Housing price forecast based on rough-set extreme learning machine

Benxue Wang*

Institute for Economic and Social Development, Quzhou University, Zhejiang, 324000, China

Received 1 March 2014, www.cmnt.lv

Abstract

The work, based on various factors affecting housing price in 31 provinces cities as research object, firstly adopted rough set theory to reduce those factors. Then, the main reduced influence factors were used as the input of extreme learning machine. On such basis, the housing price forecast model based on rough-set extreme learning machine was ultimately established. According to the simulation results, the algorithm in this work has good prediction effect, and its prediction precision is higher than that of BP neural network and RBF neural network. Therefore, this algorithm, with a certain practical and theoretical value, can be promoted to other areas for predication and classification.

Keywords: rough set theory, neural network, extreme learning machine, real estate, forecast error

1 Introduction

At present, Chinese real estate industry is booming rapidly, which is beneficial to the sustainable development of current economy and the increase of GDP. However, housing price has become the focus of government and the burden of people's life due to the high price of current commodity house. Too-high housing price becomes a huge threat to social stability and harmony. Therefore, studying the influence factors of housing prices contributes to the formulation and implementation of policies, with positive significance to the control of housing price.

Many famous experts and scholars have conducted researches on housing price. In 2002, Qiao Zhimin and his team did some researches on the fluctuation of housing price, finding out that the fluctuation of production cost affected the fluctuation of real estate price. In 2004, Ju Hong analyzed the influence of the holding cost and production cost on the price of commodity house. In the same year, Ping Xinqiao and Chen Minyan empirically analyzed the relationship of real estate investment, commodity house sale price, land price and government credit functions.

In 2004, Shen Yue and Liu Hongyu, adopting sample regression and city annul dummy variable, put forward the hypothesis that commodity house was not in conformity with the efficient market in China.

In 2007, Duan Zhongdong analyzed the relationship between real estate price and inflation output based on real estate data of China, discovering that inflation output contributed to the rising of housing price.

In 2012, Peter Wallstrom and his team applied support vector machine algorithm to real estate price evaluation, and the simulation effect was superior to the effect of other methods.

In 2013, W.K.Wong,Z.X.Guo and some other experts combined extreme learning machine with intelligent algorithm, and put forward housing price forecast model based on genetic-algorithm extreme learning machine.

2 Combined algorithm of rough set and extreme learning machine

2.1 ALGORITHM IDEA

The work combined rough set with neural network due to their strong advantage complementary. Then, extreme learning machine neural network was used to forecast housing price after the reduction of influence factors of housing price. Following that, a comparison analysis was conducted to extreme learning machine neural network and BP neural network and RBF neural network.

After the analysis of sample data, an initial information table was generated according to the known domain knowledge. Then, an appropriate discrete method was used to discretize continuous features, and rough set theory was adopted to achieve a quick feature reduction for the data. The reduced features were considered as input layer neurons. In the end, extreme learning machine neural network was applied to train and forecast the processed data. The process is shown as Figure 1.

**Corresponding author's* e-mail: wbxdx0516@sina.com

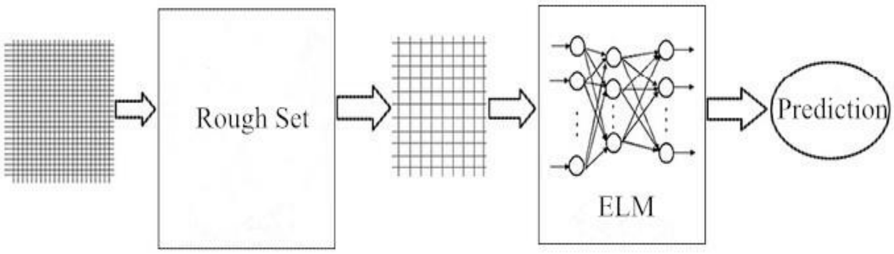


FIGURE 1 Process Diagram

2.2 ALGORITHM PROCESS

1) Discretization of Continuous Features: Continuous variables should be discretized before applying rough set method to data analysis. In essence, discretization is to divide N-dimensional space formed by condition feature into limited set of regions through selected breaking point, thus obtaining the same decision value of objects in each region.
2) Formation of Decision Tables: Quantized condition features and decision features were used to make a two-dimensional table with each line describing one object and each column showing one feature of corresponding object.
3) Reduction of Features: The reduction process of decision table feature is to eliminate unnecessary condition features of decision table system and figure out the decision principles of condition features in reduction to decision features.

