LONG SHORT-TERM MEMORY ARTIFICIAL NEURAL NETWORK WITH DENDRITIC FULL CONNECTED LAYER FOR FORECASTING

Erol Egrioglu – Eren Bas

Abstract

In the literature, deep neural networks are considered as alternative methods that provide a solution to the forecasting problem. It is always advantageous to work with nonlinear functions of lagged variables to obtain forecasts. An important class of deep neural networks is recurrent networks, and the most popular deep neural networks can be created using LSTM cells. Another important artificial neural network used for forecasting is the dendritic neuron model artificial neural network, which can perform better than the multilayer perceptron. In this study, a new deep artificial neural network architecture in which the dendritic neuron model is used in the fully connected layer and LSTM Cell in the other layers are defined specifically for the forecasting problem. In addition, a new training algorithm based on the sine cosine algorithm for the proposed new artificial neural network is given with a step-by-step algorithm. The performance of the proposed forecasting method was compared with other deep and shallow artificial neural networks in the literature by using Istanbul stock market data. It was concluded that the proposed method produced successful forecasting results and the strengths and weaknesses of the method were determined.

Key words: long short-term memory, deep learning, dendritic neuron model, forecasting

JEL Code: C45, C53

Introduction

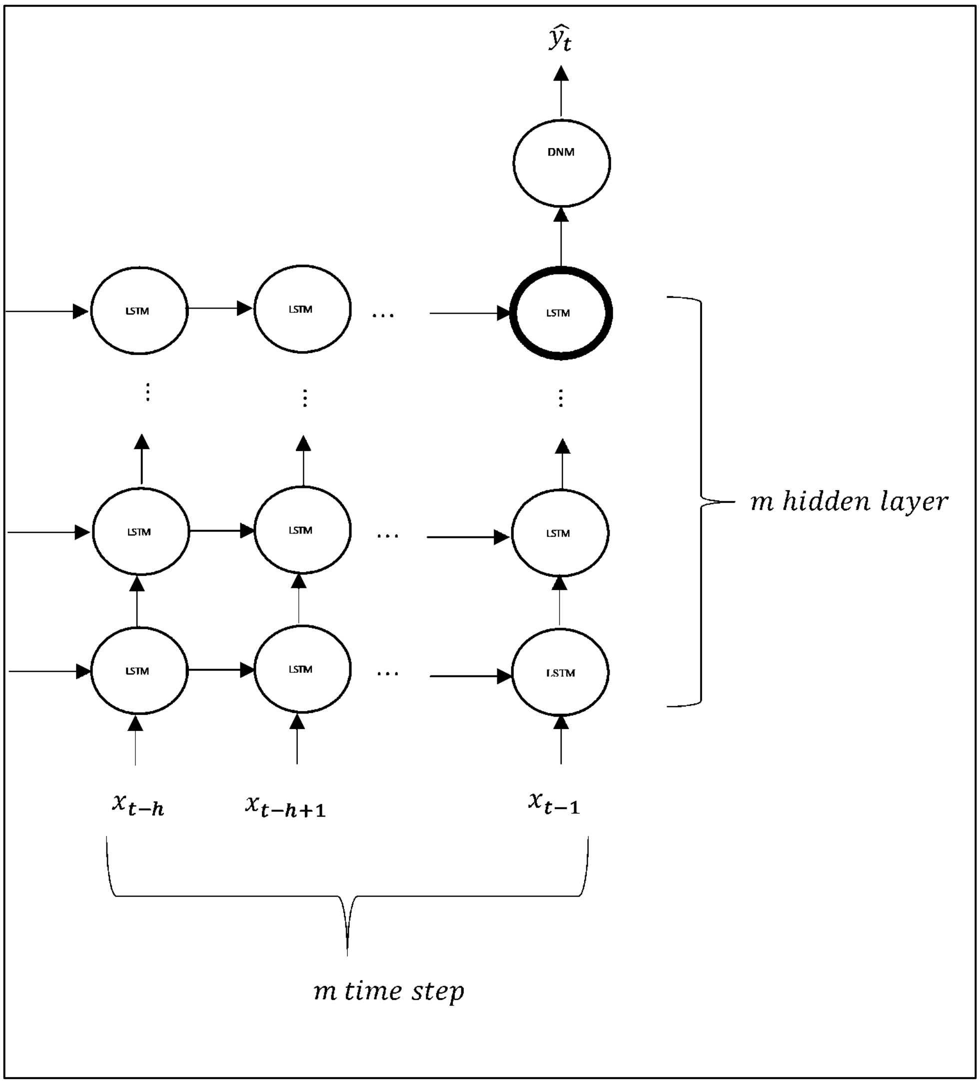
In addition to the classical time series analysis methods, artificial neural networks have become a powerful alternative for solving the forecasting problem. The inputs of artificial neural networks are the lagged variables of the time series and the target value is the simultaneous variable of the time series, in other words, the forecasting problem can be solved. The most commonly used artificial neural network for the forecasting problem is the multilayer perceptron. A multilayer perceptron is an artificial neural network that works with an input, one or more hidden and one output layer and works with neuron models using aggregation functions. Other alternative artificial neural networks used in the literature to solve the forecasting problem are artificial neural networks using the multiplicative neuron model. Multiplicative neuron model is a neuron model that uses multiplicative aggregation instead of additive aggregation. Pi-sigma neural network is a high order artificial neural network based on the multiplicative neuron model proposed by (Shin and Ghos, 1991). Pi-sigma artificial neural network is known to produce successful results in solving the forecasting problem. Another artificial neural network that produces successful results in solving the forecasting problem is the dendritic neuron model (DNM) artificial neural network proposed in (Todo et al., 2014). Ji et al. (2016), Gao et al. (2019), Al-qaness et al. (2022) and Egrioglu et al. (2023) focused on new artificial neural networks proposed based on the dendritic neuron model. Dendritic neuron model artificial neural network is a type of ANN that has been popularly studied in recent years. (Egrioglu et al., 2022) proposed a new artificial neural network that provides a new solution to the time series prediction problem by making the dendritic neuron model artificial neural network recurrent. (Gul et al., 2023) proposed a statistical learning algorithm for training a dendritic neuron model artificial neural network. In recent years, recurrent deep neural networks with a sequential architecture have been frequently used to solve the forecasting problem. The neural network that was originally used for natural language processing but offers an effective solution to the forecasting problem is the long short-term memory (LSTM) artificial neural network. The LSTM artificial neural network was proposed by (Hocreiter and Schmidhuber 1997). In solving the forecasting problem with LSTM, a hierarchical structure based on LSTM cells is created and the last part of the hierarchical structure is connected to the target value with a full connected layer similar to a multilayer perceptron. As such, the LSTM artificial neural network used for the forecasting problem can be defined as a structure that provides the process of analyzing the preprocessed data with MLP.

