Applying Unascertained Theory, Principal Component Analysis and ACO-based Artificial Neural Networks for Real Estate Price Determination

Wanqing Li
Hebei University of Engineering, Handan, P.R.China
Email: lwq9006@163.com

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

***Abstract*—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 receive the critical attention and development it deserves. In this paper, we present a multi-resolution approach for the determination of the real estate pricing. The proposed method firstly utilizes unascertained theory to describe and quantity the price indices of the real estate, then principal component analysis (PCA) were introduced in to eliminate the real estate pricing indices having the relativities and overlap information. The representative indices from principal component analysis process substitute for the primary indexes. Thus subjective random problem in choosing indices can be avoided. Finally, Using ACO-based artificial neural networks, real estate pricing was analyzed and the results show that this method is more convenient and practical compared with the traditional one.**

***Index Terms*—unascertained theory, principal component analysis, ant colony optimization, artificial neural networks, real estate, price determination**

I. INTRODUCTION

It is generally acknowledged that the price of real estate is highly complicated and is interrelated with a multitude of factors. Research on the impact of real estate prices and real estate development, real estate pricing questions can not only provide consumers with information support for the operation of real estate development projects to provide a reference. At the same time, the government can also provide the basis for macro-control[1-4].

The problem of real estate prices are the current focus of the community's concern. Because of the difference between the data based on, analysis methods and different starting point, the determination of the real estate prices and their movements are not established multi-recognition standards. Real estate prices are the price that a person or organization from another person or institution to obtain the real estate must pay. Because of the division of real estate prices are not uniform, leading to the determination of real estate prices relatively complicated. In addition, real estate prices of the impact of factors of diverse and dynamic, so that real estate pricing is a complex matter.

There are many factors affecting the accident risk of construction, but some of the factors are related and redundant. PCA[6-8] is a powerful tool for analyzing data. The goal of PCA is to reduce the dimensionality of the data while retaining as much as possible of the variation present in the original data set. In this paper, we use principal component analysis (PCA) to reduce some related or redundant factors. Artificial neural networks are a method of information processing and computation that takes benefit of today's technology. Mimicking the processes present in biological neurons, artificial neural networks are used to predict and learn from a given set of data information. At data analysis neural networks are more robust than statistical methods because of their capability to handle small variations of parameters and noise. We use artificial neural networks to determine the real estate price.

Unascertained system[5] that imitates the human brain's thinking logical is a kind of mathematical tools used to deal with imprecise and uncertain knowledge. Artificial neural network(ANN)[7]that imitates the function of human neurons may function as a general estimator, mapping the relationship between input and output. It has outstanding characteristics in machine learning, fault, tolerant, parallel reasoning and processing nonlinear problem abilities Combining of Unascertained method with neural network technology, the reasoning process of network coding can be tracked, and the output of the network can be given a physical explanation. Then unascertained neural network network was set up. It can be compared with the fuzzy network, which can find their owe advantages and shortcomings, so that we can make

doi:10.4304/jsw.6.9.1672-1679

further study on the uncertainty network, and improve the uncertainty network more complete.

Ant Colony Optimization[5,6] (ACO) is a paradigm for designing metaheuristic algorithms for combinatorial optimization problems. ACO is a class of algorithms, whose first member, called Ant System, was initially proposed by Colorni, Dorigo and Maniezzo[7,8]. The main under-lying idea, loosely inspired by the behaviour of real ants, is that of a parallel search over several constructive computational threads based on local problem data and on a dynamic memory structure containing information on the quality of previously obtained result. The collective behaviour emerging from the interaction of the different search threads has proved effective in solving combinatorial optimization (CO) problems[9-11].

II. METHODOLOGY

A. Introduction to Unascertained Number Algorithm

Unascertained mathematics, proposed by Want [1], is a tool to describe subjective uncertainty quantitatively. It deals mainly with unascertained information, which differs from stochastic information, fuzzy information, and grey information. Unascertained information refers to the information demanded by decision-making over which the message itself has no uncertainty but, because of situation constraints, the decision-make cannot grasp the whole information needed. Hence, all systems containing the behaviour factors, such as the problem of clustering have unascertained property.

