

RESEARCH PAPER

Real Estate Evaluation Model Based on Genetic Algorithm Optimized Neural Network

Yan Sun

Department of Economic Management, Sichuan College of Architectural Technology, CN
yanssuny@126.com

With the rapid development of society, the real estate economy, as an important part of Chinese economy, is showing a growing trend. But it is also the most likely to generate bubble economy, causing financial risks; it will trigger a series of social contradictions and cause social unrest in severe cases. Therefore, it is urgent to improve and optimize the real estate evaluation model. In this study, the real estate was evaluated based on the neural network model optimized by genetic algorithm. Through sorting out and summarizing the real estate data in a period of time, the corresponding model was established and the test data were obtained. The average relative error value of the genetic algorithm optimized neural network model was 3.552, which was smaller than that of the Back-Propagation (BP) neural network prediction model. The experimental conclusion that the new network model was better than the traditional model was obtained. This work opens up a new route of real estate evaluation.

Keywords: genetic algorithm; neural network; real estate; evaluation model

1. Introduction

With the acceleration of urbanization, the real estate industry is developing vigorously and irresistibly. People generally yearn for real estate consumption activities, hoping to invest more in the real estate industry and obtain their own interests. However, with the emergence and growth of these transactions, a series of social contradictions have also been aroused. The situation of the real estate industry is worrying. In order to improve that unfavorable situation and the trend of the real estate industry, it is urgent to effectively and accurately evaluate real estate price (Domian et al. 2015). In response to that problem, many experts and scholars have put forward their own views. Poursaeed et al. (2018) believed that the estimation of housing market value is very inaccurate and differ greatly from the actual value in some cases although some real estate websites use special formulas Svetlana (2017). Therefore, a new framework of automatic value assessment was proposed and applied to the new data set of real estate photos and metadata. Albu (2017) put forward that real estate appraisal activities changed with time and the physical or legal action was reasonable only when the future value was unaffected. The theory and practice of appraisal activities quantified three kinds of value methods, including sales comparison method, income method and cost method. Li et al. (2015) believed that unlike the real estate market in developed countries, the traditional mass assessment technology has been difficult to support and implement the real estate market in China. Therefore, an innovative and unified evaluation method based on the traditional mass assessment technology was proposed to deeply study the characteristics of the real estate market in China. Zhang and Ji (2010) put forward a method based on Back-Propagation (BP) neural network theory to analyze the real estate investment risk and its changing law, which improved the accuracy of evaluation and was conducive to risk avoidance. Manganelli et al. (2015) explained the relationship between real estate price and the geographical location of real estate with genetic algorithm and found the potential and limitations of genetic algorithm in the field. In this study, the simulation experiment of the real estate evaluation model based on genetic algorithm optimized neural network was carried out to effectively improve and optimize the real estate evaluation model. The accuracy and practicability of the method was tested, which is conducive to promoting the healthy development of the real estate industry.

2. Real estate assessment

Real estate assessment, also known as real estate appraisal, refers to forming application mechanism by simulating real estate market price and simulating the formation process of real estate market price using different effective evaluation methods. Real estate appraisal involves a complex economic category, including the appraisal of buildings and land prices. It is an effective means to allocate resources in the real estate market (Daniel et al. 2015). The widely used methods of real estate evaluation include market method, cost method, income method, etc. In the real estate appraisal work, it is deeply influenced by various factors, including the level of economic development, urbanization degree, social security situation, government policy, assessor's professional accomplishment, architectural style characteristics, market personnel conditions, geographical environment, economic environment, ecological environment and so on. These factors and conditions will affect the overall quality of the appraisal work (Fife 2017).

Different real estate evaluation methods have some differences in the simulation of formation of real estate market price, and the final value will be different. In practice, the evaluators usually need to adopt at least two evaluation methods for evaluation and compare and analyze the obtained evaluation data in a unified way, so as to effectively improve the accuracy of data evaluation. Moreover, in the real estate evaluation work, we should take fairness, objectivity, rigorousness and legality as the starting point and working purpose to improve the quality of the evaluation work. In the real estate appraisal work, the difference between the actual price of the real estate market and the predicted price is the risk of the real estate appraisal work. Therefore, we should strengthen the risk awareness of the appraisal staff and improve their personal professional and moral qualities. In general, we should establish and improve the real estate appraisal mechanism and try our best to minimize the risk of appraisal, improve the efficiency, and ensure the healthy development of the real estate appraisal industry.

