RIGA TECHNICAL
UNIVERSITY

Design of Experiments vs. TOPSIS to Select Hyperparameters of Neural Attention Models in Time Series Prediction

Yunus Emre Midilli[1], Sergei Parshutin[2]
[1, 2] *Riga Technical University, Riga, Latvia*

Abstract – **Attention models are used in neural machine translation to overcome the challenges of classical encoder-decoder models. In the present research, design of experiments and TOPSIS methods are used to select hyperparameters of a neural attention model for time series prediction. The configurations selected by both methods are compared with out-of-sample data in time interval between January 2020 and April 2020 when global economies were significantly impacted due to Covid-19 pandemic. Results demonstrated that both selection methods outperformed each other in terms of different output features. On the other hand, our results with more than 95 % coefficient of determination and less than 0.23 % MAPE verified that neural attention models had strong capabilities in exchange rate prediction even in extraordinary situations in global economies.**

Keywords – **Design of experiments, hyperparameter, neural attention, time series, TOPSIS.**

I. INTRODUCTION

Encoder-decoder models are used in neural machine translation. Despite their success, two drawbacks always become a challenge: (i) long-range dependency problem when encoders generate bad summarisation in longer sentences, decoder produces bad translation [1]; (ii) context vector that consists of the latest unit of encoder while decoder needs to focus on the different locations of the words in encoder. Authors of [2] have proposed neural attention mechanisms where all input words are considered in context vector and relative importance is given to each of them. Neural attention mechanisms have given successful results in neural machine translation. Additionally, attention models are used in time series prediction. However, neural attentions mechanisms have hyperparameters to be set like other deep learning algorithms. To the best of our knowledge, hyperparameter selection techniques are not compared for neural attention mechanism in time series prediction problems. In this context, contribution of this study can be listed as follows:

We present a neural attention model that can be used in time series prediction where inputs of encoders and decoders are capable of including multiple features in a specific time frame.

We compare two methods that are used in hyperparameter adjustment of neural attention mechanisms: the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and Design of Experiments (DOE).

The rest of the paper is arranged as follows. Section 2 is a literature review about the hyperparameter optimization of neural networks and applications of neural attention models in time series prediction. In Section 3, the proposed method based on three techniques is explained: neural attention models, TOPSIS and DOE. In Section 4, USD/EUR exchange rate has been forecast between January 2018 and April 2020 via a neural attention model.

II. LITERATURE REVIEW

Various neural network models are used in time series prediction, such as multi-layer perceptron [3], [4], recurrent neural networks [5], [6], long short-term memory [7], [8], gated recurrent unit [9], convolutional neural networks [10], [11] etc. Among these, neural attention models initially developed for neural machine translation have recently been used in time series prediction. In the research of [12], prediction performance of neural attention model is compared to benchmark models of multi-layer perceptron, convolutional neural networks, recurrent neural networks and long short-term memory models for a time series. Discrete wavelet transformation and fast Fourier transformation are used to convert time series data into signal. It is verified that neural attention models outperform the benchmark models. Authors of [13] have suggested a learnable Softmax function to perform normalization in the attention layer. They have applied the proposed model to predict cryptocurrency exchange rates. Prediction performance improved from 6 % to 15 % compared to benchmark models of classical encoder-decoder neural networks and RNN. Apart from the normalization function, neural attention models have various hyperparameters like other deep learning neural networks. Hyperparameter selection has been a challenge for neural networks. Most of the practitioners use a trial-and-error or grid search approach to overcome this challenge [14], [15]. There are various methodologies applied in the literature to optimise the hyperparameters of neural networks: simulated annealing [16], genetic algorithms [17], particle swarm optimization [18] etc. Apart from the evolutionary algorithms, design of experiments, which is a statistical technique, has also been used [19]. In the research of [20], exponential smoothing transition auto-regressive time series is used for prediction. In this context, the

authors used DOE to optimise a number of inputs, number of hidden layers, number of hidden neurons, activation function and window length. In addition, we have prepared a literature review where genetic algorithms and design of experiments are used as an optimization tool in exchange rate prediction [21]. The most common DOE methods that are used for hyperparameter optimization in time series prediction are Taguchi method [22], response surface methodology [23], full factorial design [24] etc. In addition to the evolutionary and statistical approaches, multi-criteria decision making techniques are also used for hyperparameter optimization in time series prediction. Authors of [25] conducted exploratory analysis for CNN, hybrid CNN-LSTM, MLP and radial basis neural network in order to predict six popular cryptocurrencies based on technical indicators. Researchers of [26] suggested TOPSIS methodology to rank and select the base classifiers in stock price movement direction prediction. Base classifiers consist of neural network alternatives such as MLP and radial basis function network. To the best of our knowledge, there is no study where hyperparameters of the neural attention model are selected via DOE and TOPSIS.

