Estimating the construction schools cost in Ho Chi Minh City using Artificial Neural Network

Tinh Dinh Cong¹ and Quang Nguyen Minh²

¹The University of Architecture Ho Chi Minh City, 196 Pasteur, District 3, Ho Chi Minh city, Vietnam
²The Ho Chi Minh City University of Technology, 268 Ly Thuong Kiet, District 10, Ho Chi Minh city, Vietnam

Abstract. Construction cost is considered as one of the most important criteria in making decision for investment, especially in the idea formation stage. The paper provides a tool to determine the construction cost of the school during the conceptual design phase, when project information is still sketchy, not detailed yet. This research use artificial neural network techniques for estimating the construction schools cost. The model is based on the weight which set by the excel algorithm and the weight optimization method of the error back propagation. The excel algorithm is tested and optimized by using Statistical Package for the Social Sciences software. The basic model achieves accuracy 90.1% and the optimal model achieves 96.6% accuracy optimization. The optimal weight is used to build the cost estimation model with the automatic calculation program by the excel spreadsheet, which allows for updating the data value limited, updating weight when were adjusted.

1. Introduction

The development in the education has led to the investment in technical facilities. The budget expenditure on construction in the education in Ho Chi Minh city sector has also increased year by year. It is mobilized many sources of capital such as budget capital, non-budget capital and social capital. Through this study, the author hopes to contribute an additional tool for estimating the cost of construction of schools in the initial design concept period based on the artificial neural network (ANN) techniques. This tool helps contractors and investors, project managements, as well as State agencies, which make reasonable decisions before implementing the school project.

2. Literature review

On the world, many studies on cost estimation models of building in general and school in particular have been carried out by many techniques such as Multiple Regression Analysis (MRA), Artificial Neural Network (ANN), Case Based Reasoning (CBR), and Support Vector Machine (SVM).

The estimation cost model of building by ANN techniques [2, 11, 12]. The MRA and CBR techniques were developed and compared with ANN techniques [8]. The CBR techniques can be combined with AHP and GA theory to develop the AHP-CBR model [1] and the GA-CBR model [4] in estimate cost of construction project.
In the field of cost estimation of schools, there are studies using SVM techniques [13] and ANN techniques [7]. The MRA techniques compares with the ANN techniques that estimated construction cost of elementary school project in Korea [3]. The MRA model was developed, and compared with the ANN and SVM models in school construction cost [9]. Moreover, [10, 14] has built a cost estimation model for educational facilities (primary, secondary and high schools) in Korea were built by BTL (Build-Transfer-Lease) using MRA, and ANN techniques.

In Vietnam, study on cost estimation models of ANN [6]. The other study on combination of cost and time factors was developed by ANN [15]. Or the study of schools cost estimates by the statistical method [16] and CBR techniques combined with GA techniques [5].

The results of these studies indicated that ANN techniques are nonlinearity of data and give better results than MRA techniques [2, 3]. In addition, comparing cost estimation methods, the results show that ANN model gives more accurate prediction results than MRA model and CBR model [8], and ANN model results predict more accurately than MRA model and SVM model [9].

3. Materials and methods

The main factors that affecting construction costs of schools are determined by preliminary survey results with 15 experts have more than 5 years of experience in the field of construction; and most of them have been taken part in the schools for 3 to 5 years (53.3%). The 5-point Likert scale is used in the survey. The results of the survey with the collected amount of 53 tables, the evaluation of experts all agree on 13 factors that affecting the construction costs of schools, the lowest rating is 2.25 points and the highest is 4.79 points. This research selects an average evaluation score of over 3.5 points.

Identify 7 factors that affect the construction costs of schools: (1) number of floors; (2) gross floor area; (3) ground floor area; (4) foundation structure; (5) roof structure; (6) number of classes; (7) form of construction.

The ANN model is simulated by Microsoft Excel spreadsheet as figure 1, in which two issues need attention as follows:

- Finding the network structure and training: in order to find out the network structure model is suitable, and ensure that the error in the training to reach the optimal value is appropriate.
- Comparison of errors during training: the error between the actual value and the predicted value of the model is adjusted accordingly after each repeated, until this error is

![Figure 1. Procedure for estimating the school construction cost model.](image-url)
in the value range allow.