Input: condition feature set $C = \{Y_{11}, Y_{12}, \ldots, Y_{53}\}$, decision feature set $D = \{d\}$; output: reduction set of one feature REDU.

Step 1: Figure out the D positive domain $POSC$ (D) for condition feature C;

Step 2: As for feature $Y_{ij} \in C$, calculate and eliminate the D positive domain $POSC\backslash\{Y_{ij}\}$ (D) of its condition features subset $C\backslash\{Y_{ij}\}$;

Step 3: If $POSC\backslash\{Y_{ij}\}$(D) $=POSC$ (D), then feature Y_{ij} is unnecessary for decision feature D. Thus, it means $C = C\backslash\{Y_{ij}\}$ at this moment.

Go back to step 2; otherwise, the reduction of output feature REDU = C.

4) Reduction of Object: Eliminate inconsistent and redundant objects in data. Inconsistent objects are those with same condition feature but different decision features. Redundant objects are those with same condition features and same decision features.
5) Determination of Neural Network Model: The work adopted extreme learning machine neural network.
6) Network Learning and Testing: Choose corresponding training data and features from continuous features decision table according to neural network model input to train network and test it with corresponding test samples.

2.3 EXTREME LEARNING MACHINE NEURAL NETWORK

For N different samples $(\boldsymbol{x}_i, \boldsymbol{t}_i)$ where $\boldsymbol{x}_i = [x_{i1}, x_{i2}, ..., x_{in}]^T \in \mathrm{R}^n$ and $\boldsymbol{t}_i = [t_{i1}, t_{i2}, ..., t_{im}]^T \in \mathrm{R}^m$, when the node number in one hidden layer is $\tilde{N}$, the unified SLFN model of activation function $g(x)$ will be:

$$\sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g_i(\boldsymbol{x}_j) = \sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g(\boldsymbol{a}_i \cdot \boldsymbol{x}_j + b_i) = \boldsymbol{t}_j, \quad j = 1, ..., N, \quad (1)$$

where $\boldsymbol{a}_i = [a_{i1}, a_{i2}, ..., a_{in}]^T$ refers to the input weight of node connecting the i-th hidden layer; b_i the bias of i hidden layer nodes; $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{im}]^T$ the output weight of node connecting the i-th hidden layer; $\boldsymbol{a}_i \cdot \boldsymbol{x}_j$ the inner product of $\boldsymbol{a}_i$ and $\boldsymbol{x}_j$. The activation function can be "Sigmoid", "Sine" or "RBF".

The matrix form of N above functions can be: $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$, where

$$\mathbf{H}(\boldsymbol{a}_1, ..., \boldsymbol{a}_{\tilde{N}}, b_1, ..., b_{\tilde{N}}, \boldsymbol{x}_1, ..., \boldsymbol{x}_N) = \begin{bmatrix} g(\boldsymbol{a}_1 \cdot \boldsymbol{x}_1 + b_1) & \cdots & g(\boldsymbol{a}_{\tilde{N}} \cdot \boldsymbol{x}_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(\boldsymbol{a}_1 \cdot \boldsymbol{x}_N + b_1) & \cdots & g(\boldsymbol{a}_{\tilde{N}} \cdot \boldsymbol{x}_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m}, \quad \mathbf{T} = \begin{bmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_N^T \end{bmatrix}_{N \times m}, \quad (2)$$

