



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

A Novel Optimized Neural Network Model for Cost Estimation using Genetic Algorithm

T. Hasangholipour and Fariba Khodayar
Faculty of Management, Tehran University, Jalale Ale Ahmad Avenue, Tehran, Iran

Abstract: This study compared the performance, stability and ease of cost estimation modeling between conventional Artificial Neural Networks (ANN) and optimized ANN using Genetic Algorithm (GA) to develop cost estimating relationships. In this study, GA is employed not only to improve the learning algorithm, but also to reduce the complexity in parameter space. The GA optimizes simultaneously the connection weights between layers and the thresholds. In addition, GA reduces the dimension of the feature space and eliminates irrelevant factors. Results showed that optimized model has advantages in compare with conventional ANN in terms of accuracy, variability, model creation and model examination. Both simulated and actual data sets are used for comparison.

Key words: Artificial neural network, cost estimation, genetic algorithm, optimization

INTRODUCTION

Cost estimation is a fundamental activity of many engineering and business decisions and normally involves estimating the quantity of labor, materials, utilities, floor space, sales, overhead, time and other costs for a set series of time periods. Smith and Mason (1996) and Drury (1992) used this estimate as inputs to deterministic analysis methods, such as net present value or internal rate of return calculations, or as inputs to stochastic analysis methods, such as Monte Carlo simulation or decision tree analysis. There has also been some interest in applying newer computational techniques, such as fuzzy logic and artificial neural networks, to the field of cost estimation. Applying fuzzy techniques to cash flow analysis has been used successfully. Musilek *et al.* (2000) discussed using fuzzy composition to estimate NPV after specifying the membership functions for future cash flows and Choobineh and Behrens (1992) compared interval mathematics and fuzzy approaches in cost estimation. A drawback of the fuzzy approach is that the relationships are developed from qualitative information of the cost estimating problem, usually elicited from a knowledgeable person. Fuzzy relationships are not primarily empirical models like regression and neural networks.

Artificial neural networks are purely data driven models which through training iteratively transition from a random state to a final model. Sexton *et al.* (1999) and Camargo *et al.* (2003) showed that they do not depend on assumptions about functional form, probability distribution or smoothness and have been proven to be universal approximations. While theoretically universal approximations, there are practical problems in neural

network model construction and validation when dealing with stochastic relationships, or noisy, sparse or biased data. Camargo *et al.* (2003) showed that ANN had some limitations in learning the patterns because cost data has tremendous noise and complex dimensionality. Liu and Setiono (1996) discussed that ANN has preeminent learning ability while it is often confronted with inconsistent and unpredictable performance for noisy data. In addition, sometimes the amount of data is so large that the learning of patterns may not work well.

Gupta and Sexton (1999) and Holland (1992) used the genetic algorithms such as search techniques based on an analogy with biology in which a group of solutions evolves through natural selection. Vafari and Jong (1998) showed that a population of randomly generated candidate solutions evolves to an optimum solution through the operations of genetic operators consisting of reproduction, crossover and mutation.

This study proposes a new hybrid model of ANN and Genetic Algorithm (GA) for model optimization. Properly data reduction can simplify the process of learning and may improve the performance of the learned results. This study uses GA to search the optimal or near-optimal the connection weights between layers and thresholds in ANN. The simulation results show that the performance of optimized model is higher than conventional ANN. In addition, model creation will be easier and faster.

ARTIFICIAL NEURAL NETWORK

An artificial neural network is modeled as a massively parallel-interconnected network of elementary processors or neurons. It has been shown that a three-layer feed

