International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 4 ǁ April 2023 ǁ PP. 671-680

Research on Prediction of Second-hand House Prices Based on BP Neural Network and Multiple Linear Regression

Chunhui Zhao
School of management ,Tianjin University of Technology,Tianjin 300384,P.R.China

Abstract: *With the development of the economy,there is less and less land available for development in the city, and the transaction of second-hand housing has become more active, and the demand for real estate valuation has also increased.Quickly and accurately predicting the price of second-hand housing is of great significance to real estate developers and home buyers.This paper collects a total of 4179 data records by crawling the price data of second-hand housing in Tianjin on the official website of Lianjia, and identifies ten factors affecting the price of second-hand housing, including: area, unit type, area, orientation, floor, total floor, decoration, construction time, traffic volume, and number of followers.Then,according to the variable type,the data is quantified.This study creates a multiple linear regression model and a BP neural network model with the goal of forecasting Tianjin's second-hand home prices.First,the R language software is used to determine the parameter estimation of the multiple linear regression model. After the model passes the test,the house price of the test set is predicted by linear regression.In the BP neural network model,the data is normalized first, and then the input layer, hidden layer, output layer and network parameters are determined, and the house price prediction result of the neural network is given.By comparing the prediction functions of the two models (for the two indicators of relative error and matching), it is found that the performance of the BP neural network model is better. Finally, this paper is summarized and the shortcomings in the study are prospected.*

Keywords: *Second-hand House Price,BP Neural Network,Multiple Linear Regression*

Date of Submission: 14-04-2023 Date of acceptance: 30-04-2023

I. Introduction

For a long time, the real estate industry has become an indispensable part of the national economic development process with its own advantages and the growth rate that cannot be underestimated.From 1995 to 1997, China has implemented the national housing project in 88 cities, promoted the real estate reform, solved the housing problems of some poor people, and won the support of the masses. At the same time, it has also led to economic development and the adjustment of the real estate industry structure. Since then, the real estate industry has experienced rapid expansion, and residential housing prices have also experienced a rapid rise stage.Especially in recent years, with the sharp decrease of land supply in key cities and the rapid growth of land transaction prices,the purchase of second-hand houses has become the choice of more and more families in modern society due to the pressure of house purchase or investment demand.Under this background, the second-hand housing market has developed rapidly.

China's second-hand housing market is in the process of vigorous development. There is less and less land available for development in the main urban areas of first-tier big cities such as Beijing, Shanghai, Shenzhen and Guangzhou.The housing sales in the main urban areas will mainly rely on second hand housing sales .Taking Tianjin as an example, according to the Century-City Price Index of the China Housing Index System, the price of used housing in Tianjin rose 0.08% month-on-month in September 2022, with the average sample price of 25007 yuan pers quare meter and the median sample price of 20039 yuan pers quare meter.After the decline in the first week of October, it was corrected, with a month-on-month increase of 4.93%.However,the price after the correction did not exceed the average price of 21880.27 yuan per square meter in the fifth week of September.

However, the current situation of second-hand housing prediction is not optimistic. Due to various factors affecting the house price, the price fluctuation of second-hand houses has great uncertainty, which brings great risks to real estate developers and buyers. Therefore, this paper aims to build a practical model, input data on factors affecting second-hand housing prices, which can accurately predict second-hand housing prices and provide a feasible method for real estate valuation.In view of this,this paper proposes the idea of predicting second-hand housing prices based on multiple linear regression models and BP neural network models.In this study, the model includes variables that could impact the cost of used homes. First, a multiple linear regression model relying on artificial assumptions is established to predict and analyze the price of used housing.At the same time,the BP

neural network model is established.Different from the traditional mark method,the neural network algorithm can fully mine the data and reduce the subjective arbitrariness of the forecast results.Using this method does not require prior assumptions about the relationship between prices and influencing factors,and avoids assumption errors.Moreover,the neural network algorithm can quickly and accurately process a large amount of input information, increasing the number of objects in the valuation model.

II. Literature Review

Housing prices have always been a topic of widespread concern.One of the topics that is explored by the general public, academics, and professionals is how to create a high-precision model for predicting house prices.Therefore, many people have established a variety of house price prediction models based on different research perspectives.

