A neural network approach for early cost estimation of structural systems of buildings

H. Murat Günaydın *, S. Zeynep Doğan 1

Faculty of Architecture (Mimarlık Fakültesi), Izmir Institute of Technology (İzmir Yüksek Teknoloji Enstitüsü), Gülbahçe Köyü, Urla, 35430 Izmir, Turkey

Received 24 June 2003; received in revised form 3 September 2003; accepted 2 April 2004

Abstract

The importance of decision making in cost estimation for building design processes points to a need for an estimation tool for both designers and project managers. This paper investigates the utility of neural network methodology to overcome cost estimation problems in early phases of building design processes. Cost and design data from thirty projects were used for training and testing our neural network methodology with eight design parameters utilized in estimating the square meter cost of reinforced concrete structural systems of 4–8 storey residential buildings in Turkey, an average cost estimation accuracy of 93% was achieved.

Keywords: Building design; Cost; Managing projects; Neural networks

1. Introduction

Cost is one of the major criteria in decision making at the early stages of a building design process. In today’s globally competitive world, diminishing profit margins and decreasing market shares, cost control plays a major role for being competitive while maintaining high quality levels. The cost of a building is impacted significantly by decisions made at the design phase. While this influence decreases through all phases of the building project, the committed costs increase. Increasing construction costs render effective and efficient decision making on cost issues a sine qua non for designers. To this end, designers use a number of cost estimating techniques and intuitive judgments by utilizing both their experience and data from previous projects. Several cost estimating methods for the different phases of a project can be observed in the literature, including: traditional detailed breakdown cost estimation; simplified breakdown cost estimation; cost estimation based on cost functions; activity based cost estimation; cost index method; and expert systems [1–3]. Traditional cost estimating procedures follow a quantity take off, while comparative cost estimating relies on parameters such as type, size, and capacity of building. While traditional cost estimating makes use of blueprints and specifications, comparative cost estimating assumes a linear relationship between the final cost and the basic design variables of the project. However the assumption about a linear relationship is questionable. Developments in computer and software technology have facilitated novel approaches for cost estimation. By the emergence of Artificial Intelligence (AI) tools (i.e., neural networks) possible multi- and non-linear relationships can now be investigated. Methods involving the new technology yield results that are both more realistic and accurate vis-à-vis real life conditions.

Current practice shows that the design of a building and the selected materials bear a significant impact on the cost of a building [3]. The cost of a building consists of several items including the structural system, the walls, the doors and windows, the mechanical system, finishings, etc. The relative weights of these items differ for different projects according to the type and usage of materials.
the buildings [4], and cost estimating might require enhanced multi-disciplinary collaboration [5]. It has been observed that the cost of building materials constitutes about 60% of the cost of residential buildings [6]. For multistory reinforced concrete residential apartment buildings, however, the structural frame system including the foundations covers about 25% of the total construction cost [7]. The overall cost of a multistory residential building may come down considerably if the structural system is designed efficiently. It is understandable that, both architects and structural engineers should exercise optimum care in making design decisions for the structural system.

The objective of this paper is to develop and test a model of cost estimating for the structural systems of reinforced concrete skeleton buildings in the early design phase via the application of artificial neural networks (ANN). An ANN model can help the designers to make informed decisions at the early phases of the design process. It should be pointed out that with an ANN model, it is possible to obtain a fairly accurate prediction, even when adequate information is not available in the early stages of the design process [23]. This approach encourages a feedback process that may help designers to reach an optimum solution.

The sample data employed for cost estimation in this paper comes from a research report of residential building construction cost analyses made in Turkey [8]. As a developing country, Turkey experiences rapid population growth, and parallel to this an increasing demand for housing. Residential building construction constitutes 72.8% of the construction market [9]. Eighty-two percent of these buildings are 4–8 storey apartment blocks with reinforced concrete structural systems [9]. With these considerations, the cost of the structural system becomes an increasingly important issue in developing countries such as Turkey.

2. Methodology

In order to achieve the above-mentioned objective of the study, a two-step research methodology was developed. In the first stage, other ANN based cost estimating models were investigated. In the second stage an ANN model was designed for early cost estimation in the design phase for 4–8 storey apartment buildings with reinforced concrete structural systems. Then the model was tested for obtaining the best-possible network configuration. These stages will be presented in the following sections.