III. METHODOLOGY

This study focuses on comparing two methods in hyperparameter selection for the neural attention model in time series prediction. Based on identified hyperparameter types, some configurations are created as a candidate solution. Criteria are the response values that depict the forecasting performance. Neural attention model is used to train and test each candidate solution. According to test results, DOE and TOPSIS methods are applied separately. The best configurations are selected for both methods. Out-of-sample (unseen) data are shown to a neural attention model under the selected hyperparameter configurations. Prediction performance of both configurations is compared via Wilcoxon signed rank test (Fig. 1).

A. Neural Attention Model

Neural attention models have been proposed to improve the performance of encoder-decoder models in neural machine translation [2], [27], [28]. Neural attention model has the attention layer in addition to its encoder-decoder networks. Encoder-decoder units are GRU units where input of them is one-hot vector representation of sentence in input and target

languages, respectively. Decoder has two inputs initially: the latest state of encoder unit and a start token. In this study, neural attention model for neural machine translation has been converted in a way that we can perform time series prediction (Fig. 2). Input for an encoder unit can be represented by, F_{t-N}^i, feature vector of input dataset:

$$F_{t-N}^i = [f_{t-N}^1, f_{t-N}^2, \dots f_{t-N}^i], \ \forall N = \{1, 2, \dots, n\}, \tag{1}$$

where i is the number of input features, n is backward window length of time stamp t. Input for decoder F_{t+M}^o is a feature vector for output dataset:

$$F_{t+M}^o = [f_{t+M}^1, f_{t+M}^2, \dots f_{t+M}^o], \ \forall M = \{0, 1, 2, \dots, m\}, \tag{2}$$

where o is the number of output features, m is forward window length of time stamp t. Encoder units consist of GRU units with γ number of hidden neurons as output space dimensionality. Hidden states of encoder and decoder units are represented with

$$H_{t-n}^e = [h_0^e, h_{t-1}^e, h_{t-2}^e, \dots h_{t-n}^e] \tag{3}$$

$$H_{t+m}^d = [h_0^d, h_t^d, h_{t+1}^d, \dots h_{t+m}^d], \tag{4}$$

respectively. Initial state of encoder, h_0^e, is a zero vector. Initial state of decoder, h_0^d, is equal to the latest state of encoder h_{t-n}^e. In this study, attention layer proposed by Bahdanau [2] is used. Bahdanau's attention layer alignment scores are:

$$\text{score}(h_{t-N}^e, h_{t+M}^d) = W_3^T \tanh(W_1 h_{t-N}^e + W_2 h_{t+M}^d),$$

$$\forall M = \{0, 1, \dots, m\}, N = \{1, 2, \dots, n\}. \tag{5}$$

Attention weights, $a_{t-N,M}$, are calculated using Softmax function:

$$a_{t-N,M} = \frac{e^{\text{score}(h_{t-N}^e, h_{t+M}^d)}}{\sum_{\dot{m}=1}^m e^{\text{score}(h_{t-N}^e, h_{t+\dot{m}}^d)}}, \forall M = \{0, 1, \dots, m\}. \tag{6}$$