In 2001, Nguyen Nghiep Cripps Al used the neural network model and the simple multiple linear regression model to build a housing price prediction model for different aggregate data, and predicted and estimated the housing price. By comparing the results of the two models, it is found that the prediction effect of the artificial neural network model is better than the regression model for the housing price under the condition that the data volume is medium or above. In 2012,in addition to using the characteristic price theory,Yang Mu also introduced the random forest theory in the machine learning theory, applied statistical analysis methods related to regression analysis,established a regression model of various factors on the characteristic price for prediction,and obtained the multiple linear regression .The prediction results obtained by the random forest are compared with the results obtained by the random forest prediction, and the conclusion that the prediction accuracy is high by using the random forest method is drawn.In 2014, Hou Puguang and others used ARIMA model and wavelet analysis to predict the house price in Shanxi Province, and the results showed that the individual model had a large error. In the same year, Gao Yuming, Zhang Renjin and others used the improved BP neural network model of genetic algorithm to improve the prediction accuracy. In 2015, Malpezzi applied a regression model to macroeconomic indicators and house price index, and the results showed that house prices did not walk randomly.In 2017, Cao Yang used a dynamic model to predict the house prices of large and medium-sized cities, taking into account the dual effects of time and variables on house prices, but the factors involved were not comprehensive and the results were not representative.In 2017, Jin Haozhi found that the characteristic variables affecting the housing price include 8 variables:decoration degree, floor area ratio, property management, parking space,floor of the house, building area, house type and living facilities.In 2018, Li Chunsheng et al. added a genetic algorithm based on the backpropagation neural network to predict housing prices. The sigmoid activation function was used in the model for optimization, and a prediction result was obtained through continuous training and parameter adjustment.Therefore, the backpropagation neural network based on genetic algo-rithm has great advantages in improving the prediction accuracy of housing prices and optimizing parameter updates.In 2020,Luo Bowei and others proposed to use a multiple linear regression statistical model to predict housing prices.In order to address the multicollinearity issue,the author used the variance inflation factor to correct it.The bench mark model is optimized by using the ratio of the variance when there is multicollinearity among the house price characteristics to the variance when there is no multicollinearity, so as to improve the accuracy of predicting house prices.

Compared with previous articles, this paper summarizes the results of previous studies, including: the selection of factors affecting house prices is more comprehensive, making the results more representative; In terms of research methods, linear regression and BP neural network are used for comparative analysis and prediction. Taking relative error and matching degree as indicators, the results of the two models are evaluated to make the research results more reliable.

Table 1: Relevant literature

Auther	Document name
Nguyen Nghiep ,Cripps Al	Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks.
Malpezzi S	A Simple Error Correction Model of Housing Prices.
Gao Yuming, Zhang Renjin	Analysis of house price prediction based on genetic algorithm and BP neural network.
Li Chunsheng, Li Xiaoye, Zhang Kejia	Prediction and Analysis of House Price Based on BP Neural Network Improved by Genetic Algorithm.
Luo Bowei, Hong Zhiyong, Wang Jinyi	Application of multiple linear regression statistical model in house price prediction.

III. Data Acquisition and Quantification

3.1 *Data Sources*

At present, there are many large-scale online housing sales agencies that provide second-hand housing transaction information,such as Anjuke,Fangtianxia,Lianjia and other powerful brands.This article uses Lianjia as the data source.Compared with other online housing sales platforms,Lianjia provides more comprehensive second-hand housing information and more comprehensive relevant indicators, which is more helpful for

accurately estimating the second-hand property sale price.In addition, Lianjia is the leader of the domestic second-hand housing market.It has the largest second-hand housing database in the country, including nearly 60 million housing data in more than 20 large and medium-sized cities in the country.Therefore,in order to study the used housing price market in Tianjin,it is representative and practical to select the used housing data of Lianjia.

Crawl the Tianjin second-hand housing price data on Lianjia's official website by region, and collect as many relevant characteristic indicators as possible.In the end,a total of 4179 data records are collected in this paper,and each record contains 11 original feature fields[1],including:unit price,region,apartment type,area,orientation,floor,total floor, decoration,building time,traffic,number of follows.

3.2 *Data Quantification*

The key variables that have an impact on the second-hand housing real market price are rather intricate.This paper constructs a more reasonable relevant index system based on the existing data indicators provided by the official website of LinkedIn,and through the literature research method,referring to some documents on the analysis of the relevant factors affecting the price of second-hand housing.Finally,through collation,this paper takes the following 10 original characteristic indicators as the relevant characteristics that have an impact on the cost of used houses.