2.1. Applications of artificial neural networks in cost estimation

The inspiration for artificial neural networks originated from the study of processes in the human brain. The network acquires knowledge through a learning process. The inter-neuron connection strengths known as synaptic weights are used to store the knowledge [10]. This learning ability of neural networks gives an advantage in solving complex problems whose analytic or numerical solutions are hard to obtain [23]. Cost estimation in the early design phase is one of those problems.

Various researchers have used neural networks as a tool for prediction and optimization previously. But in the area of cost estimating there exist only few applications. The works of Shtub and Versano [11] and Zhang and Fuh [1] in the manufacturing industry, comprise alternative ANN models for cost estimating. Shtub and Versano [11] have developed a system that was based on a neural network that was trained to interpret cost estimates when a new technology was introduced for steel pipe bending. They also found out that neural networks outperform linear regression analysis for cost estimation. Zhang and Fuh [1] designed a neural network model for early cost estimation of packaging products. In their model, they extracted and quantified cost-sensitive attributes of a product design. The correlation between these cost features and the final cost of the product was found by using a back-propagation neural network algorithm depending on historical data.

In the construction industry, Adeli and Wu [12] formulated a regularization neural network to estimate highway construction costs which were very noisy. They observed that as the number of attributes was increased, the construction cost was estimated more accurately. In another study a neural network model for parametric cost estimation of highway projects was proposed by using spreadsheet simulation [13]. Hegazy and Ayed [13] developed a very adaptable and flexible model of ANN by simply facilitating a spreadsheet program. One particular study by Harding et al. [14] constructed an ANN model, which aimed to provide an accurate comparative cost of different procurement routes. Among the 40 variables they used in their study were design specific criteria such as the frame type and gross internal floor area. Emsley et al. [15] suggested that procurement routes cannot be isolated from cost significant variables (i.e., design and site related variables, project strategic variables) in a building project. Therefore they developed Harding et al.’s [14] model one step further by using a more complete and sophisticated data set which would not be available at the early design stage. Their findings indicated 16.6% mean absolute percentage error.

Hashem Al-Tabtabai et al. [16] also developed a neural network model that could be used to estimate the percentage increase in the cost of a typical highway project from a baseline reference estimate. They used environmental, company and project specific factors.
Their model measured the combined effect of these factors on the percentage change in expected cost. The network generated outputs reaching a mean absolute percentage error of 8.1%.

Squeira [17] presented an automated cost estimating system for low-rise structural steel buildings by utilizing design variables such as area, perimeter, height, load, etc. He used a commercial software of ANN and showed that the neural network model outperformed regression. The mean absolute percentage error calculated for the neural network model and regression equation over the entire data set were 11% and 15%, respectively, for the cost estimating of structural steel framing. For the two other models (i.e., total direct cost and cost of wall panels), the mean absolute percentage errors for the neural network approach and regression were 13% vs. 21% and 18% vs. 57%, respectively. Creese and Li [18] developed a neural network application for the parametric cost estimating of timber bridges and again found that the neural network method outperformed common linear regression methods. Their study also showed that the estimation using neural networks improved when more independent variables were introduced in training. However, Bode [19] concluded in his research report that neural networks can only work with a limited number of cost drivers, and more attributes with cost effects need larger case bases for the learning algorithm to achieve satisfying accuracy.

Setyawati et al. [20] compared their results with those of Creese and Li [18] and pointed out the inappropriateness of standard statistical methods for cost estimating and suggested regression analysis based on percentage error and on combined methods for obtaining the appropriate linear regression which might outperform ANN models for cost estimating. Smith and Mason [21] also examined the performance, stability, and ease of cost estimation modeling using regression versus neural networks to develop cost estimating relationships. They reported that the cost data did not enable fitting a commonly chosen model, or did not allow the analyst to discern the appropriate cost estimating relationships; the problem of model commitment became more complex as the dimensionality of the independent variable set grew.

Lack of a cost estimation model utilizing ANN for early design stage motivated the authors and the following model is developed.