3. Genetic algorithm optimized neural networks

3.1. Genetic algorithm

Genetic algorithm achieves the goal of individual optimization in a population by following the biological evolution process of Darwin's biological evolution theory and genetic mechanism, abiding by natural evolution process and the principle of survival of the fittest, and retaining individuals with high adaptability for genetic manipulation (Metawa et al. 2017). It is an intelligent evolutionary algorithm, which is essentially a natural random search. It is suitable for solving a series of problems such as non-linearization, and can also improve the efficiency and feasibility of the algorithm (Sevinç and Coar, 2018). The proposed method has attracted the attention of scholars and experts. It is widely used because of its good macro-exploration ability and global optimization ability (Ender and Ahmet 2018). The execution of the genetic algorithm is shown in **Figure 1**.

3.2. BP neural network model

The multi-layer neural network model based on BP algorithm is generally called BP neural network. BP algorithm is the most widely used learning method at present (Liu & Shi 2018). One part of BP algorithm is the forward propagation of signals, and the other part is the reverse propagation of errors (Tang et al. 2018). Neuron is the most basic component and the basic information unit of neural network operation. In addition to neurons, it is also composed of different parts which play their respective roles. It has a strong perception function and can achieve multi-level network learning to bring readers a unique experience. However, the model still has some limitations, such as low convergence efficiency and easy to fall into local extremum, which can affect the final results of network training and the effect of real estate evaluation.

Figure 2 is the flow chart of BP neural network algorithm.

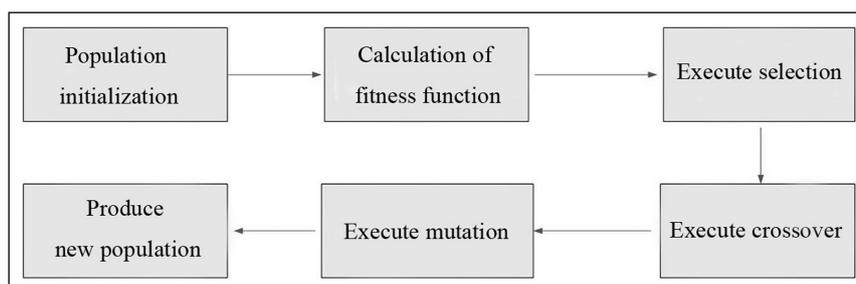


Figure 1: Execution process of genetic algorithms.

