Real estate price based on ARMA-ANFIS Research on prediction model

Kai Zhou[a, *], Hang Ji, Haoming Zhang, Ping Liu

Army Academy of Artillery and Air Defense Hefei, Anhui, 230031, China

[a]1763331669@qq.com

*Corresponding author

Keywords: ARMA, adaptive fuzzy neural network, time series

Abstract: In this paper, the principle theory of ARMA algorithm and adaptive fuzzy neural network is expounded, and a time series prediction method based on ARMA adaptive fuzzy neural network is proposed. The research shows that good prediction results have been obtained in terms of fitting and detection accuracy, which provides a new method for real estate evaluation of time series.

1. Introduction

With the rapid development of China's comprehensive economic strength, the real estate industry has also developed a rapid development period, but the real estate industry in the big economic background still shows its fragile side. In 2006, real estate prices have skyrocketed. But with the advent of the 08 year financial crisis, the national real estate price has entered a deep water period, highlighting the chill. After 2010, with the financial crisis in the past, real estate prices were still high. The national real estate price is directly related to the people's happiness index. Facing the nonlinear real estate price trend, it is an important issue to further improve the healthy development of the society to accurately assess the real estate price in China and make it develop in a better direction.

There are many factors that affect the price of real estate in China, such as national income, total economic aggregate, national policy factors and so on. There is a complex relationship between these factors and real estate. Scholars at home and abroad have adopted many forecasting methods to try to grasp the rules. In general, many methods can be divided into two categories: one is the traditional regression forecasting model, and Liu Dong, such as multi factor regression and state space model, predicts the trend of real estate. Guirguis and other methods use Kalman filtering and exponential smoothing to evaluate the price of the real estate. The two is the prediction model of modern development, such as Zhang Weicai. The real estate price is predicted by the rough set BP neural network, and the prediction of real estate is predicted by the time series prediction by Chen Jichun. Hu Xiaolong uses the Elman recurrent neural network to predict the real estate price, and Li Daying uses the rough set and the wavelet neural network model to predict the real estate price. Zhao Jun and others used Logistic regression model to predict, Ouyang Jiantao and other nonlinear gray prediction models were used to predict.

However, there are many problems in many methods, such as too high precision, BP gradient descent into local minimum and poor fault tolerance. This paper uses a logistics and adaptive fuzzy neural reasoning system (ANFIS) to predict real estate price, in order to achieve better prediction effect.

2. Prediction model related theory

2.1 An overview of the ARMA model

ARMA algorithm is an effective method to predict time series. The. ARMA algorithm can be explained by the lag term of variable U itself and random error term. Under the premise of unknown data mode, the model of. ARMA model suitable for data investigation has 3 basic types: the

autoregressive (AR) model and the moving average (MA) model and autoregressive moving average (ARMA) model.

(1) Autoregressive model AR (P)

$$u_t = c + a_1 u_{t-1} + a_2 u_{t-2} + ...a_p u_{t-p} + \varepsilon_t, t = 1,2,3....., T \tag{1}$$

C is a constant; P is the order of AR model; ε_t is a white noise sequence with σ^2 mean value of 0 and a variance of a; it is a coefficient of AR model.

(2) Moving average model MA (q)

$$u_t = \mu + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + ...b_p \varepsilon_{t-p} + \varepsilon_t, t = 1,2,3....., T \tag{2}$$

It is a constant coefficient, b is a coefficient of MA, and ε_t is a white noise sequence with a mean of 0 and a variance of σ^2.

(3) Autoregressive moving average model ARMA (P, q)

$$\begin{aligned} u_t &= \mu + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + ...b_p \varepsilon_q + \varepsilon_t + a_1 u_{t-1} + \\ & a_2 u_{t-2} + ...a_p u_{t-p}, t = 1,2,3....., T \end{aligned} \tag{3}$$

The upper form is a combination of the above two models, p=0, ARMA (0, q) =MA (q); q=0, ARMA (P, 0) =AR (P). In practical applications, it is necessary to analyze the value of P and Q with the actual data to make the model stable. The autocorrelation coefficient of ARMA (P, q) model can be regarded as a mixture of autocorrelation function of AR (P) model and autocorrelation coefficient of MA (q) model.

