Fusing Neural Networks, Genetic Algorithms and Fuzzy Logic for Analysis of Real Estate Price

Huawang Shi
School of Civil Engineering
Hebei University of Engineering
Handan, P.R.China
stone21st@163.com

Wanqing Li
College of Economics and Management
Hebei University of Engineering
Handan, P.R.China
stone21st@163.com

***Abstract*—It is generally acknowledged that the price of real estate are highly complicated and are interrelated with a multitude of factors. It will be advantageous if the parties to a dispute have some insights to some degree. This paper introduces a hybrid genetic algorithm (HGA) approach to instance selection in artificial neural networks (ANNs) for housing price determinants. ANN has preeminent learning ability, but BP training algorithm is based on the error gradient descent mechanism that the weight inevitably fall into the local minimum points. In this paper, an improved genetic algorithm was used to optimize the weights of neural network A case study was carried out on housing price determinants of a sample project using this model. The results concerning the efficiency of the proposed framework in terms of accuracy and computational time are also presented. It shows that more accurate price prediction of real estate can be acquired with the GA-ANN model.**

Keywords-artificial neural networks (ANNs); hybrid genetic algorithm (HGA); fuzzy logic; real estate price

I. INTRODUCTION

Real Estate Industry is both capital-intensive, highly related industries and industries essential to provide the daily necessities. However, the real estate pricing models and methods of research rarely receives the critical attention and development it deserves. As the real estate projects more heterogeneous, it is not possible adopted a uniform pricing model and methods as for other products. As a result, the real estate pricing model and is relatively backward. Real estate pricing model and is lagging behind. At present, more mature real estate pricing model is a cost-plus pricing model and fuzzy comprehensive evaluation pricing model, and more cutting-edge is the model Hedonic price model.

As their own characteristics of the real estate project such as the location of diversity, there can be no real estate projects that are exactly the same and always have differences in some areas. But always could be recognized that there are some projects which has a relatively small difference in some of the real estate project has been recognized in the market , which is relatively similar to the real estate projects. The usual way is to comparing with the similar types of real estate prices to determine the sales price, and the correct use of appropriate comparison is directly related to the science and rationality of pricing model.

This paper considers the determinants for housing price by using the integrated neural network genetic algorithm. It provides a new way to the complex, non-linear, polyatomic spring back problem. But the fatal drawbacks of BP neural network are the low speed of convergence and easy trapping in local optimization. Therefore, the present work intends to integrate ANN with GA[1-4] to determine properly the weights of neural network, making up for the defects of BP algorithm. The application of genetic algorithm into artificial neural network in this paper is regarded as the process of searching for optimum in the weight space. Genetic algorithm is a randomized search algorithm borrowing ideas from natural selection mechanism and genetic mechanism of living nature to acquire an optimal or sub-optimal solution. However, simple GA is difficult to apply directly and successfully to a larger range of difficult-to-solve optimization problems. The settings of the HGA[5,6] used for the values of ANNs are described in the pricing model of real estate project.

In the present paper, a pricing model of real estate project was taken as the subject investigated. Based on data of Beijing city, the prediction model of pricing was developed by neural network and improved genetic algorithm, providing theoretical guide for real estate project and tools design.

II. BRIEF INTRODUCTIONS OF FUZZY LOGIC, GENETIC ALGORITHMS, AND NEURAL NETWORKS

A. Fuzzy Logic

Zadeh first proposed Fuzzy Logic as a tool with which to describe uncertainty and imprecision. Because Fuzzy Logic imitates the high-order mode in which the human brain makes decisions in the face of uncertainty or vagueness, it provides an effective way for automated systems to describe highly complex, ill-dened, or difficult-to-analyze subjects. In general, Fuzzy Logic composed of a fuzzier, rule-base, inference engine and defuzzier. The Fuzzy Logic approach still has certain problems to overcome such as membership function conguration, composition operator determination, and application special fuzzy rule acquisition. Although the

978-1-4244-4994-1/09/$25.00 ©2009 IEEE

Fuzzy Logic parameters can be determined using the experience and knowledge of experts, determining these parameters in the absence of such experts remains difficult for particularly complex problems.

