ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМАННЯ ПРИЙНЯТТЯ РІШЕНЬ

UDC 519.8
DOI: 10.20535/SRIT.2308-8893.2021.2.03

INVESTIGATION OF COMPUTATIONAL INTELLIGENCE METHODS IN FORECASTING PROBLEMS AT STOCK EXCHANGES

Yu. ZAYCHENKO, G. HAMIDOV, A. GASANOV

Abstract. In this paper, the forecasting problem of share prices at the New York Stock Exchange (NYSE) was considered and investigated. For its solution the alternative methods of computational intelligence were suggested and investigated: LSTM networks, GRU, simple recurrent neural networks (RNN) and Group Method of Data Handling (GMDH). The experimental investigations of intelligent methods for the problem of CISCO share prices were carried out and the efficiency of forecasting methods was estimated and compared. It was established that method GMDH had the best forecasting accuracy compared to other methods in the problem of share prices forecasting.

Keywords: share prices forecasting, LSTM, GRU, RNN, GMDH.

INTRODUCTION

The problem of share prices and market indicators forecasting attracts great attention from the specialists and financial managers. Traditionally for this problem statistical methods, ARMA, ARIMA, exponential smoothing method, Kalman filters and other methods were used.

But these methods have some drawbacks and based on assumptions which usually don't fulfill in practice: financial processes are non-stationary, and nonlinear by parameters, errors are correlated and may haven't zero mean and bounded variance.

Therefore last years for forecasting financial processes at stock exchanges intelligent methods are widely used. One class of such methods are recurrent neural networks (RNN) [1–7]. They enable to detect hidden dependences in data and perform long-term forecast of time series.

Now this class of RNN includes simple recurrent networks, LSTM and GRU [1-10]. As alternative intelligent method GMDH from the other side is also widely used for forecasting share prices at stock exchanges [11; 12] and other financial processes. GMDH has some advantages over other forecasting methods: 1) it enables to construct structure of forecasting model using experimental sample and find analytical models; 2) it may work with short samples.

It's interesting to compare these alternative methods at solution of practical forecasting problems. The goal of this paper is to investigate recurrent networks and method GMDH at forecasting of share prices, compare their efficiency and find the best method for this class of problem.

LSTM AND GRU MODELS DESCRIPTION

Networks of Long Short Term Memory (LSTM) were developed by “LSTM”, Hochreiter and Schmidhuber [1; 2] LSTM — is a special type of RNN, capable to train long-term dependencies. They work well for the most problems and are constructed so that to exclude problems which usually occur with deep learning networks. LSTM enable to prevent problem of decay or explosion of gradient when training using Back Propagation algorithm.

The architecture of LSTM is presented in the Fig. 1. It has chains type structure consisting of sequence of modules (blocks).

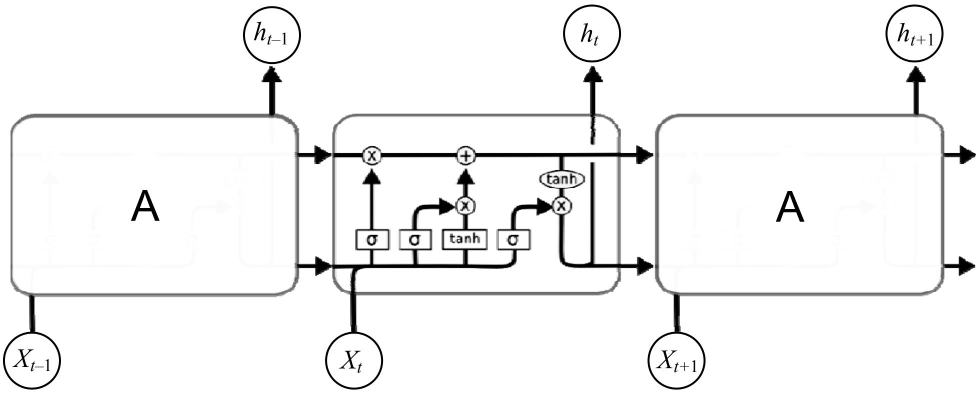


Fig. 1. The architecture of LSTM network

LSTM has capability to add or delete information which is regulated by special modules — gates (Fig 2). Gate consists of sigmoidal layer (σ) and operation of pointwise multiplication.

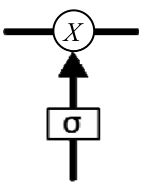
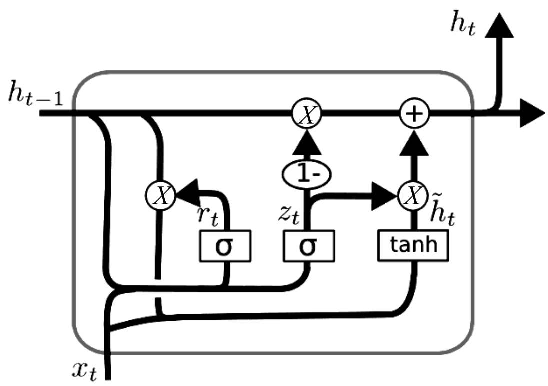


Fig. 2. Gate

Another RNN Gated Recurrent Unit (GRU) somewhat differs from LSTM. It integrates input and forgetting gates in one “update gate”.

A model GRU is therefore is more simple than conventional LSTM (see Fig. 3) and it has won popularity and wide applications owing to this property:



$$z_t = \sigma(W_z, [h_{t-1}, x_t]);$$

$$r_t = \sigma(W_r, [h_{t-1}, x_t]);$$

$$h_t = \tanh(W, [r_t h_{t-1}, x_t];$$

$$h_t = (1 - z_t) h_{t-1} + z_t \overline{h}_t] .$$

Fig. 3. Structure of GRU

Extended LSTM with forgetting gate. The extended LSTM is also two-layer recurrent network (Fig. 4.) Instead of hidden neuron a memory module is used which consists of one or more cells (Fig. 5). Forgetting gate is used to prevent uncontrolled increase of variable value in a memory cell.

