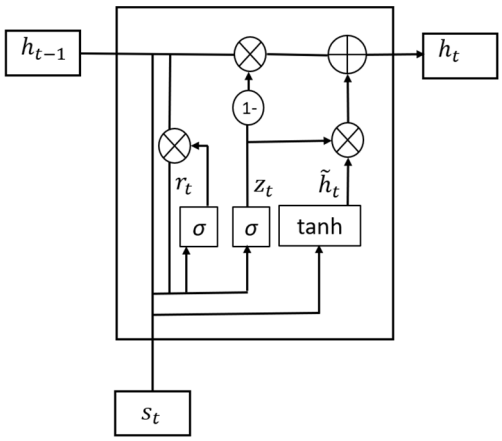

Fig. 4. GRU block scheme.

2.2.4 Projected

The projective layer is based on the LSTM. Its basis is the mode for long-term dependencies training by projected trained weights. The advantage over the original LSTM is the decrease of the number of trained parameters. This is achieved during the projecting of the matrices,

arriving to the layer, into a vector that may improve the accuracy of the SME index forecasting.

2.2.5 Dropout

The auxiliary layer. The Dropout layer sets some NN nodes to 0 in random manner. With the help of this layer, network architecture between iterations changes and the problem of NN overfitting is solved [11].

3 Experiment results

In order to form the training data, IMF Bz and SME index time series over the 5-year period were used. This resulted in the training data array composed of more than 44 million of elements. The trained networks were tested on the one-year data, which were not used in the training. The network performance was estimated on the results of calculation of the root-mean-square error (RMSE) and the mean absolute percentage error (MAPE).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \hat{z}_i)^2},$$

$$MAPE = 100\%\frac{1}{N}\sum_{i=1}^{N}\left|\frac{z_i-\hat{z}_i}{z}\right|,$$

where N is the sampling length, z_i is the SME data, $\hat{z}_i$ is the NN forecast.

NNs were formed according to a single structure: Input layer – X – Fully connected layer – Output layer. The hidden layer X was replaced by one of those described above (LSTM, BiLSTM, GRU, Projected). In case of using the Dropout layer, it was placed after the X layer. The neuron number in the hidden layer was 256 (512 neurons were used for the BiLSTM). The number of learnable properties for each network without the Dropout layer is illustrated in Table 1.

Table 1. Total learnables.

LSTM	BiLSTM	GRU	Projected
277.7*103	555.5*103	208.3*103	3.5*103

The testing results of the NNs without the Dropout layer are shown in Table 2. Analysis of Table 2 illustrates comparable error values among similar architecture of LSTM, BiLSTM, GRU. The least MAPE was obtained in BiLSTM. The GRU architecture showed the highest MAPE but the least RMSE. The NN architecture with the Projected hidden layer demonstrates significant decrease of errors compared to other networks. In this case, the Projected architecture has the least number of learnable properties (Table 1).

Table 2. Estimates of the performance of the NNs without the Dropout layer.

	LSTM	BiLSTM	GRU	Projected
RMSE	141	140	136	110
MAPE	49	44	61	32

Application of the Dropout layer made it possible to improve the quality of generalization among the architectures of LSTM, BiLSTM, GRU that is illustrated in Table 3. The architecture with the Projected layer did not show error decrease.

Table 3. Estimates of the performance of the NNs with the Dropout layer.

	LSTM	BiLSTM	GRU	Projected
RMSE	127	135	133	112
MAPE	30	32	33	33

Results of the network error estimates show that application of the BiLSTM architecture for constructing the neural network model of the SME index does not offer an advantage compared to other architectures. The BiLSTM contains twice as much learnables than the initial LSTM that affects the amount of computation. The Dropout layer decreases the effect of network overtraining and increases the capability to generalization for the data applied in the work. The architecture with the Projected layer is the best one with respect to the ratio of computational resource and the errors obtained on a test set.

4 Conclusion

In this paper, we have estimated different neural network architectures in the problem of geomagnetic SME index modeling. The architecture with a hidden Projected layer showed the least modeling error. Application of the auxiliary Dropout layer improves the quality of generalization of the networks with LSTM, BiLSTM, GRU layers. The obtained results can be used in further work on SME index modeling and forecast.

The work was supported by IKIR FEB RAS State Task (subject registration No. 124012300245-2).

References

1. S.S. Haykin, *Neural networks: a comprehensive foundation* (Prentice Hall, Upper Saddle River, N.J, 1999).
2. V. Gallego, D. Ríos Insua, Current Advances in Neural Networks. Annu. Rev. Stat. Appl. **9**, 197-222 (2022).
3. N.S. Davydov, V.V. Evdokimova, P.G. Serafimovich, V.I. Protsenko, A.G. Khramov, A.V. Nikonorov, Comp. Opt. **47**, 991-1001 (2023).
4. M. Bernaś, B. Płaczek, M. Lewandowski, Sensors **22**, 9451 (2022).
5. P.T. Newell, J.W. Gjerloev, J. Geophys. Res. **116**, 2011JA016936 (2011).
6. Y. Polozov, *Wavelet-based analysis of interplanetary magnetic field and AE-index, in SMART automatics and energy* (Springer Nature Singapore, Singapore, 2022), pp. 469-476.
7. S.G. Mallat, *A wavelet tour of signal processing: the sparse way* (Academic Press, Boston, 2009).
8. A. Marques De Souza, E. Echer, M.J.A. Bolzan, R. Hajra, Ann. Geophys **36**, 205-211 (2018).
9. S. Hochreiter, J. Schmidhuber, Neural Computation **9**, 1735-1780 (1997).
10. K. Cho, B. van Merrienboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: encoder-decoder approaches. arXiv.org. (2014). https://arxiv.org/abs/1409.1259
11. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, The Journal of Machine Learning Research **15**, 1929-58 (2014).