

House Price Prediction Model Based on Neural Network

chusheng Wang (✉ 307476117@qq.com)

Chongqing Technology and Business University

Research Article

Keywords: BP neural network, house price, influential factors, prediction model

Posted Date: September 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-805003/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

House Price Prediction Model Based on Neural Network

Chusheng Wang

Chongqing Technology And Business University, No.19, Xuefu Ave, Nanan District, Chongqing, P.R.China 400067
wangchusheng75@126.com

ABSTRACT

Through an in-depth understanding of house price prediction issues, the paper aims to establish a BP neural network model for house price prediction based on ideas and methods of the BP neural network. By the BP neural network method, the paper realizes sorting, statistics, and analysis of house prices in Chongqing for 10 years as well as the main factor analysis of house prices. Then, correlative analysis is conducted with the software on house prices and their influential factors. Significant correlations are manifested. MATLAB was used to compile relevant programs and a BP neural network model was established for prediction and verification of house prices. Through the market survey, house prices are finally predicted by software.

Key words: BP neural network; house price; influential factors; prediction model

1. Introduction

House prices concern the national economy and the people's livelihood, involving interests and resources of different aspects, the especially optimized configuration of scarce resources. Therefore, house price prediction appears to be quite important. Also, with the prosperity of the real estate industry and further perfection of the real estate market, it is necessary and urgent to realize the prediction of house prices. House price prediction exists in the whole lifecycle of real estate. In different stages such as real estate investment, development, construction, sales, distribution, leasing, sub-leasing, transfer, mortgage, gifting, expropriation, insurance, taxation, and dismantling, corresponding price prediction is needed, based on which decisions and actions can be implemented.

Also, house prediction plays an important role in keeping sustainable and healthy development of the real estate market, which is mainly manifested in aspects such as promoting the healthy development of real estate credit and loans, preventing financial risks, safeguarding legal rights and interest of relevant parties, safeguarding national and public interests, reducing management cost and providing comprehensive services for the real estate market, etc. Also, with the continuous development of the social economy, the house price prediction will be extended to more fields. Prices of real estate are influenced by many factors, while they have different influential directions, degrees, and relations on real estate prices. Some factors can be completely measured with math

models, but others cannot. Some of them can only be judged by prediction staffs based on experience. Therefore, house price prediction is an important part of huge difficulties, in need of high technologies in the real estate industry. Therefore,

it is of great significance to realize in-depth and systematic research on real estate prediction and perfection and optimization of theories and methods of house price prediction[1-3].

2. Meaning of Prediction

In actual life, the prediction is widely applied. For example, historical materials of a period in the economic system or management system can be used to predict trends of the system development. Prediction means that development trends and results of future events are deduced based on certain scientific methods and logistic deduction as well as qualitative or quantitative calculation and exploration of evolution rules of things. Prediction means the estimation of the future of one thing based on its past and present and deduction of unknown things according to known things. The accuracy of prediction completely depends on the acquisition degree of motion rules of the researched courses. The more complete and deeper acquaintance of objective rules will facilitate more accurate predictions. The accuracy degree of prediction also relies on the method used for prediction. Prediction serves decision making. It is an important manner to avoid blindness and enhance conscientiousness and scientific properties. Correct decision making is sourced from accurate information and scientific prediction. Therefore, the prediction is deemed as the premise or basis of decision making. For example market prediction is the basis of production decision making[4-5].

Prediction is also the basis for formulating development plans. For example, population prediction is the basis for formulating development plans concerning housing and highways, etc. Prediction of technological progress and scientific development can promote the updating of products and enhance the competitiveness of products. It is an important approach for operation and management. A prediction course is an acquaintance course of things. Therefore, prediction can enhance the manager's acquaintance of things, to enhance his/her foreseeability in work, and enhance operational and

management level[6].

3. Comparison of House Price Prediction Methods

Common house price prediction methods include house price prediction with regression analysis, prediction with a gray prediction mode, a prediction based on the Elman neural network, and prediction with a neural network model [1,7].

3.1 Prediction Based on Regression Analysis

Regression prediction means that a regression equation between the independent variable and the dependent variable is sought according to changing rules of historical data, and model parameters are confirmed, based on which prediction is realized. According to the number of independent variables, the regression issue can be divided into unitary regression and multiple regression. The regression analysis method is generally applicable to middle-phase prediction[8-9]. The regression analysis method is mainly characterized in that:

- (1) Technologies are relatively mature, and prediction courses are simple;
- (2) Influential factors of a prediction object are decomposed, and changes of different factors are investigated so that the quantity state of the prediction object is estimated;
- (3) The regression model has a large error and poor extrapolation. When influential factors are complicated or relevant factor data cannot be obtained, the error of the regression model cannot be corrected even though the addition of computation quantity and complexity.