Definition 1: Suppose a is arbitrary real number, $0<\alpha\le 1$,then definite $[[a,a],\varphi(x)]$ is first-order unascertained number, where

$$\varphi(x)=\begin{cases}\alpha, x=a\\ 0, x\ne a\bigcup x\in R\end{cases} \tag{1}$$

Note that $[a,a]$ express the interval of value, and $\varphi(x)=\alpha$ express belief degree of a. When α =1, belief degree of a is 1. Where α =0, belief degree of a is zero.

Definition 1: Suppose $[a,b]$ is arbitrary closed interval, $a=x_1<x_2<\cdots<x_n=b$, if

$$\varphi(x)=\begin{cases}\alpha_i, x=x_i(i=1,2,\cdots,n)\\ 0, other\end{cases} \tag{2}$$

and $\sum_{i=1}^{n}\alpha_i=\alpha$, $0<\alpha\le 1$, then $[a,b]$ and $\varphi(x)$ compose a n -order unascertained number, as follow $[[a,b],\varphi(x)]$, where α is total degree belief, $[a,b]$ is the interval of value, is $\varphi(x)$the density function.

Definition 1: Suppose unascertained number is $A=[[x_1,x_2],\varphi(x)]$, where

$$\varphi(x)=\begin{cases}\alpha_i, x=x_i(i=1,2,\cdots,k)\\ 0, other\end{cases} \tag{3}$$

$0<\alpha_i<1, i=1,2,\ldots,k$, $\alpha=\sum_{i=1}^{k}\alpha_i\le i$.

Then first-order unascertained number :

$$E(A)=\left[\left[\frac{1}{\alpha}\sum_{i=1}^{k}x_i\alpha_i,\frac{1}{\alpha}\sum_{i=1}^{k}x_i\alpha_i\right],\varphi(x)\right],$$

$$\varphi(x)=\begin{cases}\alpha, x=\frac{1}{\alpha}\sum_{i=1}^{k}x_i\alpha_i\\ 0, other\end{cases} \tag{4}$$

It is expected value of unascertained number A . When $\alpha=1$, as $E(A)$, unascertained number A is discrete type random variable. When $\alpha<1$, $E(A)$ is first-order unascertained number. Where $\frac{1}{\alpha}\sum_{i=1}^{k}x_i\alpha_i$ as expected value of A that belief degree is α .

Each unascertained number includes two parts of probable value and belief degree. So, unascertained number algorithm also includes two parts. Suppose unascertained numbers are A and B .Where

$$A=f(x)=\begin{cases}\alpha_i, x=x_i(i=1,2,\cdots,m)\\ 0, other\end{cases},$$

$$B=g(x)=\begin{cases}\beta_i, y=y_i(i=1,2,\cdots,n)\\ 0, other\end{cases}$$

$C=A\times B$ also is unascertained number. Probable value and belief degree of C is calculated as follows.

（1）Constituted multiply matrix of probable value of unascertained number A and B ,where individual is probable value number series $x_1, x_2,\ldots, x_k$ and $y_1, y_2,\ldots, y_m$ as A and B , permute from little to big.

（2）Constituted multiply matrix of belief degree of unascertained number A and B , where individual is belief degree number series $\alpha_1,\alpha_2,\ldots,\alpha_m$ and $\beta_1,\beta_2,\ldots,\beta_n$ are A , B .Suppose a_{ij} and b_{ij} individual is element of multiply matrix of probable value of A and B , here i is line of matrix, j is array of matrix. We called a_{ij} and b_{ij} as relevant position element.