### 2.2. The design of artificial neural network model

The problem presented in this paper is based on optimum design and prediction utilizing a feed-forward neural network architecture and back-propagation learning technique. An ANN software as well as a spreadsheet were used for modeling. The software adopted in this application was NeuroSolutions by NeuroDimensions Inc. [22]. Hegazy and Ayed’s [13] spreadsheet was also modified and used for comparison.

The model has been developed in three phases: the modeling phase, the training phase, and the testing phase. The modeling phase involves the analysis of data, the identification of cost estimation parameters and the selection of the network architecture and of the internal rules. The training phase requires the preparation of the data and the adoption of the learning law for the training. The testing phase evaluates the prediction accuracy of the model. The actual costs are compared with the estimated costs and the cost percentage error is calculated.

### 2.3. The modeling phase

The modeling phase includes the design of the neural network architecture. It is a complex and dynamic process that requires the determination of the internal structure and rules (i.e., the number of hidden layers and neurons and the type of activation function). The model is designed according to the type of the data and the response required by the application (Fig. 1).

The current model has been designed to include an input layer of eight processing elements (neurons) corresponding to the eight input parameters and an output layer of one processing element (neuron) as the target. One hidden layer of four processing elements was selected after several trials during the testing phase. The function of the hidden layer is to extract and remember the useful features and the sub features from the input patterns to predict the outcome of the network (values of the output layer) [23]. Therefore, an effective number of processing elements is usually determined by trials for the hidden layers, since there is no rule to determine it.
The total area of the building  
The ratio of the typical floor area to the total area of the building  
The ratio of ground floor area to the total area of the building  
The number of floors  
The console direction of the building  
The foundation system of the building  
The floor type of the building  
The location of the core of the building  
The cost of the structural system per square meter

The consoles of the buildings are analyzed to be directed either in one or two directions. The two-way console directions have an effect on the amount of reinforcement regardless of their total area, which in turn adds to the structural cost. Some of the case examples do not have any consoles; then the parameter refers to a range from no consoles to two-way consoles. The floor type of the apartments, whether reinforced concrete floor systems or precast concrete structural units is also considered to affect the structural cost.

Besides the variables considered above, there are some other important variables that have not been taken into account in the ANN modeling. Since the selection of input variables significantly impacts the accuracy of the neural network predictions [20], one may obtain different or better results if other possible important input variables are studied. The variables that could be investigated include the total height and the roof type of the building, quality classification of structural materials (concrete and steel), the ratio of the area of curtain walls to the total area of the vertical construction, the ratio of the number of secondary beams to the total number of beams in a typical storey of the building, etc. However only the variables that can easily be identified in the early design stage are considered in the current study.

The records of 30 projects in Saner’s [8] study contain data on all of the selected eight design parameters and the corresponding costs of the structural system per square meter. The range of data for these design parameters are presented in Table 1. The console direction, the foundation system, the floor type and the location of the core of the buildings included qualitative data and they were represented by values ranging from 0 to 2. While pre-processing the input data for the ANN implementation, a numerical scale should be associated to the qualitative data considered (i.e., the location of the building core in the middle was indicated by one and the location of the core at the side is indicated by 0) since the network can only handle a numerical process [22]. Zhang and Fuh [1] called this process the quantification of cost-related features. They reported that while the value of quantified features is strongly related
within the same feature, they are independent of the ones assigned to other features. Chua et al. [25] support this and assert that quantification of data improves both the learning and prediction performance of the neural network model.

The cost and design related data from 30 projects were divided into two sets: one set for training the neural network; and the second set for validating the performance of the trained network. For each training cycle, 20% of the observations (data) were selected at random for the validation set. This provided a training set containing data for 24 projects, and a test set of six projects. There are no acceptable generalized rules to determine the size of the training data for suitable training; however, the training sample should cover all spectrums of data available [20,22]. To this end, data from extreme cases (i.e., maximums and minimums) selected for training purposes. According to Zhang and Fuh [1], the training set can be modified if the performance of the model does not meet the expectations. Adding new data to the training samples and retraining the network accordingly can do this.