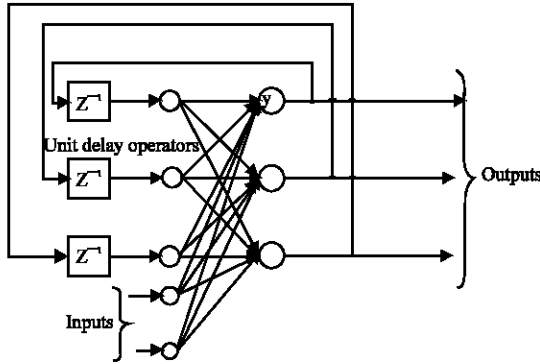


Fig. 1: The back propagation neural network

forward network can generate arbitrary complex decision regions. The multi-layered neural networks operate in two modes: Training and testing. In the training mode, a set of training data is used to adjust the weights of the network interconnections so that the network responds in a specified manner. In the testing mode, the trained network is evaluated by the test data. Rumelhart *et al.* (1986) used the backpropagation learning algorithm which is the most frequently used method in training neural. This study uses a three-layer neural network which is trained by using the error backpropagation (BP) algorithm which is shown in Fig. 1. The number of neurons in the layers, termed as input layer, hidden layer and output layer, are determined by experimentation with an object that the ANN learns and generalizes the situation. Each neuron of the ANN uses a mapping function. For the studies reported in this paper, a sigmoid transfer function for neurons between the input and middle layers and a linear transfer function for neurons between middle and output layers are used. The sigmoid transfer function maps the neuron input from the interval $(-\infty, +\infty)$ into the interval $(0, 1)$, i.e.,

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

The linear transfer function is $f(x) = x$, which maps the neuron input from the $(-\infty, +\infty)$ in to the same interval.

GENETIC ALGORITHM APPROACH

Gupta and Sexton (1999) and Holland (1992) introduced the GAs such as search techniques based on an analogy with biology in which a group of solutions evolves through natural selection. In their implementation, a population of randomly generated candidate solutions evolves to an optimum solution through the operations of genetic operators consisting of reproduction, crossover and mutation.

Here, a standard GA approach for searching the optimal or near optimal connection weight in ANN model for cost estimation problem is described.

The principal components of the optimization based on standard GA are given below:

Chromosomes: A chromosome can be taken as an array holding a candidate optimization. The connection weights and thresholds are set as elements in the chromosomes.

Fitness function: This is the evaluation function used to calculate the degree of fitness or appropriateness of the candidate solutions. The following fitness function can be used:

$$F = \frac{M}{10^{-5} + H} \quad (2)$$

where, M is a constant for amplifying the fitness value. The value of H approaches zero towards convergence. To avoid any numerical difficulty that may occur in calculating F and H is augmented by 10^{-5} .

Crossover operation: This is a genetic operation which is responsible for producing two new candidate solutions from two selected parent chromosomes. Vafari and Jong (1998) proves that in the present working, the two-point crossover method is adopted so that more diversity in the population of chromosomes can be achieved. In this method, two numbers within the length of the chromosome are randomly generated. The elements between the two numbers in the two parent chromosomes are swapped to form two new chromosomes.

Mutation operation: An element of a chromosome is randomly selected. The voltage value of the element is replaced by a value arbitrarily chosen within a range of voltage values. Using the above components, a standard GA procedure for solving the load flow problem is summarized below:

- Step 1:** Initialize S chromosomes in the population. The elements of a chromosome are the candidate modal
- Step 2:** Generate the next generation of S chromosomes in the following way
- Step 3:** The next generation formed in step 2 is now taken to be the current generation. New generations are produced by repeating the solution process starting from step 2 until the specified maximum number of generations is reached

ACTIVITY BASED COSTING SYSTEM

Innes and Mitchell (1993) divide the Activity Based Costing (ABC) system to several steps. In a first step, a company's most significant activities are identified base on the Brimson (1991). In a second step, overhead costs associated with each of these activities are determined. Then factors determining the cost of an activity are ascertained and are referred to as cost drivers which are used to describe the events or forces that are the significant determinants of the cost of these activities. Finally, overhead costs per unit cost driver (cost driver rate) are applied to cost objects. Kaplan (1990) showed that the ABC system are associated with the hierarchical structure of activities and cost drivers and consist usually of five levels:

- Unit level
- Batch level
- Product level
- Facility level
- Customer level
- Activities and cost drivers

ABC techniques have been applied to support new approaches to pricing decisions, profitability analysis and internal performance measurement and cost management.

CASE STUDY

The design processes in textile printing follows a chronological sequence of steps, starting with the concept design to product process plan definition.

