Investment Analysis of Real Estate by Using Radial Basis Probabilistic Neural Networks

X.-G. Wang[1,3], Y.-S. Ding[1,2,*], X.-F. Zhang[1], Y.-Q. You[1], and S.-H. Shao[1,2]

[1] College of Information Sciences and Technology, Donghua University, Shanghai 200051, P. R. China
[2] Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education
Donghua University, Shanghai 200051, P. R. China
[3] Urban Planning Administration Bureau of Changning District, Shanghai 200336, P. R. China
*Email: ysding@dhu.edu.cn

Abstract

Intelligent Decision Analysis System (IDAS) for real estate investment involves a great deal of factors, such as market of house property, urban construction, construction material, etc. In this paper, we consider the main factors that affect the selling price of the real estate products, and propose a selling price prediction model by using a radial basis probabilistic neural network (RBPNN). The RBPNN consists of four layers: one input layer, two hidden layers, and one output layer. Input layer has four nodes, which are four main factors: the content rate, the environment landscape, the public traffic condition, and the location of the real estate. The output is the selling price, which has 10 nodes corresponding to the 10 classes of the selling price. We totally collect 180 samples of living areas in Shanghai, 150 of which are used as the training data, and 30 of which are used to examine the model. The experiment results are fitted to the practical instances, and verify the feasibility of the RBPNN-based selling price prediction model.

Keywords: IDAS, RBPNN, selling price, real estate.

1. Introduction

With the improvement of people's living standard, and large quantity of the development activities of urban land launching, the investment development for the real estate is becoming prosperous. There are a great deal of factors, such as market of house property, urban construction, construction material, should be considered for the real estate investment, so we establish an investment decision analysis system (IDAS) for the real estate investors [1]. In our IDAS for the real estate, its overall framework is the combining mode of B/S and C/S. It can be used to increase the working efficiency, reduce the influence of the human factor, and provide more scientific basis for the decision-maker.

In the evaluation process of the economic efficacy for real estate investment, it is important to predict its selling price [2]. How to decide the base price, and then to adjust the selling price according to the important policy and the market is a very important topic.

There are many factors to affect the selling price of real estate products, where the main factors are the content rate, the location of the real estate, the environment landscape, and the public traffic condition [3-5]. The relationship of the main factors and their contributions is difficult to be made clear, so the predicting result is difficult to be expressed by a mathematical function.

In this paper, we develop a radial basis probabilistic neural network (RBPNN)-based prediction model for deciding the levels of the selling base-price. The RBPNN model proposed by Huang [6], comes from the radial basis function neural network (RBFNN) and the probabilistic neural network (PNN) [7, 8]. PNN is a neural network model based on statistic theory, whose training speed is faster than that of BP algorithm. It has better classifying capability. So the RBPNN has the advantages of the above two networks. In order to verify the feasibility of the RBPNN prediction model for the selling base-price, we totally collect 180 samples of the living areas in Shanghai, 150 of which are used as the training data, and 30 of which are used to examine the model. The experiment results verify the feasibility of the RBPNN prediction model.

2. The RBPNN-Based Selling Price Prediction Model

2.1. Main factors to affect the selling price

Many factors can affect the selling price of real estate products, but there are only four crucial factors as follows:

(1) The content rate. The content rate is an important factor that affects the selling price of the products of the real estate. If the other conditions of the land are the same, the construction area will decrease with the decreasing of the rate of develop content. As such, the average price of every building area and the total cost will rise, with result in the reducing of the expect profit of the project. What's more, the investment venture becomes larger. To the developers of the real estate, the greatest profit is their only target, so they will improve the accessories to decrease the disadvantage that is led by the decrease of content rate. At the same time, they will increase the selling price of the real estate products. That is to say, the decrease of the content rate will lead to promotion of the selling price.

(2) The environment landscape. The environment factors are the surrounding physical factors that affect the selling price, such as noise, air pollution, water pollution, vision, environment condition, and so on. These factors influence the comfort and value of the real estate, so they also affect the selling price.

(3) The public traffic condition. The public traffic condition influences the activities of the inhabitants, so it is one of the important factors when the inhabitants buy the houses.

