ISSN 2278-3091
Volume 9, No.3, May - June 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse99932020.pdf
https://doi.org/10.30534/ijatcse/2020/99932020

Predicting the Cost of Housing using Neural Networks

Boyko A.N.[1], Gladyshev A.G.[2,*], Kadyrova G.M.[3] Barmenkova N.A.[4], Zybenko A.V.[5]

[1] Financial University under the Government of the Russian Federation, 125993, Russia
boykoan@mail.ru

[2] Financial University under the Government of the Russian Federation, 125993, Russia
gladyshevag@mail.ru

[3] Financial University under the Government of the Russian Federation, 125993, Russia
kadyrova@mail.ru

[4] Financial University under the Government of the Russian Federation, 125993, Russia
barmenkova@mail.ru

[5] Financial University under the Government of the Russian Federation, 125993, Russia
zybenko@mail.ru

ABSTRACT

The transition to market relations in the economy and scientific and technological progress have greatly accelerated the pace of introduction of the latest scientific developments in the field of information technologies into all spheres of social and economic life of society. Achieving high results in the economy and gaining a place as a full partner in the world economic system largely depends on the extent to which modern information technologies will be used in all aspects of human activity, as well as on the role these technologies will play in improving the efficiency of economic relations.

With the development of theoretical approaches to create adequate models of real estate market behavior in Western countries and the United States, new intelligent computer technologies were actively introduced into the practice of making financial and investment decisions. First in the form of expert systems and knowledge bases, and then in the late 80's-neural network technologies that are an adequate tool for solving forecasting problems.

Key words: Neural network, technologies, knowledge, development.

1. INTRODUCTION

Obviously, the price of an apartment depends on many factors, such as the total and residential area, number of rooms, floor, territorial location of the house, its number of floors, condition, availability of communications, etc. Experienced realtors cope with the assessment task without effort, applying their knowledge and intuition, relying on known analogues and using associative thinking. All this knowledge and skills are among the poorly formalized, partially unconscious, so the development of an unambiguous algorithm for determining the price based on the value of the influencing factors is an extremely difficult and almost impossible task.

To model the real estate market, a neural network is created [2,3], in which the number of input neurons corresponds to the number of input factors that affect the price. The source layer will have only one neuron corresponding to the source factor-price.

The task of forecasting financial time series was and remains relevant, since prediction is a necessary element of any investment activity, because the very idea of investing-investing money to generate income in the future-is based on the idea of forecasting the future.

Recently, when powerful collection and processing tools became available

for example, the task of forecasting financial time series has also become

one of the most popular tasks for practical application of various methods of data mining. The widespread use of Data Mining methods in this area is due to the presence of complex patterns in most time series that do not turn out to be linear methods.

To date, there are many models for predicting time series: regressive and autoregressive models, neural network models, exponential smoothing models, models based on Markov chains, classification models, etc. The most popular and widely used classes are autoregression and neural network models.

Classical artificial neural networks have repeatedly shown their suitability in the field of financial analysis due to their leap-rate nature, their ability to universal approximation.

Let's analyze the real estate market in Moscow using neural network modeling. The network outputs are prices for 2-room apartments in Moscow. The following statistics are submitted for input:

1) location (distance from the center);
2) floor;
3) type of housing;
4) area;
5) housing condition.

The distance from the center was estimated by the following indicators:

1 – center;
2 – closer to the center;
3 – blime to the edge;
4 – outskirts.

The type of housing was assessed by the following indicators:

1 – krushchevka brick (buildings from 1970 to 1985 years);
2 – panel;
3 – brick;
4 – brick new building.

The floor is evaluated according to two indicators:
1 – first and last floor;
2 – all the rest.

The condition of the apartment is assessed by three indicators:
1 – requires repair;
2 – average;
3 – good;
4 – excellent.

2. MATERIALS AND METHODS

A multi-layer perceptron is used as the base model. Implementing the net - input layer with 30 neurons (the window length), the first hidden layer with 64 neurons, after the normalization layer, then activation function, then the output layer of 1 neuron (in the case of a regression task), 2 neurons (in the case of classification tasks). For a regression problem, the activation parameter must be linear at the end. Then the error functions and optimization algorithm are defined.

The length of the gradient descent step is 0.001. to normalize the loss parameter in classification problems, you need to determine the cross-entropy, and for the regression problem - the average quadratic error.

The important point is training. Learning algorithms based on such data takes longer than 50-100 epochs. If you analyze the training data, you will see that 55% of the Windows were for one pattern (increase, for example), and the remaining 45% were for another (decrease). In our case, 53% of Windows are "down" and 47% are "up", and our goal is to get accuracy above 53%.

The results of solving the classification problem showed that the error and accuracy for the test sample remain at approximately the same level all the time. When training on the training field, the error falls, and the accuracy increases, which indicates retraining. A deeper model with two layers was used for comparison, but the results were almost identical.