Estimating the school construction cost model has the main output variable is construction cost. Data source was collected from Construction of Services Information Center (COSIC) in Ho Chi Minh city. The data are determined from the basic design appraisal documents, the detailed design appraisal documents, decision approving construction design and cost estimate, decision approving construction investment project. The data of work construction expenses is converted to the same year through the construction price index. As a result, the number of data sets was determined to build the estimating model for schools in Ho Chi Minh city.

The initial estimation cost model is determined based on a set of weights established by Excel spreadsheets. Network training to find the optimal weight set is the process of optimizing value according to the error back propagation. The ANN structure is multi-layer forward network, including 01 input layer, 01 hidden layer, and 01 output layer with the structure of 10-6-1. Hidden layer transfer function, and output layer transfer function: Tangent Hyperbolic (TANH). Errors are limited to training and testing by 2% and 5% respectively. The total average weighted value after weighting optimization is 3.49%. The estimated result reached an error of 9.88%.

Estimation cost model is tested and optimized by SPSS. The network is trained in many different structures to find the network structure that is the best suited by selecting the number of neurons in the hidden layer, and choosing the transfer function. The number of neurons in the hidden layer ranges from 08 to 21 neurons (14 structures).

Result, ANN model with the number of hidden neurons is 11; TANH transfer function has the smallest error average of 0.353. As a result from SPSS, the weights are determined. This set of weights is used to test the excel algorithm, and optimize the estimation cost model. The optimal model is set on excel with an error of 3.36%. Assessing the suitability of the model calculated \( R^2 = 0.8249 \). The model is quite suitable.

4. Results
Data of 27 school projects in Ho Chi Minh City is divided into two parts. Part one: 24 data used for training. Part two: 3 data used for testing. Step by step for estimating the school construction cost model was built by an excel spreadsheets according to the following tables.

The main variables in the construction cost model, there are 7 mains input variables is the factors that affect the construction costs of schools and 1 main output variable is the construction cost. There are has four quantitative variables and three qualitative variables. The qualitative variables include: foundation structures (spread footing, strap footing, pile cap), roof structures (tole, tile, reinforced concrete), form of construction (new construction, new construction and renovation). Quantitative variables are type directly from the data sets, qualitative variables are encoded by dummy to put into use for the model.

| Code | X1  | X2  | X3  | X4  | X5  | X6  | X7  | X8  | X9  | X10 | Y    |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 6    | 1   | 4   | 8,715.00 | 2,220.00 | 1 | 0 | 1 | 0 | 62 | 1 | 0 | 10,062,982,570 |
| 7    | 2   | 3   | 1,774.00 | 591.00 | 0 | 1 | 1 | 0 | 38 | 1 | 0 | 6,771,762,663 |
| 8    | 3   | 4   | 1,958.00 | 489.50 | 1 | 0 | 1 | 0 | 20 | 1 | 0 | 3,502,997,380 |

Add the minimum value for the X1 variable (cell B37) by the formula:
\[ B37 = \text{MIN}(B6:B32) \]  

Similarly, add the maximum value for the X1 variable (cell B38) by the formula:
\[ B38 = \text{MAX}(B6:B32) \]  

Similarly, copy the formula for all cells.
### Table 2. Organizing input data.

|   | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   | K   | M   |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 36| MIN | X1  | X2  | X3  | X4  | X5  | X6  | X7  | X8  | X9  | X10 | Y   |
| 37| MAX | 5   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 3,502,997,380 |

Set up the formula in cell B46 for the X1 variable as follows:

\[
B46 = \frac{2(B6 - B37)}{(B38 - B37)} - 1
\]

As a result, the values after scaling will get the value a range of [-1 to 1].

The input parameter bias B1 is set to a value of 1.