(1)Region:Second-hand houses are mainly located in nine main districts of Tianjin:Beichen District,Dongli District, Heping District, Hebei District, Hedong District, Hexi District, Hongqiao District, Nankai District and Tanggu District.When purchasing a house, consumers usually need to consider the location of the house.Whether consumers purchase houses for rigid demand or as return on investment,the factor of region is the primary factor that must be considered.Among them,the above-mentioned nine areas are represented by numbers 1-9 respectively.

(2)Transportation:The subway is a large-scale public transportation tool for urban citizens to travel.Usually, the location of the subway station is prosperous and the population density is high, so the surrounding housing resources are often very tight, and the prices are relatively high.Therefore, it is stipulated to use 1 tore present 500 meters or less from the subway, and 0 tore present more than 500 meters to the subway.

(3)House type:The house type usually refers to the main body of the house including several rooms, halls,kitchens and bathrooms.Housing type is an important factor that many home buyers will consider when conducting second-hand housing transactions, because different housing types meet different living needs and affect the comfort of life.Therefore, the second-hand house types are divided into 1 room and 0 hall and 0bath,1 room and 0 hall and 1bath,1room and 1 hall and 0 bath,1room and 1 hall and 1 bath,1 roomand 1 hall and 2bath,1room and2 hall and 1 bath,2room and 0hall and1 bath,2 room and 1hall and 1 bath,1 room and 1 hall and 2 bath,2 room and 2 hall and 2 bath,3 room and 1 hall and 1 bath,3 room and 1 hall and 2 bath,3 room and 2 hall and 1 bath,3 room and 2 hall and 1 bath,3 room and 2 hall and 2 bath 4 rooms and 1 hall and 1 bathroom,4 rooms and 1 hall and 2 baths,4 rooms and 2 halls and 1 bathroom,4 rooms and 2 halls and 2 baths,4 rooms and 2 halls and 3 baths,4 rooms and 3 halls and 2 baths,4 rooms and 3 halls and 3 baths, 5 rooms and 1 hall and 2 baths, 5 rooms and 2 halls and 2 baths, 5 rooms and 2halls and 3 baths, and total 27 types of apartments.Among them, the above house types are represented by numbers 1-27 in sequence.

(4)House area:The building area is usually expressed in square meters.General building area is one of the main factors considered by buyers,which is closely related to the house type.

(5)Orientation:The orientation of a house is one of the criteria for many people to judge whether a house is good or bad.The orientation of a house is directly related to the lighting, ventilation, noise and other aspects of the house. Because the south-facing house can keep enough sunlight, even in winter, it can enjoy the sunshine,

which is bright and warm, so many people prefer the south-facing house.Therefore, use 1 to represent second-hand houses facing south,and use 0 to represent houses facing other directions.

(6)The floor:The floors mainly include low floors,middle floors and high floors.Different floor heights bring different living experiences.Generally, the middle floors are more popular, while the lower floors and upper floors will have more inconvenience. Among them, 1, 2, and 3 respectively represent low floors, middle floors, and high floors.

(7)Total floor:Generally,the more the total number of floors of a house,the higher the floor height,and the greater the population density of the households,the more crowded the travel is,which affects the suitability of living.

(8)Furnish:House decoration is generally divided into four types: rough decoration, simple decoration, hardcover decoration and luxury decoration.In reality, the price of hardcover houses is higher than the first two, and the price of luxury houses is generally the most expensive.Different decoration conditions bring different conveniences to house consumers. Hardcover houses can be taken in directly, saving home buyers the cost of decoration.Among them, 1, 2, 3, and 4 are used to represent the four types of rough, simple, hardcover, and luxury.

(9)Building time:The time when the house was built often determines the different living comfort. Usually,the early buildings are relatively simple in terms of unit type and structure, the facilities are relatively old,and the safety of building materials is also different from that of today.

(10)Number of followers:Often good second-hand housing sources are more favored by consumers, so the more people pay attention to it,the more they can gain the trust of consumers.The results section represents the core findings of a study derived from the methods used to collect and analyse information. It presents these findings in a logical order without author bias or interpretation, preparing readers for subsequent interpretation and evaluation in the discussion section.