In this study, a new hybrid artificial neural network is proposed for the forecasting problem by serially connecting two artificial neural networks, LSTM and DNM, which provide effective solutions to the solution of the forecasting problem. A training algorithm using the sine cosine algorithm is presented for the proposed new artificial neural network. In the second part of the paper, the proposed new artificial neural network and its training algorithm are summarized. In the third section, the application results for Istanbul stock market are presented.

1 Proposed Method

In this study, LSTM artificial neural network and DNM artificial neural network are combined in a single architecture and a new deep feedback artificial neural network is created for solving the forecasting problem. The architectural structure of the proposed LSTM-DNM deep neural network is shown in Fig. 1.

Fig. 1: Architecture of LSTM-DNM Deep Artificial Neural Netwrok



"Source: Own research"

Fig. 1, the last cell in the LSTM neural network is defined as a DNM neural network. The output of LSTM-DNM is calculated by the following formulas.

In a LSTM cell, input gate, forget gate and cell candidate outputs are calculated as in the following equations.

$$i_t = \sigma(W_i x_t + R_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma\left(W_f x_t + R_f h_{t-1} + b_f\right) \quad (2)$$

$$g_t = \sigma\left(W_g x_t + R_g h_{t-1} + b_g\right) \quad (3)$$

Cell state value is computed by using Equation (4) and the previous cell state and the outputs of these gates.

$$c_t = f_t * c_{t-1} + i_t * g_t \quad (4)$$

The output of the output gate is calculated by using the (5) formula. The hidden state value is calculated with the equation (6) and the output gate output and the cell state outputs.

$$o_t = \sigma(W_o x_t + R_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t * \tanh(c_t) \quad (6)$$

The input of a LSTM cell is lagged variable of $x_t = (y_t, y_{t-1,\dots,}y_{t-p+1})$. The output of the LSTM-DNM deep recurrent artificial neural network is one step forecast of time series. The output of the network is calculated with following formulas. The dimension of h_t is $h \times 1$ and therefore the number of inputs of the DNM output cell is h. The h_t used in these formula is the output of the LSTM cell indicated by the bold frame in Fig 1.

$$Y_{ij} = \frac{1}{1+\exp(-k(w_{ij}h_t(\mathrm{i})+\theta_{ij}))} \quad i = 1,2,\dots,h; j = 1,2,\dots,Md \quad (7)$$

$$Z_j = \left(\prod_{i=1}^{h} Y_{ij}\right); j = 1,2,\dots,Md \quad (8)$$

$$V = \sum_{j=1}^{Md} Z_j \quad (9)$$

$$\hat{y}_t = \frac{1}{1+exp(-k_{soma}(V-\theta_{soma}))} \quad t = 1,2,\dots,n \quad (10)$$

The proposed training algorithm for training the LSTM-DNM deep neural network is given in Algorithm 1.