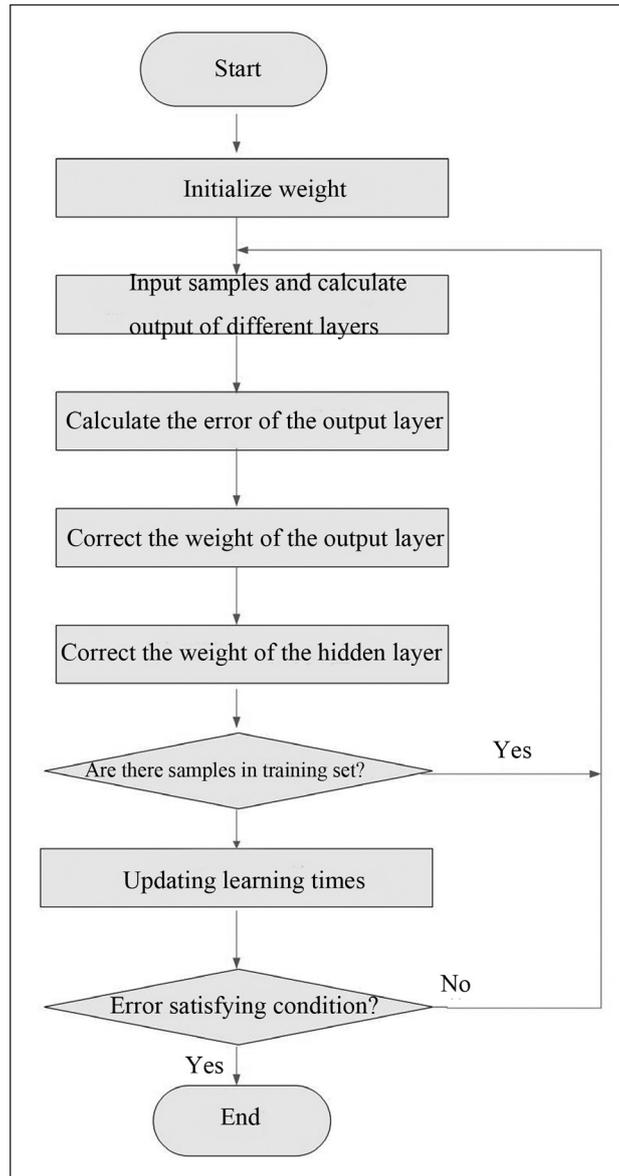


Figure 2: The flow chart of BP neural network algorithm.

3.2.1. Forward propagation of signals

$y_k = f_2(v_k)$ represents the output result of the k -th output layer node, x_l ($l = 1, 2, \dots, m$) represents the l -th input variable, ω_{lq} ($q = 1, 2, \dots, q$) represents the weight of the l -th input variable connecting to the q -th hidden layer node, b_q represents the threshold of the q -th hidden layer node, $f_1(v_q)$ represents the activation function of the hidden layer node, $f_2(v_k)$ represents the activation function of the output layer node, and φ_{qk} ($k = 1, 2, \dots, s$) represents the weight of the output variable of the q -th hidden layer connecting to the k -th output layer node.

The output of the l -th node (the hidden layer) is:

$$y_l = f_1\left(\sum_{l=1}^p \omega_{lq} x_l + b_q\right), (q = 1, 2, \dots, p). \quad (1)$$

The output of the k -th node (the output layer) is:

$$y_k = f_2\left(\sum_{q=1}^s \varphi_{qk} y_q + b_k\right), (k = 1, 2, \dots, s). \quad (2)$$

3.2.2. Back propagation of errors

η stands for learning rate, n stands for number of iterations, E stands for the network error value of the actual output, e_k stands for the error of the k -th output layer node, v_q stands for the result of weighted sum

of input variables by the q -th hidden layer node, φ_{qk} ($k = 1, 2, \dots, s$) stands for the weight of the output variable of the q -th hidden layer connecting to the k -th output layer node, and $b_k(n)$ stands for threshold value.

The total network error is:

$$E(w, b) = \frac{1}{2} \sum_{k=1}^s e_k^2. \quad (3)$$

The weight and threshold of the output layer to the hidden layer are set: the connection parameters in the system can be corrected based on gradient descent method to obtain the connection parameters of the output layer of the system. Moreover variation, $\Delta\varphi_{qk}(n)$ new weight $\varphi_{qk}(n+1)$, variation $\Delta b_k(n)$ and new threshold value $b_k(n+1)$ are obtained:

$$\Delta\varphi_{qk}(n) = \eta e_k(n) f_2'(w_k(n)) y_q(n), \quad (4)$$

$$\varphi_{qk}(n+1) = \Delta\varphi_{qk}(n) + \varphi_{qk}(n), \quad (5)$$

$$\Delta b_k(n) = \eta e_k(n) f_2'(w_k(n)), \quad (6)$$

$$b_k(n+1) = b_k(n) + \Delta b_k(n). \quad (7)$$

The weight and threshold of the hidden layer to the input layer are set: correction $\Delta\omega_{lq}(n)$, new weight $\omega_{lq}(n+1)$, correction $\Delta b_q(n)$ and new threshold $b_q(n+1)$ are obtained:

$$\Delta\omega_{lq}(n) = \eta e_k(n) f_2'(w_k(n)) \varphi_{qk}(n) f_1'(v_q(n)) x_l(n), \quad (8)$$

$$\omega_{lq}(n+1) = \Delta\omega_{lq}(n) + \omega_{lq}(n), \quad (9)$$

$$\Delta b_q(n) = \eta e_k(n) f_2'(w_k(n)) f_1'(v_q(n)), \quad (10)$$

$$b_q(n+1) = b_q(n) + \Delta b_q(n). \quad (11)$$

3.3. Evaluation of real estate based on genetic neural network

To make up the deficiencies of BP neural network, neural network is optimized by genetic algorithm. General algorithm with excellent global searching ability can effectively obtain the global optimal solution and make up the shortcoming of the neural network, i.e., the weak global searching ability. The steps of evaluating real estate using the genetic neural network are as follows.