As requested by the regression analysis method, the sample size must be large, while samples must have good distribution rules. When the prediction length is larger than the occupied original data length, prediction with this method theoretically cannot guarantee the accuracy of prediction results. Also, the quantified results may go against qualitative analysis results. Sometimes, a proper regression equation type cannot be found[10].

3.2 Prediction Based on Gray System Theories

Gray's theory was proposed by Professor Deng Julong in the 1980s. It was firstly applied in control theories. Its basic idea is to integrate known time-correlated data sets, thing sets, decision sets, and relation sets, combine them according to certain rules, constitute a dynamic or non-dynamic combination, form white modules, and solve future gray modules through certain conversion and solutions. As believed in the gray theories, the objective system features complicated manifestations and discrete and chaotic data, it has an overall function anyway and is always orderly. Hence, it must have certain internal rules[11].

Different from the traditional method of random change data sequence processing, the gray theory deems all the random variables as gray variables changing within a certain scope. Gray variables are not treated with big sample analysis and research from the perspective of statistical rules. Instead, they are generated and restored through data processing methods. The chaotic original data is sorted into generation data with strong rules for research. In other words, the gray system theory establishes a data model rather than an original data model. Through original data processing and gray model establishment, the gray prediction finds and masters system development rules and makes scientific and quantitative predictions of future states. Its prediction model is an

exponential function. If the object to be measured develops with an exponential rule, it is predicted to obtain prediction results with high accuracy[12].

Table 1 Comparison of real and predicted house prices in Chongqing in 10 years

House price	2010	2011	2012	2013	2014
Actual price	1387.56	1422.27	1542.48	1548.98	1875.98
Predicted price	1447	1587	1769	2018	2350
House price	2015	2016	2017	2018	2019
Actual price	2191.02	2868.00	3191.74	4650.17	5895.00
Predicted price	2784	3445	4263	5276	6528

In the actual application of many fields, a lot of cases succeed in model prediction but are also limited in high prediction deviations. Key factors influencing the accuracy and adaptability of model prediction include the construction of background values in a model and selection of initial values in a prediction formula. Some articles study this aspect and propose formulas for background values in new construction models as well as a selection of initial values, which to some extent have increased the accuracy of prediction models. However, the improvement is not thorough. Further studies are still needed[13].

Next, we use the GM(1,1) model to predict 10-year house prices in Chongqing through MATLAB software programming [2,7,14] (see appendix 2 for the program) and obtain prediction values. Actual values and prediction values are shown in the following table:

As shown in the comparison results, the errors between actual values and predicted values are large.

3.3 Prediction Based on Elman Neural Network

An Elman neural network is generally divided into four layers: an input layer, hidden layer, receiving layer, and output layer. The connection of the input layer, hidden layer, and output layer is similar to a feedforward network. Units in the input layer only play a role in signal transmission. Units of the output layer play a role in linear weighting. A linear or nonlinear function can be used as a transmission function of units in the hidden layer. The receiving layer is also called a context layer or a state layer. It is used to memorize the output value of units in the hidden layer at the previous moment and

return it to the network input. It can be deemed as a one-step delay operator.

The Elman neural network is characterized in that output of the hidden layer is independently connected to the input of the hidden layer through the delay and storage of the receiving layer. Through such a model of independent connection, it is sensitive to historical state data. Through the addition of the internal feedback network, the network is more capable of creating dynamic information, so that dynamic modeling is realized. Also, the Elman neural network can approach any nonlinear mapping at any accuracy. The specific form for external noise to influence the system does not need to be considered. The system modeling can be realized when the input-output data pair of the system is given[15].

Next, we predict the 10-year house prices in Chongqing with the Elman neural network. Here, house prices in 3 previous years are used to predict house prices in the fourth year. It is assumed that house prices in 2019 are unknown, then we use data of 2016, 2017, and 2018 to predict that in 2019. Through normalized processing of house prices in Chongqing during 10 years from 2010 to 2019, this result is obtained: $R=[0.0077 \ 0.0344 \ 0.0358 \ 0.1084 \ 0.1783 \ 0.3284 \ 0.4003 \ 0.7238 \ 1.0000]$. As selected, training samples are $Q=[0.0077 \ 0.0344; 0.0077 \ 0.0344 \ 0.0358; 0.0344 \ 0.0358 \ 0.1084; 0.0358 \ 0.1084 \ 0.1783; 0.1084 \ 0.1783 \ 0.3284; 0.1783 \ 0.3284 \ 0.4003]$; testing samples are $[0.3284 \ 0.4003 \ 0.7238]$. The MATLAB programming (see appendix 1) is used to obtain the prediction value of 2009: 0.7662 which has a big error compared to the actual value[16].

