Research Journal of Applied Sciences, Engineering and Technology 4(13): 2003-2008, 2012
ISSN: 2040-7467

Submitted: March 15, 2012 Accepted: April 08, 2012 Published: July 01, 2012

On the Estimate Method of Construction Engineering Cost Based on the RS-GA-NNA Model

Xie Zheng and Mo Lian-Guang
Hunan City University Yiyang, Hunan, 413000, China

Abstract: Given the low intelligent level and the low accuracy of valuation of civil architecture projects, we put forward in the study a constructional engineering assessment method based on Artificial Intelligence which taking advantage of data-calculation from rough set theory, genetic algorithm and neural network algorithm. First, the rough set theory is used to reduce the discrete attributes to optimize the input variables of BP neural network. And then use the global search feature of genetic algorithm to optimize the initial weight and the threshold value of BP neural network. The new algorithm covers both the global random search capability of genetic algorithm and the learning ability and robustness of neural network, thus the computational speed and accuracy have been more significantly improved than the traditional methods. To empirically analyze a case selected from a city in Hunan Province, the results show that the new algorithm model can rely on the engineering features, assess the construction costs scientifically and objectively and have high practical value.

Key words: Estimation, genetic algorithm, neural network, rough set

INTRODUCTION

Many factors affect the cost of the project with a complex constitute and its large ambiguity of connotation and extension, showing a highly non-linear relationship (Duan, 2007). The first foreign estimate method of construction cost, the BCIS model (Liu and Huang, 2003) was presented by the British Ministry of Engineering Cost Information Services in 1962. In 1974, Kousloulas and Koehn (1974) from Britain suggested using the regression analysis method according to the shortcomings of the BCIS model. They put forward the estimation model:

$$C = a_0 + a_1v_1 + a_2v_2 + a_3v_3 + a_4v_4 + a_5v_5 + a_6v_6$$

where, v_1 referred to the region index, v_2, the national price index, v_3, the building type, v_4, the height index, v_5 is the quality index and v_6 is the technology index. In the early 1980s, the Monte Carlo randomly simulating estimation model (Wall, 1997) based on the Monte Carfo method showed up. In our country in recent years, Duan (2007) have advised using Fuzzy Math, Gray Theory and other methods to estimate project cost. The inadequacy of these methods is not to have the dynamic nature of the cost considered. Thus the accuracy of the estimation is not good, buy subjective, resulting in the estimates are sometimes not accurate enough. Wang and Zuo (2010) pointed out the artificial neural network method. Compared with other estimation methods, the artificial neural network method is endowed with high estimate speed and accuracy, objectivity and so on. But the traditional BP neural network in estimation methods of the project cost usually has such shortcomings, like a slow convergence and being easy to fall into local minimum. Gao *et al*. (2009) built an assessment model of the rural land acquisition price with the adoption of the genetic neural network. Jing and Duan (2010) established a highway project cost estimation model with the genetic neural network as well. They both took advantage of the genetic algorithm to optimize the global fast to effectively solve the problems of the application of the complex non-linear project cost estimation.

The valuations of constructional engineering often face low accuracy for it is more complex than the valuations of highway engineering and of rural lands. In order to swiftly work out the valuations of constructional engineering with high accuracy, this essay starts with rough set theory, genetic algorithm and neural network, the basic principles. Then the construction engineering cost model is built accordingly on the basis of rough set, genetic neural network. At last, the author will deal with the practical test of the model.

FACTORS INFLUENCING THE CONSTRUCTION ENGINEERING COST

The composition of the construction engineering cost: Construction Engineering Cost is part of the fixed assets invested in the construction project, takes all the costs of investment in fixed assets from the planning to the completion and the initiation into service, the entire construction process. That is to say, it is the price of

Corresponding Author: Xie Zhengm, Hunan City University Yiyang, Hunan, 413000, China

construction and installation engineering or the total price of construction projects formed in trading activities which is expected to or actually in the land market, equipment market, technology market, labor market and contracting market to complete a project. Construction Engineering Cost regards the project as a particular commodity to exchange objects by tendering and bidding, contracting or other form of transactions and is the project cost finally formed by the market based on several estimates.

The main factors influencing the construction engineering cost: Many factors affect the project cost, studies have shown that the main factors influencing the construction cost, labor and material consumption include:

(1) Geological conditions of the engineering and hydrological geological conditions
(2) Structural features of the building
(3) Construction techniques and mechanical equipment
(4) Influence of the construction unit
(5) The construction unit, etc.