2.2 Fuzzy reasoning related theory

The fuzzy set is first introduced by the University of California at Berkeley Zadeh in 1965. The fuzzy set is corresponding to the exact set. It is easy to understand that the exact set describes a problem of whether an apple is red, and the fuzzy set describes the degree of an apple red (an element in a class of similarity). It's full-red, redder or reddish.

The elements in the fuzzy set are corresponding to a subordinate value through the membership function. If the fuzzy set Q (x) is discrete, then the set Q (x) can be expressed as:

$$(Q\ （x_i) = \sum_{i=1}^{n} \mu_Q(x_i) / x_i \tag{4}$$

Continuous can be expressed as

$$Q(x) = \int_x M_Q / x \tag{5}$$

And FLC was developed by Mamdani and Assilian in 1975. FLC modeling can be divided into two kinds of fuzzy modeling and precise fuzzy modeling, and the adaptive fuzzy neural inference system is based on the latter. Among the many methods in precise fuzzy modeling, the Takagi-Sugeno method has been widely applied [1].

2.3 Adaptive fuzzy neural system

The adaptive fuzzy neural system is a multilayer feedforward network. In general, it describes the complex relationship between the input and output, each input can be expressed by N language terms, and the language item corresponds to the membership function. We use Di to represent the bottom edge of the membership function distribution triangle (i=1, 2....N), the input of the system is M, then the rule base of the system has NM kinds of input combinations at most, and the output of each rule is as follows:

$$y^i = a_0^i + a_1^i x_1 + ... + a_n^i x_n \tag{6}$$

An adaptive fuzzy neural system is designed based on the fuzzy set theory of neurons. An ANFIS structure with two inputs with two language terms can be described as shown in the following diagram.

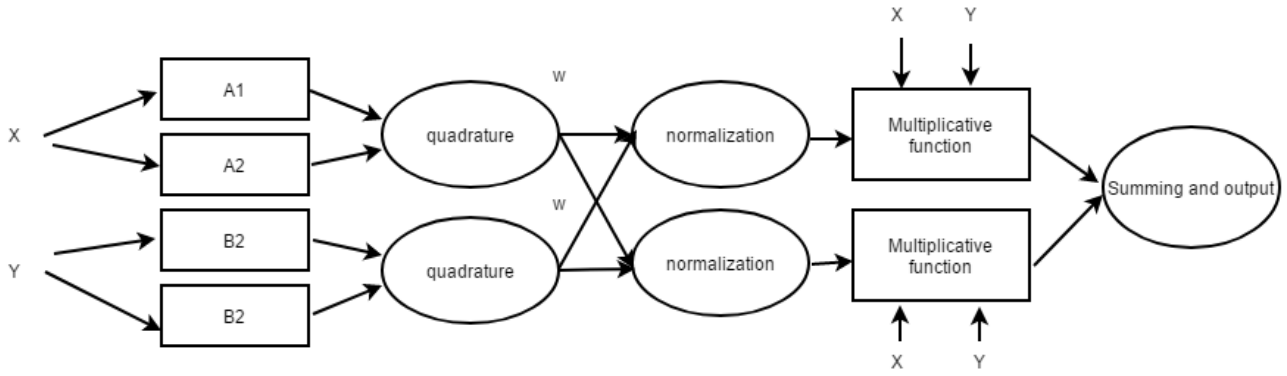


Figure 1. ANFIS

The important functions of each node are as follows:

The first level: find out the corresponding membership values of fuzzy sets after input russification and move them to the next level.

The second level: multiplying the incoming membership value 22, as the output intensity of the corresponding node, is expressed by me.

The third level: normalization of the excitation intensity that activates incoming rules.

The fourth level: the activation intensity $\overline{\omega}$ after the specification is multiplied by the output y of the corresponding rule [2].

The fifth layer: $\overline{\omega}$ y and output for each activation rule

$$F_{out} = \sum \overline{\omega}_i y^i \tag{7}$$

The mapping of input and output of an adaptive fuzzy neural system is determined by the coefficients of the bottom edge D of the membership function and the expression (3). ANFIS is based on the NN, so we can get an error function. Because the variable is the above two factor, we can train the batch training by the gradient descent or the genetic algorithm. The parameters of the training model are optimized to improve accuracy.

3. Data modeling

3.1 Training sample acquisition

The real estate price is an economic time series. According to the reliability of the data sources, the data of 2002-2017 years in Beijing are selected to be analyzed. The data are from the website of the National Bureau of statistics of China. Before data preprocessing, we need to analyze the factors that affect the price of real estate. By employing experts, the author makes a brief introduction to the factors that affect the real estate price assessment, and selects the per capita GDP, per capita disposable income and the completed area as the influencing factors. The training data is shown in the following table [3].