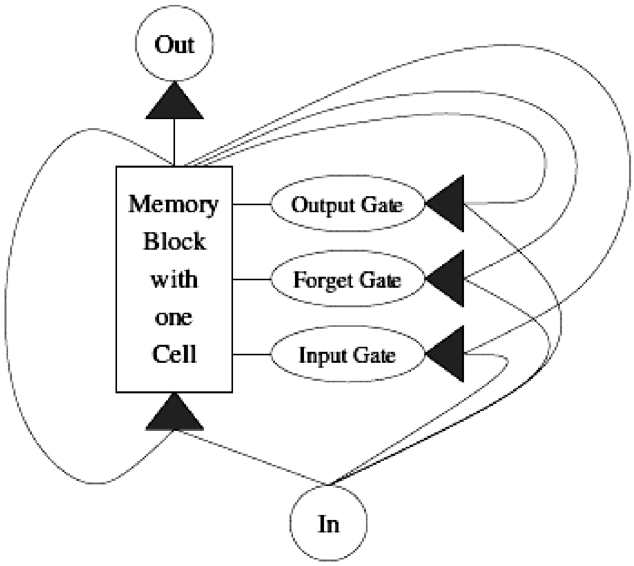


Fig. 4. LSTM with forgetting gate

Training of the extended LSTM with forgetting gate is performed by error correction method (supervised learning) in combination of Back Propagation algorithm (BPTT) and recurrent training in real time (RTRL).

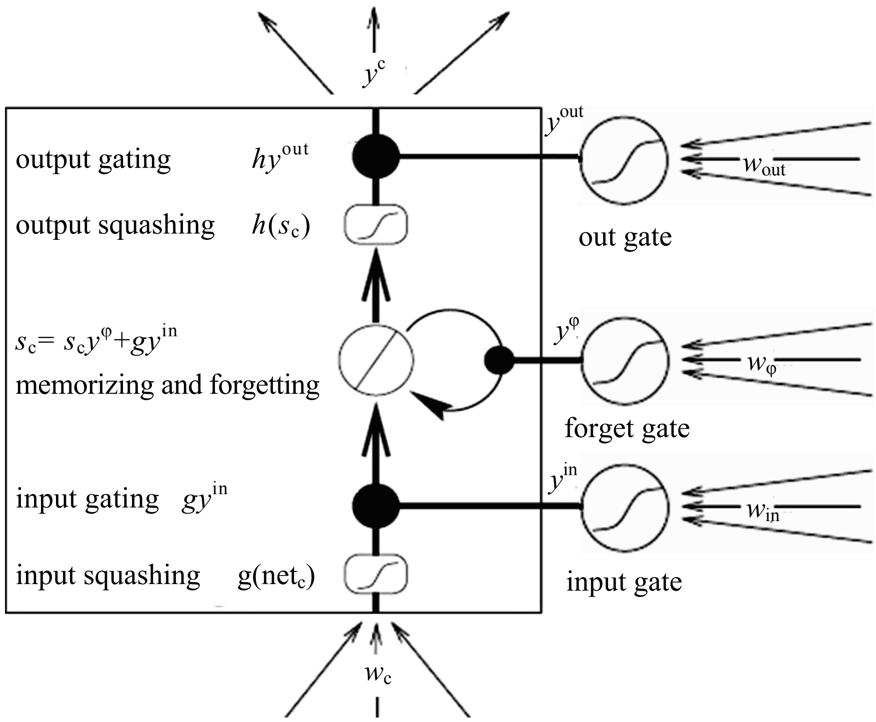


Fig. 5. Memory cell structure

Let's mark advantages of LSTM networks.

1. LSTM is universal approximator like BP networks. It may ensure global approximation of non-linear mapping of input signal into output.
2. It performs high quality generalization of input data.
3. Automatically is determined number of hidden layers (one).
4. Unlike static ANN, LSTM enables to perform adaptive filtration, forecasting, adaptive control, parametric models identification and classification of non-stationary signals.
5. Unlike simple RNN (ENN, JNN, NARNN i NARMANN) LSTM enable to work with long-term non-stationary sequences (time series).

But LSTM have also the following drawbacks.

1. Training process runs more slowly than in cases of MLP, RBFNN, PNN, Hamming RNN, Kohonen networks.
2. Automatic determination of number of neurons (memory blocks) in hidden layers and number of cells in each block is absent.
3. The model of training LSTM can't be transformed to the quadratic programming problem in convex region which has one optimal solution.

EXPERIMENTAL INVESTIGATIONS

The goal of investigations was to estimate accuracy of share prices forecasting by LSTM networks, find the best structure of recurrent networks LSTM, GRU and simple RNN and compare their efficiency with method of self-organization GMDH.

As input data share prices of CISCO at the stock exchange NYSE since 2006 till 2018 were taken. In the Table 1 daily data is presented including the fields: Open — value of open share price of current day; High — maximal daily price value; Low — minimal daily price; Close — close price value of current day; Volume — sell volume value. As forecasting data was taken the field "High".

Table 1. CISCO share prices dynamics (fragment of input sample)

Date	Open	High	Low	Close	Volume	Name
2006-01-03	17,21	17,49	17,18	17,45	55432166	CSCO
2006-01-04	17,48	17,93	17,85	17,85	80409776	CSCO
2006-01-05	17,94	18,48	17,93	18,35	118588943	CSCO
2006-01-06	18,51	18,88	18,47	18,77	122450979	CSCO
2006-01-09	18,97	19,11	18,92	19,06	78604868	CSCO

As a training sample was taken data since 2006 till the end of 2016 year and a test sample the data since 2017 till 2018 year was taken. Total size of the sample was 3019 values. The flow chart of data (share prices) is presented in the Fig. 6, where training sample is shown in blue color while test sample — in red color.