Data are generally normalized for confidentiality and for effective training of the model being developed. The normalization of training data is recognized to improve the performance of trained networks [17,26,27]. Also in this study, the input and output values were normalized for training and testing purposes.

A neuron basically computes the sum of their weighted inputs, subtracts its threshold from the sum, and transfers these results by a function. This can be explained mathematically as (Eq. (1)):

\[ y_i = f_i \left( \sum_{j=1}^{n} w_{ij}x_j - s_i \right), \] (1)

where \( y_i \) represents the output of a neuron, \( w_{ij} \) represents the weight associated with the input \( j \), \( s_i \) represents the threshold value of the neuron, and \( f_i \) presents the transformation function. The neurons are interdependent on each other via weighted connections. These weights form the power of the influence between the neurons. All neurons are connected to the other neurons in the next layer. An activation function is used because several impacts, if applied additively, might cause these quantities (i.e., target values) to fall below the lower or rise above the upper bound. Gündaydın and Arditi [28] have reported a similar case of cross-impact analysis. The function adopted for the current cost estimation problem was a hyperbolic tangent curve, given by (Eq. (2)):

Activation function (\( x \)) = \( \frac{e^x - e^{-x}}{e^x + e^{-x}} \). \( \) (2)

This hyperbolic tangent function generates output values between -1 and 1. The characteristics of the activation function are important since it defines the behavior of the network model.

### 2.5. The training phase

An important issue to be resolved when applying neural networks to a problem is to determine which training procedure to adopt [24]. There are many other alternative paradigms to choose from. The back-propagation algorithm which belongs to the realm of supervised learning is the most widely used training technique for problems similar to the current study [13,16,17,20]. This algorithm has been shown to be theoretically sound [29], performs well in modeling nonlinear functions, and is simple to code.

Additional to the activation law regulating the weight adjustment among the neurons, which is the hyperbolic tangent function explained in the modeling phase, another fundamental rule of the network is the learning law. All trial models experimented in this study were trained in a supervised mode by a back-propagation learning algorithm. Accordingly, the connection weights are modified continuously until the error between the desired output and the model output is minimized and the modeler has decided on the size of the training set and training type, learning rate and momentum coefficients, network architecture, and the number of iterations for achieving best model outputs.

The Back-propagation algorithm involves the gradual reduction of the error between model output and the target output. Hence it develops the input to output mapping by minimizing a mean square error cost function measured over a set of training examples. The mean square error is expressed as (Eq. (3)):

\[ \text{MSE} = \frac{\sqrt{\sum_{i=1}^{n} (x_i - E(i))^2}}{n}, \] \( \) (3)

where \( n \) is the number of samples to be evaluated in the training phase (i.e., \( n = 24 \)), \( x_i \) is the model output related to the sample \( i \) (\( i = 1, \ldots, n \)), and \( E(i) \) is the target output, i.e., the estimated building cost. The mean square error is a good overall measure of whether a training run was successful [16,24,30]. It was used for evaluating the performance of the model during the training process. Generally, there is no advantage to train a neural network beyond the point where its performance ceases to improve for the set of test observations. The training of all cases in a training set is called an epoch [31]. The error was measured for each run of the epoch number selected and the results were shown as a performance curve such as the one in Fig. 2, which indicated a reduction in the mean absolute error from \( \approx 0.37 \) to 0.03. Training should be stopped when the mean square error remains unchanged for a given number of epochs. This is done in order to avoid
overtraining, in which case the network memorizes the training values and is unable to make predictions when an unknown example is presented to it. In this case, epoch number 10,000 was found adequate for the final training process in a series of test runs. The performance of the network deteriorated for fewer iterations than 10,000 and the network began to memorize the output values for iterations more than 10,000.

Two other important network training parameters are the learning rate and the momentum coefficient. These constant terms are specified at the start of the training cycle and determine the speed and stability of the network [31]. The learning rate determines the amount of weight modification among the neurons during each training iteration [30]. This value ranges between 0.0 and 1.0, where a value closer to 1 indicates significant modification in weight, while a value close to 0 indicates little modification. A small learning rate of 0.05 was arbitrarily chosen for the current problem, since larger learning rates often have been found to lead to oscillations in weight changes and divergence of algorithm, which resulted in an increase in error. One way to allow faster learning without oscillations is to apply a momentum coefficient to the weight change used in the previous training iteration [16,22,30]. In this study, a coefficient of 0.1 was found to perform well.