Context vector, $c_{t-N} = \sum_{\dot{m}=0}^m a_{t-N,M} h_{t-N}^e$, is concatenated with inputs of the decoder layer. Concatenated values are used as inputs to GRU units. Start token is denoted with a zero vector $[0 \ \cdots \ 0]_{o \times 1}$. Output of decoder is passed through the dense layer to create predicted output vectors. Teacher forcing technique, where output vector is passed as the next input to decoder, is used during training. During testing, the predicted output vector is passed as the next input to decoder.
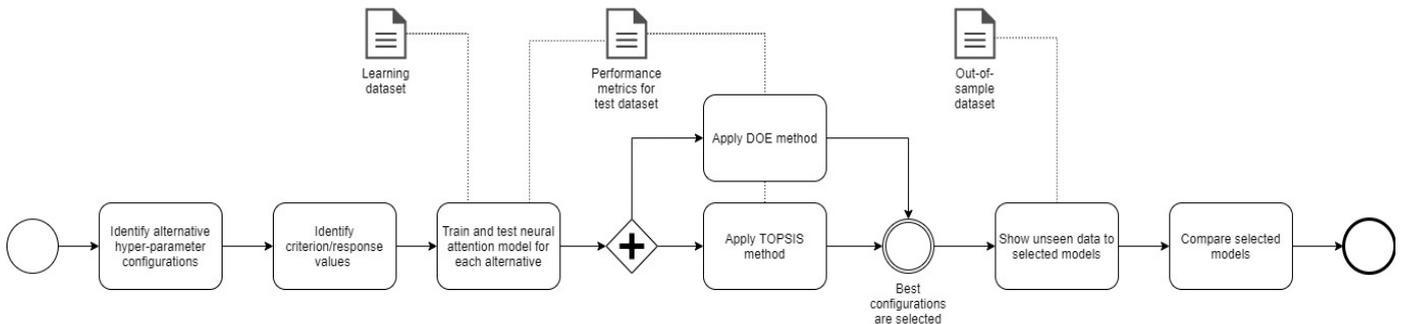
Fig. 1. The proposed model.
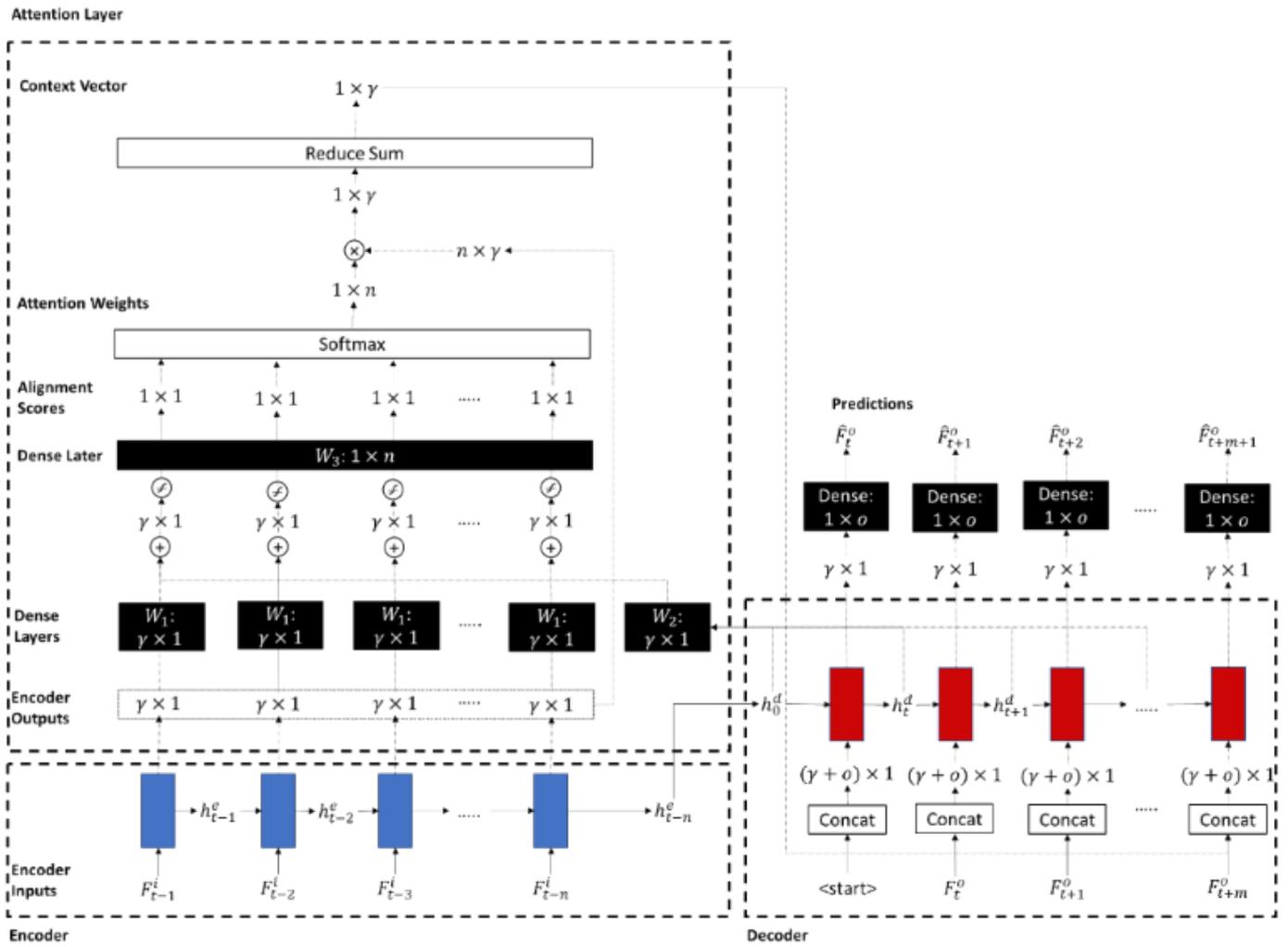
Fig. 2. The proposed neural attention model for time series prediction.

B. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

TOPSIS method is based on the concept that the chosen alternative should have the shortest Euclidian distance from the positive ideal solution (PIS) and the furthest Euclidian distance from the negative ideal solution (NIS). Alternatives are ranked with number in TOPSIS method where PIS has ranking number 1 and NIS has ranking number 0. TOPSIS method consists of eight activities [29]:

1. Establish the decision matrix – Let $X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$ be a decision matrix, where m is the number of alternatives and n is the number of criteria.

2. Determine weight of criteria – Let $W = [w_1 \quad \cdots \quad w_n]$ be a weight vector, where $\sum_{j=1}^{n} w_j = 1$.

3. Calculate normalized decision matrix – Normalization of decision matrix is as follows:

$$n_i = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}, \tag{7}$$

where $i = \{1, \dots, m\}$ and $j = \{1, \dots, n\}$.

4. Calculate the weighted normalized decision matrix – Let v_{ij} denote weighted normalized value, then $v_{ij} = w_j n_{ij}, \forall\, i = \{1, \dots, m\}$ and $j = \{1, \dots, n\}$.