The next step of the program run is data normalization. After that the training of neural networks is performed.

At the end of software work the flow charts of real and forecasted stock prices, error value and accuracy of the model were determined which are presented in Fig. 7–19. For LSTM were constructed and investigated 5 models. The forecasting results and criteria values MSE, MAE and R2 score for LSTM 1-5 are shown in the Fig. 7–11 correspondingly.

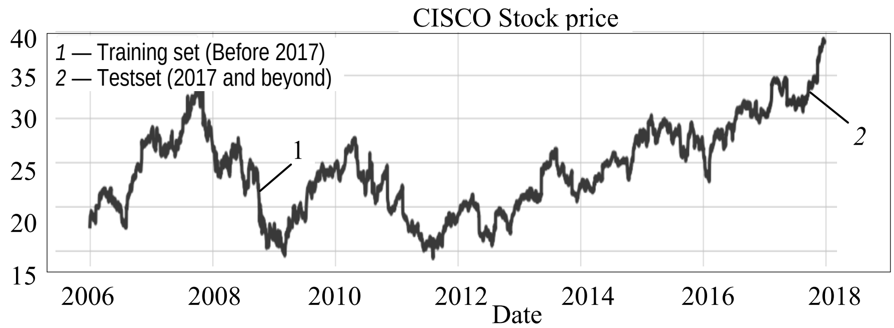


Fig. 6. Flow chart of share prices of the whole sample

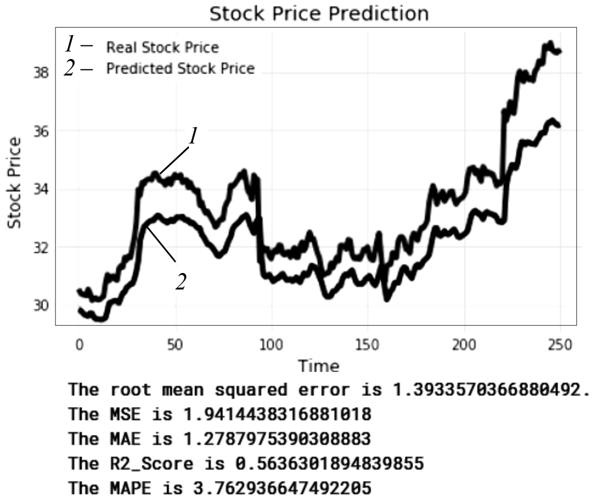


Fig. 7. LSTM-1

(LSTM-1 has 4 layers, each of which consists of 100 neurons, at each eve n layer Dropout — 0,2; uneven Dropout — 0,3)

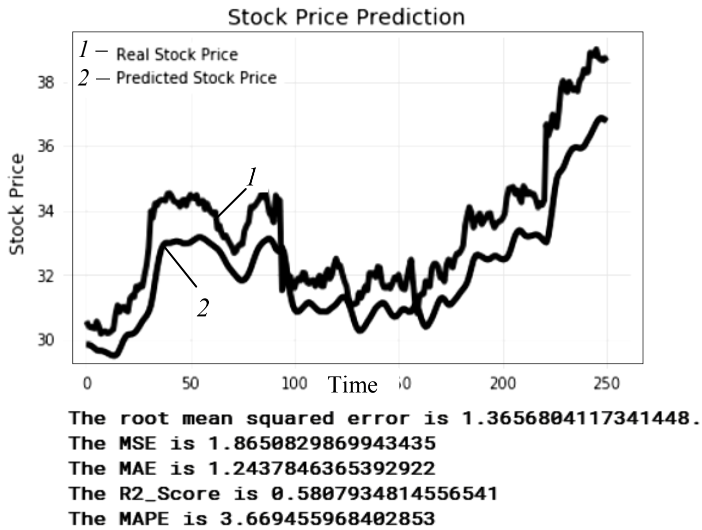
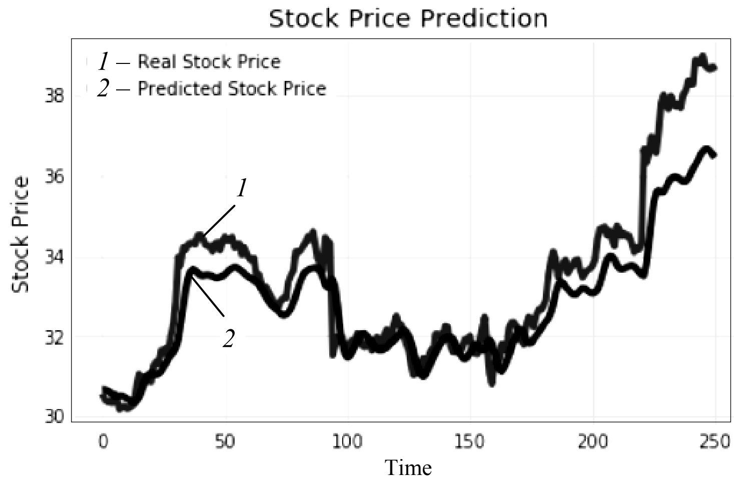