2.6. The testing phase

The models’ performance is measured by using the cost percentage error (CPE) formula (Eq. (4); Fig. 4) [1]:

$$\text{CPE} = \frac{E(i) - T(i)}{T(i)} \times 100\%.$$  \quad (4)

The effect that each of the network inputs have on the network output can be analyzed. This provides a feedback as to which input parameters are the most significant. For achieving this, a sensitivity analysis is carried out. It is a method for extracting the cause and effect relationship between the inputs and outputs of the network. This will reduce the size of the network and in turn the complexity of the model and the training times. During sensitivity analysis, the network learning is disabled, so the network weights are not affected. The basic idea is that the inputs to the network are altered every time and the corresponding change in the output is reported as a percentage in a figure. In this model, the ratio of the typical floor area to the total area of the building and the ratio of the ground floor area to the total area of the building are found to be the most effective design parameters. However, the usage of only these two design parameters reduced the average prediction accuracy of the model from 93% to 90%. This finding may suggest that even the small clues (i.e., attributes) provided by the other parameters may enhance the model’s prediction capability (i.e., learning capability) as in the case of the problem solved by Adeli and Wu [12] (Fig. 3).
Using a neural network for cost estimation of the structural system at the early stages of the building design process. The approach was shown to be capable of providing accurate estimates of building cost per square meter by using eight parameters available at the early design phase. This model establishes a methodology that can provide an economical and rapid means of cost estimating for the structural system of future building design processes.

The major drawbacks of conventional cost estimates include the need for detailed project information, uncertainties regarding project development, changes in some design parameters, etc. Linear regression analysis shows little or no success at all when early design parameters are used. However, our model has been developed for early cost estimation of structural systems of buildings by applying the principles of supervised learning of neural networks. This model proved that neural networks are capable of reducing uncertainties related to the cost of a structural system of a building.

The ANN capability of seizing knowledge by examples and not by rules represents a very interesting and innovative factor in terms of cost estimating. ANN approach broadens the possibility of building a full-scale knowledge-based model for predicting the total building cost. This approach might solve the complex non-linear mapping for predicting the total building cost at any phase of the design and construction processes. ANN can also be utilized in the other areas of project management such as quality, productivity, constructability, value engineering, scheduling, etc.

### 2.7. Testing results and discussion

Data from six projects were used for testing purposes. Results show 93% average accuracy (Fig. 4 and Table 2) with a mean square error of $\text{MSE} = 0.038$. These figures are considered to be good cost estimation at the early design stage. These findings also outperformed Saner’s [8] study which had resulted in a linear regression coefficient of only 0.47; whereas the current model reached 0.99. The maximum deviation of the cost estimate from the actual cost is about 12% while the average deviation is about 7% (Table 2). The model is also tested through another ANN application (the modified spreadsheet application of ANN) developed by Hegazy and Ayed [13], and similar results were obtained (i.e., 92.5% average accuracy).

The accuracy of the cost estimates can be close to an expert system developed by Mohamed and Celik [3], where estimated total costs were found to be between the range of 2% and 4% of the actual cost. Of course, one should not forget that an expert system uses more accurate and complete information about the building design. At the early design stage for the design professionals, the accuracy of the reported cost estimation model has invaluable benefits for better decision making.

### 3. Conclusions

Neural networks learn from examples, and so the performance of a neural network model of cost estimation strongly depends on the quality and the quantity of examples. The more examples there are, the less the prediction error is. Thus, to study modeling and prediction methods, and construct an accurate prediction model of building costs, there is a need for reliable, high-quality, full-scale cost data of buildings of various types and conditions.

Though the data selected for this model was limited in scope, the results are encouraging for further research of expanded data sets. The main objective of this work was to introduce a new and alternative approach of using a neural network for cost estimation of the structural system at the early stages of the building design process.
[17] Squeira I. Neural Network-Based Cost Estimating, Master’s Thesis. Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Quebec, Canada; 1999.