Setting the initial value of the system: the evaluation staff determines the selection conditions of samples and input the qualified data information into the network as learning samples; standardized operation is performed on learning samples; the expected error is determined; the maximum training times of BP neural network is set; the crossover probability, mutation probability and the maximum iteration times of genetic algorithm are calculated.

Obtaining the optimal neural network: M groups of data are randomly generated in $[-1, 1]$ (M means the scale of initial population); the M groups of data are coded using binary coding; the training samples are input into BP neural network, decoded, and the initial weights and thresholds of BP neural network are given; the corresponding fitness function is selected to calculate the global error value of the network, and operation is done to obtain the most appropriate initial weights and thresholds of BP neural network, and the optimal BP neural network is obtained for prediction.

Real estate evaluation: The optimal BP neural network model obtained in the above steps is used to evaluate the real estate; the evaluator inputs the data information of the subject to be evaluated into the acquired model to obtain the corresponding evaluation results for reference.

The genetic algorithm has the ability of macroscopical exploration and global optimization, through which the optimal solution can be obtained to optimize the connection weight and threshold of BP neural network. BP neural network has good perception ability and can realize multi-level network learning. In order to implement high-quality real estate evaluation, genetic algorithm and BP neural network are optimized and combined to absorb the advantages of both and to achieve the goal of remedying the limitations and shortcomings of GA and BP neural network, and the genetic neural network model (Tan et al. 2018) can be obtained. Through this way, a reasonable network structure is constructed, which can effectively

guarantee the accuracy of real estate evaluation data. **Figure 3** is the flow chart of the genetic neural network algorithm.

Binary coding is as follows. Suppose that the range of a parameter o is $[\min, \max]$, which is expressed as m -bit binary number. Suppose the decimal number corresponding to the m -bit binary number as a , then we have:

$$o = \min + \frac{a}{2^m - 1}(\max - \min). \quad (12)$$

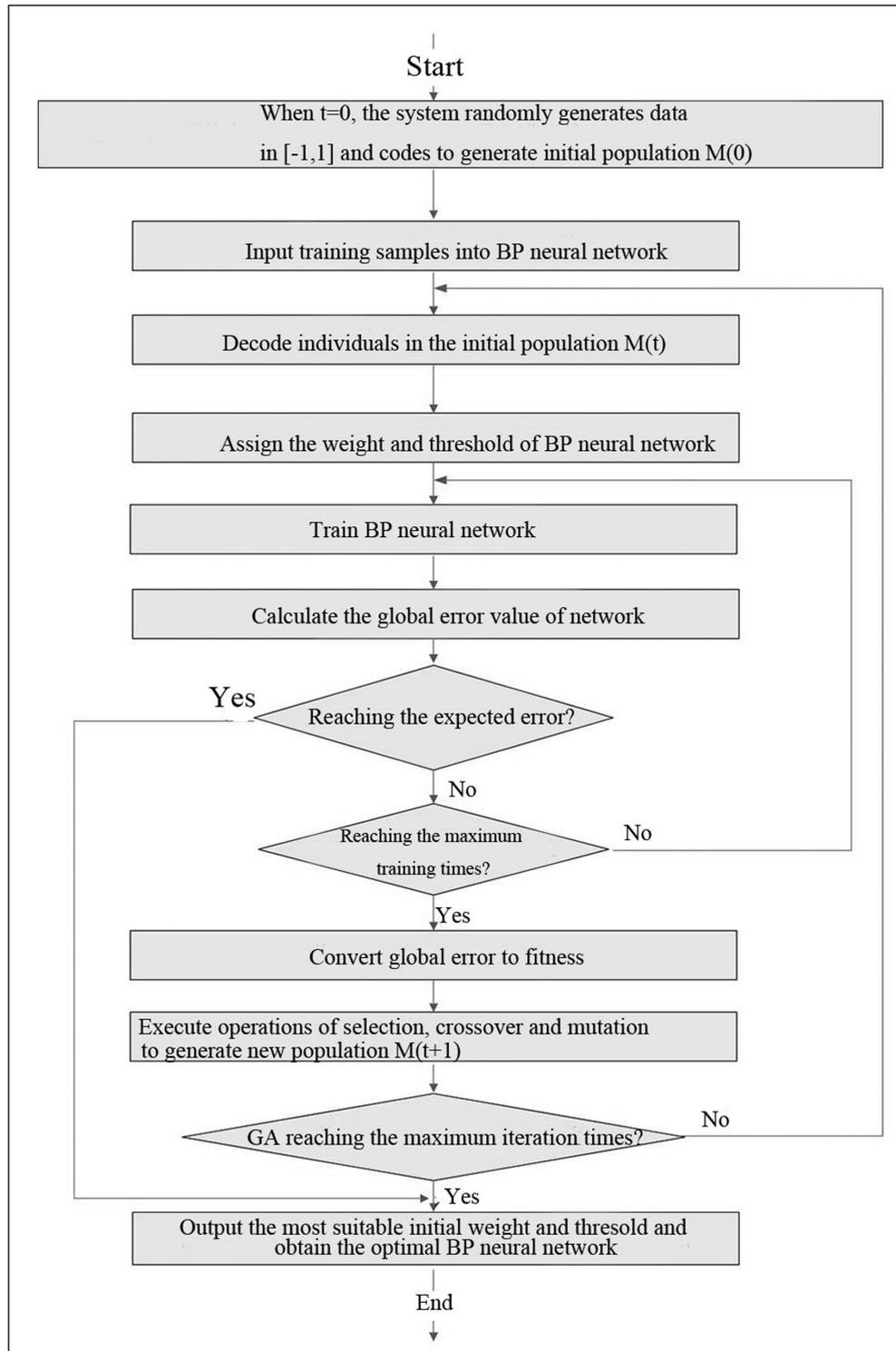


Figure 3: The flow chart of the genetic neural network algorithm.

The error value of the output result of the neural network is obtained by setting corresponding fitness function, and it was expressed as E . The fitness function indicated that the smaller the prediction error value of the neural network, the larger the corresponding fitness function, i.e., the poorer the fitness, satisfying the objective expectation:

$$f = \frac{1}{E}. \quad (13)$$

The crossover process follows crossover probability p_c . The difference value of gene in this simulation experiment is the sum of the absolute value of the difference between the genetic value of two individual coding strings corresponding to the same gene position. Through trial calculation method, it is found that the average relative error is 0.04, 0.05, 0.09 and 0.07 respectively when the crossover probability is 0.06, 0.1, 0.15 and 0.2, i.e., the error is the smallest when the crossover probability is 0.06. Therefore 0.06 was selected as the crossover probability. Moreover, an adaptive mutation probability is adopted; G stands for the total number of evolution generations, m stands for the current evolution generation number, F stands for the current fitness function of individuals, A and B stand for adaptation coefficient, F_{\min} stands for the minimum value of fitness function of the current population, and F_{\max} stands for the maximum value of fitness function of the current population:

$$p_M = A * B^{\frac{-m}{G}} * (F_{\max} - F) / (F_{\max} - F_{\min}). \quad (14)$$

Activation function mainly refers to the use of non-linear function to retain and map the features of the activated neurons. Activation function is needed in neurons; therefore threshold unit function is mainly used in this paper. Its expression can be defined as:

$$f(s) = \begin{cases} 1 & s \geq 0 \\ 0 & s < 0 \end{cases}. \quad (15)$$

Moreover, gradient descent method is used as an optimizer which is needed in the algorithm. It is assumed that the parameter of the model which needs learning and training is M , the cost function is $J(M)$, the learning rate is η_t , the model parameter at time t is M_t , then the relevant gradient is obtained, $\Delta J(M)$. Then the updated parameter after the using the gradient descent method is:

$$M_{t+1} = M_t - \eta_t \Delta J(M_t). \quad (16)$$

The normalized standardized method is used, x stands for the standardized data, \min stands for the minimum value of data information of the same type in every group of input data information, \max stands for the maximum value of data information of the same type in every group of input data information, and the input data information is limited in $[0,1]$. The specific algorithm is as follows:

$$p = \frac{x - \min}{\max - \min}. \quad (17)$$

4. Prediction of real estate assessment by simulation experiments

4.1. Experimental methods and parameters

The real estate evaluation was simulated by MATLAB 7.0. Twenty groups of real estate data information obtained from the real estate market survey activities of Q city from October 2018 and January 2019 were selected. Quantitative and standardized operations were performed.