Fig. 8. LSTM-2

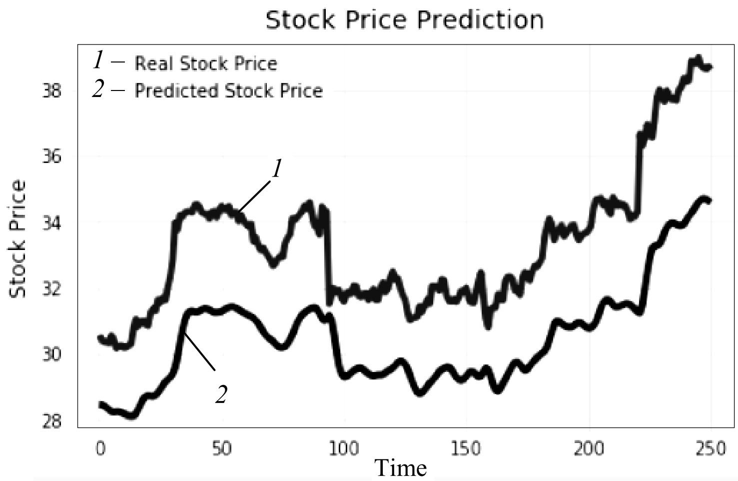
(LSTM-2 has 4 layers, each layer consists of 50 neurons and Dropout is 0,4 at all layers)




```
The root mean squared error is 0.9392288028778943.
The MSE is 0.8821507441554426
The MAE is 0.6915106432181433
The R2_Score is 0.8017228483303791
The MAPE is 1.9872419136689634
```

Fig. 9. LSTM-3

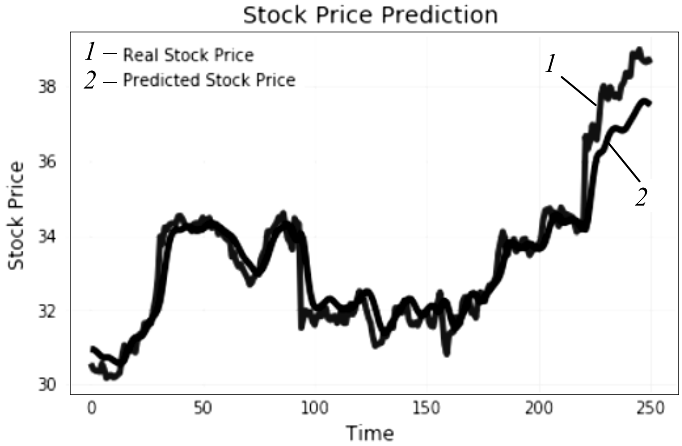
(LSTM-3 model has 4 layers, each of which has 30 neurons and Dropout — 0,1 at each layer)




```
The root mean squared error is 0.6851362567918017.
The MSE is 0.4694116903706816
The MAE is 0.5497780894853204
The R2_Score is 0.8944923942492073
The MAPE is 1.6779161468172838
```

Fig. 10. LSTM-4

(This model has 4 layers, each of which has 30 neurons and Dropout — 0,2 at each layer)



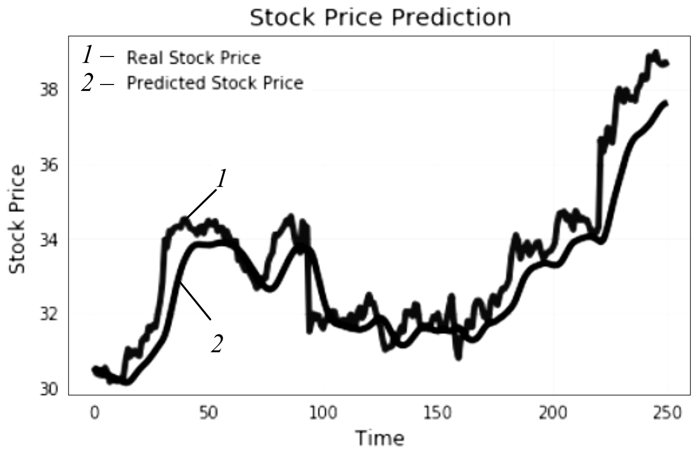

```
The root mean squared error is 0.6495040583547091.
The MSE is 0.42185552181923747
The MAE is 0.4552895151571448
The R2_Score is 0.905181385566364
The MAPE is 1.3343542934218122
```

Fig. 11. LSTM-5 (the best one)

(This model has 5 layers, each of which has 50 neurons and Dropout — 0,4 at each layer)

As it follows from presented results the best one is model 5 LSTM which consists of 5 LSTM layers, each layer has 50 neurons and each layer uses Dropout — 0,4) and one output layer. Training time takes about 11 minutes.

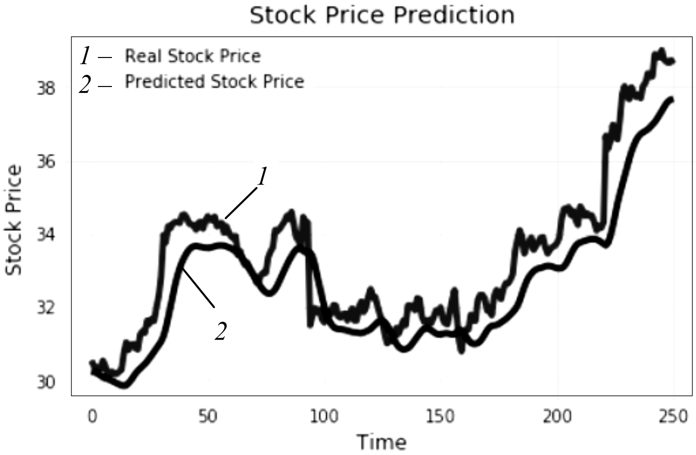
At the second stage of investigations 5 different GRU models were constructed and investigated. The forecasting results are presented in the Fig. 12–16. The criteria values — MSE, MAE, MAPE, R2 are also presented.




```
The root mean squared error is 0.9262376590860856.
The MSE is 0.8579162011092716
The MAE is 0.7061408899314852
The R2_Score is 0.807169940223738
The MAPE is 2.0611404261833535
```