(3)-(5) can not be considered at the design stage.

Analyzed according to principles of construction project budget, it is known that the factors affecting the project cost are the levels of the construction engineering, structure type, building area, storey height, building layers, foundation engineering, concrete supply methods, pile types, external walls works, interior walls works, masonry works, decoration, ground works, doors and windows works, ceiling decoration, plane combination and so on.

THE ESTABLISHMENT OF THE MODEL

The rough set theory:

- **Knowledge, knowledge base and classification:** In rough set theory, "knowledge" is considered to be the capabilities of classifying the study object. Let U be the discourse domain and $u \neq \phi$.
  Knowledge: Any subset of U is known as a concept of U and knowledge refers to any concept family of U.
  Partition:

  If $r = \{x_1, x, ..., x_n\}, x_i \subseteq u, x_i \neq \phi, x_i \cap x_j \neq \phi$

  for I ≠ j, $\bigcup_{i=1}^{n} x_i = u$, r is a partition of u.

  A set made up of all equivalence classes U/R of equivalence relation R is a partition of U, forming knowledge.
  A partition family of U is a knowledge base of U.

- **The knowledge representation system and decision table:** In form, the quadruple S = (U, A, V, f) is a knowledge representation system, in which U: the non-empty finite set of the object is called the discourse domain; A: the non-empty finite set of the object is called the set of attributes:

  V: v= $V = \bigcup_{a \in A} V_a$ Va is an attributive range of a

  f: U × A → V is an information function, which gives every attribute of each object an information value, i.e., $\forall a \in A$, x $\in U$, $f(x, a) \in Va$
  Decision table is a special and important knowledge representation system. Most decisive problems can be expressed by the decision table.
  The decision table can be presented according to the definition of the knowledge representation system as follows:

Knowledge representation system: S = (U, A, V, f), of which, A = CUD, C ∩ D ≠ ϕ, C is called the condition attribute set. D, the decision-making attribute set. The knowledge representation system with the condition attributes and decision attributes is called decision tables:

- **Indiscernibility relation:** In the knowledge base K = (U, R), PR and P ≠ ϕ, then ∩ P (the intersection of all the equivalence relations in P) is also an equivalence relation, named as the indiscernibility relation of P, is abbreviated to IND (P). From this definition, it is evident that indiscernibility relation is an equivalence relation as well, which is constituted by the intersection of equivalence relation family. Indiscernibility relation is the equivalence relation in the discourse domain U when species is represented by the attribute set P, namely:

  IND (P) = {(X, Y), X, Y ∈ U: ∀a ∈ P, f (X, a) = f (Y, a)} (1)

  where, f (X, a) is the value of the discourse domain element X ∈ U on the attribute a and where indiscernibility relation, IND (P) constitutes a category of the discourse domain U, denoted by U/IND (P).

- **Knowledge reduction:** Knowledge in the knowledge Base (attribute) is not of the same importance. Some of the knowledge even is redundant. Knowledge reduction is to delete the irrelevant or unimportant knowledge while maintaining the same ability to classify the knowledge base.
  Let R be a family of equivalence relation, r ∈ R. If IND (R) = IND (R-{r}), r is the knowledge that can be reduced in R. And if P = R-{r} is independent, P is a reduction of R.

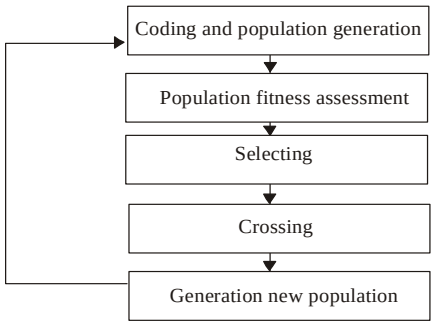


Fig. 1: The basic process of genetic algorithm

GA: GA (Genetic Algorithm) is a global optimization algorithm designed with the help of biological group evolution and natural selection mechanism. The result obtained by the use of the genetic algorithm is not necessarily the optimal solution. But GA has its own advantages and is a powerful tool to solve complex problems. GA process presents in Fig. 1.

BASIC THEORY OF BP NEURAL NETWORK

Artificial neural network, ANN is a pioneering and interdisciplinary subject that develops rapidly in recent year, featured by self-organization, self-learning, association, fault tolerance, anti-interference and nonlinear dynamic processing. ANN can realize a nonlinear mapping relation between network input elements and network output targets. Neural network can reveal the neural network which is contained in data sample and the diagnosis module is in charge of fault samples' network training as well as the accomplishment of fault inference reasoning.