B. Artificial Neural Networks

Artificial Neural Networks (ANNs) are composed of simple elements that imitate the biological nervous systems. In the last few decades, significant research has been reported in the field of ANNs and the proposed ANN architectures have proven the efficiency in various applications in the field of engineering. Artificial Neural Networks focus primarily on computing and storing information with in a structure composed of many neurons. Because NN imitate the human brain in terms of learning, recall and generalization, they are usually designed to solve non-linear or ill-structured problems. The NN model is shown as Figure 1.

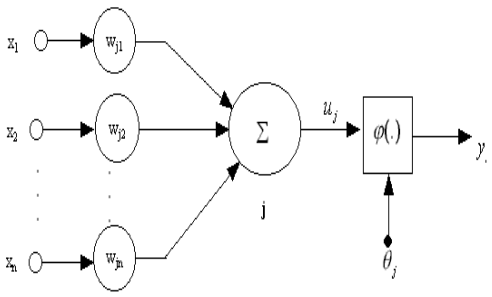

Figure 1. Artificial neural model

The NN model frequently used is multilayer perceptron learning with error back-propagation. In the present research work, the sequence with which the input vectors occur for the ANN straining is not taken into account, thus they are static networks that propagate the values to the layers in a feed-forward way. The training of the neural networks is performed through a back-propa-gation algorithm. In generals, the back-propagation algorithm is a gradient-descent algorithm in which the network weights are moved along the negative of the gradient of the performance function.

C. Genetic Algorithms

Genetic Algorithms (GAs), which imitate parts of the natural evolution process, were first proposed by Holland [9]. Genetic algorithm does not require a gradient of the objective function as a search direction, it can automatically acquire and accumulate knowledge on search space and adaptively control the searching process, so as Gas are stochastic search approaches inspired by natural evolution that involve crossover, mutation, and evaluation of survival fitness[13-18]. Gas out perform the efficiency of conventional optimization techniques in searching non-linear and non-continuous spaces, which are characterized by abstract or poorly understood expert knowledge. Furthermore, to the contrary with the standard algorithms, Gas generate at each iteration a population of points that approach the optimal solution by using stochastic and not deterministic operators. As a result, the search can be deployed without being trapped in local extremes. Based on its merits, the potential of using GA in optimization techniques has been in tensively studied [3,4]. However, simple GA is difficult to apply directly and successfully to a larger range of difficult-to-solve optimization problems. The settings of the GA used for the values of ANNs are fully described in the next paragraphs.

III. ESTABLISHING THE PRICE MODEL USING THE INTEGRATED NEURAL NETWORK GENETIC ALGORITHM

A. Neural network based on HGA

From the BP neural network algorithms and genetic algorithms speak its own characteristics, BP training algorithm is based on the error gradient descent mechanism that the weight inevitably fall into the local minimum points; GA is good at global searching, and search for precision appears to be partial capacity Inadequate. So, in this paper, the GA was used to optimize the weights of neural network. Before the genetic algorithm is conducted, a group of solutions were generated at random. Among them, the individuals with higher fitness were selected according to the principle of survival of fittest to do selection, crossover and mutation[10]. After evolution from generation to generation, it will converge to the fittest individual at last. That is the solution of the problem

B. Simulation

BP algorithm with the classic, traditional genetic algorithms and genetic with momentum improvement of hybrid neural network algorithm to train the network, the learning rate was 0.01, and expectation error was 0.001. Based on GA with momentum, the network training parameters are as follows:

GA parameters:

$$popsize = 40$$

$$W = [-1,+1]$$

$$pc = 0.8$$

$$pm = 0.04$$

$$epoch = 2000$$

BP algorithm parameters

$$\eta = 0.1$$

$$\gamma_1 = 1.05$$

$$\gamma_2 = 0.7$$

$$\phi = 0.5$$

The results of the simulation algorithm comparing with classical GA and of GA With momentum are shown as Figure 3 and Figure 4.