Algorithm 1. Training Algorithm for LSTM-DNM

Step 1. The initial parameters are determined.

p: The number of inputs for LSTM cells in the first layer of LSTM-DNM

h: The number of time steps

m: The number of hidden layers in LSTM-DNM

Md: The number of dendrites in LSTM-DNM

Ps: Population size in sine cosine algorithm

$mitr$: Maximum iteration number in sine cosine algorithm

Step 2. Initialize the population of SCA.

Step 3. Calculate the fitness function value of each position. The mean square error (MSE) calculated from the outputs of the LSTM-DNM network was chosen as the fitness function.

Step 4. Obtain the best position (P_{best}) in the population

Step 5. Obtain new positions in the population by using following equations.

$$\mathrm{P}_\mathrm{i}^{\mathrm{t+1}} = \begin{cases} \mathrm{P}_\mathrm{i}^\mathrm{t} + r_1 \times sin(r_2) \times |r_3 P_{best} - P_\mathrm{i}^\mathrm{t}| & r_4 < 0.5 \\ \mathrm{P}_\mathrm{i}^\mathrm{t} + r_1 \times cos(r_2) \times |r_3 P_{best} - P_\mathrm{i}^\mathrm{t}| & r_4 \geq 0.5 \end{cases} ; i = 1,2,\cdots pn \qquad (11)$$

$$r_1 = \alpha \left(1 - \frac{t}{maxitr}\right) \qquad (12)$$

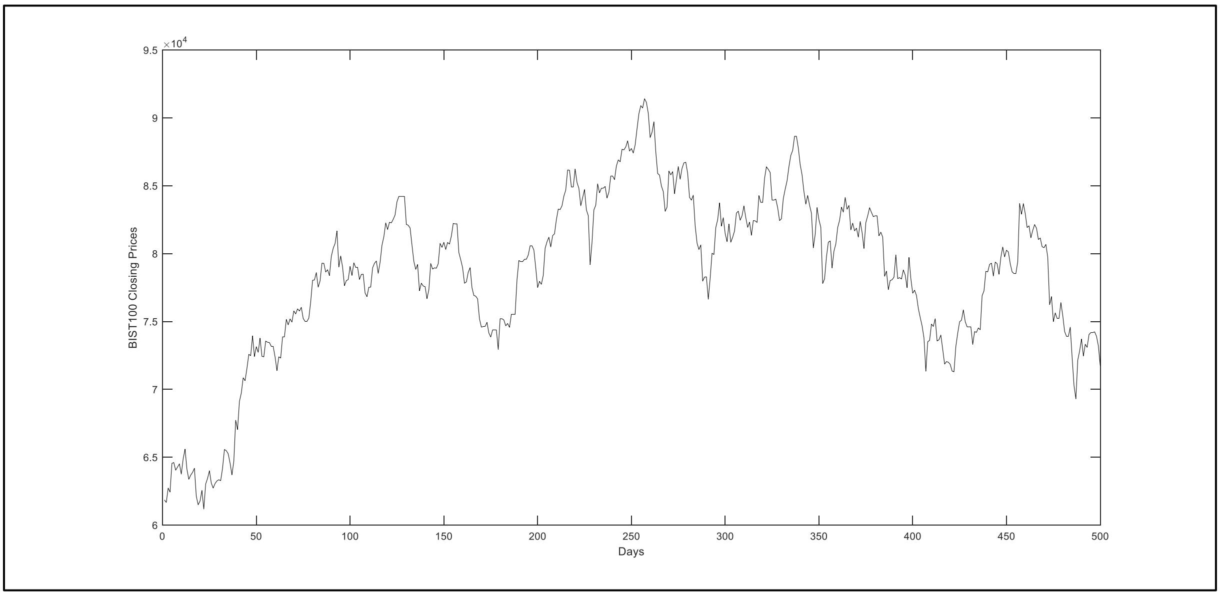
Step 6. Update P_{best}.

Step 7. Check the stopping conditions. If the stopping condition is not met, go back to step 5, otherwise terminate the algorithm.

2 Applications

In the application, 500-day time series of Borsa Istanbul BIST100 index between 31.01.2014 and 31.12.2015 were used. The graph of the time series is given in Fig 2.

Fig. 2: BIST100 closing price time series



"Source: Own research"

In practice, the time series were analysed by LSTM deep artificial neural network, Pi-sigma high-order artificial neural network and bootstraped hybrid ANN (B-HANN) methods proposed in (Egrioglu & Fildes, 2022). In addition, the naive method and Holt's linear exponential smoothing methods are used as classical time series forecasting methods.