Fig. 12. GRU-1

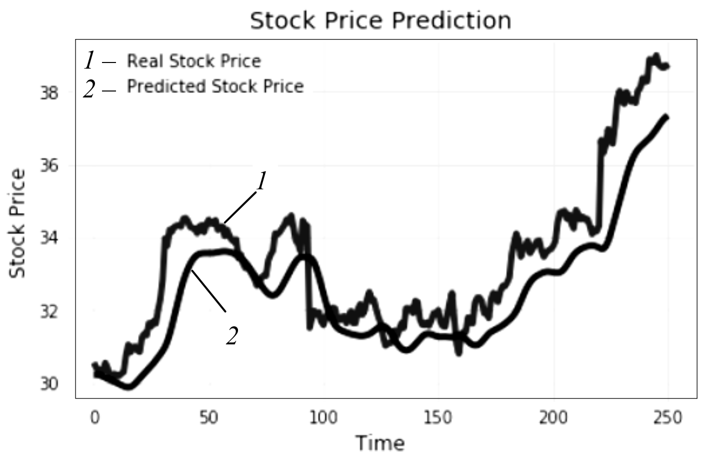
(This model has 5 layers, each of which has 50 neurons and Dropout — 0,4 at each layer)



The root mean squared error is 1.0282586133250005.
The MSE is 1.0573157758770528
The MAE is 0.8442865078382759
The R2_Score is 0.7623517728175077
The MAPE is 2.48163666434782

Fig. 13. GRU-2

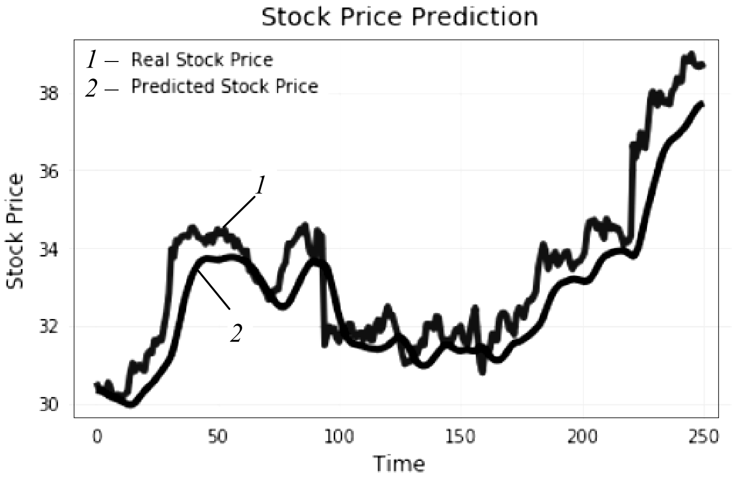
(This model has 4 layers, each hash of which has 100 neurons and Dropout — 0,4 at each layer)



The root mean squared error is 1.192612444843173.
The MSE is 1.4223244435948101
The MAE is 0.9703879158050417
The R2_Score is 0.6803103763222991
The MAPE is 2.841359595077174

Fig. 14. GRU-3

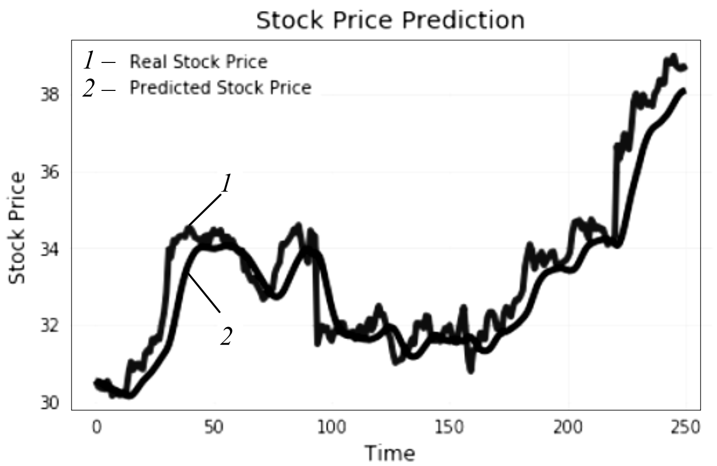
(This model has 5 layers, each of which has 60 neurons and Dropout — 0,2 at each layer)




```
The root mean squared error is 0.9958534566875114.
The MSE is 0.9917241071964651
The MAE is 0.7934728417263563
The R2_Score is 0.7770945243544862
The MAPE is 2.326249600133111
```

Fig. 15. GRU-4

(This model has 4 layers, each layer consists of 80 neurons and Dropout — 0,5 at each layer)

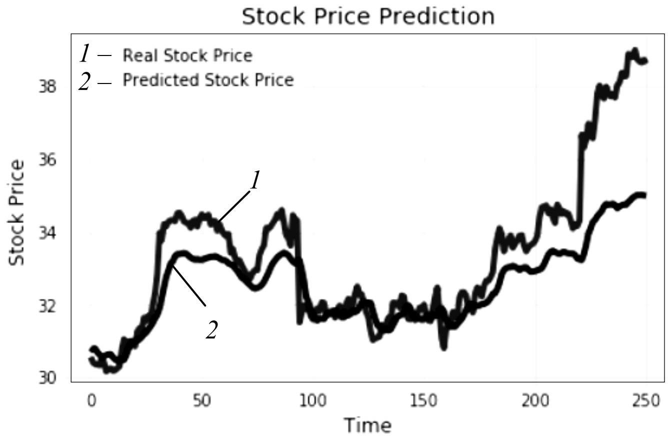



```
The root mean squared error is 0.8195130279208084.
The MSE is 0.6716016029319317
The MAE is 0.6098577865661378
The R2_Score is 0.8490470548618269
The MAPE is 1.7913809698036365
```

Fig. 16. GRU-5

(This model has 4 layers, each of which consists of 120 neurons with Drop-out — 0,2 at each layer)