5. Determine the PIS and NIS – PIS represents the extreme performance on each criterion while NIS represents reverse extreme performance on each criterion. Let positive ideal solution and negative ideal solution be denoted with A^+ and A^-, respectively,

$$A^+ = (v_1^+, \dots v_n^+,) =$$
$$\left(\left(max_i(v_{ij}) | j \in \right), \left(min_i(v_{ij}) | j \in J \right) \right) \tag{8}$$

and

$$A^- = (v_1^-, \dots v_n^-,) =$$
$$\left(\left(max_i(v_{ij}) | j \in \right), \left(min_i(v_{ij}) | j \in J \right) \right), \tag{9}$$

where I is associated with benefit criteria, J with cost criteria and $i = \{1, \dots, m\}$ and $j = \{1, \dots, n\}$.

6. Calculate the separation measures from the PIS and NIS for each alternative – Let Euclidian separation of each alternative from PIS and NIS be denoted with d_i^+ and d_i^-, respectively for each $i = \{1, \dots, m\}$:

$$d_i^+ = \sqrt{\left(\sum_{j=1}^n (v_{ij} - v_j^+)^2\right)} \qquad (10)$$

$$\text{and } d_i^- = \sqrt{\left(\sum_{j=1}^n (v_{ij} - v_j^-)^2\right)} \qquad (11)$$

7. Calculate the relative closeness to PIS for each alternative – Let R_i be relative closeness of i th alternative with respect to PIS of it. R_i is calculated as: $R_i = \frac{d_i^-}{d_i^- + d_i^+}$, where $0 \le R_i \le 1$ for each $i = \{1, \dots, m\}$.

8. Rank the preference order to select the alternative close to 1 – Relative closeness to PIS values are ranked descending. The alternative that has the highest relative closeness to PIS value is selected.

In this study, alternatives of decision matrix are the different configurations of hyperparameters for the neural attention model. Neural attention model is trained and tested with the same data for each alternative. Performance metrics of test results is noted for each alternative. Criterion for decision matrix is average ($\bar{X}$) and range (R) values of each output feature within the forward window length. Once decision matrix is built, the above-mentioned TOPSIS method is applied. The best alternative is selected to be simulated with out-of-sample data.

C. Design of Experiments

A design of experiments (DOE) is a set of trials, in which change in response is observed systematically under different factor levels [30]. DOE is mostly used in production and service industries for the process optimization purposes. In this study, selecting some hyperparameters of a neural attention mechanism is considered a process optimization objective. Procedure for performing DOE consists of the following steps:

1. Recognise and state the problem – Hyperparameter selection is an unknown procedure where practitioners mostly apply trial and error methodology. Trial and error method enables one to monitor the change in response based on the impact of single factors. Second or higher order interaction between factors might have a significant effect on the response variables.

2. Choose factors and levels – In this study, factor levels are hyperparameters of a neural attention mechanism. Learning from literature and previous experiments can be used to identify the levels of hyperparameters.

3. Select a response variable – Response variables are the forecasting performance of neural attention mechanism. In time series analysis, there can be multiple output features to be predicted. In this paper, average and range values of output features within the future window length are considered response values.

4. Choose experimental design – There are various experimental designs: 2-factor experimental design, full factorial design, Taguchi design, Box-Behnken design, central composite design etc. In this study, full factorial design is used.

5. Perform the experiment – According to chosen experimental design, different configurations of hyperparameters are created. Each configuration is run. Configurations can be run more than one time as well. In this study, each configuration is run only once due to computation costs.

6. Perform data analysis – Analysis of variance (ANOVA) is applied to identify the factors that have a significant effect on response values. In case there are multiple response values, ANOVA is applied to each of them separately. ANOVA can be applied in a stepwise approach by eliminating the effects that are not significant. According to ANOVA, mathematical models are created. These mathematical models are an objective function of optimization model. Multiple response, weights and importance values are given according to a desirability function [31]. Desirability is an objective function that ranges the goal between zero and one. Response values are weighted between zero and one. According to the output of response optimization procedure, the best configuration is selected.

IV. APPLICATION

In order to apply the proposed method in a time series, hourly exchange rate prices of EUR/USD are gathered from MetaTrader trading platform [32] for the time period between 1 January 2018 and 31 December 2019 (Fig. 3).
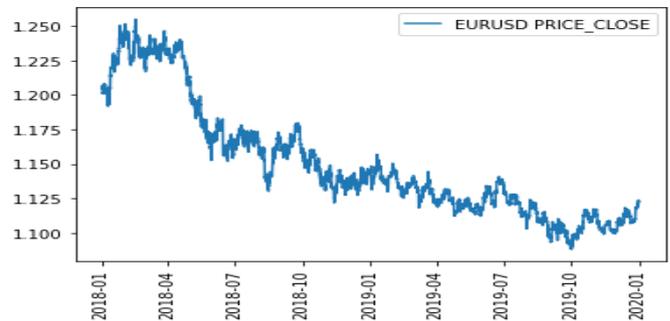
Fig. 3. Closing price of EUR/USD.

Input features for encoder are day, weekday, and hour of time stamp, opening price and closing price. Output features for decoder are opening price and closing price. Table I demonstrates the descriptive statistics of sample data.