IV. Multiple Linear Regression Model

The source of the data and the attribute indicators that will affect the price of second-hand housing are determined above.The following data and variables will be used to establish a model of multiple linear regression.Based on the existing data,the multiple linear regression model can establish the mathematical relationship between the response variable and multiple explanatory variables, analyze the direction and magnitude of the influence of the explanatory variables on the response variable,and can make a point on the response variable under the condition of the given explanatory variable.It is feasible and reasonable to apply this model to the valuation of used housing by forecasting and interval forecasting,.

4.1 *Model Building*

From the previous section, there are a total of 4179 observation data in this paper, of which 3999 are randomly selected as the training set, and the remaining 180 are the test set data.Now use the training data to build a multiple regression prediction model for second-hand house prices.Second-hand home prices are the dependent variable in the multiple linear regression model, while other factors that influence housing prices are the explanatory variables.

Therefore,this paper will first construct a linear regression model for statistical analysis,and the model structure is as follows:

$y = \beta_0 + \beta_{1x1} + \beta_{2x2} + \beta_{3x3} + \beta_{4x4} + \beta_{5x5} + \beta_{6x6} + \beta_{7x7} + \beta_{8x8} + \beta_{9x9} + \beta_{10x10}$.

The coefficient of this formula is determined as follows.

4.2 *Model Parameter Estimation*

According to the above training set data, multiple linear regression analysis is carried out in R language software,and the regression model is shown in the figure below.

```
> fm=lm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10,data=data)
> fm

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 +
    x10, data = data)

Coefficients:
(Intercept)           x1           x2           x3           x4
  1586190.6       -626.2        779.6       -106.6      -3209.9
         x5           x6           x7           x8           x9
   -2005.7        479.8       4975.1       -785.1       7279.8
        x10
    -408.5
```

Figure 1: Results of multiple linear regression

Therefore, the multiple linear regression model for second-hand housing prices in Tianjin is y=1586190.6−626.2x_1+779.6x_2−106.6x_3−3209.9x_4−2005.7x_5+479.8x_6+4975.1x_7−785.1x_8+7279.8x_9−408.5$_{10}$.

4.3 *Model Inspection*

4.3.1 T-Test

Based on the above experimental results,the multiple linear regression model of used housing prices in Tianjin is obtained, and the t test is performed on this model to determine whether its coefficient is significant.Enter the corresponding code in the R language software to get the result as shown in Figure 2.

The t-test findings demonstrate that the p-values of the coefficients in the model are far less than 5%, and it can be considered that the explanatory variables affecting second-hand housing prices are

significant.Among them, the coefficients of region, area, orientation, floor, building time, and number of followers are negative, indicating that they have a negative impact on the dependent variable second-hand housing prices.

4.3.2 Model ANOVA

In R language,anova function is used for variance analysis,and the results are shown in Figure 3.

Figure 3 demonstrates that the p-values of most variables are far less than 5%, and the null hypothesis is rejected at the level of α =0.05.In this instance, the regression coefficient is statistically significant.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1586190.57   74846.75  21.193  < 2e-16 ***
x1             -626.18      99.21  -6.312 3.06e-10 ***
x2              779.59     108.40   7.191 7.62e-13 ***
x3             -106.61      14.91  -7.149 1.03e-12 ***
x4            -3209.92     594.60  -5.398 7.11e-08 ***
x5            -2005.74     313.14  -6.405 1.68e-10 ***
x6              479.76      31.31  15.324  < 2e-16 ***
x7             4975.09     426.91  11.654  < 2e-16 ***
x8             -785.08      37.58 -20.889  < 2e-16 ***
x9             7279.81     527.22  13.808  < 2e-16 ***
x10            -408.54      76.76  -5.322 1.08e-07 ***
```

Figure 2: Multiple linear regression coefficient t-test