Back propagation artificial neural network, BPANN is a typical multilayer feed-forward network, which consists of input layer, hidden layer and output layer. The three layers are connected together by all connecting method and the units in the same layer are not connected. The basic thought of BP network algorithm is to adjust and rectify the connection weight of network consistently through the back propagation of output error, thus the network errors are minimized. The training process of BPANN consists of feed-forward calculation and error back propagation. The input signals firstly propagate forward to hidden layer, after the calculation of action function, the output information is propagated from hidden layer to output layer. If the output layer fails to get the expected output, then the error signals will return along the original path. After correcting the weights of neurons in every layer, the error signals are minimized. The node action function of BP neural network is generally "S" function.

Common activation function f(x) is derivable Sigmoid function:

$$f(x) = \frac{1}{i + e^{-x}} \tag{2}$$

Error function R is:

$$R\frac{\sum (Y_{mj} - Y_j)^2}{2} \text{(j = 1, 2, …, n)} \tag{3}$$

In this formula, Y_j is expected output; Y_{mj} is actual output; n is sample length.

The uniform expression of weight modified formula of BP algorithm is:

$$W_{ij}(t+1) = W_{ij}(t) + \eta\delta_{pj} O_{pj} \tag{4}$$

In this formula, W_{ij} is the connecting weight of neurons; η is networks learning rate; O_{pj} is the output of sample; δ_{pj} is error correction value.

Integration of rough set theory-genetic algorithm-neural network algorithm: The specific steps of RS-GA-BP eural network algorithm:

The mathematical description of optimization of GA-BP neural network goes as follows:

$$\left\{\min E(w_i, w_o, \theta, r) = \frac{1}{2}\sum_{t=1}^{N_1}\sum_{k=1}^{n}(y_k(t) - \hat{y}_k(t))^2 = \min R\right.$$

$$\text{s.t } w_i \in R^{m\times p},\ w_o \in R^{p\times n},\ \theta \in R^p,\ r \in R^n \tag{5}$$

In the formula, $\hat{y}_k(t)$ is the output of the network; y_k (t) is the target; w_i is the weights from hidden layer to output layer; w_o is the weights from hidden layer to output layer; θ is the threshold value of hidden layer; r is the threshold value of the output layer.

Specific implementation steps are as follows:

- Establishing the index set of the construction engineering estimate
- Discretization of all attribute values
- Formation of decision table
- Attribute reduction
- Initialize the population

Determine the scope and encode the length of the BP network weights and threshold value. According to population size, i.e., popsize, chromosome length and the range of genes, randomly generating popsize one-dimensional arrays with the length of S, thus forming the first population oldpop, the initial population.

W_{11}	W_{12}	…	$W_{S1:}$	W_{21}	W_{22}	…	W_{OS1}	B_{11}	…	B_{21}	…

Fig. 2: Genetic constitution of chromosome chart

The determination on the individual encoding and the length of chromosome:

First, determining the parameter of the genetic algorithm should have the problem domain represented as chromosome.

This essay adopts the real code to combine all the weights and threshold values, changing them into the chromosome in the genetic space. In the case study, three-layer BP neural network is employed and I and O respectively represent the input and output dimensions. S_1 stands for the hidden layer nodes, IW for the connection weight matrix (denoted by W_1) from the input layer to hidden layer and LW for the connection weight matrix (abbreviated as W_2) from the hidden layer to output layer. The first part of Coding is IW, followed by LW, then following the threshold value B1, B2. Therefore the length of chromosome should be $L = I\times S_1+S_1\times O+S_1+O$, namely, a chromosome is constituted by genes with the number of L, just as what shows in Fig. 2:

- **Defining the fitness function:** The performance of BP neural network is measured by R, the squared error between the network output and target output value. R indicates that the smaller it is, the better the network performs. Therefore, in the optimization process of genetic algorithm, the maximum of 1/R is worked as its fitness function to define the fitness function of genetic algorithm:

$$F(w_i w_o \theta, r) = 1/(\frac{1}{2}\sum_{t=1}^{N_1}\sum_{k=1}^{n}(y_k(t)-\hat{y}_k(t))^2) = 1/R \qquad (6)$$

- **Implementation of genetic algorithm:** Put the individual fitness of population in the descending order. And then drive out the low-fitness individuals by a certain percentage (45%). Evenly cross the rest of individuals, i.e., the better individual to generate new individuals, which add to the populations, thus ensuring that the populations remain unchanged. Finally, the mutation operation will be performed to form sub-populations
- Decoding and encoding the fittest individual to have it resolved into the weights and threshold value corresponding to the BP network, to which the early gained weights and thresh hold value will give. Then re-train the network by the training sample
- Store the parameter of BP network training which reaches the training accuracy. Put in the testing data set and put out the simulation value.