The data were divided into two parts: training samples and testing samples. The number of nodes in the input layer of neural network was the number of categories of real estate data information. The number of node of the output layer was 1. The number of nodes in the hidden layer was determined by trial-and-error method. Firstly, the number of nodes in the hidden layer was set as 3, and then it was adjusted constantly using data set. It was found that the result was the best when the number of nodes in the hidden layer was 8. Through repeated tests, it was found that the experimental error was the smallest when the population scale was 60. Therefore the population scale was set as 60, individuals of the initial population was set as the random numbers in $[-1, 1]$, and the number of iterations was 7.

4.2. Experimental results

It was found from **Figure 4** that with the increase of training times, the mean square error of the two models decreased, and the overall trend was downward. With the increase of training times, the accuracy was improved. However, under the same number of training, the mean square error of the optimal BP neural network obtained by the genetic neural network model was smaller than that obtained by the BP neural network model. Therefore, the accuracy of genetic neural network in real estate data evaluation was much higher than that of BP neural network. Moreover the convergence speed of the training of the genetic neural network model was much faster than that of the BP neural network model. The reason was that the genetic neural network model had screened out the most appropriate weights and thresholds, and the optimal BP neural network was obtained, which greatly improved the training convergence speed.

It was found from **Figure 5** that there was a large difference between the real price and the price evaluated by BP neural network model, the trend of them was different, and the overall fitting degree was low. The difference between the real estate price and the price evaluated by the genetic neural network was small, the trend of them was basically similar, and the overall fitting degree was high. Therefore, it was concluded that the genetic neural network model was more accurate for real estate evaluation data and had a higher overall fitting degree.

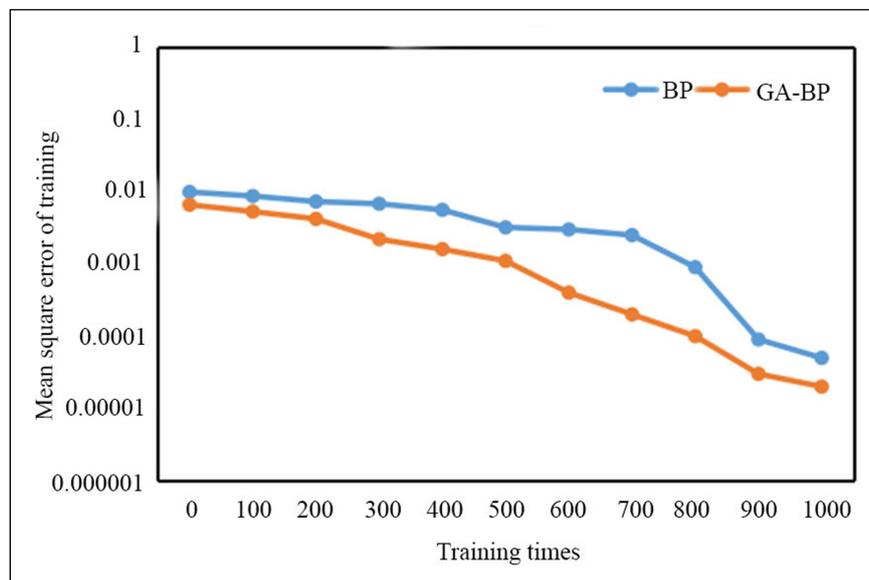


Figure 4: Variation of training mean square errors of the two models.