```
Response: y
             Df     Sum Sq    Mean Sq  F value    Pr(>F)
x1            1 1.8458e+09 1.8458e+09   7.8241   0.00518 **
x2            1 6.1038e+09 6.1038e+09  25.8738 3.812e-07 ***
x3            1 3.5523e+10 3.5523e+10 150.5789 < 2.2e-16 ***
x4            1 2.7375e+10 2.7375e+10 116.0426 < 2.2e-16 ***
x5            1 1.0103e+10 1.0103e+10  42.8247 6.747e-11 ***
x6            1 8.1556e+09 8.1556e+09  34.5712 4.446e-09 ***
x7            1 2.8661e+10 2.8661e+10 121.4936 < 2.2e-16 ***
x8            1 1.3568e+11 1.3568e+11 575.1595 < 2.2e-16 ***
x9            1 4.5685e+10 4.5685e+10 193.6569 < 2.2e-16 ***
x10           1 6.6827e+09 6.6827e+09  28.3277 1.081e-07 ***
Residuals 3988 9.4079e+11 2.3591e+08
```

Figure 3: Model ANOVA results

4.4 *Model Prediction*

The training set data has been modeled above and passed the relevant test, so the model can be used for regression prediction of the test set data.This paper uses the R language to predict the price of second-hand houses from 180 sample data in the test set,and some results are shown in the following table.

Table 2: Comparison between predicted value and true value of multiple linear regression model

Serial number	Actual price	Predicted price
13	11849	8410
37	17129	18027
54	42471	41901
71	25130	21050
130	21160	26004

V. BP Neural Network Model

5.1 *Introduction to BP Neural Network*

In the above,the multiple linear regression model is used to fit and predict the second-hand house price in Tianjin,but the result error is large,so the more complex BP neural network model is used.In this paper,a total of 4,179 pieces of second-hand housing data are crawled, and 180 observations are extracted through stratified sampling for prediction, and the remaining 3,999 observations are used for training the model.Therefore, this section will use these 3999 training data for neural network modeling, among which 3500 pieces of data are used as training sets and 499 pieces of data are used as test sets.

5.2 *Model Building*

1.Data normalization

Generally, the initialization formula is adopted:

$$x_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

2.Input layer and output layer

This paper mainly selects 10 factors that affect the transaction price, that is, there are 10 input neurons in the BP neural network, so the number of input layer units of the network is 10. The prediction result of price is the output index,that is,the predicted price is the only output neuron.

3.Hidden layer

Finding the appropriate number of nodes in hidden layers requires multiple experiments based on experience.Choosing the number of hidden layer nodes using a standard empirical formula is:

$$m = \sqrt{i} + j + n.$$

The formula uses the letter m for the number of nodes in the hidden layer, i for nodes in the input layer, j for nodes in the output layer, and n represents a constant value between 1 and 10.

According to the above formula,the enumeration method has been used for many times,and finally the number of nodes in the hidden layer is 6.

4.Activation function

In the three-layer BP neural network model, it is necessary to select an excitation function for the calculation between hidden layer and output layer neurons.Typically, the excitation function is the sigmoid function.

5.Network parameter setting

Set the parameters of the BP neural network model to be trained.Among them, the number of training times is set to 1000 times,the learning rate is set to 0.01,and the minimum error of the training target is set to 0.001.

5.3 *Result Analysis*

5.3.1 Training Results

According to the above network settings, 10 input neurons and 1 output neuron are selected, the hidden layer node is 6, the number of network iterations is set to 1000, the learning rate is set to 0.01, and the error is controlled at 0.001.The fundamental steps of network training are depicted in the following figure[5].

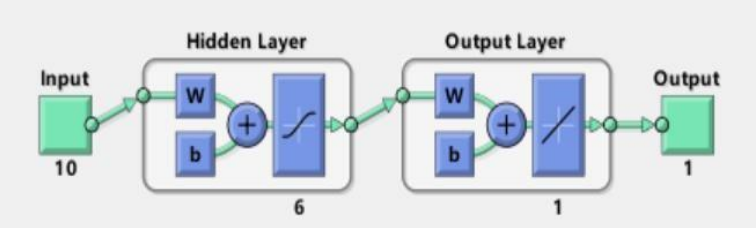

Figure 4: Neural network training diagram