Tab. 1: RMSE Statistics of Methods

Methods	Mean	Standard Deviation	Random Walk	Holt Linear Trend
LSTM	1223,6421	15,1430	1207,0821	1210,3548
PSGM	1214,5322	8,1635		
B-HANN	1212,3883	1,8548		
LSTM-DNM	1178,5634	0,7495		

"Source: Own research"

In the analysis of all methods, the time series is divided into 3 blocks as training, validation and test set. While the optimal weights of the network were determined with the training data, the hyperparameters of the methods were selected with the validation set. The test set was used to compare the performance of the methods. For the best hyperparameters, all ANN methods were run with 30 different random initialisations and the RMSE mean and standard deviation statistics of the methods were calculated and given in Tab. 1.

$$RMSE = \sqrt{\frac{1}{ntest}\sum_{t=1}^{ntest}(y_t - \hat{y}_t)^2} \qquad (13)$$

The best hyperparameter values in LSTM are; number of inputs 2, number of hidden layers 1 and number of time steps 2. The best hyperparameter values in PSGM are 4 inputs and 4 hidden layer units. The best hyperparameter values in B-HANN are 4 inputs and 5 hidden layer units. In the proposed method, the best hyperparameter values are determined as number of inputs 5, number of hidden layers 1, number of time steps 5 and number of dendrites 2. Tab 1 shows that the proposed method has lower mean and standard deviation than all ANNs. It is also seen that the average performance of the proposed method is better than the classical prediction methods.

Conclusion

In this study, LSTM-DNM as a new deep artificial neural network is proposed for solving the forecasting problem. A training algorithm based on the sine cosine algorithm is proposed for the proposed LSTM-DNM. The performance of the proposed method is compared with some popular ANNs and classical time series methods for Borsa Istanbul stock exchange time series. It is observed that LSTM-DNM is able to produce more successful prediction results than other methods for the analysed time series. In future studies, it is planned to apply

LSTM-DNM to a large number of time series and evaluate the results obtained by using statistical hypothesis tests.

References

Al-qaness, M. A., Ewees, A. A., Elaziz, M. A., & Samak, A. H. (2022). Wind Power Forecasting Using Optimized Dendritic Neural Model Based on Seagull Optimization Algorithm and Aquila Optimizer. *Energies*, 15(24), 9261.

Egrioglu, E., & Fildes, R. (2022). A New Bootstrapped Hybrid Artificial Neural Network Approach for Time Series Forecasting. Computational Economics, 59(4), 1355–1383.

Egrioglu, E., Bas, E., & Chen, M. Y. (2022). Recurrent dendritic neuron model artificial neural network for time series forecasting. *Information sciences*, 607, 572–584.

Egrioglu, E., Bas, E., & Karahasan, O. (2023). Winsorized dendritic neuron model artificial neural network and a robust training algorithm with Tukey's biweight loss function based on particle swarm optimization. *Granular Computing*, 8(3), 491–501.

Gao, S., Zhou, M., Wang Y. , Cheng J., Yachi H., Wang J. (2019). Dendritic Neuron Model with Effective Learning Algorithms for Classification, Approximation, and Prediction, IEEE Transactions on Neural Networks and Learning Systems, 30(2), 8409490, 601–614.

Gul, H. H., Egrioglu, E., & Bas, E. (2023). Statistical learning algorithms for dendritic neuron model artificial neural network based on sine cosine algorithm. *Information Science*, 629, 398–412.

Ji J., Gao S., Cheng J., Tang Z., & Todo Y. (2016). An approximate logic neuron model with a dendritic structure. Neurocomputing, 173, 1775–1783.

S. Hocreiter, J. Schmidhuber, Long Short-Term Memory, *Neural Computation* 9(8) (1997) 1735–1780.

Shin, Y., Gosh, J. (1991). The Pi-Sigma network: an efficient higher-order neural network for pattern classification and function approximation. *In: Proceedings of the International Joint Conference on Neural Networks*, Seattle, 13–18.

Todo, Y., Tamura, H., Yamashita, K., & Tang, Z. (2014). Unsupervised learnable neuron model with nonlinear interaction on dendrites. *Neural Networks,* 60, 96–103.

Contact

Erol Egrioglu

Giresun University

Gure Campus, Faculty of Arts and Science, Department of Statistics, Giresun, Türkiye, 28200.

erol.egrioglu@giresun.edu.tr

Eren Bas

Giresun University

Gure Campus, Faculty of Arts and Science, Department of Statistics, Giresun, Türkiye, 28200.

eren.bas@giresun.edu.tr