3.4 Prediction Based on Neural Network

Theoretical research of ANN (Artificial Neural Networks) is an emerging marginal and interdisciplinary science. It can represent nonlinear relations, learning, or the like, providing new ideas and new methods for many actual problems with complicated uncertainties and time-dependent features. In the field of prediction, Lapeds and Farber firstly used neural networks for prediction in 1987, initiating prediction of ANN. Most economic models established based on econometrics are linear models. Theoretically, a neural network can approach any nonlinear function and can be adjusted randomly, so that the non-linear prediction problems can be solved effectively. The learning function of ANN is used to train neuron networks with a lot of samples. Its connecting value and the threshold value are adjusted. Then, the confirmed model can be used for prediction. A neural network can automatically learn previous experience from data samples without complicated queries or expressions. It automatically approaches the best function which scores sample data rules. Despite forms of these functions, the feature of the neural network is more obvious when the function form manifested by the considered system is more complicated.

The basic idea of error backpropagation (BP algorithm) is to adjust and amend the connecting weight and threshold of the network through the backpropagation of network errors to minimize the errors. Its learning course includes forward computation and error backpropagation. By using a simple three-layer ANN model, it can realize any complicated nonlinear mapping relation from input to output. As theoretically proved by Kolmogorov already, a neural network with one hidden layer (there are enough hidden nodes as assumed) can express any continuous function at any accuracy;

K.Hor et al. also verified that the neural network can randomly approach a large category of functions and can reveal nonlinear relations hidden in data samples[17].

In recent years, as a nonlinear model, the neural network model has been used to study prediction issues. Because of its features, the neural network model is a data-driven method. At present, the neural network model has been successfully applied to many fields, such as many economic fields like economic forecasts, fiscal analysis, loan mortgage evaluation, and bankruptcy prediction. It is applied to the prediction field.

3.5 Analysis of Advantages of Neural Network Prediction Method

Various applications of neural networks are available [18-21]. These applications exist from diverse perspective.

(1) A different modeling method is used. In general, the quantitative prediction is a prediction method established based on statistical analysis, which requests complete and definite original data. In an actual system, statistical data is often incomplete and fuzzy. A neural network is driven by data and realized by black-box modeling, without prior (statistical knowledge) information. In a complicated data environment without accurate and complete information sources, it can extract data features through adjustment of its structure and then make an effective prediction of the future.

(2) General prediction methods cannot realize learning and pattern recognition of data samples in modeling. Its modeling is an abstracting course of original data. This abstract course is completed in numerical computation. The abstract results can be completed by a complete math analysis formula. The neural network model is an "image" memory course. By the learning course of the network model, it completes the "memory" of internal relations with the data. The trained neural network model is not meant for the explanation of input and output relations.

(3) Once being established, traditional prediction models have high structural stability. They are limited in processing scales. The actual prediction environment is complicated and varying. In the face of a complicated and varying prediction environment, traditional prediction models cannot be adapted to new relations of actual system variables. As a varying structure mode, the neural network model is capable of self-adaptation and parallel processing, so that it has a strong data processing ability. Also, through network learning of new samples, its internal structure can be adjusted, so that it can be adapted to changes in system variables. Hence, as for the non-linear high-dimensional and high-order issues, the neural network can play a more capable role.

(4) General prediction methods do not have a learning course of samples, so the fitness for samples is relatively low. The learning feature of a neural network model decides its maximum fitting with samples. Also, the neural network model has a fault-tolerant ability. Through the training of the neural network model, the system variable relations can be highly fitted, so that correlated influences between system variables can be analyzed. Therefore, rules of relations between system variables can be found and can be applied to the decision making of actual issues.

(5) During the establishment of a neural network model, a network model can be trailed under multiple virtual variables to recognize features of system variables in different development stages based on changes in its internal structure.

To correctly reflect expression forms of a system in different stages, a traditional prediction model must set different virtual variables for model correction. The ANN is advantageous in that the structure and functions such as information processing and retrieval of a brain neural system can be simulated based on different degrees and layers. It is highly self-adaptive to a lot of non-structural and non-accurate rules and is characterized by information memory, independent study, knowledge deduction, and optimized computation. It prevails over conventional algorithms and expert system technologies in unique owning of self-learning and self-adaptation functions.

In general, a learning algorithm is constituted of forwarding propagation and backpropagation. It is assumed the input vector u of a known network has n dimensions, while the output vector y has m dimensions, and the length of the given input/output sample pair is L . The BP learning algorithm is implemented by the following steps:

- (1) An initial weight system $W(0)$ is set, which is a small random non-zero value;
- (2) An input/output sample pair is given to compute the output of the network.

Asset, the p -th group of sample input: $u_p = (u_{1p}, u_{2p}, \dots, u_{np})$, output: $d_p = (d_{1p}, d_{2p}, \dots, d_{mp})$.