Figure 5 shows the convergence results of the model network.The figure demonstrates that the network converges after 71 times of training, and the network error is 0.0051,which essentially satisfies the network error setting's requirements.The dynamic effect of network error is shown in the figure, where the ordinate is the error.We can clearly see that as the number of training increases,the difference in error between the training set and the test set steadily gets smaller, and both sets' trends are steady.When the network is trained to 71 times, the three curves essentially merge into one, and the error is stable.At this time,the model training effect is excellent.

It can be seen from Figure 6 that the training sample's goodness of fit is 91.50%,the goodness of fit of the test set is 91.40%, and the samples' overall quality of fits also reaches 91.07%.This degree of fit proves that the network training effect is excellent. This training model has high accuracy and can explain about 91% of the data samples.Excluding the influence of individual data samples,we can use this model to accurately predict house prices.

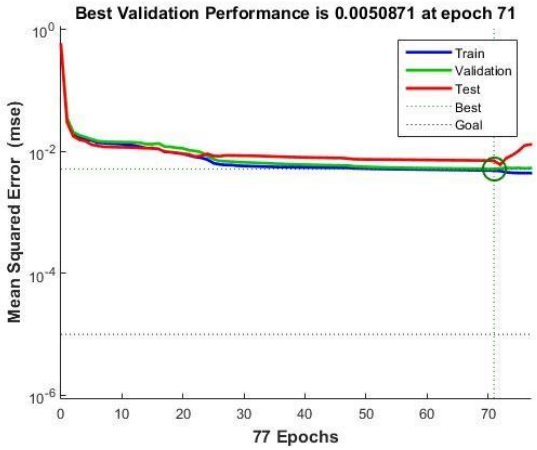

Figure 5: Neural network iteration diagram

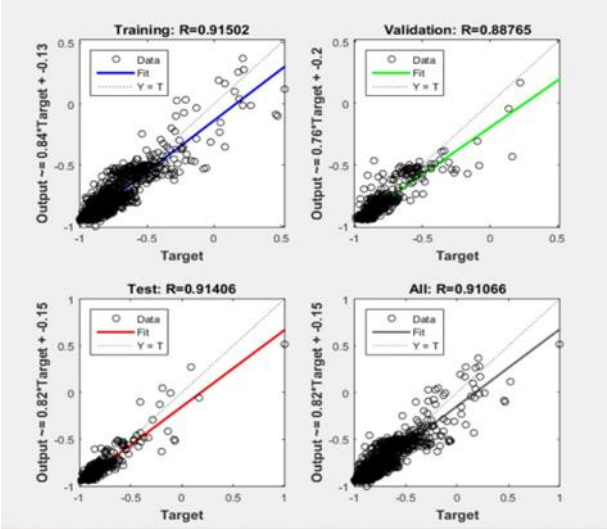

Figure 6: Neural network fitting diagram

5.3.2 Predicted Results

In the neural network model, 3999 pieces of second-hand housing data crawled are used for training the model,of which 3500 pieces of data are used as training sets and 499 pieces of data are used as test sets.The predicted value and the actual value of the test set for the BP neural network are compared in Figure 7.It can be seen from the figure that most of the errors between the predicted value and the actual value are concentrated near zero,and the fluctuations are not large,and the error is basically stable within a range, indicating that the model training results are relatively good and can be used for the next data prediction.

In order to verify the effectiveness of the above BP neural network model for predicting second-hand housing prices in Tianjin, this paper captures another 180 pieces of second-hand housing price data in Tianjin.After certain data preprocessing, the trained BP neural network model is used to predict the data set.Table 2 is a comparison between randomly selected part of the predicted data and the real data,as well as their relative error and absolute error.

After experimental research and comparison, it is found that the second-hand housing price prediction model in Tianjin based on BP neural network has a certain predictive ability,and the error interval is basically within 10%.

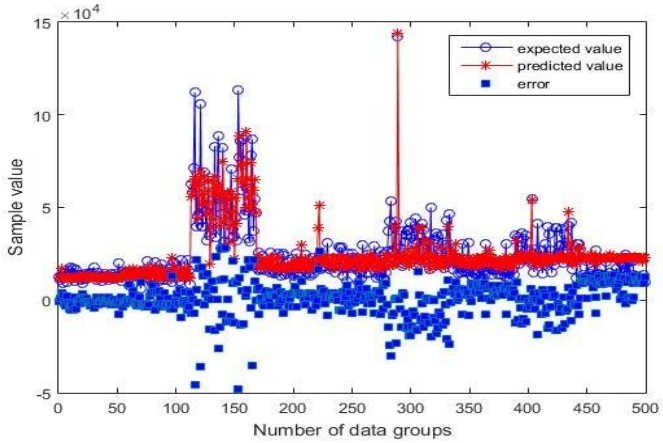

Figure 7: Comparison chart of predicted value and actual value of BP neural network test set

Table 3: Comparison chart of predicted value and actual value of BP neural network test set

Serial number	Actual price	Predicted price	Absolute error	Relative error
11	11761	12410	649	5.52%
24	12930	12990	60	0.46%
52	49630	49580	50	0.10%
75	23272	25050	1778	7.64%
94	19696	18630	1066	5.41%
118	37774	33660	4114	10.89%
128	26263	26660	397	1.51%
154	28434	27000	1434	5.04%
173	13716	13810	94	0.69%

VI. Comparative Analysis of Model Prediction Results

The prediction result of the model can be as close as possible to the real value of used housing, butit can not be completely equal,so the prediction error must exist.At present,there is no strict regulation to explain the error standard in the used house price prediction.In some literature, the margin of error is specified as ±10%.Although such a determination is relatively extensive,it is important and necessary to determine a reasonable error range in the housing price forecasting work.The above-mentioned valuation error range is worth learning.

In this paper, two indicators are selected to compare and evaluate the prediction accuracy of the two models.Part of the comparison results are shown in Figure 8.

1.Relative error.