（3）$\bar{x}_1,\bar{x}_2,\ldots,\bar{x}_k$ result from multiply matrix of probable value of unascertained number A and B , which permute from little to big. And an equal element is one element of relevant position element in multiply matrix of belief degree. Suppose $\bar{r}_1,\bar{r}_2,\ldots,\bar{r}_k$ is relevant position element permutation. Where

$$C=\varphi(x)=\begin{cases}\bar{r}_i, x=\bar{x}_i(i=1,2,\cdots k)\\ 0, other\end{cases}$$

Suppose $C=\varphi(x)$ is arithmetic product of unascertained number A and B . Where

$$C=A\times B=f(x)\times g(x)=\begin{cases}\bar{r}_i, x=\bar{x}_i(i=1,2,\cdots k)\\ 0, other\end{cases}$$

B. Introduction to PCA

PCA was invented in 1901 by Karl Pearson[2]. Now it is mostly used as a tool in exploratory data analysis and for making predictive models. PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute. Principal component analysis (PCA) is a useful statistical technique that has found application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension. Principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

Problems arise when performing recognition in a high-dimensional space (e.g., curse of dimensionality). Significant improvements can be achieved by first mapping the data into a lower-dimensionality space. The goal of PCA is to reduce the dimensionality of the data while retaining as much as possible of the variation present in the original data set.

Supposing n samples, each sample has m target factors, x_j (j = 1,2, ..., m), derived from observation values x_{ij} (i=1,2,...,n), constitute the raw data matrix X=(x_{ij})$n\times m$, shown as bellow:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{5}$$

The target factor is often relevant, thus increasing the internal complexity of the samples. Principal component analysis is to have a correlation between a number of factors into a set of mutually independent factor of a few General methods. These will be the original general factor target factor in the overlapping information removed, to the original contains only significant difference between the target and reflect the original main target factor information purposes. That is, without changing the original data provided by the basic information on more focused and typically show the characteristics of the study. Principal component - the specific algorithm for cluster analysis are as follows.

(1) Original data will be standardized (Z-Score Standardization)

Class and quantity in order to eliminate the impact of different dimension, first of all original data on the standardization of treatment (standardized value of the post-treatment x_{ij}*

$$x_{ij}^{*} = \frac{x_{ij} - \overline{x_j}}{S_j} \tag{6}$$

Where: $\overline{x_j}$ and S_j , respectively, are the mean and standard deviation of the jth target sample, and

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{7}$$

$$S_j = \left[\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2\right]^{1/2} \tag{8}$$

(2) Calculation of correlation between the matrix

Based on the standardized data matrix $X^{*} = (x_{ij}^{*})$, calculated the correlation coefficient matrix $R = (r_{ij})m\times m$. Where, r_{ij} are the correlation coefficient between the x_i and x_j target factor .

$$r_{ij} = \frac{1}{n-1}\sum_{k=1}^{n} x_{ki}^{*} x_{kj}^{*} = \frac{\sum_{k=1}^{n}\left(x_{ki} - \overline{x_i}\right)\left(x_{kj} - \overline{x_j}\right)}{\sqrt{\sum_{k=1}^{n}\left(x_{ki} - \overline{x_i}\right)^2 \left(x_{kj} - \overline{x_j}\right)^2}} \tag{9}$$

Where, i, j=1,2,…,m.

(3) Solving eigenvalue of the correlation matrix and eigenvectors

Calculating the characteristic equation $\left|R - \lambda I\right| = 0$, obtained all of the eigenvalue $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$, and the corresponding Tikhonov unit eigenvector t_j=（t_{1j}，t_{2j}，…，t_{mj}）

$$Y_j = \sum_{k=1}^{m} t_{kj} \bullet x_k^{*} \tag{10}$$

Where: x_k* is the standardized sample matrix.

(3)To determine the number of principal components

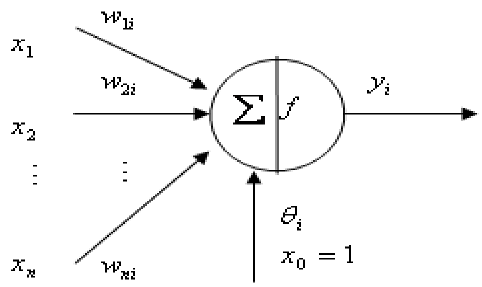
Selecting r principal components in the m principal components that have been identified to finally realize the evaluation analysis. In general, the contribution rate of variance $e_j = \lambda_j / \sum_{k=1}^{m} \lambda_k$ could explain that principal component Y_j reflects the amount of information size. R is determined by the principle that accumulated contribution value $G(r) = \sum_{k=1}^{r} e_k$ is large enough (typically more than 85%). K is kth measured values of the ith and jth factor , k=1,2,…,r.