When the node i is input in the p -th group of samples, the output is y_{ip} .

$$y_{ip}(t) = f[x_{ip}(t)] = f\left[\sum_j w_{ij}(t)I_{jp}\right] \quad (1)$$

Where: I_{jp} denotes the j -th input of the node i when it is input in the p -th group of samples; (1) is a differentiable S-type action function formula $f(x) = \frac{1}{1+e^{-x}}$. The output of the network output node can be computed from the input layer via the hidden layer to the output layer.

- (2) The network target function J is computed.

It is assumed E_p is the target function of the network in case of input of the p -th group of samples. Here, the L2 norm is selected, so that:

$$E_p(t) = \frac{1}{2} \|d_p - y_p\|_2^2 = \frac{1}{2} \sum_k [d_{kp}(t) - y_{kp}(t)]^2 = \frac{1}{2} \sum_k e_{kp}^2(t) \quad (2)$$

Where: when $y_{kp}(t)$ is subject to the input of the p -th group of samples, the network output is adjusted after t weight adjustments; k is the k -th node of the output layer.

Total target function of the network:

$$J(t) = \sum_p E_p(t) \quad (3)$$

It is taken as the evaluation of network study effects.

- (3) As judged, the algorithm is ended when $J(t) \leq \varepsilon$; otherwise, the step (5) is started. In the formula, ε is determined in advance, $\varepsilon > 0$.

- (4) Backpropagation computation

Back computation is conducted in the output layer according to J with a gradient descent method. Weights can be adjusted layer by layer.

According to $a_{k+1} = a_k - \eta_k \frac{\partial J(a_k)}{\partial a_k}$, a constant is taken as the step length, so that adjustment equation of $t+1$ times of the connecting weight from neuron j to neuron i can be obtained:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial J(t)}{\partial w_{ij}(t)} = w_{ij}(t) - \eta \sum_p \frac{\partial E_p(t)}{\partial w_{ij}(t)} =$$

$$w_{ij}(t) + \Delta w_{ij}(t) \quad (4)$$

Where: η is the step length, which is called a learning operator here.

Specific algorithm:

$$\frac{\partial E_p}{\partial w_{ij}} = \frac{\partial E_p}{\partial x_{ip}} \cdot \frac{\partial x_{ip}}{\partial w_{ij}} \quad (5)$$

$$\text{Asset: } \delta_{ip} = \frac{\partial E_p}{\partial x_{ip}} \quad (6)$$

Where: δ_{ip} is the sensitivity of the state x_{ip} of the i -th node to E_p

As shown in Formula (5) and Formula (6):

$$\frac{\partial E_p}{\partial w_{ij}} = \delta_{ip} \cdot I_{jp} \quad (7)$$

δ_{ip} can be computed based on the following two cases.

1. If i is an output node, namely $i=k$, it can be known according to Formula (2) and Formula (6):

$$\delta_{ip} = \delta_{kp} = \frac{\partial E_p}{\partial x_{kp}} = \frac{\partial E_p}{\partial y_{kp}} \frac{\partial y_{kp}}{\partial x_{kp}} = -e_{kp} f'(x_{kp}) \quad (8)$$

Formula (8) is substituted into Formula (5):

$$\frac{\partial E_p}{\partial w_{ij}} = -e_{kp} f'(x_{kp}) I_{jp} \quad (9)$$

2. If i is not an output node, namely i is not equal to k , so the Formula (6) is then converted to:

$$\delta_{ip} = \frac{\partial E_p}{\partial x_{ip}} = \frac{\partial E_p}{\partial y_{ip}} \frac{\partial y_{ip}}{\partial x_{ip}} = \frac{\partial E_p}{\partial y_{ip}} f'(x_{ip}) \quad (10)$$

Where

$$\begin{aligned} \frac{\partial E_p}{\partial y_{ip}} &= \sum_{m_1} \frac{\partial E_p}{\partial x_{m_1 p}} \cdot \frac{\partial x_{m_1 p}}{\partial y_{ip}} = \\ &\sum_{m_1} \frac{\partial E_p}{\partial x_{m_1 p}} \cdot \frac{\partial}{\partial y_{ip}} \sum_j w_{m_1 j} I_{jp}^* = \\ \sum_{m_1} \frac{\partial E_p}{\partial x_{m_1 p}} w_{m_1 j} &= \sum_{m_1} \delta_{m_1 p} w_{m_1 i} \quad (11) \end{aligned}$$

In the formula, m_1 is the m_1 -th node of the layer in succession to the node i ; I_{jp}^* is the j -th input of the node m_1 . When $i=j$,

$$y_{ip} = I_{jp}^*$$

Formula (10) and Formula (11) are substituted into Formula (5), then:

$$\begin{aligned} \frac{\partial E_p}{\partial w_{ij}} &= f'(x_{kp}) I_{jp} \sum_{m_1} \frac{\partial E_p}{\partial x_{m_1 p}} w_{m_1 i} \\ &= f'(x_{kp}) I_{jp} \sum_{m_1} \delta_{m_1 p} w_{m_1 i} \quad (12) \end{aligned}$$

As found, the Formula (9) and Formula (12) can be used for weight adjustment computation of Formula (4).