The application of the construction estimate model, the RS-GA-BP model pre-disposition of rough set in the construction estimate process:

- **The determination of condition attribute set and decision attribute set:** Many factors affect the construction engineering valuation and the relationship between them is very complex. According to above analysis of the main factors affecting the construction engineering valuation, the condition attribute set is composed of the main factors, while the decision attribute set is formed by the cost of each unit area. In this essay, condition attribute set C contains 20 attributes, namely, engineering type (C1), structure type (C2), story height (C3), layers (C4), basic type (C5), concrete supplying form (C6), pile type (C7), outer wall decoration (C8), interior decoration (C9), masonry works (C10), ground floor engineering (C11), windows and doors project (C12), ceiling decoration (C13), plane composition (C14), equipment engineering (C15), hydropower projects (C16), beam construction (C17), circulation space (C18), building area (C19), structure area index (C20), Decision attribute set D belongs to single attribute set, that is, unilateral cost (D1).
- **Data collection of construction engineering estimate:** The source of sample data comes from 40 cases of general civil construction engineering valuation in a city in Hunan Province. 36 samples will be included in the training set for training BP neural network to get the construction engineering model. The test set, composed of 4 samples, will be used to test the accuracy of BP neural network model, which has finished being trained.
- **Data discretization:** When the rough set theory is applied to deal with decision-making table, it is requested that the value of the decision-making table be expressed as the discrete data. If the range of the condition attribute or the decision attribute is a discrete value, it is accordingly assigned. If the range of the condition attribute or the decision attribute is a continuous value, it must be discrete in handling before processing. This study adopts the method of equidistant division to discretely proceed the continuous data (Table 1).
- **Reduction of condition attributes:** After discretization, an improvement heuristic applied to the rough set attribute reduction is used to do the attribute reduction. The minimum attribute reduction is {C1, C2, C3, C4, C5, C8, C10, C11, C16, C20} and they are all included in the attribute core.

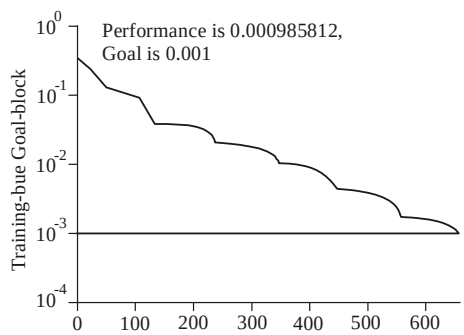
To determine the parameter of RS-GA-BP: The three-layer BP neural network is in the application and the transfer function from the input layer to the hidden layer and from the hidden layer to the output layer is log sigmoid function. Following the use of attribute reduction,

Table 1: Data discretization

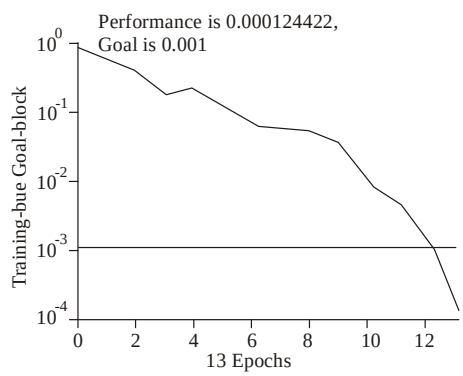
	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	c_{16}	c_{17}	c_{18}	c_{19}	c_{20}	c_{21}	D_1
1	1	3	3	4	2	1	2	3	3	3	3	2	4	3	1	3	5	2	1	3	3	3
2	2	3	4	3	2	3	4	3	4	1	4	3	2	3	3	3	3	3	2	2	4	4
3	2	2	2	3	5	3	1	3	4	2	2	4	3	4	5	1	2	2	3	3	2	2
4	1	3	5	1	4	2	4	1	1	2	4	3	2	1	4	5	4	3	5	3	5	5
5	1	2	2	3	2	3	3	5	3	2	2	4	3	4	3	2	3	2	4	5	2	2
6	2	1	1	3	2	2	4	1	1	2	1	1	2	1	2	1	2	1	6	3	1	1
7	1	3	2	5	1	2	1	2	1	2	2	2	5	1	1	3	1	2	3	7	1	2
8	1	1	2	2	1	2	1	3	2	2	2	3	2	3	2	2	5	2	2	4	2	2
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Table 2: Estimation results by BP network