Table 1. Training data

	year	per capita income disposable	GDP	floor space completed	average price
1	2002.00	12463.90	4315.00	9697.70	3241.00
2	2003.00	13882.60	5007.20	11262.20	4532.00
3	2004.00	15637.80	6033.20	13457.80	6343.00
4	2005.00	17653.00	6970.00	14391.00	7812.00
5	2006.00	19977.50	8117.80	14356.10	9865.00
6	2007.00	21988.70	9846.81	14535.58	11553.26
7	2008.00	24724.90	11115.00	14145.21	12418.00
8	2009.00	26738.50	12153.03	14302.94	13799.00
9	2010.00	29072.90	14113.58	15572.10	17782.00
10	2011.00	32903.00	16251.93	18041.56	16851.95
11	2012.00	36468.80	17879.40	19306.84	17021.63
12	2013.00	43910.00	19801.00	35292.40	18553.00
13	2014.00	43910.00	21331.00	13095.00	18833.00
14	2015.00	48458.00	23015.00	13089.80	22633.00
15	2016.00	52530.00	25669.00	13089.80	27497.00
16	2017.00	57230.00	28000.00	12608.60	66165.00

3.2 Forecast process

The prediction process of this paper is mainly. After data preprocessing to get the above data, three factors are predicted by ARMA regression. As a sample, the value of all factors in 2017 is predicted before 2016, and the adaptive fuzzy neural network is trained by the whole data from 2002 to 2016. Then the predicted three factors are input to the adaptive fuzzy system and the predicted values are compared with the actual values in 2017. The flow chart of the algorithm is as follows:

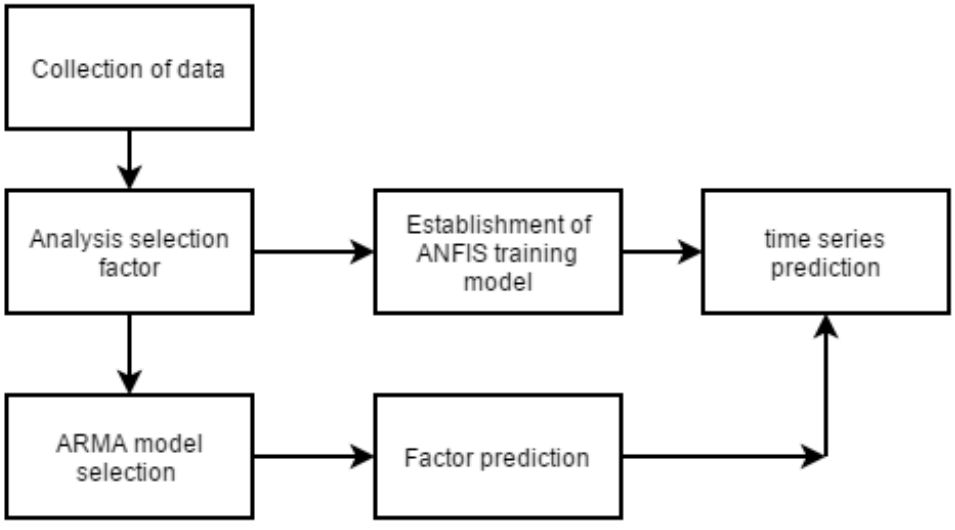


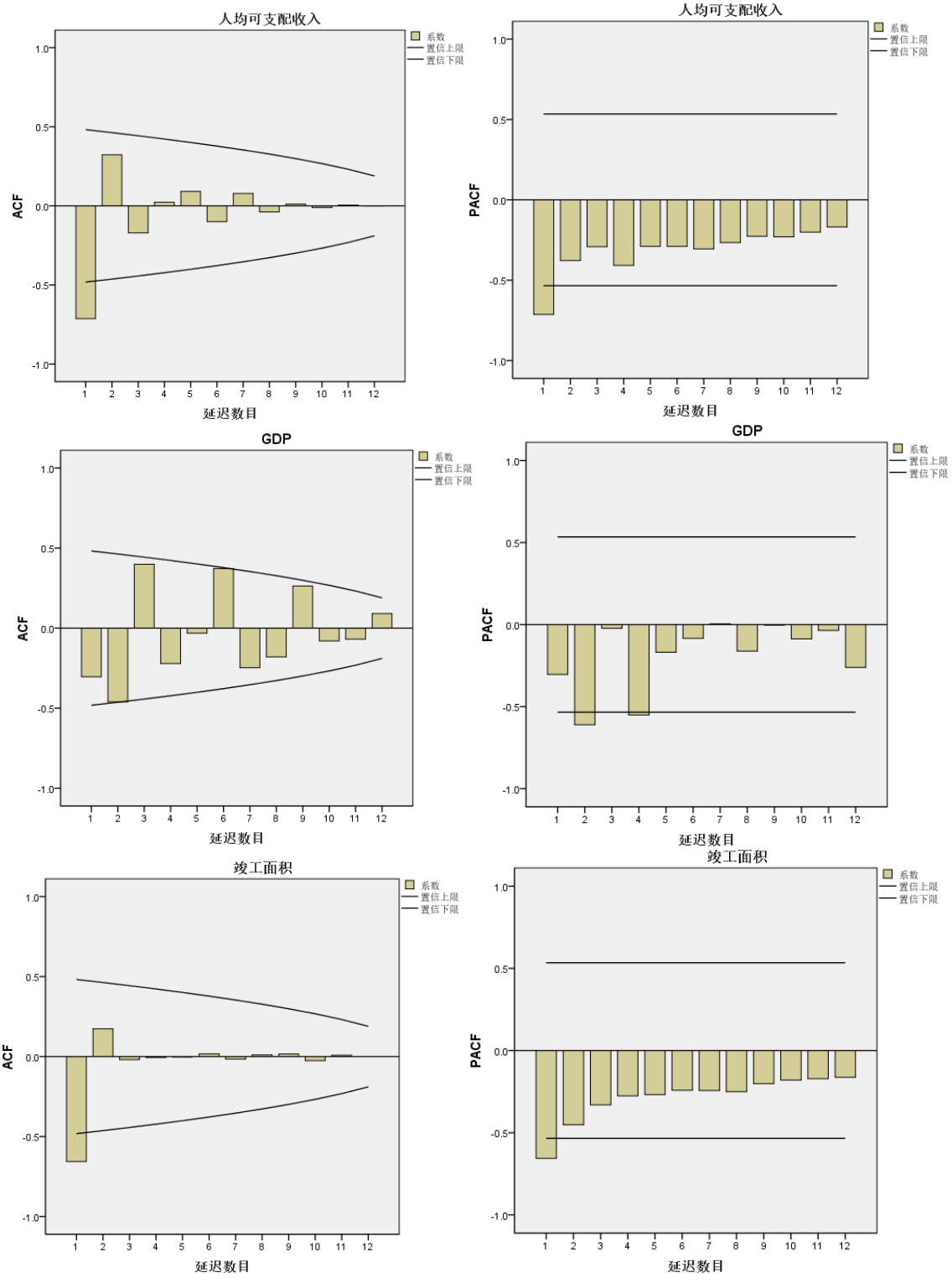
Figure 2. Algorithm flow chart

3.3 Model building

(1) First, we determine the ARMA model. From the characteristics of the time series model, we can see that the AR, MA and ARMA model should be a stationary random sequence with zero mean value. However, the actual modeling object often includes both the stationary random part and the determined nonrandom component [4]. Therefore, in the time series modeling, first of all, we need to test the sequence of the observation data to make the nonstationary data sequence into a stationary random sequence with zero mean. From table 1, we can see that three influence factors have an upward trend and a series of unstable characteristics. A non-stationary sequence with a trend item can

always change to a stationary sequence after several times of difference. Through the above analysis product test, it is proved that the three influence factors and the average price can be achieved by two order differential convergence.

An important symbol of the ARMA model is that the autocorrelation and partial correlation function converge to zero in negative exponential velocity, so the autocorrelation and partial correlation function of the sample should also tend to zero, and the determination of the P and Q values depends on the convergence rate of the autocorrelation function and the biased function [5]. The autocorrelation function and partial correlation function of the three influencing factors and the average price are analyzed as follows.