Thus, this dynamic process starts with a set of ideas proposed by the designer that are classified by a complex evaluation system before the product arrives to the market as well explained by Moxey and Studd (2000). The first designer inspiration sources are for example, the intent to create new forms or the use of elements, coming from the social or natural environment. At this stage the designer decisions depending mainly on the below factors:

- **Aesthetical factors:** Color, texture, brightness, touch and pattern
- **Functional factors:** Isolation, chemical resistance, heat transfer and dissipation etc.
- **Commercial factors:** Delay, quality and price

The classical idea is that the designers make choices based only on aesthetical parameters. But in practice the product definition process takes into account the economical and functional constraints. For a designer the

paradigm of aesthetic conventions that determine creative solutions within printed fabric design are constrained by technological and market factors. The creative design must be new related to the previous collections and must be validated by the members of the system. For the specific case of the textile printing industry, we have carried a research study about the process of selecting design. The results show it is highly speculative and could take and the financial investment is extremely high in comparison to the potential product commercial success.

Unfortunately, all the esthetical paradigms and customer requirements are explicit only in part until the product freeze point, when a lot of time has passed it could take between three and four months. Thus, makes very important to have reliable economical evaluation of the product changes in the phase when the product is being defined. For the manufacturers in that kind of highly dynamic environment the product development implies to react as soon as possible to customer requirements and at the same time to optimize the capital investment. In fact, the product total cost depends on several components related to the direct and indirect cost. The time and resources spent in the product development process, the technological capabilities (production infrastructure, human knowledge and skills) and of course the product features. The combination of those factors results in a specific product cost.

EXPERIMENTAL RESULTS

Following an experimental research and the considerations in for the textile printing industry, we have treated a database in order to obtain an accurate cost model but more easily interpretable than models obtained from others modeling techniques. Moxey and Studd (2000) introduced the methods of creativity in the Development of Fashion Textiles. In the textile industry unfortunately, all the esthetical paradigms and customer requirements are explicit only in part until the product freeze point, when a lot of time has passed it could take between three and four months. Thus, make very important to have reliable economical evaluation of the product changes in the phase when the product is been defined. For the manufacturers in that kind of highly dynamic environment the product development implies to react as soon as possible to customer requirements and at the same time optimize the capital investment.

The results show that the accuracy of the simplified model (BPLT) remains acceptable compared with the optimized neural network model. In this case, we define 4 input parameters to estimate the cost of product.

Table 1: Definition of the auxiliary parameters

Auxiliary		
R parameters	Definition	
1	TP	The No. of positive class predicted correctly as positive class
2	FP	The No. of negative class predicted wrongly as positive class
3	FN	The No. of positive class predicted wrongly as negative class
4	TN	The No. of negative class predicted correctly as negative class

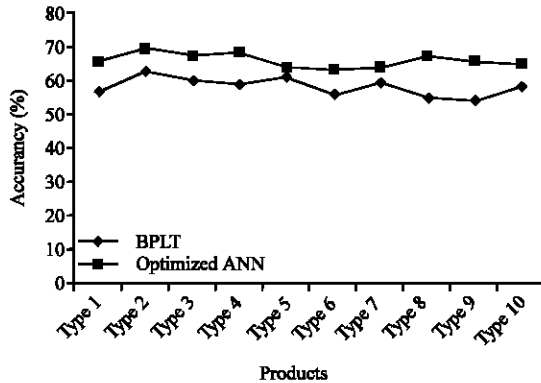


Fig. 2: The average predictive accuracy for BPLT and optimized ANN models

- No. of pattern
- No. of fabric color
- No. of pattern color (complexity)
- Size

Linear transformation with the back propagation neural network (BPLT) and the linear transformation with ANN trained by GA is simulated using Matlab software. We defined four auxiliary parameters to calculate the performance of each model. Daniel-Ramirez *et al.* (2004) introduced auxiliary parameters which are shown in Table 1.

To compare the performance of the models, we calculate the below parameters:

$$A = (TP + TN) / (TP + FP + TN + FN) \tag{3}$$

$$P = TP / (TP + FP) \tag{4}$$

$$R = TP / (TP + FN) \tag{5}$$

Where:

A = Accuracy

P = Precision

R = Recall rate or sensitivity

Two models are compared according to the methods of determining the connection weights and feature transformation. Figure 2 describes the average prediction accuracy of each model for 10 different types of products.