Data are pre-processed by shifting the operation of backward window length (n) and forward window length (m). Backward window length for input dataset is 3 hours that depicts t-1, t-2 and t-3. Forward window length for the output dataset is 2 hours that depicts t, t+1 and t+2. Input and output matrices are partitioned to 80 % and 20 % as training and test datasets after the shuffling operation. Train and test datasets consist of 9915 and 2479 observations, respectively. Train and test datasets are standardised between 0 and 1.

TABLE I

DESCRIPTIVE STATISTICS OF OUTPUT FEATURES

	OPENING PRICE	CLOSING PRICE
COUNT	12394	12394
MEAN	1.15043	1.15043
STD	0.04155	0.04155
MIN	1.08804	1.08804
25 %	1.11974	1.11975
50 %	1.13746	1.13745
75 %	1.16924	1.16923
MAX	1.25489	1.25490

In this study, hyperparameters of epoch size, batch size, number of hidden layers and backward window length are to be identified via DOE and TOPSIS. In this context, alternative configurations are prepared based on predefined minimum, maximum and step values (Table II).

TABLE II

LEVELS OF HYPERPARAMETERS

HYPERPARAMETER	MINIMUM	STEP	MAXIMUM
EPOCH SIZE (A)	10	45	100
BATCH SIZE (B)	16	48	112
NUMBER OF HIDDEN NEURONS (C)	20	30	80
BACKWARD WINDOW LENGTH (D)	4	4	12

Each configuration is run once. Decision matrix is built as noting the MAPE test results of each alternative (Table III). Each criterion has equal weight. TOPSIS method is applied for a decision matrix. Rankings after TOPSIS method show that the best configuration is where epoch size is 10, batch size is 112, the number of hidden neurons is 80 and backward window length is 4.

Stepwise selection ANOVA method is applied to a decision matrix for each criterion. Final ANOVA outputs with confidence level of 0.15 are represented in Table IV. Final ANOVA outputs demonstrate that there are two-way interaction and square effects of factors. According to significant effects, response surface optimization is applied as considering weights of each response equal based on mathematical models. Optimum architecture is found where

epoch size is 10, batch size is 112, the number of hidden neurons is 20 and backward window length is 4. The best selection of DOE is the second best selection of TOPSIS (Table III).

TABLE III

DECISION MATRIX (81 ALTERNATIVES ARE CHECKED)

ALTERNATIVE	OPENING PRICE		CLOSING PRICE		RANK
	$\bar{X}$	R	$\bar{X}$	R	
(10-112-80-4)*	0.1018	0.0077	0.1149	0.0325	1
(10-112-20-4)	0.0896	0.0509	0.1089	0.0146	2
(10-64-20-4)	0.1078	0.0305	0.1166	0.0524	3
(10-112-20-8)	0.1129	0.0244	0.1282	0.0173	4
(10-64-50-8)	0.1288	0.0859	0.1257	0.0299	5
(10-112-80-12)	0.1131	0.0492	0.1442	0.0307	6
(10-64-80-8)	0.1461	0.0472	0.14	0.0172	7
(10-112-50-12)	0.1338	0.0902	0.1358	0.0589	8
(10-112-50-8)	0.1277	0.1043	0.1351	0.0653	9
(10-112-80-8)	0.1729	0.1535	0.1306	0.0342	10
...	…	…	…	…	…

Both configurations that are identified via TOPSIS and DOE are run in a simulation 30 times with unseen data. Out-of-sample data consist of 2433 observations between 2 January 2020 and 22 May 2020. Wilcoxon signed rank test is applied to test the difference between TOPSIS selection and DOE selection in terms of R^2 and MAPE (Table V). For both time intervals $t+1$ and $t+2$, TOPSIS selection outperforms on opening and closing price predictions. Table VI shows the mean and standard deviation of coefficient of determination and MAPE results for each output feature. It can be seen that prediction performance is decreasing when it moves from t to $t+2$. However, there is not such a trend for the deviations. Both selections verify that the neural attention model is successful for time series prediction. Out-of-sample time interval is Covid-19 pandemic period when global economies are significantly impacted. This study verifies that neural attention models are strong enough even the prediction period is an extraordinary situation.

TABLE IV

FINAL ANOVA OUTPUTS OF RESPONSES

a) Opening Price ($\overline{X}$)

SOURCE	DF	ADJ SS	ADJ MS	*F*-VALUE	*P*-VALUE
Model	8	22.4529	2.8066	17.70	0.000
Linear	3	18.3179	6.1060	38.51	0.000
A	1	13.6642	13.6642	86.18	0.000
B	1	1.1343	1.1343	7.15	0.009
D	1	3.5194	3.5194	22.20	0.000
Square	3	2.7270	0.9090	5.73	0.001
A*A	1	1.9086	1.9086	12.04	0.001
B*B	1	0.4660	0.4660	2.94	0.091
D*D	1	0.3524	0.3524	2.22	0.140
2-Way Interaction	2	1.4079	0.7040	4.44	0.015
A*B	1	0.8917	0.8917	5.62	0.020
A*D	1	0.5163	0.5163	3.26	0.075

b) Opening Price (R)