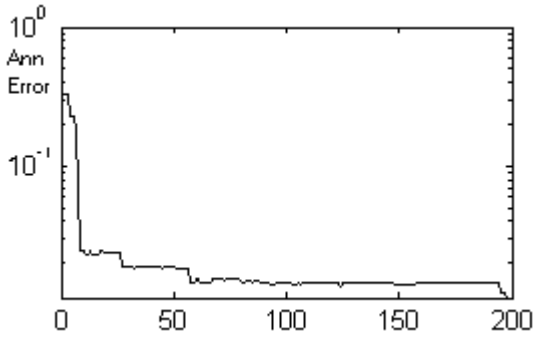

Figure 2. The simulating chart of the classical GA

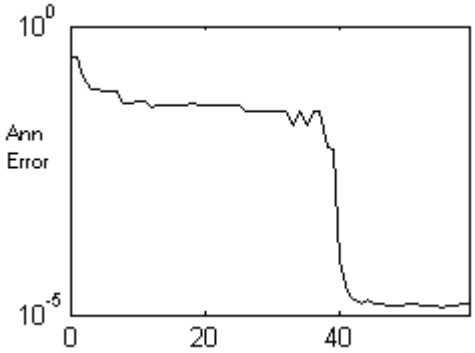

Figure 3. The simulating chart of GA With momentum

IV. CASE STUDY

A. Indexes Dealing with Fuzzy

Before the comprehensive evaluation, the first step is to quantify the qualitative indicators within unity [0-1], that is, all qualitative indexes value are to deal with fuzzy mathematic.

Assuming construction projects comprehensive price evaluation index system of the universe:

$$U_i = \{u_{i1}, u_{i2}, ...u_{i5}\}$$

Where: $u_{i1} \sim u_{i5}$ presents {best, better … bad}.

Its subordinated vector is $r_i = (r_{i1}, r_{i2}, ..., r_{i5})$, Assuming $B = (B_1, B_2, ...B_5)^T$ where B_i presents the i-level corresponding scale. Quantitative the housing price factors value, the standard scale used for is :

Price indexes: $B = [0.9, 0.8, 0.6, 0.4, 0.1]^T$

Non-Price indexes: $B = [0.9, 0.7, 0.5, 0.3, 0.1]^T$

B. Forecasting Model

1) Selecting Indicators: Rough set (RS) theory, proposed by Pawlak[11] , has been proved to be a powerful tool for handling various types of data. It has important applications to artificial intelligence and economic and financial prediction, as a tool for dealing with vagueness and uncertainty of facts, and in classification. The set about objects is represented in classification the form of an information system. The rows of the information system are labeled by objects, where as columns are labeled by attributes and entries of the information system are attribute values.

We classify the factors that would influence the Chinese in real estate market into nine categories based on the literature review. The first category includes economic, political and social factors, which reflect the economic, political and social situation and developing trends. The second one covers regional factors developing trends, characterizes the overall regional environment. The third one is related to living conditions of urban residents, which are etc. The afford ability for real housing demand. The forth one is reflecting the characteristics of housing such as quality Standard, units Designing, The intelligence level and so on. The fifth category contains public facilities. The sixth category contains factors describing the environment of housing, which may influence the Chinese housing price. The final category covers the internal factors of developers.

There are 18 indicators which influence Chinese real estate market to various extents. We select these indicators as condition attributes and the national commercial housing sales price index as a decision attribute to construct the decision information system. Then, we disperse all of the attribute values discretely with equal frequency method, and reduce the attributes. Finally, we use the reduction as the main indicators to analyze and measure the housing price. Data of the commercial housing sales price index is from Beijing city. We analyze the decision information system with rough set theory, and then calculate all the reduction. According to the important degree, we make these indicators in descending order and choose the top 8 important indicators as Table1 shows.

TABLE I. THE SAMPLES OF INPUT AND OUTPUT IN THE NETWORK

Project No.	Input node						Output node
	Appreciation potential	*Location*	*Transportation*	*Decoration Standard*	*Peripheral services*	*Surrounding landscape*	*Average price*
1	0.8	0.41	0.667	0.586	0.135	0.766	16000
2	0.9	0.471	0.367	0.568	0.665	0.610	12000
3	0.7	0.472	0.983	0.812	0.54	0.645	18800
4	0.7	0.625	0.683	0.772	0.64	0.766	17000
5	0.9	0.361	0.333	0.812	0.52	0.781	23000
6	0.75	0.361	0.633	0.678	0.61	0.751	18800

7	0.95	0.568	0.95	0.76	0.49	0.711	13500
8	0.9	0.41	0.683	0.723	0.41	0.723	20000
9	0.95	0.625	0.4	0.658	0.79	0.689	13000
10	0.8	0.41	0.7	0.745	0.52	0.732	6300
11	0.8	0.41	0.667	0.699	0.711	0.683	7800
12	0.9	0.41	0.65	0.710	0.608	0.721	11000
13	0.75	0.41	0.65	0.865	0.812	0.766	10000
14	0.9	0.41	0.7	0.87	0.766	0.812	11000
15	0.9	0.41	0.95	0.845	0.688	0.845	16000