4. House Price Prediction Model of BP Neural Network

House price is influenced by factors of various aspects, so a lot of factors need to be considered during the establishment of a house price prediction model. The BP algorithm manifests strong processing capability during the processing of a multi-input nonlinear system, so here the BP algorithm is used for prediction.

4.1 Principles of BP Neural Network Prediction Model

In applications such as system modeling, discrimination, prediction, and the like, as for a linear system, the black box input/output model can be effectively expressed in transfer function matrices in the frequency domain; in the time domain, based on various parameter estimation methods, the ARMA model can give systematic descriptions of input/output, so that the problem of linear system prediction is solved perfectly. As

for a nonlinear system, a non-linear self-regression based moving average model is generally used for prediction. However, it is not easy to find a proper parameter estimation method for this model. Hence, it is hard to realize the discrimination of traditional nonlinear systems in theoretical research and actual applications.

Neural networks manifest obvious superiorities in this aspect. Capable of approaching any non-linear mapping by learning, neural networks can be applied to modeling and discrimination of nonlinear systems. They can get rid of the limits of nonlinear models, so that learning algorithms easy to implement in engineering can be given conveniently.

Through learning and training of a lot of dispersing experimental data, field knowledge of neural networks can be extracted and expressed as sizes and distribution of network connecting weights. A system model reflecting the internal rules of an actual course can be established so that effective solutions can be provided for technological and information modeling. A prediction modeling course based on a BP network is shown in the following diagram.

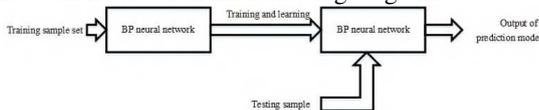


Fig.1 Flow chart of the predictive modeling system

As shown in Fig.1, in the predictive modeling system, the sample data set obtained in the experiment is taken as a training sample to train the neural network. After training, the model representing the neural network as well as weights and thresholds of knowledge parameters are acquired. Then, the requested testing data is taken as the input of the neural network in the neural network training of acquired weights and thresholds, so that the requested prediction modeling output is acquired.

As an effective intelligent information processing technology, neural networks can realize modeling according to internal associations of data and have the good nonlinear approaching capability and the ability for comprehensive processing of chaotic information, which makes them become a powerful tool for exploring the secrets of AI. They've achieved remarkable achievements in many fields such as signal processing, pattern recognition, and control and have been applied widely. The neural network has a very strong nonlinear approaching capacity and features such as self-learning and self-adaptation so that it does not need the establishment of display relations and math models of a complicated nonlinear system. Many limits and difficulties of traditional quantitative prediction methods can be overcome, while the influences of many manual factors can be avoided. Hence, neural network technology is applied to house price prediction. It enjoys unique advantages in rationality and applicability of construction of a house price prediction model.

A neural network model features massive parallelism, storage distribution, structural variability, high-degree nonlinearity, self-learning, and self-organization, etc. It can approach functions which can most effectively depict sample data rules, despite forms of these functions. The extensive adaptability, learning ability, and mapping ability of the neural network manifest certain superiority in prediction through study and

mastering of inter-data dependence relations. It prevails over traditional methods which are very accurate because of dependence on the deduction of math models and parameter optimization but are also limited due to the same reason.

4.2 Basic Steps for Application of BP Neural Network in House Price Prediction

House price prediction with the BP neural network can be divided into training and prediction of the neural network. Specific steps are as follows:

(1) Training sample data is selected to construct the training sample. House prices are influenced by many economic factors. Also, future trends of house prices may be influenced by manual roles, governmental regulation, and the like. Therefore, it is necessary to select proper house price sample data, otherwise, the network prediction ability will be reduced by improper data.

(2) House price sample data is pre-processed. Before network prediction, to avoid the neural network paralysis caused by overlarge original data, we normalized the house price sample data. Given the large changing range, the predicted values should not be taken as the network input, so they were also normalized. The data is normalized into $[-1,1]$, so that data smoothness could be realized and the noise of prediction results could be eliminated.

(3) Training samples are constructed. House price prediction is based on a lot of historical data. Here, we mainly selected GDP, land price, RMB exchange rate (USD to RMB), urban population, and per capita disposable income of urban residents which could influence house prices during 2000-2009. Data from 2000-2008 was taken as the model training sample. The data of 2009 was taken as the testing sample.

(4) A BP neural network with a 3-layer structure is used for prediction model establishment.

(5) The network model is tested. The testing sample is used for testing the trained model.