If there is a re-training effect, you need to add regularization. During retraining, a model is built that "remembers" training data and does not allow you to generalize knowledge to new data. In the process of regularization, certain restrictions are imposed on the weights of the neural network, so that there is not a large spread in the values and, despite a large number of parameters (i.e., network weights), some of them turn to zero for simplification. In terms of the error function, such a neural network learns better, but accuracy still suffers [9].

It is worth adding even more regularization using the popular dropout technique in recent years (a method for solving the problem of retraining in neural networks) - this is the accidental "ignoring" of certain weights in the learning process to avoid co-adaptation of neurons (so that they do not learn the same characteristics). The results of using the Dropout technique showed that error and accuracy graphs already have better results.

If you stop training the network a little earlier, you can get 58% accuracy in predicting the price movement.

Another interesting and intuitive point of forecasting financial time series is that fluctuations in the next day are random, but when we look at the charts, we can still notice a trend for the next 5-10 days.

We need to check whether our neural network can handle this task. I predict the price movement in 5 days with the latest successful network architecture and will teach the model for more epochs. If you stop training early enough (overfitting occurs over time anyway), you can get 60% accuracy, which is a very good result.

For the regression problem, let's take our latest successful architecture for classification (it has already shown that it can study the necessary features), remove Dropout, and teach it in more iterations. Also, in this case, we can look not only at the error value, but also visually assess the quality of the forecast.

Modeling of the real estate market using neural networks was carried out in several stages [4]:

1. Forming a training sample. At this stage, the type of representation of historical and forecast data is determined, and a block of representative (training) samples is formed.
2. Training a neural network using a block of training samples formed at the first stage. The quality of training was characterized by a learning error, defined as the total square deviation of the values at the outputs of the neural network in the training sample from the actual values obtained at the outputs of the neural network.
3. The third stage is testing of neural network. The quality of forecasting is determined when 4.0-5.0 % of sets from the training sample are submitted for input. The experiment is successful if the relative confidence is at least 80.0 %.
4. at the fourth stage, a trial prediction is performed. At the input of the neural network-sets that were not included in the training sample, but whose result (forecast) is known.

During the formation of a representative sample, appropriate functional transformations were applied to some of the factors. Prices and floor space were prologarithmated [5]. After transformation re-sampling in Statistica NN was randomly divided into training (80%), verificated (10%) and test (10%).

After the study, the top 10 neural network structures were selected (table 1)

Table 1: Search results for the best neural network structures

	Type	Error	Inputs	Hidden	Performance
01	RBF	0.1247329	1	3	1.01068
02	RBF	0.1207868	3	3	1.021653
03	RBF	0.1197812	5	4	1.01826
04	RBF	0.1197633	5	3	0.9738516
05	RBF	0.1185169	2	3	1.012003
06	MLP	0.09789	1	1	0.863804
07	MLP	0.09641	1	3	0.8482065
08	MLP	0.09416	1	5	0.8370729
09	MLP	0.09187	3	3	0.8168508
10 *	MLP	0.07401	3	6	0.6126456

Figure 1 shows the best network of the multi-layer perceptron (MP) type, which contains 3 inputs, 6 hidden layers, the modeling error of which is 0.07 with a performance of 0.613, the correlation coefficient is 0.89564. The inputs to this network do not include indicators of the condition of the apartment and the type of house.

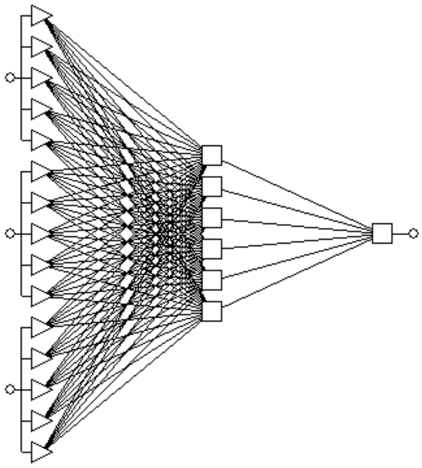

Figure 1: Neural network for predicting prices in the real estate market.

In order to clearly demonstrate the results and the average modeling error, figure 2 shows the ratio between the verified and modeled sample.

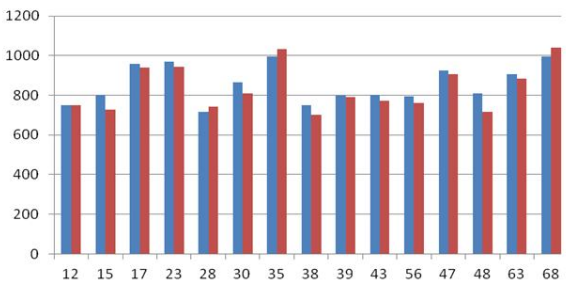

Figure 2: Results of neural network modeling of apartment prices

Despite the fact that the modeled values do not completely coincide with the practical ones, we can say that the trend of changes in the desired values is accurately reflected. For final confirmation of the models ' performance, testing data was submitted to the network input (figure 3).