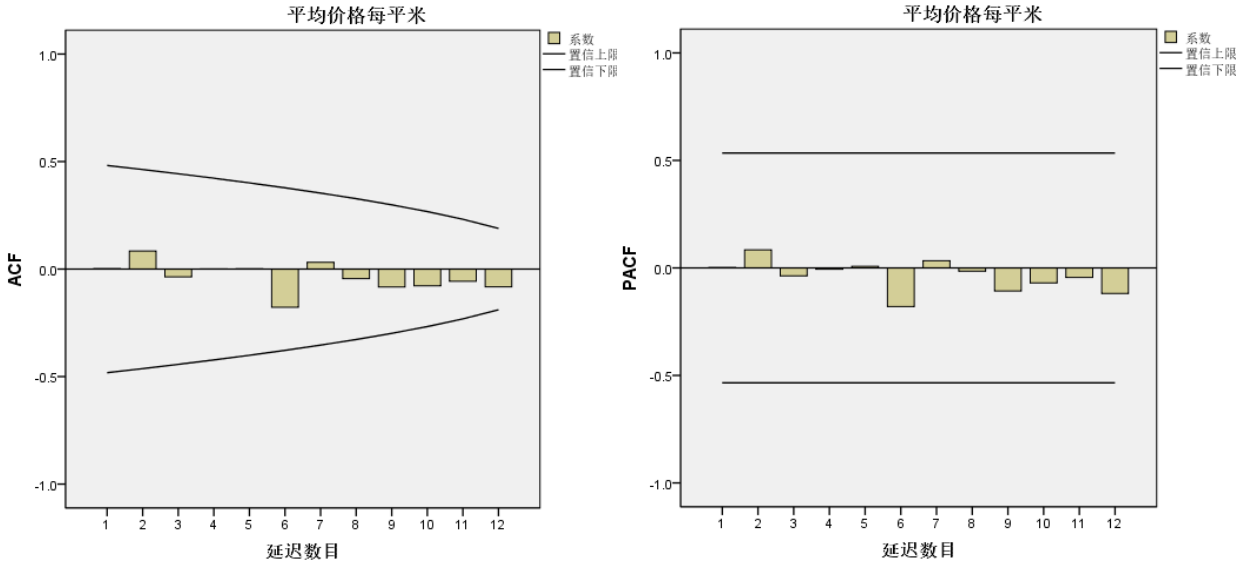


Figure 3. Autocorrelation and partial correlation graphs

The autocorrelation and partial correlation graph of the sample data shows that the autocorrelation function value AC of per capita disposable income attenuates to a smaller value after 4 steps, and the partial correlation function value PAC attenuates to a smaller value after 2 steps, and selects ARIMA (4, 2, 2) through the order of AIC to predict. Similarly, GDP selects ARIMA (5, 2, 3), the completion area is ARIMA (5, 2, 2), and the average price selects ARIMA (5, 2, 5) for prediction [6].