(4) The location of the real estate. The location of the real estate is not movable, and it directly influences the economic benefit and satisfaction to the inhabitants. Hence, different locations of the real estate have different selling prices, and sometimes the difference between them is very large. Because the location also shows the grade and the degree of economical prosperity, and the inhabitants always consider the factors such as the surrounding environment, public traffic, quiet, the distance to the city center, and so on.

In addition, the location is only related with the cost of the land. From the market of the real estate in Shanghai, the price of the real estate in the city center is higher than that in the suburb, and the price of the real estate in traditional top grade living area is higher than that in less top grade living area.

2.2. RBPNN-based prediction model

We analyze the main factors that affect the selling price of the products of the real estate, and design a RBPNN-based selling price prediction model as shown in Fig.1. The selling price achieved from the RBPNN-based prediction model, is only a base price as a reference for decision-making. The practical price needs to be adjusted according to the practical situations.

The RBPNN consists of four layers: one input layer, two hidden layers, and one output layer. Input layer has four nodes: content rate x_1, the environment landscape x_2, the public traffic condition x_3, and the location of the real estate x_4. The first hidden layer is a nonlinear processing layer, consisting of the selected hidden centers determined by the input training set. The nodes of the first hidden layer are the same as the quantities of the input samples. The second hidden layer corresponds with the first hidden layer, which has the same size as the output layer. The second hidden layer makes the selection on the first hidden neurons and sums the selected first hidden outputs for its own output. The output layer is the classes of the selling price. Here, the second hidden layer and the output layer both have 10 nodes corresponding to the 10 classes of the selling price, i.e. the 10 pattern classes, as shown in Table 1.

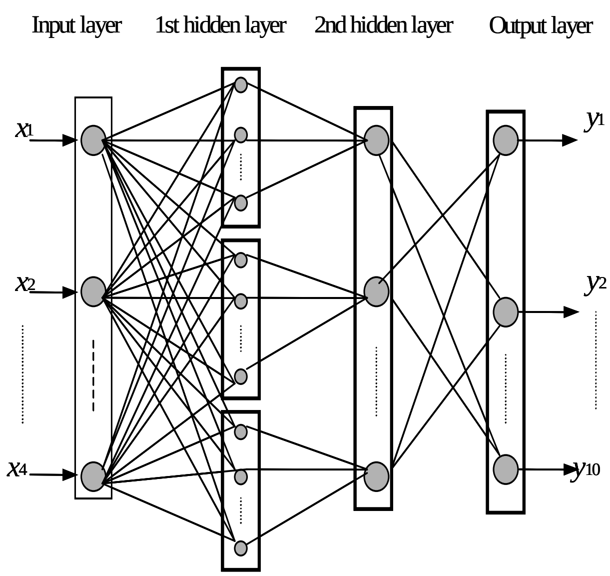

Fig. 1: The RBPNN-Based Selling Price Prediction Model.

Table 1: Inputs and Outputs of RBPNN-Based Selling Price Prediction Model.

	Input (x)		Output (y)		
x_1	Content rate	y_1	<=4000	y_6	8000~9000
x_2	Environment landscape	y_2	4000~5000	y_7	9000~10000
x_3	Public traffic conditions	y_3	5000~6000	y_8	10000~11000
x_4	Location of the real estate	y_4	6000~7000	y_9	11000~12000
		y_5	7000~8000	y_{10}	>=12000

2.3. Selection for the smooth factor σ

In mathematic terminology, the *i*th output of RBPNN can be written as

$$y_i = \sum_{k=1}^{m} w_{ki} H_k(x) \quad i = 1,2,...,m \tag{1}$$

$$H_k(x) = \sum_{i=1}^{n_k} \phi_i(x, c_{ki}) = \sum_{i=1}^{n_k} \phi_i(\|x - c_{ki}\|_2) \;\; k = 1,2,...,m \tag{2}$$

Where x is a n-dimensional input vector; m is the sort number of matching, which equals the number of output neuron; $H_k(x)$ is the output vector of the second hidden layer; w_{ki} is the weight from the second hidden neuron k to the output neuron i; $\phi_i(\cdot)$ is the nonlinear mapping function (or kernel function) of the first hidden layer; c_{ki} is the hidden center vector corresponding to the *k*th neuron of the second hidden layer; n_k is the area number of the *k*th class [9].