$E(\mathbf{W})$ is the square sum of error between expected value and actual value. The problem solution is to figure out the optimal weight $\mathbf{W} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\beta})$, making cost function $E(\mathbf{W})$ minimum. Its mathematical model can be expressed as:

$$\begin{aligned} &\underset{\mathbf{W}=(\boldsymbol{a},\boldsymbol{b},\boldsymbol{\beta})}{argmin}\, E(\mathbf{W}) = \underset{\mathbf{W}=(\boldsymbol{a},\boldsymbol{b},\boldsymbol{\beta})}{argmin} \|\boldsymbol{\varepsilon}\|^2 \\ &s.t. \sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g(\boldsymbol{a}_i \cdot \boldsymbol{x}_j + b_i) - \boldsymbol{t}_j = \boldsymbol{\varepsilon}_j, \quad j = 1, ..., N \end{aligned}, \quad (3)$$

where $\boldsymbol{\varepsilon}_j = [\varepsilon_{j1}, \varepsilon_{j2}, ..., \varepsilon_{jm}]$ is the error of the j-th sample.

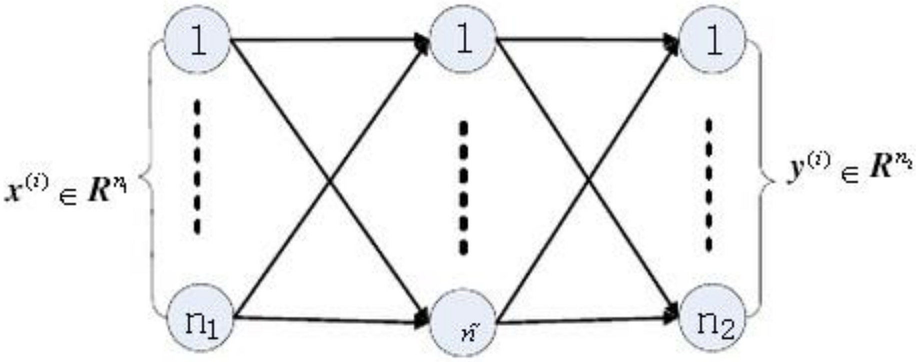


FIGURE 2 Extreme Learning Machine Schematic Diagram

3 Data sources

Data in this work, from *China Statistical Yearbook 2012*, cover all influence factors of housing price in 31 provinces, including building completion cost, land purchase expense, per capita disposable income, per capita GDP, population density, construction cost, consumer price index, land transaction price index and some other indexes.

There are various factors affecting housing price. Thus, this work firstly reduced those factors through rough set theory. Then, the reduced influence factors as the input and housing price as the output of extreme learning machine were applied to establish training model and forecast housing price. There are two sets of samples to validate the predication ability of extreme learning machine network: 31 training samples and 5 testing samples.

After rough-set reduction, the main factors affecting housing price refer to building completion cost, land purchase expense, per capita disposable income, per capita GDP and population density.

Figures 3-5 showed the predication results of extreme learning machine. Figure 3 represented the housing price forecast result of training samples, indicating a good predication and strong generalization ability. Figure 4 showed the forecast results of testing samples, mainly validating the effectiveness and correctness of extreme learning machine. It could be seen from the predication-error graph that the algorithm in this work has good prediction effect.