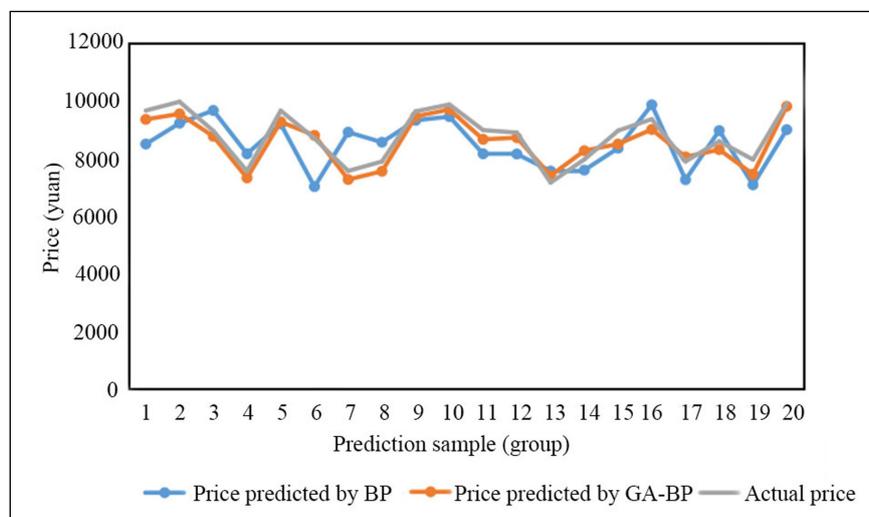
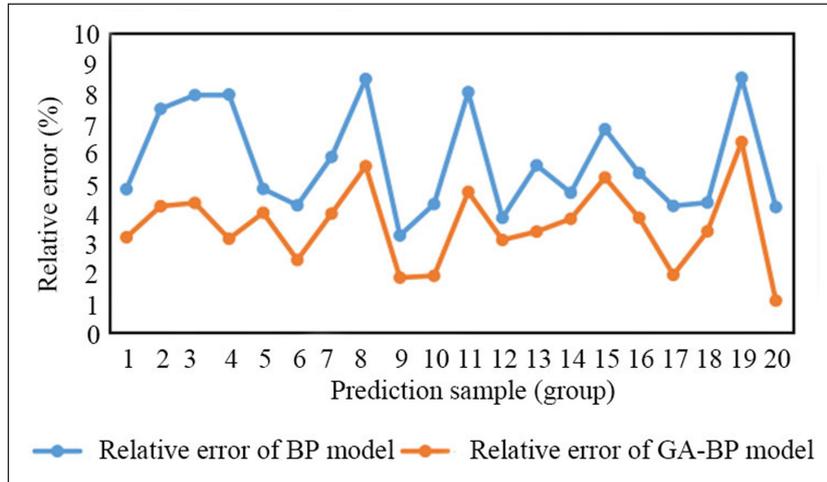


Figure 5: Comparison of the predicted value and actual value of the real estate evaluation by the two models.

Table 1: Error analysis.

	Average relative error (%)
BP Neural Network Model	5.722
Genetic Neural Network Model	3.552

**Figure 6:** Relative errors of two real estate assessment models.

4.3. Comparison of errors

By analyzing the data of different models, **Table 1** is obtained.

It was found from **Figure 6** and **Table 1** that the relative errors of the two models showed upward and downward fluctuations. The average relative error of the genetic neural network model was 3.552, and that of the BP neural network model was 5.722. Compared with the latter, the former has a smaller average relative error, which was close to the real estate price and the trend, and the overall fitting degree was high.

Overall, the evaluation data obtained by the BP neural network model were more accurate. It was concluded that the genetic neural network model had better prediction performance for real estate evaluation. The reason was that the genetic neural network model input the most appropriate weights and thresholds to obtain the optimal BP neural network, which improved the training accuracy and enhanced the experimental effect. Moreover, the prediction performance of the BP neural network model needs to be greatly improved and optimized to improve the quality of real estate evaluation works.

5. Conclusion

In this study, the problem of real estate evaluation was studied, and a real estate evaluation model based on genetic algorithm optimized neural network was put forward. The experimental results showed that the average relative error value of the genetic neural network model for real estate evaluation was 3.552, the evaluation value was close to the actual value, and the overall fitting degree was high, which indicated that the evaluation value was more accurate. Therefore, it was concluded that compared with the traditional BP neural network model, the real estate evaluation model based on genetic algorithm optimized neural network had more remarkable performance, the prediction error was smaller, and the fitting degree with the experimental data was higher. The reason for the result was that the genetic neural network selected the most appropriate weight and threshold, the selection effects of the weight and threshold were better than methods proposed in Ying et al. (2017) and Chen et al. (2015); the optimal BP neural network was obtained, which improved the prediction performance, and the prediction performance was superior to the imperialist competitive algorithm (Fathy and Rezk, 2017). The work proved the feasibility and practicability of the proposed method and provides a new way for the evaluation of real estate. There are some deficiencies in this study although some achievements are obtained, for example, the insufficient analysis of real estate market and small size of sample. In the future work, the size of training samples needs to be measured more precisely, and more factors influencing the price of real estate should be included in the study.

Competing Interests

The author has no competing interests to declare.