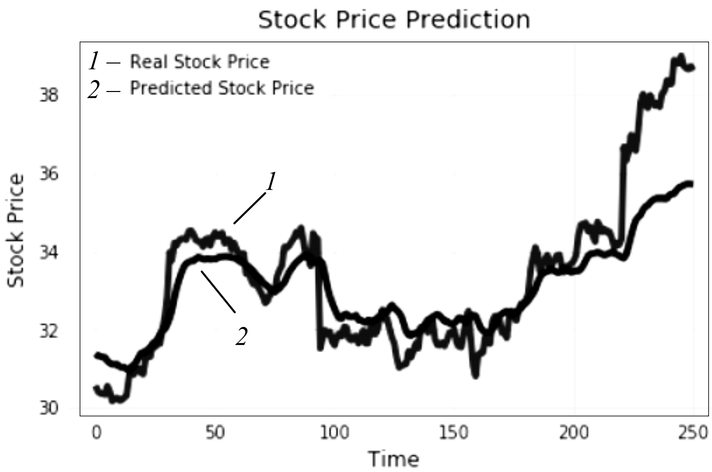
As it follows the best one is the 5-th network which consists of 4 GRU layers (each layer has 120 neurons with Dropout 0,2 at each layer) and one output layer. As training algorithm was used Stochastic Gradient Descent (SGD). Training time was approximately 9 minutes. For conventional recurrent neural networks (RNN) three models were constructed, forecasting results and criteria values are presented below in the Fig. 17–19.




```
The root mean squared error is 1.3598444216911991.
The MSE is 1.8491768512046718
The MAE is 0.9394346959087478
The R2_Score is 0.5843686337970668
The MAPE is 2.670376617454158
```

Fig. 17. Simple RNN-1

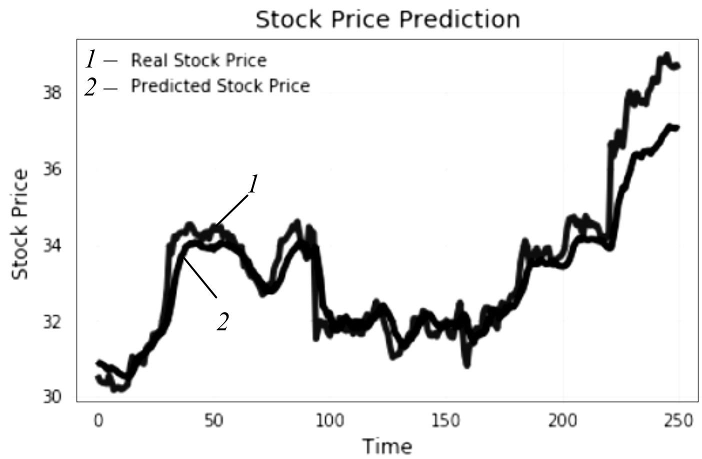
(This model has 4 layers, each layer has 50 neurons with Dropout — 0,15 at each layer)




```
The root mean squared error is 1.120834789285415.
The MSE is 1.2562706248724804
The MAE is 0.7718656559674388
The R2_Score is 0.7176335644715632
The MAPE is 2.227058208726607
```

Fig. 18. Simple RNN-2

(This model has 4 layers, each layer has 100 neurons with Dropout — 0,3 at each layer)



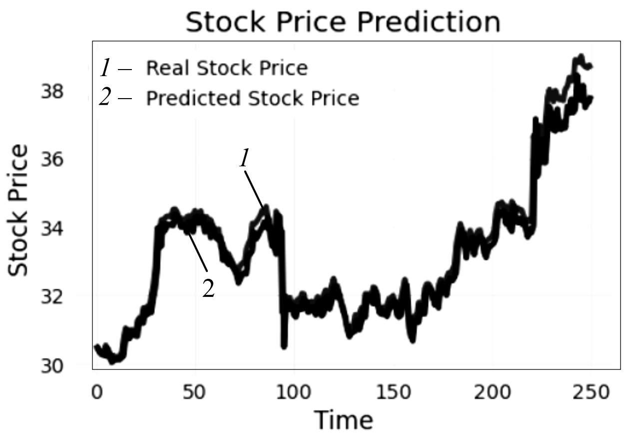

```
The root mean squared error is 0.7670696860625308.
The MSE is 0.5883959032760695
The MAE is 0.5369296562529181
The R2_Score is 0.8677488348464224
The MAPE is 1.550921348443134
```

Fig. 19. Simple RNN-3

(This model has 4 layers, each layer has 60 neurons with Dropout — 0,2 at each layer)

The best RNN model appeared to be the last one which consists of 4 layers with 60 neurons and one output layer. Each layer uses Dropout — 0,2. As a training algorithm was used Adam. Training time is about 8 minutes.

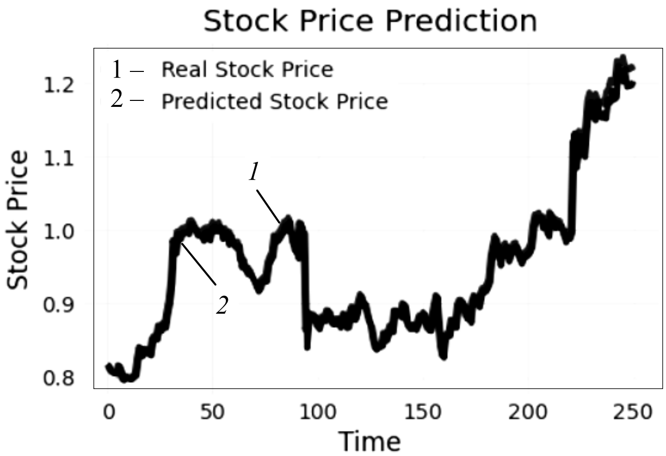
At the next experiments algorithm GMDH was used. Two models were constructed, trained and investigated. The forecasting results and criteria values for models are presented in the Fig. 20, 21.




```
The root mean squared error is 0.5053921554019186.
The MSE is 0.25542123074179707
The MAE is 0.3613253517353678
The R2_Score is 0.9425900908172803
The MAPE is 1.05461391041376
```