C. Artificial Neural Networks

An Artificial Neural Network is data information processing paradigm that is inspired by the way biological nervous systems such as the brain process information. The important element of this paradigm is the novel structure of the information processing system.

It is created of a large number of highly interconnected processing elements known as neurons working in unison to solve specific problems. ANN is similar to people which learn by example. An ANN is defined for a specific application such as pattern recognition or data classification through a learning process. Learning in biological systems includes adjustments to the synaptic connections that exist between the neurons.

An artificial neuron is a device with many inputs and one output. The neuron has two modes of operation; the training mode and the using mode. In the training mode, the neuron can be trained to fire (or not), for particular input patterns. In the using mode, when a taught input pattern is detected at the input, its associated output becomes the current output. If the input pattern does not belong in the taught list of input patterns, the firing rule is used to determine whether to fire or not. The neural model was shown bellow:



Its input-output relationship can be described as:

$$\begin{cases} I_i = \sum_{j=1}^{n} w_{ji} x_j - \theta_i \\ y_i = f(I_i) \end{cases} \tag{11}$$

Where, $x_j (j = 1,2,\cdots,n)$ are the output signal of the neurons before, θ_i is the threshold value, w_{ji} expresses the connection weights from neuron j to neuron i , $f(*)$ is the transfer function. In general transfer function $f(*)$ is S -type function:

$$f(x) = \frac{1}{1+e^{-x}} \tag{12}$$

D. A Hybrid Approach with Artificial Neural Networks and Ant Colony Optimization

In recent years, with the rise of evolutionary computation study of fever, people gradually evolutionary computation and artificial neural networks combined, using a variety of evolutionary methods to train neural networks. As the evolutionary algorithm has strong global convergence ability and strong stability, the characteristics of the problem and does not require use of information, such as derivative and so on gradient information. Therefore, a combination of both, not only can play the generalization mapping ability of neural networks and neural networks can improve the convergence speed and learning ability. Evolutionary computation for neural network optimization has two main aspects: First, for the network training, that is, the connection weights between the layers; the second is to optimize the network topology. Specific research methods have the following three ways:

(1) By optimizing the connection weights to train the neural network

Specific neural network, listing all of the neurons, and neurons that may exist for all the connection weights encoded into a digital string of binary code, or indeed that individual, randomly string of groups of these codes were generated, according to conventional methods implementation of the evolutionary operations. Decoding the newly formed string of code constitutes a neural network to calculate the average error that all the training samples generated by this neural network, in order to determine the fitness of each individual. This method is a simple idea, but the large amount of calculation for optimization, especially in large-scale optimization solutions to complex problems neural network, along with how many neurons, the total number of connection weights also will be increased, resulting in the increasing search evolutionary computation space.

(2) Optimize the neural network structure and learning rules

(3) Optimizing neural network structure and connection weights at the same time

Algorithm flow:

The process using ant colony algorithm optimization neural network is as follows: (1)Establishment of feed-forward neural network models, including the identification of the network layers, each layer of nodes until the optimal weight value range, as well as samples; (2) Initialization of all the ant path. At the beginning, without the guidance of pheromones, ants randomly from the range of each parameter, select a value, and build a complete path. Sometimes, random selection can lead to excessive concentration of the selected value, will directly affect the results after the optimization. Recommended parameters uniformly discrete for discrete points of the path after the initialization; (3) When all ants complete understanding of the building, the input sample. According to equation (4) for pheromone update. Ants return to the starting point; (4) According to equation (5) calculation of all the parameters of the probability distribution function; (5) All of the ants, according to the probability distribution function, in turn selecting from the n-parameters within the selection of specific values, constitute a complete the solution. Run(3) ~ (5) step until to meet the termination condition.