4.3 Implementation of House Price Prediction with BP Neural Network

The selection of influential factors on house prices should be scientific and rational and must of representative and practical significance. There are a lot of influential factors on house prices. It is not practical or meaningful to research and analyze all the influential factors. Also, some of the influential factors on house prices must be qualitatively analyzed, such as house-dependent factors, house supply-demand factors, some national macro-regulation policies; others can be quantitatively analyzed, such as GDP, land price, interest rate, and per capita disposal income. During the selection of influential factors on house prices, it is necessary to consider factors which can influence house prices, which must be easy for quantified processing.

It is not objective to find all the influential factors on house prices. Analysis of all the influential factors on commercial housing prices is not realistic. Here, we only research house prices in Chongqing under factors of GDP, land price, RMB exchange rate (USD to RMB), urban population, and per capita disposal income of urban residents (see table 2 for relevant data).

Table 2 studies the real estate price of Chongqing under various factors

particular year	GDP (100 million yuan)	low price	Urban population (ten thousand people)	RMB rate (US dollar against RMB)	Per capita disposable income of urban residents (element)	mean price (Yuan / m ²)
2010	0	0	0	1.0000	0	0
2011	0.0561	0.0339	0.0629	0.9990	0.0435	0.0344
2012	0.1110	0.0677	0.1449	0.9990	0.1202	0.0344
2013	0.1749	0.2376	0.2796	0.9990	0.1809	0.0358
2014	0.2251	0.5005	0.4059	0.9989	0.2493	0.1084
2015	0.3317	0.5903	0.5908	0.9398	0.3370	0.1783
2016	0.4507	0.8250	0.7499	0.7871	0.4668	0.3284
2017	0.6306	1.0000	0.8085	0.5317	0.6540	0.4003
2018	0.8114	0.7712	0.8871	0.0742	0.8441	0.7238
2019	1.0000	0.8233	1.0000	0	1.0000	1.0000

Year GDP (RMBx100million) Land Price (RMB) Urban Population (x10000) RMB Exchange Rate (USD to RMB) Per Capita Disposable Income of Urban Residents (RMB) Average House Price (RMB/m²)

Data source: Dynamic Monitoring of Land Prices in Chinese Cities (www.landvalue.com.cn)

Chongqing Statistics (vwww.cdstats.Chongqing.gov.cn/)

Here, a BP neural network model form with five inputs, one hidden layer, and one output is established. Too many nodes of the hidden layer will lead to extra-long learning time. If there are only a few nodes in the hidden layer, the fault tolerance will be low, so the ability to recognize unlearned samples is poor. Hence, the quantity of the hidden layer is generally determined with the following formula based on previous experience: $n = \sqrt{n_i - n_o} + a$, where n is the quantity of nodes in the hidden layer; n_i is the number of input nodes; n_o is the number of output nodes; a is a constant between 1-10;

Data from 2010 to 2018 was taken as the training set sample for the training of the neural network so that the neural network model needed was obtained. Here, it is assumed that the house price in 2019 is unknown. Then, influential factor variables of 2019 were input into the neural network model we trained, so that the predicted value was obtained. Then, whether the error between the predicted value and the actual value could reach the expected error scope is confirmed.

At first, the original data should be standardly normalized (processes shown in the appendix. The processed data is as follows:

Table 3. processed data

particular year	GDP (100 million yuan)	low price (element)	Urban population (ten thousand people)	RMB rate (US dollar against RMB)	Per capita disposable income of urban residents	mean price (Yuan / m ²)
2010	0	0	0	1.0000	0	0
2011	0.0561	0.0339	0.0629	0.9990	0.0435	0.0344
2012	0.1110	0.0677	0.1449	0.9990	0.1202	0.0344
2013	0.1749	0.2376	0.2796	0.9990	0.1809	0.0358
2014	0.2251	0.5005	0.4059	0.9989	0.2493	0.1084
2015	0.3317	0.5903	0.5908	0.9398	0.3370	0.1783
2016	0.4507	0.8250	0.7499	0.7871	0.4668	0.3284
2017	0.6306	1.0000	0.8085	0.5317	0.6540	0.4003
2018	0.8114	0.7712	0.8871	0.0742	0.8441	0.7238
2019	1.0000	0.8233	1.0000	0	1.0000	1.0000

Year	GDP (RMBx100million)	Land Price (RMB)	Urban Population (x10000)	RMB Exchange Rate (USD to RMB)	Per Capita Disposable Income of Urban Residents (RMB)	Average House Price (RMB/m ²)
2000	1312.99	3456	426.06	8.2784	7649	1387.56
2001	1492.04	3594	432.02	8.2770	8128	1422.27
2002	1667.10	3727	439.80	8.2770	8972	1542.48
2003	1870.80	4424	452.57	8.2770	9641	1548.98
2004	2031.07	5495	464.54	8.2768	10394	1875.98
2005	2371.01	5861	482.07	8.1917	11359	2191.02
2006	2750.48	6817	497.15	7.9718	12789	2868.00
2007	3324.40	7530	502.71	7.6040	14849	3191.74
2008	3901.00	6598	510.16	6.9451	16943	4650.17
2009	4502.60	6810	520.86	6.8383	18659	5895.00