2.Suitability:The suitability is equal to the predicted value divided by the real value.

This indicator can reflect the matching between the predicted value and the actual value.The closer the matching degree is to 1,the better the prediction effect of a single sample.

Through screening, it can be found that the relative error of the multiple linear regression model falls within 5%, accounting for 23%, and 34% between 5% and 10%, while the BP neural network model is 32% and42% respectively.For the matching degree, the proportion of multiple linear regression model falling between 0.9-1.1 and 0.8-1.2 is 19.7% and 46.7% respectively, while that of BP neural network model is 32% and 69% respectively.This shows that compared with the multiple linear regression model, the BP neural network algorithm can quickly and deeply discover the laws between the data and make accurate predictions.At the same time, it also has a good expansion effect, which is a method worthy of promotion and application.

test sample serial number	practical value	MLR		BP	
		relative error	suitability	relative error	suitability
11	11761	8.01%	1.31	5.52%	1.06
24	12930	1.00%	1.12	0.46%	1.00
52	49630	0.28%	0.98	0.10%	0.99
75	23272	4.89%	0.72	7.64%	1.08
94	19696	11.27%	1.31	5.41%	0.95
118	37774	4.52%	0.95	10.89%	0.89
128	26263	3.29%	1.22	1.51%	1.02
154	2843	7.68%	1.12	5.04%	0.95
173	13716	6.24%	0.63	0.69%	1.01

Figure 8: Comparative analysis of the prediction results of the two models

VII. Discussion

7.1 *Implications*

1.Theoretical Implications
A single model cannot accurately predict a complex system, so this paper uses two models for comparative analysis and prediction. The results of the two models are evaluated with relative error and matching degree as indicators. In addition,this paper takes the large second-hand house website of LinkedIn as the crawling object, and the number of data records collected is large enough, and according to the literature and website entries, it determines several factors that affect the second-hand house price.
2.Managerial Implications
The regulation mechanism of China's real estate market is not perfect. The market is greatly affected by national policies and banks, and the research on the second-hand housing price index is lacking. Only with the constant adjustment of China's real estate market mechanism and the constant stability of national policies can the housing price index system gradually enter a stable stage.The model established in this paper can better realize the prediction of second-hand house prices in Tianjin, provide scientific evaluation prices for developers to develop real estate and residents to purchase houses, and ensure that both sides of real estate transactions can effectively promote the industry, which has practical application value.

7.2 *Outlook*

There are still many problems in the research of this paper, so this paper makes the following outlook:
1.What is more important in housing price prediction is whether the dimension of the data set is large enough.Crawling a sample with a larger data capacity for prediction may make the prediction result more refined.
2.In terms of variable selection:The cost of used housing is influenced by numerous factors.In addition to the listing data information on the website, a batch of survey data (such as property management,community cultural construction, community environment, etc.) and a batch of socioeconomic data (such as economic development expectations,housing policy,consumer psychology,etc.)should be added.

In the future, with the continuous liberalization of the birth policy, the number of new born population will continue to grow, and the number of families with the need to buy and change houses will continue to increase. Therefore, the development space of second-hand housing market is relatively large. In the future, the paper research should focus on the demand, supply, operation characteristics, acquisition ability, industrial chain and value chain of the second-hand housing industry, integrate the multi-level data and information resources of the industry, market, enterprise and user, provide customers with in-depth market research reports of the second-hand housing industry, discover investment value and investment opportunities, and improve management and operation capabilities.