SOURCE	DF	ADJ SS	ADJ MS	*F*-VALUE	*P*-VALUE
Model	7	76.546	10.9351	21.85	0.000
Linear	3	66.010	22.0032	43.97	0.000
A	1	49.485	49.4847	98.88	0.000
B	1	5.036	5.0364	10.06	0.002
D	1	11.489	11.4885	22.96	0.000
Square	2	6.958	3.4789	6.95	0.002
A*A	1	5.624	5.6236	11.24	0.001
D*D	1	1.334	1.3342	2.67	0.107
2-Way Interaction	2	3.578	1.7892	3.58	0.033
A*B	1	1.688	1.6876	3.37	0.070
A*D	1	1.891	1.8909	3.78	0.056

c) Closing Price ($\overline{X}$)

SOURCE	DF	ADJ SS	ADJ MS	*F*-VALUE	*P*-VALUE
Model	8	22.9368	2.8671	19.65	0.000
Linear	3	19.2470	6.4157	43.96	0.000
A	1	14.7916	14.7916	101.36	0.000
B	1	1.1417	1.1417	7.82	0.007
D	1	3.3137	3.3137	22.71	0.000
Square	3	2.4548	0.8183	5.61	0.002
A*A	1	1.5921	1.5921	10.91	0.001
B*B	1	0.4224	0.4224	2.89	0.093
D*D	1	0.4402	0.4402	3.02	0.087
2-Way Interaction	2	1.2350	0.6175	4.23	0.018
A*B	1	0.6784	0.6784	4.65	0.034
A*D	1	0.5566	0.5566	3.81	0.055

d) Closing Price (R)

SOURCE	DF	ADJ SS	ADJ MS	*F*-VALUE	*P*-VALUE
Model	11	84.574	7.6886	19.01	0.000
Linear	4	68.864	17.2159	42.56	0.000
A	1	51.628	51.6275	127.63	0.000
B	1	5.070	5.0704	12.53	0.001
C	1	1.285	1.2847	3.18	0.079
D	1	10.881	10.8809	26.90	0.000
Square	3	7.423	2.4742	6.12	0.001
A*A	1	4.788	4.7883	11.84	0.001
B*B	1	0.981	0.9808	2.42	0.124
D*D	1	1.653	1.6535	4.09	0.047
2-Way Interaction	4	8.288	2.0720	5.12	0.001
A*B	1	1.294	1.2944	3.20	0.078
A*C	1	1.689	1.6893	4.18	0.045
A*D	1	2.011	2.0108	4.97	0.029
B*C	1	3.293	3.2935	8.14	0.006

TABLE V

WILCOXON SIGNED RANK TEST FOR DIFFERENCE BETWEEN TOPSIS SELECTION AND DOE SELECTION

OUTPUT FEATURES	R^2 (CI FOR H)	MAPE (CI FOR H)
Opening Price (t)	(−0.0006655, 0.0008595)	(−0.0079796, 0.0031764)
Closing Price (t)	(−0.000373, 0.0008695)	(−0.0086915, 0.000551)
Opening Price ($t+1$)	(0.000409, 0.0046065)	(−0.030903, −0.0047)
Closing Price ($t+1$)	(0.000745, 0.005122)	(−0.0329715, −0.0075945)
Opening Price ($t+2$)	(0.0002485, 0.0045935)	(−0.0292595, −0.004484)
Closing Price ($t+2$)	(0.00068, 0.005236)	(−0.031853, −0.007039)

TABLE VI

DESCRIPTIVE STATISTICS OF TOPSIS SELECTION AND DOE SELECTION WITH OUT-OF-SAMPLE DATA

OUTPUT FEATURES	TOPSIS SELECTION		DOE SELECTION	
	R^2	MAPE	R^2	MAPE
OPENING PRICE (t)	0.990 ± 0.0017	0.091 ± 0.0142	0.990 ± 0.0017	0.094 ± 0.0128
CLOSING PRICE (t)	0.983 ± 0.0013	0.118 ± 0.0108	0.983 ± 0.0017	0.122 ± 0.0112
OPENING PRICE ($t+1$)	0.982 ± 0.0024	0.124 ± 0.0171	0.979 ± 0.0065	0.146 ± 0.0320
CLOSING PRICE ($t+1$)	0.975 ± 0.0022	0.146 ± 0.0160	0.971 ± 0.0067	0.169 ± 0.0303
OPENING PRICE ($t+2$)	0.975 ± 0.0024	0.147 ± 0.0161	0.971 ± 0.0067	0.168 ± 0.0311
CLOSING PRICE ($t+2$)	0.968 ± 0.0022	0.165 ± 0.0150	0.964 ± 0.0070	0.188 ± 0.0302