Year GDP (RMBx100million) Land Price (RMB) Urban Population (x10000) RMB Exchange Rate (USD to RMB) Per Capita Disposable Income of Urban Residents (RMB) Average House Price (RMB/m²)

Design and training of network: the network is designed as a BP neural network with a hidden layer. There are 5 neurons in the input layer of the network. There is 1 neuron in the output layer. According to previous experience, we finally set 10 as the number of neurons in the hidden layer. After standard normalization, the input vector scope of the network is [0,1]. Hence, transit is used as the transfer function of neurons in the hidden layer, while logs are used as the transfer function of neurons in the output layer, which exactly satisfies the requirement that network output is within [0,1].

Next, the MATLAB software programming is used for design and training of the BP neural network. The processes are as follows:

```
R=[0 0 0 1.0000 0;0.0561 0.0339 0.0629 0.9990-0.0435;0.1110 0.0677 0.1449 0.9990 0.1202;0.1749 0.2376 0.2796 0.9990 0.1809;0.2251 0.5005 0.4059 0.9989 0.2493;0.3317 0.5903 0.5908 0.9398 0.3370;0.4507 0.8250 0.7499 0.7871 0.4668;0.6306 1.0000 0.8085 0.5317 0.6540;0.8114 0.7712 0.8871 0.0742 0.8441];
```

```
Q=[0;0.0077;0.0344;0.0358;0.1084;0.1783;0.3284;0.4003;0.7238];
```

```
R=R';
```

```
Q=Q';
```

```
net=newff (minmax (R), [ 13, 1], {'tansig', 'logsig'}, 'trainlm');
```

```
net.trainparam.show=50;% display the changing curve of training error for every other 50 steps
```

```
net.trainparam.epochs=2000;% times of training
```

```
net.trainparam.goal=0.0001;% network performance indexes
```

```
LP. lr=0.05;% learning rate
```

```
net=train (net, R, Q);
```

```
TRAINLM, Epoch 0/2000, MSE 0.299205/0.0001, Gradient 2.86554/1e-010
```

```
TRAINLM, Epoch 4/2000, MSE 5.60795e-005/0.0001, Gradient 0.0137246/1e-010
```

```
TRAINLM, Performance goal met shown in figure 2.
```

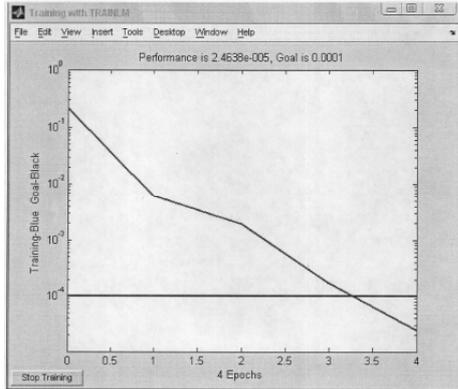


Figure 2. Simulation result

```

y=sim(net, R)
Y=
Columns 1 through 7
0.0107    0.0154    0.0225    0.0434    0.1108    0.1847
0.3300
Columns 8 through 9
0.4090    0.7261
R_test=[1.000 0.8233 1.0000 0 1.0000]';
Y_test=sim(net,R_test)
Y_test=
0.9991

```

The predicted value obtained by the BP neural network model we established is 0.9991, with the accuracy reaching 99.91%, fully demonstrating that the house price prediction result obtained by the BP neural network is relatively ideal. The prediction result is obtained through Matlab programming. It is thus the clear prediction of the house price with the BP neural network is highly reliable and relatively accurate.

5. Conclusions

ANN (Artificial Neural Network) is a network-based on a large-scale parallel structure and distributed storage information, with good fault tolerance and robustness. The BP neural network is the most widely applied network among ANN. Capable of realizing self-organization and self-learning, it has been widely applied in prediction, simulation, and other aspects.

The paper elaborates status quos of foreign and domestic house prices as well as theoretical knowledge of ANN and BP neural networks and analyzes in detail the influential factors on house prices. We conduct statistics, analysis, and processing of 10-year house prices in Chongqing and influential factors and then used the BP neural network model for house price prediction. The result verifies that it is feasible and highly accurate to predict a house price with the BP neural network.

The house price prediction method based on the BP neural network adopted in the paper is helpful to some extent for guiding house prices. Also, with the continuous progress of modern science in China and the continuous deepening of theoretical knowledge, the occurrence of new prediction methods will lay a solid theoretical basis and provide a better math model for prediction. It is convinced the prediction accuracy will see a remarkable leap with the deepening in

house price prediction studies.

Acknowledgment:

Chongqing Higher Education Scientific Research Subject No. : CQJ19B44

Conflict of Interest

The authors declare no conflict of interest regarding this paper.