III. APPLICATION CASE

A. Real Estate Price Factors

There are 17 indicators which influence Chinese real estate market to various extents. The 17 indicators are including in such as Regional economic I_1 , Currency supply and demand I_2 , Appreciation potential I_3 . Location I_4 , Architecture category I_5 , Type design I_6 , Design I_7 , Landscape Architecture I_8 , Floor area ratio I_9 . Energy-saving system I_{10} , Intelligent I_{11} , Decoration standard I_{12} Traffic convenient I_{13} , Surrounding landscape I_{14} Surrounding Service I_{15} , Developer background I_{16} and Cooperation agencies I_{17} .

B. Indexes Quantification Based on UM

The indices for instance Intelligent, is described with residential intelligent, subdistrict intelligent and long-range control. In this paper, we use unascertained method to quantity the indexes just as Regional economic I_1 , Currency supply and demand I_2 , Appreciation potential I_3 . Location I_4 , Architecture category I_5 , Type design I_6 , Design I_7 , Landscape Architecture I_8 , Floor area ratio I_9 . Energy-saving system I_{10} , Intelligent I_{11} , Decoration standard I_{12} Traffic convenient I_{13} , Surrounding landscape I_{14} Surrounding Service I_{15} , Developer background I_{16} and Cooperation agencies I_{17} .

(1)To determine the subjective weight of indicators

The subjective weight of indicators can be determined though one or more subjective weight determining methods. Supposing the weight of subjective indicators are as follows:

$$a = \left(a_1, \quad a_2, \quad \cdots, \quad a_m\right)^T \tag{2}$$

Where, $\sum_{j=1}^{m} a_j = 1$, $a_j > 0 \ \left(j = 1,2,\cdots,m\right)$

(2)To determine the objective weight of indicators

The objective weight of indicators can also be determined though one or more subjective weight determining methods. Supposing the weight of objective indicators are as follows:

$$\beta = \left(\beta_1, \quad \beta_2, \quad \cdots, \quad \beta_m\right)^T \tag{3}$$

Where, $\sum_{j=1}^{m} \beta_j = 1$, $\beta_j > 0 \ \left(j = 1,2,\cdots,m\right)$

In this paper, information entropy [7-9] was introduced for determining the objective weight of indicators. Following the method of determining the index's identification weight by using the information entropy will be introduced. For the discrete stochastic variables, their information entropy is

$$H(\chi) = -\sum_{i=1}^{m} p(x_i) \log p(x_i) ,$$

$$0 \le p(x_i) \le 1, \sum_{i=1}^{n} p(x_i) = 1 \tag{13}$$

In this paper

$$H(j) = -\sum_{k=1}^{K} \mu_{ijk} \cdot \log \mu_{ijk} \tag{14}$$

$$\gamma_j = 1 - \frac{1}{\log K} H(j) = 1 + \frac{1}{\log K} \sum_{k=1}^{K} \mu_{ijk} \cdot \log \mu_{ijk}$$

(3)

Where $w_j = \gamma_j \Big/ \sum_{j=1}^{d} \gamma_j$,

$$w = (w_1, w_2, \cdots, w_d) \tag{15}$$

Obviously, $0 \le w_j \le 1$ and $\sum_{j=1}^{d} w_j = 1$.

(3)To determine the synthesis weight

Assuming the synthesis weight of each indicator is as follows:

$$W = \left(\omega_1, \quad \omega_2, \quad \cdots, \quad \omega_m\right)^T$$

In order to make full use of the subjective and objective weight determining method to reach the objective and subjective unity, the synthesis weight is carried out by the following formulas:

$$W = (\mu a_1 + (1-\mu)\beta_1, \mu a_2 + (1-\mu)\beta_2,$$
$$\cdots, \mu a_m + (1-\mu))$$

Where μ ($0 < \mu < 1$) is the preference coefficient that reflects the preference level of decision maker for subjective weight and objective weight determining method.