### Table 3. Scaling input data.

|   | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   | K   | L   | M   |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 45| Code | X1  | X2  | X3  | X4  | X5  | X6  | X7  | X8  | X9  | X10 | B1  | Y   |
| 46| 1   | 0   | -0.3029 | -0.2078 | 1 | -1 | 1 | -1 | 0.7857 | 1 | -1 | 1 | -0.852112 |
| 47| 2   | -1  | -1.0000 | -0.9535 | 1 | 1 | 1 | -1 | -0.0714 | 1 | -1 | 1 | -0.926309 |
| 48| 3   | 0   | -0.9815 | -1.0000 | 1 | -1 | 1 | -1 | -0.7143 | 1 | -1 | 1 | -1.000000 |
| ...| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 72| 27  | 0   | 0.4829 | 0.8792 | 1 | -1 | 1 | -1 | 0.1786 | 1 | -1 | 1 | 1.000000 |

The Tanh transfer function is used at the hidden layer in the artificial neural network model:

\[
B94 = \text{TANH} (\text{SUMPRODUCT}(B46:L46, B78:L78))
\]

\[
B120 = \text{TANH} (\text{SUMPRODUCT}(B72:L72, B78:L78))
\]

### Table 4. Hidden layer data.

|   | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   | K   | L   | M   |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 93| Code | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  |
| 94| 1   | -0.987 | 0.288 | 1.000 | -0.943 | -0.943 | -0.721 | -0.850 | -0.943 | -0.843 | -0.977 | 0.998 | 1   |
| 95| 2   | -0.992 | -0.938 | 1.000 | -0.795 | -0.799 | -1.000 | 0.599 | -0.799 | -0.853 | -0.999 | 0.999 | 1   |
| 96| 3   | -1.000 | -0.385 | 1.000 | -0.997 | -0.997 | -1.000 | -0.995 | -0.997 | -1.000 | -1.000 | 0.983 | 1   |
| ...| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 120| 27  | 0.8710 | 0.999 | 1.000 | 0.683 | 0.683 | 1.000 | 0.976 | 0.683 | 0.751 | 0.972 | 1.000 | 1   |

The output parameter bias B2 is set a value of 1.

The Tanh transfer function at the output layer is used.

\[
N132 = \text{TANH} (\text{SUMPRODUCT}(B94:M94, B126:M126))
\]

### Table 5. Output layer data.

|   | A   | N   |
|---|-----|-----|
| 131| Code | Y (ANN) |
| 132| 1   | -0.847575 |
| 133| 2   | -0.929363 |
| 134| 3   | -0.998421 |
| ...| ... | ... |
| 158| 27  | 0.896051 |
Table 6. Scaling output data and errors.

| Code | Y (Actual)   | Y (Prediction) | Errors (%) |
|------|--------------|----------------|------------|
| 196  | 10,062,982,570 | 10,264,242,245 | 2.00       |
| 197  | 6,771,762,663  | 6,636,327,416  | 2.00       |
| 198  | 3,502,997,380  | 3,573,057,328  | 2.00       |
| ...  | ...           | ...            | ...        |
| 222  | 92,218,821,630 | 87,607,880,542 | 5.00       |

At this step, the values of the predictive scaling back from data of final ANN output model according to the formula below:

\[ N_{196} = (N_{132}+1)*($M$38 - $M$37)/2+$M$37 \]  

The optimal weight is set by the method of training and testing combined with the error back propagation. Errors for training are limited up to 2% and testing are limited up to 5%.

Table 7. The weight matrix input layer - hidden layer.