References

[1]. Candas E, Kalkan S B, Yomralioglu T.Determining the factors affecting housing prices[C].FIG Working Week,2015: 17-21.
[2]. San Ong T.Factors affecting the price of housing in Malaysia[J].J Emerg Issues Econ Financ Bank,2013,1:414-429.
[3]. Qian Ying, Fang Xiunan.Multiple linear regression model and its application[J].China Science and Technology Information,2022 (04): 73-74.
[4]. Xu X, Zhang Y.Second-hand house price index forecasting with neural networks[J].Journal of Property Research,2021:1-22.
[5]. Liu L, Chen J, Xu L.Realization and application research of BP neural network based on MATLAB[C].International Seminar on Future BioMedical Information Engineering IEEE,2008:130-133.
[6]. Luo Bowei,Hong Zhiyong,Wang Jinyi.Application of Multiple Linear Regression Statistical Model in Housing Price Prediction[J].Computer Age,2020(06):51-54.
[7]. Xu X,Zhang Y.House price forecasting with neural networks[J]. Intelligent Systems with Applications,2021,12:200052.
[8]. Yuzhao Jiang,Chen Liang.Research on Factors Affecting Housing Prices in Tianjin Based on Gray Relational Analysis[J].Modern Marketing (Early Journal),2022(10):94-96.

[9]. Zhou Lin,Hao Renjie,You Quansheng.Housing Price Prediction Model Based on Multilayer Feedforward Neural Network Algorithm[J].China Collective Economy,2022(23):42-44.
[10]. Li Jiaxuan.Research on second-hand housing transaction price prediction based on machine learning model[D].Yanbian University,2022.
[11]. Zhao Shizhe.Empirical Research on Second-Hand Housing Price Evaluation in Wuhan Based on Big Data Analysis[D].Central China Normal University,2022.
[12]. Tu Jin,Jiang Wanchen,Leng Zhengxing.Differentiation Research on Factors Influencing Urban House Prices in my country——Based on the Analysis of Big Data in Second-Hand Housing Market in Chengdu[J].Price Theory and Practice,2021(10):75-78.
[13]. Zhou Hao.Price Forecast of Second-Hand Housing in Beijing and Analysis of Influencing Factors[D].Beijing University of Technology, 2021.
[14]. Xia Qiuyue,Lu Jie,Liu Chaojie,Jia Ximei.Research on the characteristic price of second-hand housing in Zhengzhou Zhongyuan District under the background of big data[J].Regional Research and Development,2020,39(01):83-88.
[15]. Cao Yafei,Guo Yaping.Improved BP Neural Network House Price Prediction Based on Genetic Algorithm——Taking Chongqing as an Example[J].Shanxi Agricultural Economics,2018(14):35.
[16]. C. Zhou, X. Li, Y. Ren, et al. How do fairness concern and power structure affect competition between e-platforms and third-party sellers? IEEE Transactions on Engineering Management, 2023, DOI: 10.1109/TEM.2023.3262318.
[17]. C. Zhou, M. Leng, Z. Liu, et al. The impact of recommender systems and pricing strategies on brand competition and consumer search. Electronic Commerce Research and Applications, 2022, 53: 1-15.
[18]. C. Zhou, W. Tang, R. Zhao. Optimal consumption with reference-dependent preferences in on-the-job search and savings. Journal of Industrial and Management Optimization, 2017, 13(1): 503-527.
[19]. C. Zhou, W. Tang, R. Zhao. Optimal consumer search with prospect utility in hybrid uncertain environment. Journal of Uncertainty Analysis and Applications, 2015, 3(6): 1-20.
[20]. P. Chen, R. Zhao, Y Yan, et al. Promoting end-of-season product through online channel in an uncertain market. European Journal of Operational Research, 2021, 295(3): 935-948.
[21]. J. Yu, J. Zhao, C. Zhou, et al. Strategic business mode choices for e-commerce platforms under brand competition. Journal of Theoretical and Applied Electronic Commerce Research, 2022, 17(4): 1769-1790.