ISBN 978-952-5726-11-4
Proceedings of the Third International Symposium on Electronic Commerce and Security Workshops(ISECS '10)
Guangzhou, P. R. China, 29-31,July 2010, pp. 151-155

The Application of Genetic Algorithm and Neural Network in Construction Cost Estimate

WenFeng Feng[1], WenJuan Zhu[1], and YuGuang Zhou[2]
[1]School of computer science and technology, Henan Polytechnic University, Jiaozuo, China
cbfwq3006@163.com
[2]China Overseas Holdings Limited, Hongkong, China

***Abstract*—During the process of construction cost estimate, BP (back propagation) neural network has a great application. Genetic algorithm optimizing BP has been proposed to aim at handling locality minimum and low convergence speed. The method based on analyzing the basic fundamental states that how to use genetic algorithm to improve the ability of BP. After optimizing, the GABP model has been built up. 18 project cases and 2 testing samples are put into the model to observe generalize ability. Comparing with general BP model, the result of GABP model can get lower forecast error and iterations. For these reason, GABP model is appropriate for construction cost estimation.**

***Index Terms*—neural network, BP algorithm, genetic algorithm, optimization, construction cost estimate.**

I. INTRODUCTION

Cost estimates are essential to all project-related engineering and greatly influence construction management. Such estimates allow owners and planners to evaluate project feasibility and control costs effectively in detailed project design work. Within the National Aeronautics and Space Administration (NASA) and other government agencies[1], cost overruns are a major problem, especially with today's emphasis on tight budgets. Overruns may lead to cancellation of a project. In some cases, a potential overrun may result in modifying a project to a design-to-cost task. Our goal was to develop a model that shows the relationship between the estimated cost of a project and the amount that should be spent on doing an estimate. We hope that such a model may help prevent or at least reduce overruns due to inaccurate cost estimates.

Researchers have worked to develop cost estimators that maximize the practical value of limited information available in order to improve cost estimate accuracy and reliability, which should improve the suitability of resultant designs and subsequent project execution work. Traditionally, cost estimating models have been developed using statistical methods. Regression analysis represents a traditional alternative, an inherent disadvantage of which is its requirement of a defined mathematical form for cost functions. In addition, traditional methods are hampered in estimating accurate project costs due to the large number of significant variables and the interactions thereof. Thus, traditional methods have limited applicability. Many previous researches have proved that the neural network cost

estimation model is superior to the traditional regression estimation model.[2]

The purpose of this study was to apply genetic algorithms (GAs) to handle generalize ability problem for improving the accuracy of cost estimation and the performance and validity of optimizing both the neural network size and its parameters using GAs.

II. THEORY OF BP NEURAL NETWORK AND GENETIC ALGORITHM

A. BP neural network

BP neural network is similar to multilayer perceptron on structure. It is a multilayer feed-forward neural network. Its name derives from adjusting training algorithm which is error counter-propagation algorithm in the training network. [3]The whole name is artificial neural network based on error counter-propagation. It is an excessive mapping function that has counter-propagation and error amendment. Memory connecting in the mind can be expressed after the learning of input and output parameters.

- BP algorithm definition

BP algorithm, which has been proposed by Rumelart in 1986, is a supervised learning algorithm. Henceforth, BP neural network has a great application because of the simple structure, multiple adjustable parameters, many kinds of training algorithm and well manipulation. According to statistics, 80%-90% neural networks adopt BP network or the transformation of BP. BP network is the hard-core of the feed-forward network. It reflects the perfect content of neural network. BP algorithm has been made up of two sections: forward-propagation of information and counter-propagation of error. During the process of forward-propagation, the input information transmits into output layer through input layer and hidden layer. Neural in every layer can affect the neural condition in the next layer. If the expected output result can not get in the output layer, the error alteration should be worked out. Then the process must add counter-propagation. Error signal counter transform into the neural in every layer along the primary connection in order to get the expected goal.

The main ideology of BP can be concluded as follows: input the samples, forward-propagate the information,

This study was supported by the National Science & Technology Pillar Program (No. 2007BAF23B0505).

adjust the weight and error through counter-propagation. It makes the input data approach the expected data as soon as possible. The training can be over when the error sum-of squares is lower can the specified error in the output layer. In the end, the weight and error should be saved.

Concrete steps can be stated as follows[4]:

(1) Initialization. Weight [w] [v]and threshold value [θ] [r] can be given in a random way.