Fig. 20. GMDH-1

(This model differs from the other by the window size was of 60 points).




```
The root mean squared error is 0.01913318487947485.
The MSE is 0.00036607876363216503
The MAE is 0.0124975643606496
The R2_Score is 0.9666910950343344
The MAPE is 1.2815932223919728
```

Fig. 21. GMDH-2

The second model GMDH turned out to be better than the first one, it has the higher accuracy and less error. The window size for this model was 30 points, freedom choice value is 7, regularization parameter L2 — 0,5. Training time was about 6 minutes.

In the next experiment the comparison of the best models of different classes was performed. Firstly the error change versus number of epochs was compared for different models. The results are presented in the Fig. 22, 23.

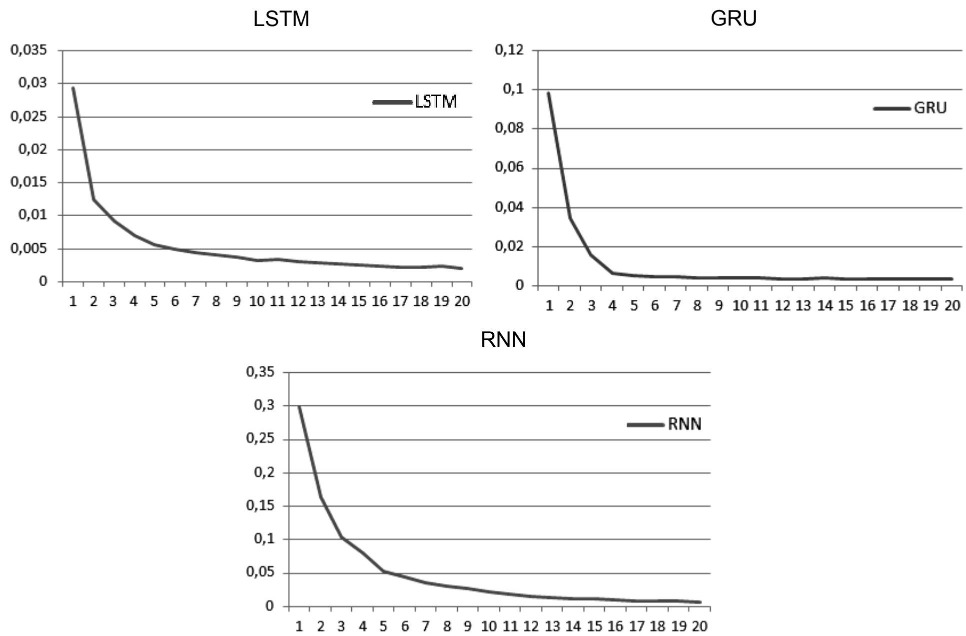


Fig. 22. MAPE value versus number of epochs for all models

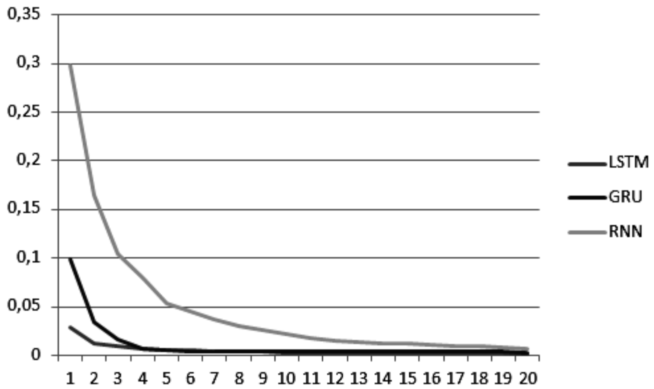


Fig. 23. Error values (MSE) dependence on epochs number for the best models

In the table 2 the results of training the best models are presented.

Table 2. Forecasting results for the best models at the test sample

Models	**Methods**					
	RMSE	**MSE**	**MAE**	**R2_Score**	**MAPE**	**Training _Time**
LSTM	0,649	0,421	0,455	90,5%	1,33	11 min
GRU	0,819	0,671	0,609	84,9%	1,79	9 min
Simple RNN	0,767	0,588	0,536	86,7%	1,55	8 min
GMDH	0,0191	0,0003	0,012	96,6%	1,28	6 min

As it follows from the presented results the best one is method GMDH by all criteria. Besides it takes the least time for training. The second one is LSTM network and the worst forecasting results were shown by GRU and simple RNN.

CONCLUSIONS

In this paper investigations of different types of recurrent networks LSTM, GRU, simple RNN and GMDH in the forecasting problem at stock exchange NYSE were carried out.

For each class of RNN several structures were investigated and the best structure was selected.

The forecasting efficiency and training time of different recurrent networks and GMDH were estimated and compared.

After experimental investigations it was determined that the best forecasting accuracy by different criteria has method GMDH, besides it took the least training time.

REFERENCE

1. S. Hochreiter and J. Schmidhuber, “LONG SHORT-TERM MEMORY”, *Neural Computation*, no. 9, pp. 1735–1780, 1997.
2. Josef Hochreiter, *DIPLOMARBEIT IM FACH INFORMATIK. Untersuchungen zu dynamischen neuronalen Netzen*. Munchen, Germany: Technische Universi at Munchen, 1991, 74 p.
3. J.J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities”, *Proc. of the National Academy of Sciences USA*, vol. 79, pp. 2554–2558, 1982.