(4)Synthesis appraisal system

As it is known[2] that $\mu_{ijk} = \mu(x_{ij} \in c_k) \quad 1 \le i \le n, 1 \le j \le m$ is the unascertained measure and μ_{ijl} is unit factor's measure appraisal matrix of x_i , in which, μ_j^i means x_{ij} makes x_i has c_k grade in j row.

$$(\mu_{ijk})_{m \times K} = \begin{pmatrix} \mu_{i11}, \mu_{i12}, \cdots, \mu_{i1K} \\ \mu_{i21}, \mu_{i22}, \cdots, \mu_{i2K} \\ \vdots \quad \vdots \quad \cdots \quad \vdots \\ \mu_{im1}, \mu_{im2}, \cdots, \mu_{imK} \end{pmatrix} \tag{16}$$

Where,

$$(i = 1,2,\cdots,n \quad j = 1,2,\cdots,m \quad k = 1,2,\cdots,K)$$

If the single factor measure appraisal matrix above is known, the each factor's classification vector about x_i is (6), and then (7) got as follow:

$$W^i = (w_1^i, w_2^i, \cdots, w_m^i) \tag{17}$$

$$\mu^i = W^i \cdot (\mu_{ijk})_{m\times K}$$

$$= (w_1^i, w_2^i, \cdots, w_m^i)\begin{pmatrix} \mu_{i11}, \mu_{i12}, \cdots, \mu_{i1K} \\ \mu_{i21}, \mu_{i22}, \cdots, \mu_{i2K} \\ \vdots \quad \vdots \quad \cdots \quad \vdots \\ \mu_{im1}, \mu_{im2}, \cdots, \mu_{imK} \end{pmatrix}$$

$$= \left(\sum_{j=1}^{m} w_j^i \cdot \mu_{ij1}, \sum_{j=1}^{m} w_j^i \cdot \mu_{ij2}, \cdots, \sum_{j=1}^{m} w_j^i \cdot \mu_{ijK} \right)$$

(18)

So μ^i is x_i's appraisal vector.

(5)Principle of identification

Because the classification of the comment ranks is orderly. e.g., c_k is "better" than c_{k+1}, the identification principle of "maximum measure'' is not available. The credible identification principle is needed. Let the credible identification be $\lambda, (\lambda > 0.5)$, and it is always adopt 0.6 or 0.7.

$$k_0 = \min_k \left[(\sum_{l=1}^{k} \mu_{il}) \geq \lambda, k = 1,2,\cdots K \right]$$

(19)

then x_i belongs to the rank c_{k_0}. It means that when x_i is not lower than c_k, the fiducially degree is λ, or in other words lower than c_k is 1-λ.

C. Real Estate Price Analysis Based on PCA

(1) First, the original data in Table 1 was processed for the standardization, and by 5, the correlation coefficient matrix was calculated.

(2) By the correlation coefficient matrix eigenvalue calculation, as well as all the main components of the contribution rate and the cumulative contribution rate. The principal component: $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7$ of the cumulative contribution rate were up to 89.7%, so just find the 7 principal component.

(3)We classify the factors that would influence the Chinese 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. Characterizing the overall regional environment. The third one is related to living conditions of urban residents, which are ect. the affordability for real housing demand. The forth one is reflecting the characteristics of housing such as quality standard, units designing, 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.