(2) forward-propagation of information. The output data in every layer neural should be computed by input and output specified model.

$$b_j = f(\sum_{i=1}^{n} w_{ij} x_i - \theta_j)$$

$$c_t = f(\sum_{i=1}^{n} v_{jk} b_j - r_t)$$

As on the above, b_j is the actually output result of the jth neural in the hidden layer, c_t is the actually output result of neural t in the output layer, w_{ij} is the connecting weight from input layer to the hidden layer, v_{jk} is the connecting weight from the hidden layer to the output layer.

(3) Counter-propagation of error. The error counter-propagation in the process of BP network learning is realized by the following equation.

$$E = \frac{1}{2}\sum_{k=1}^{m}(u^k - y^k)^2$$

Among the expression, y_k is the expected output data of the network.

- Drawback of BP algorithm

Due to BP network should be inferred based on gradient method and requires continuous derivative of goal function, the convergence speed is slow during the learning process and the result can easily get locality optimal which can not find global optimum and cause oscillation effect. And as a result of random chooses value of BP network's weight and threshold level, initial value in every time is totally different. This can affect the consequence of prediction after training. Owing to fortissimo macroscopic search ability of genetic algorithm, and because of biggish probability finding global optimum, so the genetic algorithm can accomplish prophase searching by overcoming the defect of BP algorithm. Genetic algorithm is a global search algorithm. BP neural network organic integrate with genetic algorithm in order to give them the ability which can have global search capability of genetic algorithm and local search capability of BP neural network. [5, 6]Randomness fault of Weight and threshold value can be made up by genetic algorithm (GA), and a better forecast result can be got by it[7]. With the combination of GA and BP, the result can be seen that it is fit for forecast, and it is better than BP.

B. genetic algorithms

Genetic algorithm is a global search algorithm which derives from imitating evolution of the creature. The circulation process consists of reproduction operator, permutation operator, and mutation operator. This process which can solve problem to approach optimum is the essence of GA.GA considers probable separate in the question domain as an entity or chromosome of the population. Each entity encoded into symbol cluster simulates heredity selection and nature obsoletes process of Darwinian. This can make the population repeatedly operate which include hereditary, crossover, and mutation. Each entity has got an evaluation according to scheduled goal fitness function. In the light of evolving regulation, superior group can gradually get. Simultaneously, with the method of global parallel search, optimum entity can be found in the optimize group, and then it can be used in the process of asking for content result[8].

- Genetic algorithm definition

Genetic algorithms were first proposed by Holland (1975) and are well described in text books (Back, 1996; Goldberg, 1989; Michalewicz, 1996; Mitchell, 1996). Genetic algorithms are theoretically and empirically proven to provide a robust search in complex spaces, thereby offering a valid approach to problems requiring efficient and effective searches.

When GAs have been used to solve the problems, the model structure and parameters should be encoded generally by character strings. Solving a particular optimization task using GA, requires the designer to address the five following issues[9, 10].

(1) A genetic representation of candidate solutions,

(2) a way to create an initial population of solutions,

(3) an evaluation function which describes the quality of each individual,

(4) genetic operators that generate new variants during reproduction, and

(5) values for the parameters of the GA, such as population size, number of generations and probabilities of applying genetic operators[11].

- Drawback of genetic algorithm

There still has a great many of problems should be researched, and the GAs also have many kinds of shortages. One of the insufficient of the GAs is to drop the convergence speed when the variable quantity is a great many, short-cut process scope is large, or the scope can not be given. Besides, the result can approach the optimum; however, it can not confirm the position of optimum exactly. [12]Finally, GAs' parameters can not be chosen in a quantify method. If GAs can be connected with BP in the model, the performance of the model can get lower error and exacter prediction.

III. MATLAB REALIZATION OF GENETIC ALGORITHM AND BP IN CONSTRUCTION COST ESTIMATE

This study investigates the use of genetic algorithm in the design and implementation of neural network controller. Features of construction project have been chosen in using matlab to realize cost estimate. According to building operations technology which was written by Xi Zhang in Mach, 2008, 7 features were chosen as

classified standards in construction such as base type, architecture form, number of plies, door and window, siding ornamental, wall, and plane assemble. The quantify description can be stated as follows. m kinds and building project samples have been got. Each sample has 7 features. So a network input model can be fixed as follows.