Assuming that the total number of the training samples is N, based on (1), we can write the vector-matrix form of some output neurons as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_N \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & ... & h_{1m} \\ h_{21} & h_{22} & ... & h_{2m} \\ ... & ... & & ... \\ h_{N1} & h_{N2} & ... & h_{Nm} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ ... \\ w_m \end{bmatrix} \tag{3}$$

In Fig. 1, $x = \{x_1, x_2, \cdots, x_n\}^T$, n = 4, x is the input neurons. It also provides the first hidden neurons with the input values. In the first hidden layer, each neuron completes the dot product of weight vector w_j of the given class and vector x of the first hidden layer, that is $z_j = x \cdot w_j$, and at the same time completes the non-liner operation to z_j before the neuron outputs to the second hidden layers. The BRPNN transfer function adopts the following formula,

$$g(z_j) = \exp[(z_j - 1)/\sigma^2]. \tag{4}$$

The formula (4) is used to replace the S-shape function used by BP neural network. Each neuron of the second hidden layer receives the outputs from a certain class of the first hidden layer. The output neurons belong to competitive neurons and they receive each class' probability density functions of the output from the second hidden layer. The output value of one neuron is 1 if the probability density function value of the neuron is maximum. This class is the sample pattern class to be recognized and all other neuron output values are 0. The second hidden layer is provided with distinguish capability and high default tolerance, combining statistic information with the network. For instance, the *k*th output of the second hidden neurons is

$$H_k(x) = \sum_{i=1}^{n_k} z_{ki} = \sum_{i=1}^{n_k} \exp[(x \cdot w_{ki} - 1)/\sigma^2]. \tag{5}$$

Here, w_{ki} is weight vector, x is the first hidden vector and a 4-dimensional input vector; σ is smooth factor; and here k is 10.

Assuming that the training samples are certain, the value of smooth factor σ ensures the effect degree among training sample and is related to the change of probability density distributed function. Generally, the network only requires a smooth factor by experience, whereas each eigenvector's sensitivity responded by distance scale is very different and it is difficult to require the optimum value by experience because a distance scale is adopted. Therefore it is necessary to optimize and estimate the smooth factor σ in order to improve the integrated performance of RBPNN.

In the prediction model for the selling base-price, the model class is adopted to estimate σ, which can preferably exhibit the relationship between the main factors and the selling price. The prediction model also adopts an adaptive method to optimize σ, which improves greatly the recognition precision.

The smooth factor σ is calculated by the following formula,

$$\sigma_i(k) = \sigma_i(k-1) + \eta\sigma_i(k) + \xi\big(\sigma_i(k-1) - \sigma_i(k-2)\big) \tag{6}$$

where, η is the learning step and $\eta = 0.05$ ξ is the momentum factor and $\xi = 0.0005$.

3. Experiment Results

We write a program of the RBPNN model by using Matlab software and use it to analyze the predicting results. Here, we take some living areas in Shanghai as examples, and the data from the selected living areas are considered to be the training samples and examining data. In order to confirm the advantage of the RBPNN, we totally collect 180 samples of living areas in Shanghai, 150 of which are used to train the RBPNN. After training, the left 30 samples are used to examine the feasibility of the RBPNN model. The results of the left 30 samples are as showed in Table 2.

From Table 2, we can see that the predicted selling price of the RBPNN model matches better with the practical average selling price. However, the market of the real estate varies times and times, and the selling price fluctuates very much, so there is still a little difference between the predicted results and the ultimate selling price. Actually, the predicted selling price by the RBPNN model is the base price, and the practical average selling price should be adjusted by considering the other factors, such as the influence of the policy and the market.

Table 2: The Results of 30 Examining Samples.