V. CONCLUSION

In this study, a stepwise approach was presented to select hyperparameters of neural attention model for time series prediction. The proposed model was applied to predict EUR/USD exchange rate between January 2018 and April 2020. Epoch size, batch size, the number of hidden neurons and backward window length of the neural attention model were identified via two techniques: TOPSIS, multi-criteria decision-making technique and DOE (i.e., a statistical technique). Selection from both methods consisted of epoch size 10, batch size 112 and backward window length of 2 hours. While TOPSIS selection had 80 hidden neurons, DOE selection had 20 hidden neurons. Both selections were simulated 30 times with unseen data between January 2020 and April 2020. Results depicted that TOPSIS selection outperformed DOE selection. As further steps of this study, the boundaries of hyperparameters can be extended for both methods. However, it is still not known how much they should be extended. On the other hand, existing setup is still strong enough to perform prediction in a volatile time series even under extraordinary situation such as Covid-19 impact on global economies. Additionally, unstructured data (e.g., news, social media etc.) and technical encoders can be used in encoder of the neural attention model. There are various types of attention models such as local attention, hard attention, soft attention etc. These models can also be used in time series prediction problems. They have additional hyperparameters that can be optimised with TOPSIS and DOE techniques.

REFERENCES

[1] K. Cho, B. v. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," arXiv, 2014. https://doi.org/10.3115/v1/D14-1179

[2] D. Bahdanau, K. Cho and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," in *International Conference on Learning Representations*, San Diego, 2015.

[3] R. Adcock and N. Gradojevic, "Non-fundamental, non-parametric Bitcoin forecasting," *Physica A: Statistical Mechanics and its Applications,* vol. 531, Article no. 121727, 2019. https://doi.org/10.1016/j.physa.2019.121727

[4] J. Henríquez and W. Kristjanpoller, "A combined Independent Component Analysis–Neural Network model for forecasting exchange rate variation," *Applied Soft Computing Journal,* vol. 83, no. 105654, 2019. https://doi.org/10.1016/j.asoc.2019.105654

[5] Z. Berradi and M. Lazaar, "Integration of Principal Component Analysis and Recurrent Neural Network to Forecast the Stock Price of Casablanca Stock Exchange," *Procedia Computer Science*, vol. 148, pp. 55–61, 2019. https://doi.org/10.1016/j.procs.2019.01.008

[6] D. C. Mallqui and R. A. Fernandes, "Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques," *Applied Soft Computing Journal,* vol. 75, pp. 596–606, 2019. https://doi.org/10.1016/j.asoc.2018.11.038

[7] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Systems With Applications,* vol. 103, pp. 25–37, 2018. https://doi.org/10.1016/j.eswa.2018.03.002

[8] Y. Liu, C. Yang, K. Huang and W. Gui, "Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network," *Knowledge-Based Systems,* vol. 188, art. no. 105006, 2019. https://doi.org/10.1016/j.knosys.2019.105006

[9] J. L. Minqi Jiang and C. L. Lu Zhang, "An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms," *Physica A: Statistical Mechanics and its Applications,* vol. 541, no. 122272, pp. 1–16, 2020. https://doi.org/10.1016/j.physa.2019.122272

[10] H. Gunduz, Y. Yaslan and Z. Cataltepe, "Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations," *Knowledge-BasedSystems,* vol. 137, pp. 138–148, 2017. https://doi.org/10.1016/j.knosys.2017.09.023

[11] E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," *Expert Systems With Applications,* vol. 129, pp. 273–285, 2019. https://doi.org/10.1016/j.eswa.2019.03.029

[12] S. Mishra, C. Bordin, K. Taharaguchi and I. Palu, "Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature," *Energy Reports,* vol. 6, suppl. 3, pp. 273–286, 2019. https://doi.org/10.1016/j.egyr.2019.11.009

[13] S. Totaro, A. Hussain and S. Scardapane, "A non-parametric softmax for improving neural attention in time-series forecasting," *Neurocomputing,* vol. 381, pp. 177–185, 2020. https://doi.org/10.1016/j.neucom.2019.10.084

[14] X. Li, W. Shang and S. Wang, "Text-based crude oil price forecasting: A deep learning approach," *International Journal of Forecasting,* vol. 34, no. 4, pp. 1548–1560, 2019. https://doi.org/10.1016/j.ijforecast.2018.07.006

[15] H. Abbasimehr, M. Shabani and M. Yousef, "An optimized model using LSTM network for demand forecasting," *Computers & Industrial Engineering,* vol. 143, no. 106435, pp. 1–13, 2020. https://doi.org/10.1016/j.cie.2020.106435

[16] C.-W. Tsai, C.-H. Hsia, S.-J. Yang, S.-J. Liu and Z.-Y. Fang, "Optimizing hyperparameters of deep learning in predicting bus passengers based on simulated annealing," *Applied Soft Computing Journal,* vol. 88, no. 106068, pp. 1–9, 2020. https://doi.org/10.1016/j.asoc.2020.106068

[17] H. Cui and J. Bai, "A new hyperparameters optimization method for convolutional neural networks," *Pattern Recognition Letters,* vol. 125, pp. 828–834, 2019. https://doi.org/10.1016/j.patrec.2019.02.009