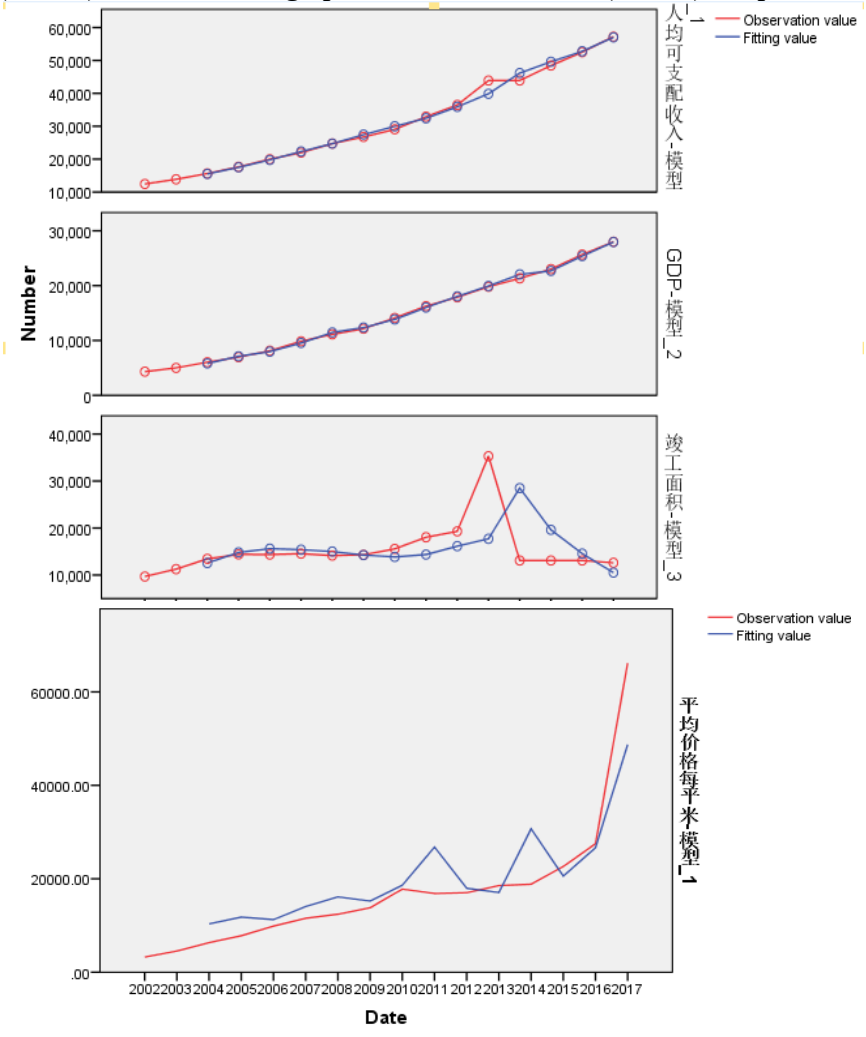


Figure 4. Fitting diagram

Table 2. Predictive value

percapitainco medisposabl e	GDP	floorspaceco mpleted
57230.00	29050.00	10028.60

(2) Using the above fuzzy neural network to forecast the real estate price in China. A fuzzy neural network system with 3 inputs and 1 outputs is defined. Its input parameters are per capita disposable income, GDP and completion area, and the output parameters are average prices. The raw data from 2002 to 2017 are selected as training sample signals and test data. In order to eliminate the influence of dimension, normalization of raw data is made to change between -1 and 1. When the simulated annealing method was iterated 100 times, the error reached 0.02. The following diagram is a membership function and a network structure diagram.