$P_k=(P1_k,P2_k,\ldots,Pn_k)$ k=1,2,…,m, n=7

It can be seen that m kinds of vectors of building project samples have been built up. There are 7 features in every sample. And then quantify description of any constructional engineering can be given. It can be shown as $T_i=(t_{i1}, t_{i2}, \cdots, t_{ij})$. Ti can be stated as the serial-number of the ith project. t_{ij} (j = 1,2,⋯,7) indicates quantify values of jth feature in project i. Taking a project for example, if the project has 7 features such as brick foundation, 5 floors, timber door and aluminum alloy window, siding rock dash, standard brick, three chambers and one hall. So the quantify description can be expressed in T_i = (1, 1, 2, 3, 2, 2, 3). We can choose 20 premises from a constructional operations company in China as these descriptions. 7 features can be considered as samples. On the basis of the above method, training samples can be stated as table 1.

TABLE I. TRAINING SAMPLES

N	input data							output
	x1	x2	x3	x4	x5	x6	x7	y1
1	1	1	2	1	1	2	2	498
2	3	1	2	3	3	2	4	525
3	2	1	1	1	2	2	2	493
4	1	1	1	1	1	1	2	487
5	1	1	1	3	2	2	3	506
6	2	1	2	3	3	2	4	538
7	3	1	1	1	2	2	4	542
8	4	1	2	3	3	2	5	562
9	2	2	4	3	3	3	4	897
10	3	2	5	3	3	3	3	989
11	4	2	6	3	3	3	4	1045
12	5	2	4	2	4	3	4	876
13	5	4	6	3	4	2	4	857
14	5	2	4	3	3	3	4	923
15	6	2	3	3	3	3	4	948
16	6	3	4	3	3	3	3	747
17	6	2	4	3	4	3	4	689
18	6	4	6	3	4	2	3	936

The data of table 1 can be loaded into matlab, and then we can use BP algorithm and GA-BP algorithm to learn the neural network with the purpose of predicting test values. There still has other examples of construction cost estimate in [13].

A. Matlab realization of BP in Construction Cost Estimate

BP network has three layers including input layer, hidden layer, and output layer. The weight from input layer to hidden layer can be set as w1, and the threshold of the hidden layer can be set as v1. The weight from hidden layer to output layer can be set as w2, and the threshold of output layer can be set as v2. Because of the singularity matrix in the sample, the first step should take the samples normalization.

There have 7 input layer neurons in the code. If we want one output result, the output layer neuron is 1. According to kolmogorov theory, the neuron number in the hidden layer is among 4 to 14.

In our model we set the hidden layer number is 10 with the method proposed by reference [15], deliver neuron function is tansig in the hidden layer, and the output layer's deliver neuron function is purelin. BP network taining function has been set at trainlm. We plan to train 1000 times, and the learning rate is 0.1. The goal value is 0.002.Then we can build up our BP network. The concrete code can be written as the following:

```
net=newff(minmax(input_train),[10,1],{'tansig','purelin
'},'trainlm');
```

The input_test value that we want to predict should be put into the network which has been build up. Input_test has two vectors, [2 1 1 3 2 2 2.5] and [5 4 7 3 4 3 4.1]. The concrete code can be written as the following:

```
an=sim(net,input_test);
```

The result is shown as figure 1.

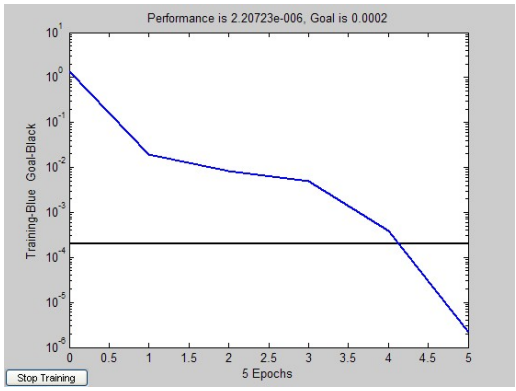

Figure1. Purelin error alteration

Errors are 0.1075, and 0.1261. If delivering neuron function of output layer is logsig and training function is traingdx, while the other parameters is not changed, then the result can be seen as figure 2.

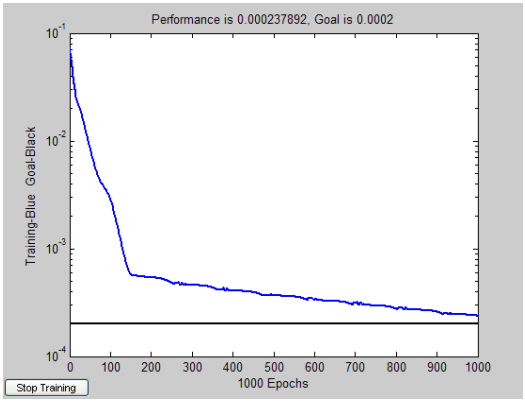

Figure2. Logsig error alteration

From the figure 2, we have seen that BP network has been sank into local minimum. How can we get the goal value in 1000 steps.