Series No.	Actual value Yuan/m^2	Estimates Yuan/ m^2	Relative error (%)
1	864	885.1231	2.44
2	885	855.4356	-3.34
3	928	885.6532	-4.56
4	1012	1035.4876	2.32



(a) BP algorithm (700 times) learning curve



(b) Optimized BP algorithm learning curve

Fig. 3: Simulation result

the union set of the minimal set of attribute reduction becomes the input variables of BP neural network. The nodes of the input layer of BP neural network are 10, the output layer is 1 and the project cost per square meter is the output. According to the empirical formula M = int (I + O) 1 / 2 + a, a = 1 ~ 10, take 13 hidden layer nodes. In order to ensure the stability of BP neural network during training, the network input values will be normalized, i.e., each of the input index value is normalized to [0, 1].

Genetic algorithms optimizing the initial connection weights of the network: The size of network weights and threshold value is limited in the interval [-5, +5]. Let the total number of the group be 50, the crossing rate be 0.68, the mutation rate be 0.005 and the iterations of the genetic algorithm be 400. Determine the initial connection weights and threshold value of BP network through the optimization of genetic algorithm.

Training of BP network: BP neural network partly adopt the method of the momentum learning rate for training. Take the momentum factor as 0.85, the initial learning efficiency as 0.072, the increasing ratio of learning efficiency as 1.05 and the decreasing ratio as 0.85. Use 37 completed samples similar to project ones in a city in Hunan Province to train the BP network.

By applying Matlab software, the simulation result of this network training is presented in Fig. 3.

It is shown in Fig. 3b that convergence rate of the network is very fast. After 12 times of learning, errors meet requirements. In Fig. 3a, however, the samples are not fuzzified. Only after 700 times of network learning can errors be reduced to 0.0018343. This can prove the necessity of applying the algorithm optimized traditional BP neural network to the construction engineering estimate.

Result of model assessment: Test the BP network that has been trained in the model, where the input of 4 samples, setting aside in advance has been well trained. The genetic neural network automatically shows the construction project cost not assessed. Compared with the actual cost, it is sure that the BP network model, based on genetic algorithm, can accurately estimate the cost of construction engineering and the error of the assessment result is relatively stable. The comparison between the estimate result and the actual cost presents in Table 2.

CONCLUSION

Factors of the construction engineering cost are a complex non-linear process. Screen the influencing factors with the rough set theory and then well integrate the genetic algorithm and BP neural network. The actual test results show that the integration has achieved a satisfying result for estimate of the construction engineering cost. The new algorithm established model

obtained from the empirical results can better overcome the shortcomings of slow convergence and easily falling into local minimum of BP neural network and can better solve the non-linear problems of multiple inputs and outputs. The convergence speed and accuracy of the new model are greater than the traditional BP neural network model, having a strong adaptation ability to solve problems in the complex non-linear system. It is a rapid method and more accurate in construction engineering estimate with a powerful practical value.

REFERENCES

Duan, X.C., 2007. Comprehensive Investment Control Theory and Methods of Government Investment Projects Study. Science Press.

Gao, Y.N., D.L. Zhu, Y.Q. Chen, etc., 2009. Price assessment of rural land acquisition based on the genetic neural network model. Systemat. Eng. Theory Pract., 48(04): 103-110.

Jing, C.G. and X.C. Duan, 2010. Estimate methods of project cost based on genetic neural network. J. Shijiazhuang Railway Univ. (Social Science Edition), 42(04): 11-17.

Kousloulas, V. and E. Koehn, 1974. Predesign, cos-estimation function for buildings. J. Constr. Division ASCE, 28(4): 589-604.

Liu, F. and Z. Huang, 2003. The comparative analysis of fixed assets investment estimate from home and abroad. Infrastruct. Optim., 26(8): 56-58.

Wang, C.J. and X.H. Zuo, 2010. Prediction model of construction engineering valuation based on BP neural network. J. Shanxi Finance Univ., 36(11): 327-330.

Wall, D.M., 1997. Distributions and correlations in Mon-Carlo simulation. Constr. Manage. Econ., 45(15): 241-258.