|     | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   | K   | L   |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 77  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | bias 1 |
| 78  | 1   | 0.561 | -0.671 | -0.775 | 0.512 | 0.580 | -0.022 | 0.090 | -0.322 | -0.065 | 0.405 | 0.670 |
| 79  | 2   | -0.319 | -0.685 | 0.237 | -0.001 | -0.560 | 0.174 | 0.262 | -0.068 | 0.612 | -0.657 | 0.319 |
| 80  | 3   | 0.464 | -0.008 | 0.181 | -0.425 | -0.047 | 0.645 | -0.317 | -0.070 | 0.412 | 0.260 | 0.679 |
| 81  | 4   | 0.325 | -0.389 | 0.598 | -0.050 | 0.145 | 0.218 | -0.090 | -0.049 | 0.119 | 0.208 | 0.095 |
| 82  | 5   | -0.203 | -0.480 | -0.205 | 0.323 | -0.157 | -0.279 | -0.802 | 0.152 | -0.276 | -0.381 | 0.451 |
| 83  | 6   | 0.300 | -0.628 | -0.557 | 0.488 | 0.231 | -0.336 | 0.315 | -0.133 | -0.166 | -0.021 | -0.197 |
| 84  | 7   | 1.557 | 0.370 | -0.237 | -0.262 | 0.244 | 0.444 | 0.867 | -0.153 | -0.501 | 0.452 | -0.488 |
| 85  | 8   | 0.102 | -0.212 | 0.209 | 0.424 | 0.438 | -0.126 | -0.708 | -0.003 | 0.278 | -0.660 | -0.398 |
| 86  | 9   | 0.235 | 0.609 | 0.606 | -0.552 | 0.379 | -0.342 | -0.450 | 0.263 | -0.214 | 0.172 | 0.253 |
| 87  | 10  | -0.645 | -0.461 | -0.135 | 0.605 | -0.049 | 0.147 | 0.134 | -0.453 | 0.324 | -0.047 | 0.190 |
| 88  | 11  | 0.670 | -0.292 | 0.156 | -0.875 | -0.257 | 0.622 | 0.193 | 0.523 | -0.249 | -0.067 | 0.416 |

Table 8. The weight matrix hidden layer - output layer.

|     | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   | K   | L   | M   |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 125 | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | bias 2 |
| 126 | 1   | -0.826 | -0.389 | 0.184 | -0.456 | 0.053 | -0.430 | 1.142 | -0.202 | 0.382 | 0.043 | -0.854 | 0.308 |

Table 9. Final test results of ANN model.

|                  | Model 1 (basic) | Model 2 (case 1) | Model 3 (case 2) | Model 4 (optimal) |
|------------------|-----------------|------------------|------------------|-------------------|
| Prediction       | 83,111,803,680  | 83,482,563,805   | 84,961,012,259   | 89,121,837,430    |
| Error            | 9.88%           | 9.47%            | 7.87%            | 3.36%             |

Note:
Model 1: The basic ANN model was based on the Excel algorithm
Model 2: Test Excel algorithm with SPSS (case 1)
Model 3: Test Excel algorithm with SPSS (case 2)
Model 4: The optimal ANN model was set on Excel algorithm

### Table 10. Program of estimation.

| Code | Main variables          | Data            |
|------|-------------------------|-----------------|
| 1    | Number of floors       | 3               |
| 2    | Gross floor area (m²)  | 16,000          |
| 3    | Ground floor area (m²) | 4,500           |
| 4    | Foundation structure   | pile cap        |
| 5    | Roof structure         | tole            |
| 6    | Number of classes      | 45              |
| 7    | Form of construction   | new construction|

**Construction school cost (VND)** 89,121,837,430

5. Discussion
The estimation cost model get high accuracy (over 90%), the study adds a new tool applied in construction cost management field, the software is relatively convenient and easy for users. This model is limited to a number of factors, not yet fully showing the factors that affecting the construction costs of school buildings, or education projects. The data sets used to build for the model is limited.

6. Conclusions
The study has identified the main factors affecting the cost of construction of school in Ho Chi Minh City. The cost estimation model was established by excel algorithm based on artificial neural network theory. Excel algorithm model is tested and optimized by Statistical Package for the Social Sciences software. As a result, the basic estimation cost model reached the error rate of 9.88%. The optimal estimate cost model reached the error rate of 3.36%. The optimal of weights was used to set up the automatic calculation program.