B. Matlab realization of GA-BP model in Construction Cost Estimate

Genetic algorithm applied in neural network on the one hand is used for optimizing the structure of ANN, on the other hand is to lean the weight of ANN. i.e., genetic algorithm has taken over traditional algorithms. The concrete steps can be written as the following:

(1) Coding. The first problem we should solve is that decision variables must be changed into string structure data, because actually decision variables can not be handled in genetic algorithm.

```
w1(i,k)=x(inputnum*(i-1)+k);
w2(i,k)=x(hiddennum*(i-1)+k+inputnum*hiddennum);
B1(i,1)=x((inputnum*hiddennum+hiddennum*outputn
um)+i);
B2(i,1)=x((inputnum*hiddennum+hiddennum*outputn
um+hiddennum)+i);
```

(2) Initialize the population. We set the number of population is 80.

(3)Compute fitness of individual in present population. The goal value can be stated as follows.

$$E = \frac{1}{2}\sum_{k=1}^{m}(u^k - y^k)^2$$

It can be considered as the count of fitness value.

(4) Selection. In the first step, the selection mechanism computes a selection probability according to which the sampling algorithm generates an intermediate population of parents in order to generate new variants by recombination and mutation. The main method is put the current population onto a roulette wheel, such that the slot size of each chromosome corresponds to its selection probability. The intermediate population is generated from identical copies of chromosomes sampled by spinning the roulette wheel N times.Here we use genetic tool box.

(5)Recombination. Instead of every parent chromosome, only a fraction of parent population selected at random according to the crossover probability Pc is subjected to recombination. Genetic tool box has the concrete code.

(6) Mutation. In GAs, mutation plays the role of a background operator which arbitrarily inverts one or more bits of a parent chromosome to increase the structural variability of the population. Mutation provides a means to restore lost or unexplored genetic material and prevent the GA from premature convergence to sub-optimal solutions. In this sense, mutation performs an exploration of the search space. The realization process can be seen as follow picture.

Due to the steps that we have stated on the above, we can get the result by the parameters which we have chosen from the above. It is easily to sink into local minimal value while use the BP only, even if the parameters are totally different. However, the combination of GA and BP can handle this problem which can be seen as figure 5.

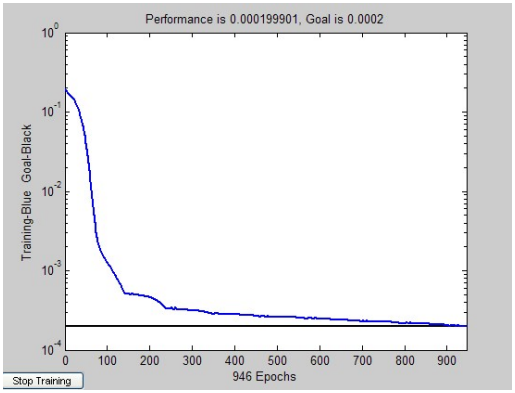

Figure3. Logsig error alteration by GA optimization

It can be seen from figure 3 that genetic algorithm can speed up convergence and handle local minimum problem. The errors are 0.0069, and 0.0263. At this moment, comparing with BP, the error is lower and the generalize ability is better. Therefore, GA-BP model is fit for Construction Cost Estimate. Even though the parameters of figure 1 have been used for comparing, we can see that like figure 4.

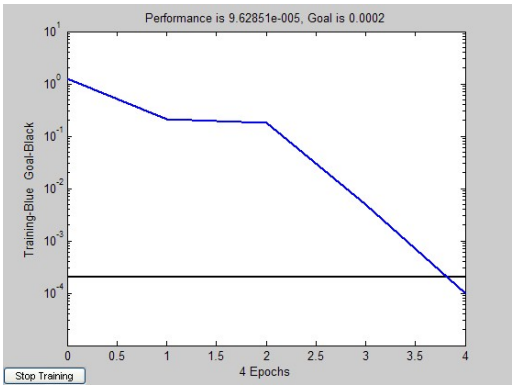

Figure4. Purelin error alteration by GA

Error curve in BP network descends quickly, while error curve in GA-BP network descends gently. But after finding the optimum, the error curve in GA-BP descends rapidly, and the error is lower than BP network. The errors are 0.0492, and 0.1264.Genetic algorithm optimizing figure is as figure 5.