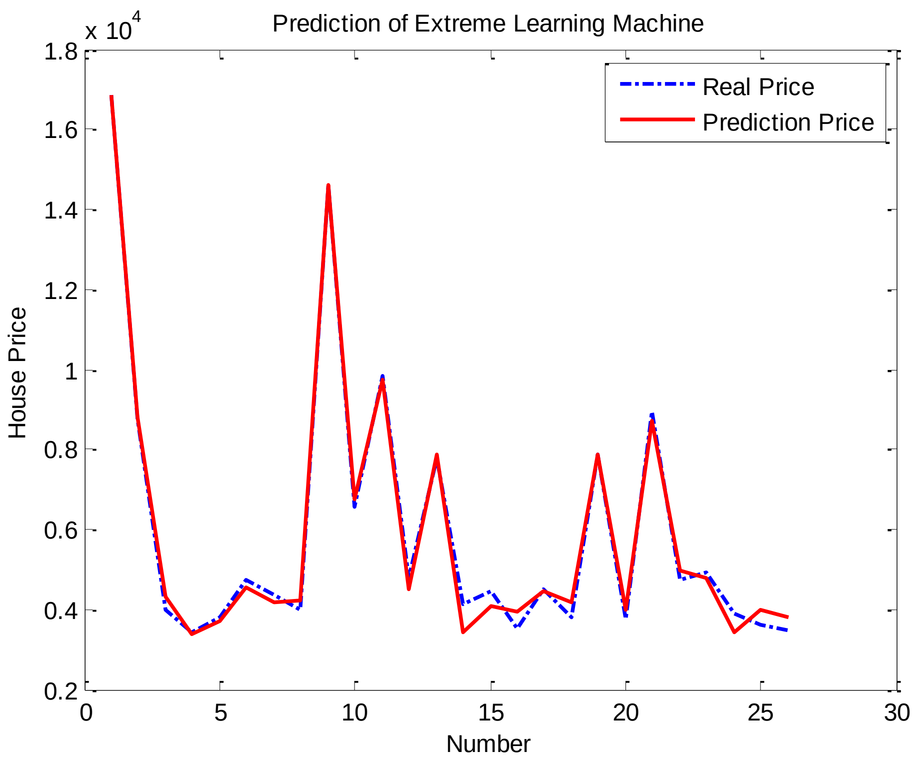


FIGURE 3 Housing Price Forecast Results of Training Samples Based on Extreme Learning Machine

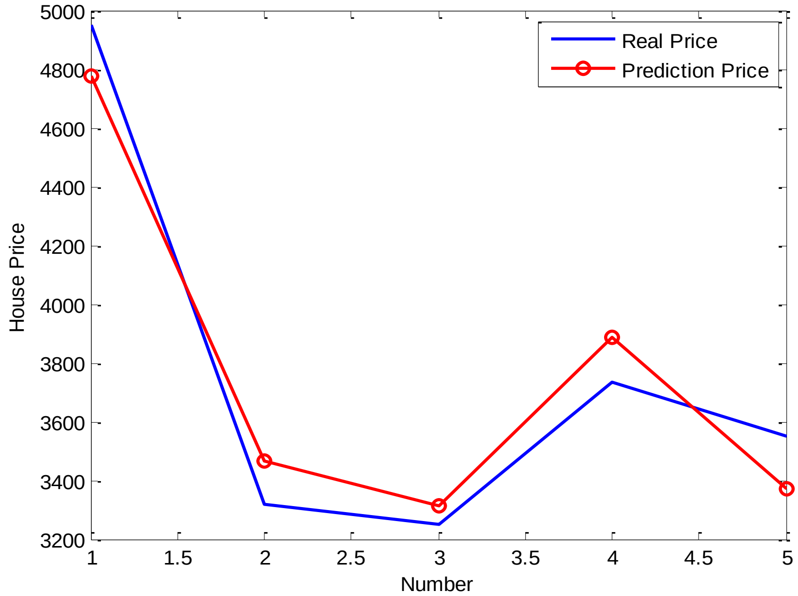
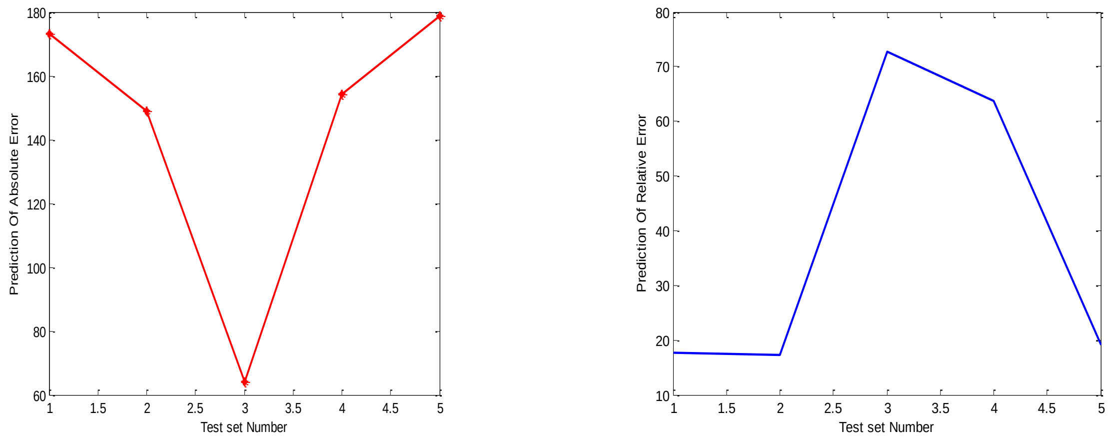