References
[1] An, Sung-Hoon., Kim, Gwang-Hee., Kang, Kyung-In. (2007), “A case-based reasoning cost estimating model using experience by analytic hierarchy process”, Building and Environment, Vol. 42, Iss. 7, pp. 2573-2579.
[2] Arafa, Mohammed., Alqedra, Mamoun. (2011), “Early Stage Cost Estimation of Buildings Construction Projects using Artifical Neural Networks”, Journal of Artifical Intelligence, Vol. 4, Iss. 1, pp. 63-75.
[3] Cho, Hong-Gyo., Kim, Kyong-Gon., Kim, Jang-Young., Kim, Gwang-Hee. (2013), “A Comparison of Construction Cost Estimation Using Multiple Regression Analysis and Neural Network in Elementary School Project”, Journal of the Korea Institute of Building Construction, Vol. 13, No. 1, pp. 66-74.
[4] Ji, Sae-Hyun., Park, Moonseo., Lee, Hyun-Soo. (2011), “Cost estimation model for building projects using case-based reasoning”, Canada Journal of Civil Engineering, Vol. 38, Iss. 5, pp. 570-581.
[5] Khanh, Le Xuan (2016), “Estimate the construction schools cost by case-based reasoning (Case-based reasoning - CBR)”, Master thesis in Construction Management, Ho Chi Minh City University of Technology.
[6] Khoa, Phan Van (2006), “Estimating the investment cost of apartment project using Neural
Networks”, Master thesis in Construction Management, Ho Chi Minh City University of Technology.

[7] Kim, Chung-Yung., Son, Jea-Ho. (2006), “A Study on the Model of Artificial Neural Network for Construction Cost Estimation of Educational Facilities at Conceptual Stage”, Korea Journal of Construction Engineering and Management, Vol. 7, No. 4, pp. 91-99.

[8] Kim, Gwang-Hee., An, Sung-Hoon., Kang, Kyung-In. (2004), “Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning”, Building and Environment, Vol. 39, Iss. 10, pp. 1235-1242.

[9] Kim, Gwang-Hee., Shin, Jae-Min., Kim, Sangyong., Kim, Yoonseok. (2013), “Comparison of School Building Construction Cost Estimation Methods Using Regression Analysis, Neural Network, and Support Vector Machine”, Journal of Building Construction and Planning Research, Vol. 1, No. 1, pp. 1-7.

[10] Kim, Jin-Won., Lee, Baek-Rae., Kim, Ju-Hyung., Kim, Jae-Jun. (2011), “A Study on the Construction Cost Estimation Model According to the Spatial Planning of Educational Facilities Using Regression Analysis – For the BTL Project in the Gyeonggi-do Region”, Journal of the Architectural Institute of Korea (Planning and Design), Vol. 27, No. 10, pp. 103-110.

[11] Luu, Van Truong, Kim, Soo-Yong. (2009), “Neural Network Model for Construction Cost Prediction of Apartment Projects in Vietnam”, Korean Journal of Construction Engineering and Management, Vol. 10, Iss. 3, pp. 139-147.

[12] Roxas, Cheryl Lyne C., Ongpeng, Jason Maximino C. (2014), “An artificial neural network approach to structural cost estimation of building projects in the Philippines”, Presented at the DLSU Research Congress, De La Salle University, Manila, Philippines, March 6-8, 2014.

[13] Shin, Jae-Min., Kim, Gwang-Hee. (2012), “A Study on Predicting Construction Cost of Educational Building Project at Early Stage Using Support Vector Machine Technique”, Journal of Korean Institute of Educational Environment, Vol. 11, Iss. 3, pp. 46-54.

[14] Son, Jea-Ho., Kim, Sung-Kyum., Kim, Jae-On. (2008), “A Study on the Analysis and Estimation of the Construction Cost Artificial Neural Network in the BTL Projects for Educational Facilities”, Journal of Architectural Institute of Korea, Vol. 24, No. 6, pp. 135-142.

[15] Tuan, Nguyen Anh (2007), “Evaluate fluctuations of costs and time in construction projects using Neuron Network (ANN)”, Master thesis in Construction Management, Ho Chi Minh City University of Technology.

[16] Tuan, Truong Anh (2007), “Evaluate overestimated cost of school projects using control method by statistic”, Master thesis in Construction Management, Ho Chi Minh City University of Technology.