TABLE I.
TYPE SIZES FOR CAMERA-READY PAPERS

No.	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11
I_1	1	1	1	1	1	1	1	1	1	1	1
I_2	0	0	1	1	0.68	0.6	0.63	0.66	0.85	0.98	0.24
I_3	0.89	0.83	0.55	0.35	0.79	0.7	0.70	0.63	0.43	0.36	0.14
I_4	0.89	0.97	0.23	0.65	0.86	0.5	0.69	0.98	0.29	0.83	0.39
I_5	0.27	0.27	0.45	0.98	0.62	0.5	0.71	0.72	0.31	0.56	0.18
I_6	0.62	0.558	0.87	0.89	0.84	0.6	0.72	0.90	0.30	0.35	0.13
I_7	0.39	0.42	0.56	0.78	0.78	0.4	0.61	0.69	0.57	0.26	0.14
I_8	0.48	0.51	0.87	0.58	0.79	0.7	0.82	0.79	0.43	0.66	0.86
I_9	0.42	0.43	0.36	0.66	0.61	0.9	0.88	0.084	0.71	0.54	0.55
I_{10}	0.43	0.67	0.80	0.58	0.85	0.6	0.83	0.86	0.83	1	0.63
I_{11}	0.73	0.53	0.53	0.35	0.75	0.6	0.71	0.71	0.28	3	0.18
I_{12}	0.13	0.13	0.36	0.89	0.65	0.5	0.80	0.79	0.36	1	0.51
I_{13}	0.93	0.46	0.79	0.96	0.76	0.6	0.60	0.80	0.45	1	0.14
I_{14}	0.72	0.80	0.69	0.66	0.88	0.7	0.79	0.83	0.25	0.56	0.37
I_{15}	0.35	0.6	0.38	0.33	0.73	0.6	0.70	0.81	0.67	0.67	0.16
I_{16}	1	1	1	0.5	1	0.5	1	1	1	0.5	1
I_{17}	0.75	0.70	0.95	0.85	0.65	0.5	0.65	0.72	0.52	0.88	0.14

C. Price Analysis Based on ANN

A total of 500 input-output data pairs were obtained for the training of the ANN for real estate prices. Due to the low dimensionality of the parameters pace and the limited range of variation in the parameters, such number of data reasonably covers the set of different possible operating points. The available data set is randomly partitioned into a training set and a checking set. The training set provides desired input-output pairs used during the training stage, while the checking set is used for testing the generalization capability of the ANN. In this case, the network structure of ANN was 7-12-1 for input layer, hidden layer and output layer respectively. The learning rate was 0.01, and expectative error was 0.001.Then the neural network was programmed by software Matlab6.1. The average variance EMS was 1.21151×10-5. The result forecasted is in good agreement with the actual values, and have been very accurate and meet the actual needs

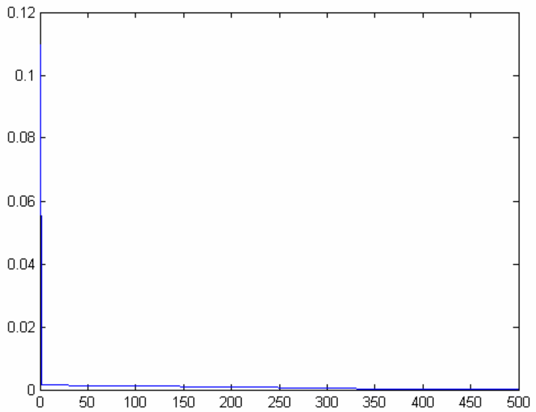


Figure 1. The error curve.

The weights and threshold value are:

0.324952154464723 9.71571109098555E-02

0.250124547364634 -0.641014173975237

0.323861331381674 -0.22303049355771

0.158178715614366 0.138666917209908;

-.217172769041793, 0.483179329404957,

-0.632219274161639, -6.61140301785335E-02

-0.642135795874909 0.300454626407174

-0.381552104035324 0.457174774042434;

-0.327169812669617 0.66131193996221E-2 0

.431956337091504 0.481893867846426

7.68916913142152E-02 0.149784863913473

-.604997180724839 -0.231748600731591;

-6.53070169637845E-02 -0.33315086601662

0.716427534372304 0.497863008634847

0.462439269702574 1.28549224252183E-02

-0.385565130504393 -0.203421285261043;

0.32113971364911 -0.153055722351439

- 0.283893698971471 0.13024766721176

-0.380003067855271 -0.355357183554511

0.106975373455426 -0.577105570461497;

0.586287795904347 -0.599069856205693

0.315075286792014 -0.243907586827217

0.189766951640581 0.627086040748795

0.658890715376864 0.768364870657052;

The threshold value of hidden layer:

6.10447100206110E+46 2.09931818094379E+50

4.28573401055569E+51 3.64734731339217E+49

4.41279072591486E+50 1.44425624918475E+51 -

2.66307449628766E+49 -4.38607598062732E+47

The values of hidden layer to output layer:

-1.92487276697165 -8.91930555655871E-02 -

1.08530381798722 -3.39624991673842

6.25721600162265 0.457364461830099 -

3.4866740222996 6.76692283529659E-02;

The threshold value of output layer:

-1.3653560217906

The running times comparison of two algotithm

Algotithm	Least times	maximum times	average times
ANN	34458	58915	47345
ANN-ACO	1236	2674	1746

Ⅳ. CONCLUSIONS

This paper utilizes the principal components analysis method of multi-dimensional statistical analysis. Using PCA, we classify the factors that would influence the Chinese real estate market into 7 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.