FIGURE 4 Forecast Results of Testing Samples Based on Extreme Learning Machine



a) Absolute Errors

b) Relative Errors

FIGURE 5 Predication Errors

The predication effects of extreme learning machine and that of BP neural network and RBF neural network were compared to further prove the superiority of the algorithm. Figure 6 and 7 showed the comparison results. It can be seen that extreme learning machine has the best prediction results; RBF neural network has relatively inferior prediction results; BP neural network has the worst. The comparison of these different algorithms can effectively validate the effectiveness and correctness of the algorithm in this work. Therefore, this algorithm, with certain practical application value, can be applied to classify and predict problems.

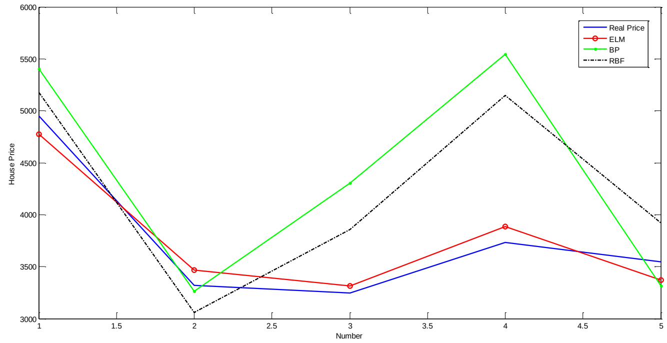


FIGURE 6 Comparison of the Predication Results of ELM, BP and RBF

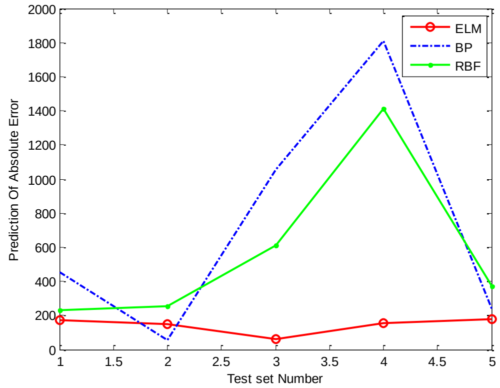


FIGURE 7 Comparison of the Predication Error of ELM, BP and RBF

4 Conclusions and analysis

With the combination of rough set theory and extreme learning machine, the work firstly reduced factors affecting housing price through rough set theory, thus obtaining the major influence factors. Then, the reduced influence factors (as the input of extreme learning machine) and housing price (as data) were used to establish the model to forecast housing price. Besides, The comparison of different algorithms – ELM, BP and RBF – proves the effectiveness and correctness of the algorithm in this work. Therefore, this algorithm, with certain practical application value, can be applied to the classification and predication of problems/

References

[24] Huang F 2007 Application of SPSS Correlation Analysis and Linear Regression Analysis *Foreign Language Teaching and Research* (10) 52-3 *(in Chinese)*

[25] Shen Y, Liu H 2004 Empirical Analysis of 1995-2002 Chinese Housing Price and Economy in 14 Cities *Economic Research Journal* (6) 42-9 *(in Chinese)*

[26] Shi W, Zhang K 2009 Political Analysis of Influence Factor of Housing Price in China *Northern Economy* (12) 25-7 *(in Chinese)*

[27] Yan J 2006 Housing Price and Land Price in China: Theoretical, Political and Policy Analysis *The Journal of Quantitative & Technical Economic*, (1) 17-26 *(in Chinese)*

[28] Xu H, Tang Z, Ba S 2010 Empirical Analysis of Periodical Fluctuation of Rising Housing Price in China *Journal of Shanghai University of Finance and Economics* **12**(4) *(in Chinese)*

Author

Benxue Wang, born in August 1969, Enshi, Hubei Province, China

Current position, grades: associate professor.
University studies: Master's degree in computer science and technology.
Scientific interests: regional economic research and computer software.
Publications: 22 papers.