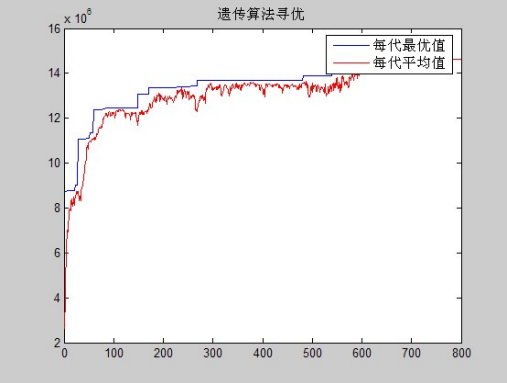

Figure5. Genetic algorithm optimizing

Blue line stands for optimum path, and red line stands for mean value path of each generation. It states that each generation constantly realizes the value that survival of the fittest, consequently the effect of optimal would be obtained.

IV. CONCLUSIONS

An optimization method for construction cost estimate is presented. The method is called GA-BP. GA-BP model adopts the algorithm that chooses the optimized dot from multiply dot in solution space simultaneously. Then a modified BP algorithm is used to search the result. This method not only can avoid local minimum and bad generalize ability problem, but also can overcome the weakness of long training time when GA find the optimum in a random way. It is an effective and efficient method. The simulate result states that, the modified model has a great improve in generalize ability, and it can give a reasonable response when the new input data is coming. Therefore GA-BP model has got a higher reliability, and it can be used in construction cost estimate. However, this method still has shortcoming. It runs a long time, and the error has not attained a satisfied degree. How to realize speediness and efficientness cost estimate is a problem that we need to investigate.

ACKNOWLEDGEMENTS

This study was supported by the National Science & Technology Pillar Program (No. 2007BAF23B0505).

REFERENCES

[1] H. R. B. Donald S. Remer, “Estimating the cost for doing a cost estimate,” Int. J. Production Economics, vol. 66, no. 1, pp. 101-104, 20000.

[2] J.-E. Y. Gwang-Hee Kim, Sung-Hoon An, Hun-Hee Cho, Kyung-In Kang, “Neural networkmodel incorporating a genetic algorithm in estimating construction costs,” building and evironment, vol. 39, no. 3, pp. 1333-1340, 2004.

[3] Zhe-Xue Ge, “neural network and realization by matlab 2007[M],” Electronic Industry Press, 2007.

[4] Jin Hai, and Shi-Wei Ye, “neural network theory,” machine Industry Press, 2004.

[5] M. Zhang, Y. Hu, T. Wang et al., “An optimal predicting method based on improved genetic algorithm embedded in neural network and its application to peritoneal dialysis,” Sheng Wu Yi Xue Gong Cheng Xue Za Zhi, vol. 26, no. 6, pp. 1186-90, Dec, 2009.

[6] R. E. D. Randall S. Sextona, John D. Johnsonb, “Toward global optimization of neural networks: A comparison of the genetic algorithm and backpropagation ” Decision Support Systems, vol. 22, no. 2, pp. 171-185, 1998.

[7] Jing-Wen Tian, “artificial neural network algorithm research and application[M],” Beijing Institute of Techmology press, 2006.

[8] Ying-Jie Lei, “matlab genetic algorithm tool-box and application[M],” Xidian University, 2006.

[9] F. H. Oscar Cordon, Frank Hoffmann,Luis Magdalena, “GENETIC FUZZY SYSTEMS: EVOLUTIONARY TUNING AND LEARNING OF FUZZY KNOWLEDGE BASES,” Advances in Fuzzy Systems — Applications and Theory vol. Vol. 19, pp. 488 pages 2001.

[10] Xiao-Ping Wang, “genetic algorithm--theory, applicationa, and software realization,” Xi'an Jiaotong University, 2002.

[11] L. D. David J. Montana, “Training Feedforward Neural Networks Using Genetic Algorithms,” Machine Learning, vol. 89, no. 1, pp. 762-767, 1989.

[12] F. G. O. Cordón, F. Herrera,F. Hoffmann,and L. Magdalena, “Ten years of genetic fuzzy systems: current framework and new trends ” Fuzzy Sets and Systems, vol. 141, no. 1, pp. 5-31, 2004.

[13] M. J. S. Wen-der Yua, “A neuro-fuzzy computational approach to constructability knowledge acquisition for construction technology evaluation ” Automation in Construction, vol. 6, no. 5, pp. 539-552, 1999.