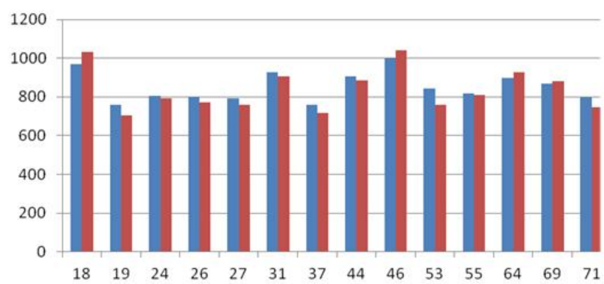

Figure 3 : Results of neural network modeling of a testing sample of apartment prices.

The predicted values of prices for apartments in the neural network differ from the real ones by 5%. This error is less than the network training error, which indicates a high accuracy of modeling the process of forming real estate prices in Moscow. The built model makes it possible to increase the efficiency of managing real estate complexes on the scale of a city or a large Corporation and make this mechanism more transparent [6-8].

So neural network technologies, in contrast to expert systems, are designed to solve poorly formalized problems. This type of technology is used to recognize such events or items. You can use them to recreate multiple relationships between multiple objects. The fundamental difference between artificial neural networks and conventional software systems, such as expert systems, is that they do not require programming. They configure themselves, i.e. they learn, and this is exactly what the user needs. [5]

3. CONCLUSION

This article shows the possibility of using Kohonen neural networks in the process of predicting prices in the real estate market. For this purpose, a neural network binarization method was proposed, combining the Kohonen network with a local binarization method (Singh's method). Using the network, the most representative shades of the prices are highlighted. This allows you to get rid of non-essential information on the image, as well as reduce the amount of pixel information for further processing.

Regarding the practical advantages of the proposed method, its main advantage is the acceleration of binarization of the same type of prices. In this case, the network should only be trained on one price to binarize the others. Thus, this method allows you to get rid of the main drawback of local binarization methods, while maintaining an acceptable quality of the final result. Despite this, a significant amount of time is spent on training the Kohonen network and calculating the winning neuron for each price per object, which is the basis for further improvement of the proposed method.

REFERENCES

1. S. Dutta and S. Shekhar, **Bond-rating: A non-conservative application of neural networks**, in Proc. IEEE Int. Conf. Neural Networks, IEEE Computer Press: Los Alamitos, CA, 2016, pp. 142–150.
2. G. Jang, F. Lai, B. Jiang, and L. Chien, **An intelligent trend prediction and reversal recognition system using dual-module neural networks**, in Proc. First Int. Conf. Artif. Intell. Appl. on Wall Street, IEEE Computer Press: New York, 2011, pp. 42–51.
3. L. Salchenberger, E. M. Cinar, and N. Lash, **Neural networks: A new tool for predicting thrift failures**, Decision Sci., vol. 23, no. 4, pp. 899–916, 2016. https://doi.org/10.1111/j.1540-5915.1992.tb00425.x
4. D.A. Bens, V. Nagar and M.H. Franco Wong. 2002. **Real investment implications of employee stock option exercises**. Journal of Accounting Research, 40(2): 359–393. https://doi.org/10.1111/1475-679X.00052
5. F. AitSahlia, M. Goswami and S. Guha. **American option pricing under stochastic volatility: An empirical evaluation**. Comput. Manag. Sci., 2010, 7(2), 189–206. doi: 10.1007/s10287-008-0083-2
6. Chicago Board Options Exchange (CBOE) Holdings, Inc. CBOE Market Statistics 2019. Chicago Board Options Exchange (CBOE), pp. 62–72, 2019

7. J.M. Binner, R.K. Bissoondeeal, T. Elger, A.M. Gazely and A.W. Mullineux. (2005). **A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia**. Appl. Econ. 37, 665–680. doi: 10.1080/0003684052000343679Azman, N., Ghani, M. K. A., Wicaksono, S. R., & Salahuddin, L. (2019). **The development of IoT Tele-insomnia framework to monitor sleep disorder**. International Journal of Advanced Trends in Computer Science and Engineering, 8(6), 2831–2839. https://doi.org/10.30534/ijatcse/2019/25862019
8. Babker, A. M., Altoum, A. E. A., Tvoroshenko, I., & Lyashenko, V. (2019). **Information technologies of the processing of the spaces of the states of a complex biophysical object in the intellectual medical system health**. International Journal of Advanced Trends in Computer Science and Engineering, 8(6), 3221–3227. https://doi.org/10.30534/ijatcse/2019/89862019
9. Badawi, B., Aris, T. N. M., Mustapha, N., & Manshor, N. (2019). **A fuzzy multi-layer color qr code decoder algorithm.** International Journal of Advanced Trends in Computer Science and Engineering, 8(1.4 S1), 131–137. https://doi.org/10.30534/ijatcse/2019/2081.42019