References

- Chen, JY, Yao, PY, Wang, B and Shui, DD.** 2015. Dynamic threat assessment based on structure entropy and IGSO-BP algorithm. *Systems Engineering & Electronics*, 37(5): 1076–1083.
- Daniel, S, Laurenju, D and Droj, G.** 2015. Usage of Location Analysis Software in the Evaluation of Commercial Real Estate Properties. *Procedia Economics & Finance*, 32: 826–832. DOI: [https://doi.org/10.1016/S2212-5671\(15\)01525-7](https://doi.org/10.1016/S2212-5671(15)01525-7)
- Domian, D, Wolf, R and Yang, HF.** 2015. An assessment of the risk and return of residential real estate. *Managerial Finance*, 41(6): 591–599. DOI: <https://doi.org/10.1108/MF-07-2013-0195>
- Fathy, A and Rezk, H.** 2017. Parameter estimation of photovoltaic system using imperialist competitive algorithm. *Renewable Energy*, 111: 307–320. DOI: <https://doi.org/10.1016/j.renene.2017.04.014>
- Fife, A.** 2017. A comparative assessment of the factors influencing the valuation and market pricing of fractional interests in real estate. *Indagationes Mathematicae*, 26(3): 455–467. DOI: <https://doi.org/10.1016/j.indag.2015.01.005>
- Liu, M and Shi, J.** 2018. A cellular automata traffic flow model combined with a BP neural network based microscopic lane change decision model. *Journal of Intelligent Transportation Systems*, 1–34. DOI: <https://doi.org/10.1080/15472450.2018.1462176>
- Li, Y and Wang, M.** 2015. Research on an Application Mass Appraisal Technology to Real Estate Markets in China. In: *Iccrem*. ASCE, 2015. DOI: <https://doi.org/10.1061/9780784413777.136>
- Manganelli, B, Mare, GD and Antonio, N.** 2015. Using Genetic Algorithms in the Housing Market Analysis. *Lecture Notes in Computer Science*, 9157: 1–10. DOI: https://doi.org/10.1007/978-3-319-21470-2_3
- Metawa, N, Hassan, MK and Elhoseny, M.** 2017. Genetic algorithm based model for optimizing bank lending decisions. Pergamon Press, Inc. DOI: <https://doi.org/10.1016/j.eswa.2017.03.021>
- Poursaeed, O, Matera, T and Belongie, S.** 2018. Vision-based real estate price estimation. *Machine Vision and Applications*. DOI: <https://doi.org/10.1007/s00138-018-0922-2>
- Sevinç, E and Coar, A.** 2018. An Evolutionary Genetic Algorithm for Optimization of Distributed Database Queries. *The Computer Journal*, 54(5): 717–725. DOI: <https://doi.org/10.1109/ISCIS.2009.5291839>
- Svetlana, A.** 2017. Technical Assessment of Constructions in Real Estate Appraisal. *Advanced Engineering Forum*, 21: 639–646. DOI: <https://doi.org/10.4028/www.scientific.net/AEF.21.639>
- Tang, MC, Zhou, CC, Zhang, NN, et al.** 2018. Prediction of the Ash Content of Flotation Concentrate Based on Froth Image Processing and BP Neural Network Modeling. *International Journal of Coal Preparation and Utilization*, 1–12. DOI: <https://doi.org/10.1080/19392699.2018.1458716>
- Tan, X, Ji, Z and Zhang, Y.** 2018. Non-invasive continuous blood pressure measurement based on mean impact value method, BP neural network, and genetic algorithm. *Technology and health care: Official Journal of the European Society for Engineering and Medicine*, 26(6): 1–15. DOI: <https://doi.org/10.3233/THC-174568>
- Ying, Z, Wang, HW, Wang, YB and Guo, PC.** 2017. A novel optimization algorithm for BP neural network based on RS-MEA. In: *International Conference on Image*, Chengdu, China on 2–4 June 2017.
- Zhang, XF and Ji, W.** 2010. Study on the Risk Assessment of Real Estate Project Based on BP Neural Network. In: *IEEE International Conference on Information & Financial Engineering*, Chongqing, China on 17–19 Sept 2010.

How to cite this article: Sun, Y. 2019. Real Estate Evaluation Model Based on Genetic Algorithm Optimized Neural Network. *Data Science Journal*, 18: 36, pp. 1–9. DOI: <https://doi.org/10.5334/dsj-2018-036>

Submitted: 26 March 2019 **Accepted:** 03 July 2019 **Published:** 23 July 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 