References:

- [1] Sun Tingting, Shen Yi & Zhao Liang. (2019). A housing price forecasting model based on BP neural network. *Computer knowledge and Technology* (28), 215-218 DOI:10.14004/j.cnki.ckt.2019.3615.
- [2] Zeng Tingting (2020). Research on housing price forecasting model based on machine learning (Master's thesis, Southwest University of Science and Technology) <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202002&filename=1020046072.nh>
- [3] Gao Wen. (2019). Research on house price prediction based on BP neural network optimized by genetic algorithm (Master's thesis, Yan'an University) <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201902&filename=1019857116.nh>
- [4] Wang Gecheng. (2020). Research on housing price prediction based on neural network. *National circulation economy* (03), 128-130 DOI:10.16834/j.cnki.issn1009-5292.2020.03.057.
- [5] Qiao Weide. (2019). Research on commercial housing price prediction based on BP neural network model. *Journal of Shijiazhuang University* (06), 127-133 DOI:10.13573/j.cnki.sjzxyxb.2019.06.018.
- [6] Zhang Qian (2019). Research on housing rent forecasting model based on the Stochastic Forest regression model (Master's thesis, Northeast Normal University) <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201902&filename=1019841897.nh>
- [7] Liu Yang. (2019). Research on house price forecasting method based on BP neural network and Baidu map API (Master's thesis, Inner Mongolia University) <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201902&filename=1019884527.nh>
- [8] Ren Ziming. (2019). Research on the prediction of housing prices in Beijing urban area based on the grey prediction and regression model. *Modern commerce and trade industry* (10), 101-102 DOI:10.19311/j.cnki.1672-3198.2019.10.047.
- [9] Zhang Russian. (2019). Improvement of housing price forecasting model based on multiple linear regression. *Electronic production* (04), 98-100 DOI:10.16589/j.cnki.cn11-3571/tm.2019.04.041.
- [10] Zhang shuangni. (2019). Fitting house price forecasting model based on the stepwise regression model. *Modern agriculture research* (02), 124-125 DOI:10.19704/j.cnki.xdnyyj.2019.02.056.
- [11] Li Zeyang. (2018). Research on comprehensive calculation model of housing price data based on prediction interval theory (Master's thesis, Wuhan University of Science and Technology) <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201802&filename=1018202726.nh>
- [12] Yang Bowen & Cao Buyang. (2017). House price forecasting model based on ensemble learning. *Computer knowledge and Technology* (29), 191-194 DOI:10.14004/j.cnki.ckt.2017.3256.
- [13] Wang Zhichao. (2017). House price forecasting and analysis based on data mining. *Computer knowledge and Technology* (23), 7 DOI:10.14004/j.cnki.ckt.2017.2365.
- [14] Rodin. (2017). Interactive research on second-hand housing in Beijing based on machine learning and gam model (Master's thesis, Taiyuan University of Technology) <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201801&filename=1017832162.nh>
- [14] Xu Ying & Huang Suzhen. (2017). Research on real estate price forecasting model in the era of big data. *China business theory* (04), 134-137 DOI:10.19699/j.cnki.issn2096-0298.2017.04.068.
- [15] Zhou Qin. (2016). House price prediction in Putian City Based on the BP algorithm of artificial neural networks. *Journal of Chifeng University (NATURAL SCIENCE EDITION)* (10), 180-182

- DOI:10.13398/j.cnki.issn1673-260x.2016.10.075.
- [16] Xu Fuqiang. (2011). Neural network combination forecasting model and its application in housing price trend analysis (Master's thesis, Anhui University)
<https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD2012&filename=1011241437.nh>
- [17] Zhang Er Wei (2006). Design of real estate market research and analysis system based on GIS (Master's thesis, Southeast University)
<https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD2007&filename=2007030346.nh>
- [18] Z. Gu, S. Nazir, C. Hong, and S. Khan, "Convolution neural network based higher accurate intrusion identification system for the network security and communication," *Security and Communication Networks*, 2020.
- [19] Z. Gu, S. Nazir, C. Hong, and S. Khan, "Convolution Neural Network-Based Higher Accurate Intrusion Identification System for the Network Security and Communication," *Security and Communication Networks*, vol. 2020, pp. 1-10, 2020/08/28 2020, doi: 10.1155/2020/8830903.
- [20] T. ahmad, Y. Ma, M. Yahya, B. Ahmad, and S. Nazir, "Object Detection through Modified YOLO Neural Network," *Scientific Programming*, pp. 1-10, 2019, doi: 10.1155/2020/8403262.
- [21] S. Asadi, R. H. Abdullah, M. Safaei, and S. Nazir, "An Integrated SEM-Neural Network Approach for Predicting Determinants of Adoption of Wearable Healthcare Devices," *Mobile Information Systems*, 2019.