2) Housing price estimate: In this case, the network structure of ANN was 6-9-1 for input layer, hidden layer and output layer respectively. The learning rate was 0.01, and expectation error was 0.001.Then the neural network was programmed by software Matlab6.1. The average variance EMS was 1.36521×10-5, and training time was only 0.566 second. The network model's training result was shown in Fig.2. After the group treated was shown in Table1. of 15 samples of data input to the neural network, we use the first 10 samples as a study group, the last 5 samples as a tested group, the network error is expected to 0.001. By using the trained neural network above, the price of new construction projects were predicted. The results forecasted is in good agreement with the actual values as shown in Table 2, and have been very accurate and meet the actual needs of engineering.

TABLE II. COMPARE BY PREDICTED RESULT AND FACT

Project No.	Actual value	Expected value	Relative error	The overall average error	Overall absolute error
11	0.3	90	1	85	0.16
12	0.3293	87.8773	1.4197	82.7738	0.17015
13	9.7826	-2.3585	41.9665	-2.6190	0.1600
14	1.15611	3.31655	11.0772	2.86078	-2.8608
15	4.95814	5.20026	11.3558	4.2121	4.2021

V. CONCLUSIONS

It is generally acknowledged that the price of real estate was highly complicated and was interrelated with a multitude of factors. It will be advantageous if the parties to a dispute have some insights to some degree. This paper introduces an hybrid genetic algorithm (HGA) approach to instance selection in artificial neural networks (ANNs) for the price of real estate. From the above discussion, the following conclusion can be made:

(1)The factors affecting the housing price were quantified with fuzzy sets and reduced by RS to the inputs of neural network.

(2) The genetic algorithm was adopted to optimize the weights of neural network. The established HGA-ANN model is capable of accurate determinants for housing price with less time and better convergence.

(3) In Simulation tests, the relative error of HGA-ANN models smaller than the classical BP, the classical GA and of GA With momentum model. Thus, the proposed HGA-ANN model is capable of more accurate prediction on price.

REFERENCES

[1] Jesus Fraile-Ardanuy, P.J.Zufiria. Design and comparison of adaptive power system stabilizers based on neural fuzzy networks and genetic algorithms. Neuro computing 70 (2007) 2902-2912.

[2] Lee Te-won, et al. independent component analysis using an extended infomax algorithm for mixed subgaussion and supergaussion sources. Neural computation, 1999, 11:417-441.

[3] Masato Sasaki, Mitsuo Gen, Fuzzy multiple objective optimal system design by hybrid genetic algorithm, Applied Soft Computing3 (2003) 189-196.

[4] H.W.Shi, W.Q. Li, W.Q. Meng. A New Approach to Construction Project Risk Assessment Based on Rough Set and Information Entropy. 2008 International Conference on Information Management, Innovation Management and Industrial Engineering.Dec 2008:187-190.

[5] Zadeh LA. Fuzzy sets. Inform ctrl 1965;8:338-53

[6] Cox E. The fuzzy systems hand book . 2nd ed. New York: AP Professional, 1999.

[7] Eom,K.B.(1999).Fuzzy clustering approach in supervised sea-ice classication. Neurocomputing,25,149-166

[8] Lee,H.S.(2005).A fuzzy multi-criteria decision making model for the selection of the distribution center. Lecture notes in articial intelligence, 3612,1290-1299.

[9] Michael Skinner. Genetic Algorithms Overview http://geneticalgorithms.ai-depot.com/Tutorial/Overview.html

[10] Pawlak Z. Rough sets. Int J Comput Sci 1982;11(5):341-356.

[11] Wu K. Y. , Chen X. J. (2002), A fuzzy comprehensive evaluation model and its appfication for the learning capability of enterprise. Forecasting, 21(2) : 71−74. (in Chinese)

[12] SHI Huawang. The Risk Early-warning of Hidden Danger in Coal Mine Based on Rough Set-neural network. Proceeding of the 2nd International Conference on Risk Management and Engineering Management.November 4-6,2008.314-317