Characterizing the overall regional environment. The third one is related to living conditions of urban residents, which are ect. the affordability for real housing demand. The forth one is reflecting the characteristics of housing such as quality standard, units designing, 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. Then, based on historical data and artificial neural networks, a new real estate pricing models was established. The experiment results show that this method is effective and precise.

In this paper, Ant colony algorithm is used to learn neural network. It overcomes the shortcomings of traditional BP algorithm. Because ant colony algorithm is used for combination optimization, the basic ant colony algorithm is modified. Both the discrete pheromone matrix and probability matrix are extended to continuous function. So the searching field is also extended to a continuous one accordingly. This new way has the merits of both ant colony algorithm and neural network. Its validity and speed are tested through an example.

REFERENCES

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

[2] K.M.Saridakis, A.J.DentsorasJ. Integration of fuzzy logic, genetic algorithms and neural networks in collaborative parametric design. Advanced Engineering Informatics 20 (2006) 379-399

[3] Chau, K. W., Ng, F F. & Hung, E.C.T.. Developer's good will assignificant in fluence on apartment unit prices[J]. Appraisal J ournal. 2001b,vol69,pp.26-34.

[4] Wenjuan Liu, Qiang Liu, FengRuan, Zhiyong Liang, Hongyang Qiu. Method for Housing Price Forecasting based on TEI@I Methodology. Systems Engineering-Theory&Practice. Volume27, Issue7, July2007.

[5] K.M.Saridakis, A.C.Chasalevris, C.A.Papadopoulos, A.J.Dentsoras. Applying neural networks, genetic algorithms and fuzzy logic for the identification of cracks in shafts by using coupled response measurements. Computers and Structures 86 (2008) 1318-1338.

[6] A. Gorban, B. Kegl, D. Wunsch, A. Zinovyev (Eds.), Principal Manifolds for Data Visualisation and Dimension Reduction, LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Wang Guangyuan. Unascertained information and unascertained process [J]. Journal of Harbin University of engineering, 1990(4):1-9(in Chinese).

[9] Liu K D, Wu H Q, Pang Y J. Process and Application of Uncertain Information [M]. Beijing: Science Press, 1999(in Chinese).

[10] Wang, GY, Mathematics treatment of unascertained information. Journal of Civil Engineering Institute of Haerbin 23(4):l-9. l990. (in Chinese)

[11] A. Gorban, B. Kegl, D. Wunsch, A. Zinovyev (Eds.), Principal Manifolds for Data Visualisation and Dimension Reduction, LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007

[12] Greenacre, Michael (1983). Theory and Applications of Correspondence Analysis. London: Academic Press.

[13] Jolliffe I.T. Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus.

[14] Greenacre, Michael (1983). Theory and Applications of Correspondence Analysis. London: Academic Press.

[15] Jolliffe I.T. Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus.

[16] Fen Z R, Safety systems engineering[M]. Beijing: Measurgy Industry Press，1987(in Chinese)

[17] Shi Hua-wang Li Wan-qing Meng Wen-qing. Coal Mine Safe Control Model Based on RS-ANN[J]. Systems Engineering-Theory&Practice,2009,29(1):174-180..

[18] 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.pp314-317.

[19] Altman,E.I.,Marco,G.,&Varetto,F.(1994).Corporate distress diagnosis comparisons using discriminant analysis and neural networks(the Italian experience). Journal of Banking and Finance, 18,505-529.

[20] J.H. Holland, Outline for a logical theory of adaptive systems, J. Assoc. Comput. Mach. 3 (1962) 297–314.