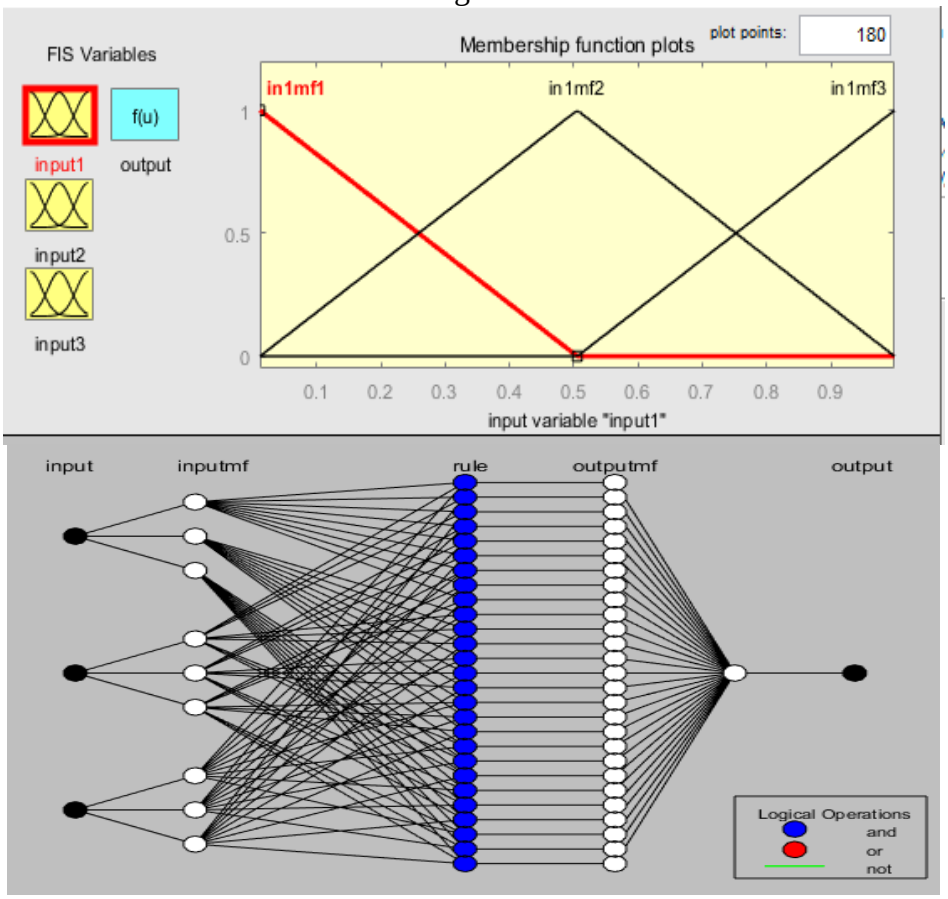


Figure 5. Membership function and network structure

4. Predictive argumentation

After establishing the ANFIS model, the data of the training set is input and checked to get Figure 5. In the input table 2, the worth to the predicted results and the actual value are compared to the table 4, and then the value input prediction model of the three predictors will be obtained. It is found that the difference between the outputs is 67204.35 and the actual value is small, and the accuracy of the ARMA prediction is higher than that of the traditional simple prediction. It can be seen that the combined prediction method proposed in this paper is very effective.

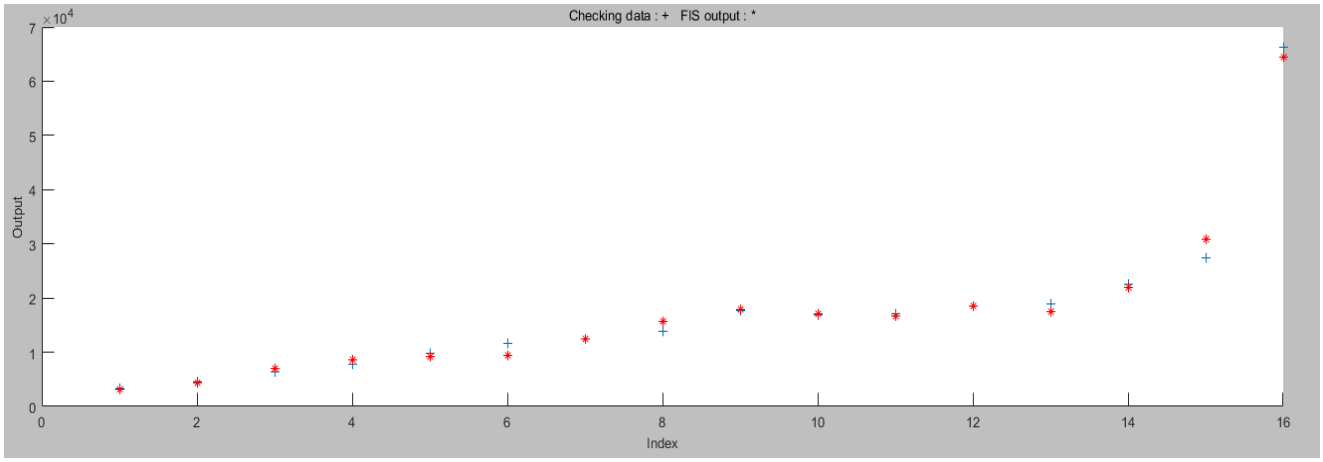


Figure 6. Comparison of predicted values

Table 3. Comparison

	Actual	ARMA	ARMAANFIS
1	66165.00	55015.99	67204.35

5. Summary

This paper skillfully combines ARMA time series prediction with adaptive neural network to predict the price of real estate and compare it with the actual value. As a model of time series prediction, it not only provides a new method for real estate price prediction, but also plays an important role in the prediction of other social economic series.

References

[1] Hossein Mohammad Khanlou. “Prediction of Fouling in Condenser Based on Fuzzy Stage Identification and Chebyshev Neural Network” [J] De Gruyter 2013-2 10.2478/msr-2013 - 0017.

[2] Junhui Luo,“ Lane-changes prediction based on adaptive fuzzy neural network IOP 2018-3 10.1088/1755 - 1315/108/3/032034

[3] D.K. Pratihar. “Soft computing” [M]. sciencep, 2008.5

[4] Zhou Hongbao. “Application of Neural Network based on rough set in pattern recognition” [J]. Computer Engineering and Design 2007 28 / 22: 5464 - 5467.

[5] Lai Hui-cheng. “An improved BP neural network algorithm and its application” [J]. Computer simulation: 2007: 75 - 111.

[6] Zhang Changshui. “Artificial Neural Network and Simulation Evolutionary Computation”. Beijing: Tsinghua University Press, 2000.