No.	x_1	x_2	x_3	x_4	RBPNN's output	Practical average price
1	2	0.7	0.7	0.8	7000~8000	6600
2	3	0.8	0.72	0.45	6000~7000	5300
3	2.0	0.78	0.83	0.83	7000~8000	8500
4	2.80	0.85	0.75	0.75	5000~6000	5430
5	4.6	0.95	0.95	0.98	8000~9000	9000
6	2.56	0.6	0.45	0.45	3000~4000	3530
7	2.93	0.88	0.88	0.88	7000~8000	8400
8	2.10	0.78	0.65	0.65	5000~6000	6000
9	5.5	0.8	0.82	0.82	6000~7000	6300
10	2.5	0.78	0.75	0.72	5000~6000	5500
11	3.14	.0.78	0.8	0.78	5000~6000	5100
12	5.2	0.8	0.75	0.75	6000~7000	6500
13	1.2	0.88	0.35	0.25	3000~4000	3500
14	3	0.88	0.80	0.80	7000~8000	7500
15	3.1	0.82	0.85	0.85	8000~9000	8000
16	4	0.82	0.88	0.88	8000~9000	8000
17	2.5	0.68	0.35	0.2	3000~4000	3175
18	2.04	0.72	0.7	0.4	4000~5000	3200
19	2.61	0.85	0.88	0.88	8000~9000	7500
20	1.30	0.72	0.28	0.30	3000~4000	3500
21	1.68	0.88	0.75	0.75	6000~7000	7000
22	1.20	0.88	0.55	0.55	5000~6000	6000
23	2.0	0.78	0.55	0.55	5000~6000	5000
24	5.0	0.75	0.71	0.70	6000~7000	5900
25	2.45	0.78	0.70	0.70	5000~6000	6200
26	3.28	0.75	0.75	0.75	6000~7000	6500
27	1.45	0.78	0.78	0.45	5000~6000	5500
28	2.0	0.82	0.45	0.45	4000~5000	5000
29	2.3	0.72	0.55	0.65	5000~6000	5000
30	4.7	0.78	0.85	0.88	8000~9000	8800

4. Conclusions

In this paper, we develop the RBPNN-based selling price prediction model by considering four main factors that affect the selling price of the real estate products. The four main factors are: the content rate, the environment landscape, the public traffic condition, and the location of the real estate. The selling price is classified into 10 classes. We totally collect 180 samples of living areas in Shanghai, 150 of which are used as the training data, and left 30 of which are used to examine the model. The experiment results are fitted to the practical instances, and verify the feasibility of the RBPNN prediction model. The RBPNN-based selling price model is useful for the real estate products to meet the current practical requirements. It has also been embedded into our IDAS for the real estate investors.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China (No. 60474037 and 60004006), Program for New Century Excellent Talents in University, and Specialized Research Fund for the Doctoral Program of Higher Education from Educational Committee of China (No. 20030255009).

References

[1] Y.-Q. You, X.-G. Wang, Y.-S. Ding, S.-H. Shao, An Internet-based investment decision analysis system for real estate, *Computer Engineering and Applications*, 2003, 39(35): 189-191, 224.

[2] X.-G. Wang, Y.-S. Ding, Y.-Q. You, S.-H. Shao, Comprehensive assessment index system and its model for the environmental quality of inhabited region, *System Engineering*, in press.

[3] Michael P. Johnson, A spatial decision support system prototype for housing mobility program planning, Geographical Systems, 2001, 3: 49-67.

[4] L. X. Lu, Analysis of relation between land price and real estate price in the down town of Chengdu, *Journal of Sichan Normal University*, 2002, 25(1): 91-93.

[5] M. Z. Tu, Z. J. Yang, and X. F. Zhou, Real estate investment and evaluation, *Press of Shanghai Jiaotong University*, Shanghai, 1996.

[6] D. S. Huang, Radial basis probabilistic neural networks: Model and application, *Int. J. Pattern Recognition and Artificial Intelligence*, 1999, 13(7): 1083-1101.

[7] D. F. Specht, Probabilistic neural networks, *Neural Networks*, 1990, 1(3): 109-118.

[8] P. P. Raghu, B. Yegnanarayana, Supervised texture classification using a probabilistic neural network and constraint satisfaction model, *IEEE Trans. Neural Networks*, 1998, 9(3): 516-522.

[9] W. B. Zhao, D. S. Huang, The structure optimization of radial basis probabilistic neural networks based on genetic algorithms, *The 2002 IEEE World Congress on Computational Intelligence*, IJCNN02, Honolulu, Hawaii, May 12-17, 2002, 1086-1091.