4. Y. Cheung, "A new recurrent radial basis function network", *Neural Information Proceeding, ICONIP'02*, vol. 2, pp. 1032–1036, 2002.
5. V. Baier, "Motion Perception with Recurrent Self-Organizing Maps Based Models", *Proc. of IJCNN'05, Monreal, Canada, 2005, July 31–Aug. 4,* pp. 1182–1186.
6. Y.-P. Chen and J.-S.Wang, "A Novel Neural Network with Minimal Representation for Dynamic System Identification", *Proceeding of IJCNN'04, Budapest, 2004*, pp. 849–854.
7. Y. Bengio, P. Siamard, and P. Frasconi, *Learning Long-Term Dependencies With Gradient Descent is Difficult*, 1994.
8. S. Preeti, R. Bala, and R. Singh, "Financial and Non-Stationary Time Series Forecasting using LSTM Recurrent Neural Network for Short and Long Horizon", *10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*.
9. T. Fischer and C. Krauss, "Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions", *European Journal of Operational Research*, no. 270, p. 654–669, 2018.
10. W. Wei and P. Li, "Multi-Channel LSTM with Different Time Scales for Foreign Exchange Rate Prediction", *Proceedings of the international conference on Advanced Information Science and System, 2019.*
11. Yu. P. Zaychenko, *Fundamentals of Intelligent systems design*, (in Ukrainian). Kiev: Publ. house "Slovo", 2004, 352 p.
12. M. Zgurovsky and Yu. Zaychenko, *Fundamentals of computational intelligence-System approach*. Springer, 2016, 275 p.
13. S. Nikolenko, A. Kadurin, and E. Arhangelskaya, *Deep learning. Involving into the world of neural networks*, (in Russian). Saint- Petersburg: "Peter", 2018, 480 p.
14. S. Heyken, *Neural networks, Full course*, (in Russian). Moscow: Publ. House "Williams", 2006, 1104 p.
15. S. Osovsky, *Neural networks for information processing*, (in Russian). Moscow: Finance and Statistics, 2002, 344 p.
16. *Recurrent neural networks*. [Online]. Available: https://neerc.ifmo.ru/wiki/index.php?title=Рекуррентные_нейронные_сети
17. *Understanding LSTM Networks*. [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/
18. O.M. Riznyk, "Dynamic recurrent neural networks", (in Ukrainian), *Mathematical machines and systems*, no. 3, pp. 3–26, 2009.
19. F.M. Gafarov and A.F. Galimyanov, *Artificial neural netwoorks and their applications*, (in Russian). Kazan: printed in Kazan University, 2018.
20. F. Gers and J. Schmidhuber, *Recurrent nets that time and count*, 2000.
21. K. Yao et al., *Depth-Gated Recurrent Neural Networks*, 2015, 5 p.
22. J.Koutn´ık, K. Greff, F. Gomez, and F. Schmidhuber, *A Clockwork RNN*. Switzerland, 2014, 9 p.
23. R. Jozefowicz, W. Zaremba, and I. Sutskever, *An Empirical Exploration of Recurrent Network Architectures*, 2015, 9 p.

Received 21.04.2021

INFORMATION ON THE ARTICLE

Yuriy P. Zaychenko, ORCID: 0000-0001-9662-3269, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: zaychenkoyuri@ukr.net

Galib Hamidov, “Azershig”, Azerbaijan, e-mail: galib.hamidov@gmail.com

Aydin Gasanov, ORCID: 0000-0002-5821-0751, Dragomanov National Pedagogical University, Ukraine, e-mail: 0677937631@ukr.net

ДОСЛІДЖЕННЯ МЕТОДІВ ОБЧИСЛЮВАЛЬНОГО ІНТЕЛЕКТУ У ПРОБЛЕМІ ПРОГНОЗУВАННЯ НА РИНКАХ ЦІННИХ ПАПЕРІВ / Ю.П. Зайченко, Г. Гамідов, А. Гасанов.

Анотація. Розглянуто проблему прогнозування курсів акцій на ринку цінних паперів NYSE. Для її вирішення запропоновано та досліджено альтернативні методи обчислювального інтелекту: мережі LSTM, графічні рекурентні модулі (GRU), прості рекурентні мережі і метод групового урахування аргументів (МГУА). Проведено експериментальні дослідження інтелектуальних методів в проблемі прогнозування цін акцій і порівняльну оцінку ефективності альтернативних методів прогнозування. З’ясовано, що метод МГУА забезпечує найвищу точність в розглянутій проблемі прогнозування цін акцій.

Ключові слова: прогнозування цін акцій, рекурентні мережі LSTM, GRU, RNN, МГУА.

ИССЛЕДОВАНИЕ МЕТОДОВ ВЫЧИСЛИТЕЛЬНОГО ИНТЕЛЛЕКТА В ПРОБЛЕМЕ ПРОГНОЗИРОВАНИЯ НА РЫНКАХ ЦЕННЫХ БУМАГ / Ю.П. Зайченко, Г. Гамидов, А. Гасанов.

Аннотация. Рассмотрена проблема прогнозирования курсов акций на рынке ценных бумаг NYSE. Для ее решения предложены и исследованы альтернативные методы вычислительного интеллекта: сети LSTM, графические рекуррентные модули (GRU), простые рекуррентные сети и метод группового учета аргументов (МГУА). Проведены экспериментальные исследования интеллектуальных методов в проблеме прогнозирования цен акций и сравнительная оценка эффективности альтернативныхметодов прогнозирования. Установлено, что метод МГУА обеспечивает наиболее высокую точность в рассмотренной проблеме прогнозирования цен акций.

Ключевые слова: прогнозирование цен акций, рекуррентные сети LSTM, GRU, RNN, МГУА.