[18] S. Cheng, F. Lu, P. Peng and S. Wu, "Multi-task and multi-view learning based on particle swarm optimization for short-term traffic forecasting," *Knowledge-Based Systems,* vol. 180, pp. 116–132, 2019. https://doi.org/10.1016/j.knosys.2019.05.023

[19] Y. E. Midilli and S. Elevli, "Optimization of Neural Networks with Response Surface Methodology: Prediction of Cigarette Pressure Drop," in *60th International Scientific Conference on Information Technology and Management Science of Riga Technical University*, Riga, 2019. https://doi.org/10.1109/ITMS47855.2019.8940643

[20] Suhartono, N. Suhermi and D. D. Prastyo, "Design of Experiment to Optimize the Architecture of Deep Learning for Nonlinear Time Series Forecasting," *Procedia Computer Science*, vol. 144, pp. 269–276, 2018. https://doi.org/10.1016/j.procs.2018.10.528

[21] Y. E. Midilli and S. Parshutin, "Review for Optimisation of Neural Networks with Genetic Algorithms and Design of Experiments in Stock Market Prediction," *Information Technology and Management Science,* vol. 22, pp. 15–21, 2019. https://doi.org/10.7250/itms-2019-0003

[22] L.-F. Hsieh, S.-C. Hsieh and P.-H. Tai, "Enhanced stock price variation prediction via DOE and BPNN-based optimization," *Expert Systems with Applications,* vol. 38, no. 11, pp. 14178–14184, 2011. https://doi.org/10.1016/j.eswa.2011.04.229

[23] S. Sakarya, M. Yavuz, A. D. Karaoglan and N. Ozdemir, "Stock Market Index Prediction with Neural Network During Financial Crises: A Review on BIST-100," *Financial Risk and Management Reviews,* vol. 1, no. 2, pp. 53–67, 2015. https://doi.org/10.18488/journal.89/2015.1.2/89.2.53.67

[24] A. Lasfer, H. El-Baz and I. Zualkernan, "Neural Network Design Parameters for Forecasting Financial Time Series," in *5th International Conference on Modeling, Simulation and Applied Optimization*, IEEE, 2013. https://doi.org/10.1109/ICMSAO.2013.6552553

[25] S. Alonso-Monsalve, A. L. Suárez-Cetrulo, A. Cervantes and D. Quintana, "Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators," *Expert Systems With Applications,* vol. 149, no. 113250, pp. 1–15, 2020. https://doi.org/10.1016/j.eswa.2020.113250

[26] R. Dash, S. Samal, R. Dash and R. Rautray, "An integrated TOPSIS crow search based classifier ensemble: In application to stock index price movement prediction," *Applied Soft Computing Journal,* vol. 85, no. 105784, pp. 1–14, 2019. https://doi.org/10.1016/j.asoc.2019.105784

[27] J. Cheng, L. Dong and M. Lapata, "Long Short-Term Memory-Networks for Machine Reading," *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, 2016. https://doi.org/10.18653/v1/D16-1053

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention Is All You Need," arXiv:1706.03762, 2017.

[29] C.-L. Hwang, S.-J. Chen and F. Hwang, *Fuzzy Multiple Attribute Decision Making: Methods.* Berlin: Springer, 1992.

[30] D. C. Montgomery, *Statistical quality control: A modern introduction*. New Jersey: John Wiley & Sons Singapore Pte. Ltd., 2013.

[31] G. Derringer and R. Suich, "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology,* vol. 12, no. 4, pp. 214–219, 1980. https://doi.org/10.1080/00224065.1980.11980968

[32] MetaQuotes, "Metatrader 5," MetaQuotes Software Corporation, 2000. [Online]. Available: https://www.metatrader5.com/en/trading-platform. [Accessed 8 5 2019].

Yunus E. Midilli graduated from the Industrial Engineering Department of Istanbul University in 2016. He continued his educational career in Ondokuz Mayis University as a Master degree student in intelligent systems engineering. Since 2018, he has been a PhD student at the Department of Information Technology of Riga Technical University. Yunus E. Midilli worked as a System Supervisor at the Quality Assurance Department of tobacco manufacturing company between 2016 and 2018. During this period, he elaborated his Master Thesis regarding the optimization of neural networks with a case study in the tobacco industry. After relocating to Riga for PhD purposes, he has been continuing his business career as an Automation Developer at Nordic bank. His PhD research deals with the optimization of neural networks in exchange rate prediction systems. His research interests are neural networks, data mining, optimization, quality improvement.
E-Mail: yunusemremidilli@gmail.com
ORCID iD: https://orcid.org/0000-0001-5373-5877

Sergei Parshutin, Dr. sc. ing., Assistant Professor of Riga Technical University, Faculty of Computer Science and Information Technology, Institute of Information Technology. In 2011, he successfully defended his Doctoral Thesis at Riga Technical University. His research interests include data mining, machine learning, decision support systems, fuzzy logic and fuzzy systems, neural networks, intelligent agent technologies and automation control. He is the author and co-author of 24 publications, including two monographs and teaching aids.
E-Mail: sergejs.parsutins@rtu.lv
ORCID iD: https://orcid.org/0000-0002-8689-3043