In Fig. 2, the optimized model has higher prediction accuracy than BPLT by 8~11% for the training data. It is a mistake to compare the prediction accuracy between the training data and holdout data. There is a wide difference between the training data and the holdout data for the two models. This result may be caused by the fact that the globally searched discretization simplifies the learning process and eliminates the irrelevant patterns. This prevents the network from falling into the problem of over fitting. We also find that the average prediction performances of BPLT and Optimized model are similar. The reasons for this result may be summarized in two points. First is that there is a generic limitation of global search algorithms. Although, a global search is more desirable than a local search for learning ANN, sometimes a local search is also needed. The other factor may be a problem is dimensionality in data. The GA is a global search algorithm; however, financial data including the stock market data is too complex to be searched easily. It is necessary to reduce the dimensionality of data and irrelevant factors before searching.

CONCLUSION

In this research, a novel ANN model is proposed and simulated. The connection weights and thresholds are optimized with genetic algorithm. The GA searches for the optimal or near-optimal solutions of connection weights in the learning algorithm. Experimental data for the textile printing industry have provided a database in order to obtain an accurate cost model. Four auxiliary parameters were defined to calculate the performance of models. The performance of the optimized model is higher than BPLT and conventional models. The optimized model has higher prediction accuracy than BPLT by 8~11% for the training data.

ACKNOWLEDGMENTS

The authors would like to acknowledge the active participation and financial support of the Management Faculty of Tehran University.

REFERENCES

Brimson, I., 1991. Activity Accounting: An Activity Based Costing Approach. John Wiley and Sons, New York.

Camargo, M., B. Rabenasolo, J.M. Castelain and A.M. Jolly-Desodt, 2003. Application of the parametric cost estimation in the textile supply chain. J. Text. Apparel Technol. Manage., 3: 1-12.

- Choobineh, F. and F. Behrens, 1992. Use of intervals and possibility distributions in economic analysis. *J. Operat. Res. Soc.*, 43: 907-918.
- Daniel-Ramirez, A., E. Israel-Trujillo, M. Juan and G. Gomez, 2004. Choosing variables with a genetic algorithm for econometric models based on neural networks learning and adaptation. <http://repec.org/sce2004/up.6312.1077917368.pdf>.
- Drury, C., 1992. *Management and Cost Accounting*. Chapman and Hall, London.
- Gupta, J.N.D. and R.S. Sexton, 1999. Comparing backpropagation with a genetic algorithm for neural network training. *Omega*, 27: 679-684.
- Holland, J.H., 1992. *Adaptation in Natural and Artificial System*. The University of Michigan Press, Ann Arbor, MI., USA.
- Innes, J. and F. Mitchell, 1993. A Review of Activity-based Costing Practice. In: *Management Accounting Handbook*, Drury, C. (Ed.). Butterworth-Heinemann Ltd., Oxford, UK., pp: 36-63.
- Kaplan, R.S., 1990. Contribution margin analysis: No longer relevant. *J. Manage. Account. Res.*, pp: 2-15.
- Liu, H. and R. Setiono, 1996. Dimensionality reduction via discretization. *Knowledge-Based Syst.*, 9: 67-72.
- Moxey, J. and R. Studd, 2000. Investigation creativity in the development of fashion textiles. *J. Text. Inst.*, 91: 174-192.
- Musilek, P., W. Pedrycz, G. Succi and M. Reformat, 2000. Software cost estimation with fuzzy models. *ACM SIGAPP Applied Comput. Rev.*, 8: 24-29.
- Rumelhart, D.E., G.E. Hinton and R.J. Williams, 1986. Learning Internal Representations by Error Propagation. In: *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Rumelhart, D.E. and J.L. McClelland (Eds.). MIT Press, Cambridge, UK., pp: 318-362.
- Sexton, R.S., R.E. Dorsey and J.D. Johnson, 1999. Optimization of neural networks: A comparative analysis of the genetic algorithm and simulated annealing. *Eur. J. Operat. Res.*, 114: 589-601.
- Smith, A.E. and A.K. Mason, 1996. Cost estimation predictive modeling: Regression versus neural network. *Eng. Econ.*, 42: 137-162.
- Vafari, N.K. and D. Jong, 1998. Feature space transformation using Genetic algorithms. *IEEE Intell. Syst.*, 13: 57-65.