Model for predicting cost of rural road projects in Thailand

Wai Phyo Lin$^{1,2}$ and Wasaporn Techapeeraparnich$^{1,2}$

$^1$Civil and Environmental Engineering Department, Faculty of Engineering, Mahidol University, Salaya Campus, Thailand
Email: wphyolin09@gmail.com; wasaporn.tac@mahidol.ac.th

Abstract. Nowadays, construction cost plays an essential role in various projects, which are buildings, roads, railways, and bridges projects. Conceptual cost estimation in feasibility study is require high accuracy and less validation error especially in construction projects at early stages. The more improvement of estimation techniques, it would lead to the lesser problems of cost overrun, and extra expenses. This paper developed a cost estimation model for new constructed rural road projects. Considering the estimation methods for predicting the cost model, parametric method based on regression learner and NN method are applied. Previously, many researchers studied the cost applicable model by using various computer applications, so that this paper differed to compare these methods based on MATLAB. Accordingly, 44 road projects were compiled from DRR database, after that identifying the effective cost parameters referred on collected data. Subsequently, the data implementation process was focusing with regression learner based on automated regression training, and lastly, with NN toolbox concentrated on nftool for simulating the cost models. Another 7 road projects are tested with holdout validation process for these models. Due to this validation, we compare the predicted cost models within these methods. Finally, the developed models are reliable for not only the conceptual phase of future rural road projects but also the related construction fields can be recovered about the cost model creation.

1. Introduction

With a growing population in Thailand, it is difficult to create a well-designed urban public transportation system that meets the needs of the entire population for improving transportation corridors. So, Thailand’s government investment in infrastructure has increased each year especially with horizontal infrastructure projects such as roads, railways, and bridges which will lay a foundation for growth and well transportation [1]. Literally, road construction is supporting national development strategies such as tourism, logistics, and agriculture. For these developments, there should be considered not only in urban but also in rural regions as well. Rural roads are mostly built to serve for both economic and social activities that are accessible to local schools, hospitals, and others emergency [2].

Rural roads are a fundamental element in the provision of access in rural areas, so that roads are crucial to economic and social development. It is surprising that the data on rural roads in Thailand region are not difficult to find, but it is hard to find a reliable statistic [3]. Due to the problems of rural roads work with increased funding uncertainties and shrinking budgets, it is more important than ever to ensure proper allocation of funds for highway projects in the preconstruction phase [4]. Although extremely important and often overlooked, project cost estimating has many obstacles to overcome throughout the process [5]. For the above reasons, the following attempts will be considered
different from previous studies by developing the reasonable cost models with high accuracy. So, the aim of this research represents:

- Description of road cost influential parameters for predicting cost model, and
- Development of road cost prediction model based on regression learners, and neural network methods

2. Identification factors

According to engineering and management, the crucial part of a road construction project represents its financial demands. Bill of quantities (BOQ) in construction project represents the most effective parameter in the construction project cost [6]. However, precise and feasible studies of construction projects are not simple to do since there have been many factors varying from one project to another such as project size, type, location, the season of the year, material type, structure, and method of construction [7].

In addition, between the several independent variables in highways’ projects, the size of the project presents an enormous variation [8]. For instance, location of the project represents important factor correlated to a populated area, ground condition and profile, and environmental impacts. Material alternatives such as several types of asphalt pavements, concrete slabs, and other types of materials contain variation in the value of project costs [9]. Referring to these important variables, the newly constructed project in this paper would be focusing on not only the engineering aspects of the bill of quantities as input factors but also manpower and their costs as outputs.

Consequently, the identification of cost parameters distinguished into six input factors for cost model development in terms of investigation of rural road projects which is highly influenced parameters summarized as below table 1.

| No | Input factors | Descriptions |
|----|---------------|--------------|
| 1  | Length        | Project road distance (km) |
| 2  | Width         | Road width of each lane (m) |
| 3  | Number of lanes | Number of lanes per direction |
| 4  | Pavement type | Asphalt concrete structure, prime coat, tack coat, coating, base, and subbase |
| 5  | Earthwork     | Embankment, and excavation |
| 6  | Miscellaneous | Drainage, guardrail, lighting, precast, retaining wall, seeding, signalization, signage, planting and extra expenses |

The above-selected input factors are used in expressing the cost model for the rural road constructed project BOQ data used in this research.

3. Data analysis and cost modelling

A cost database of 51 rural road projects was collected from DRR only for new constructed ones. An analysis of collected data from DRR can be considered into two steps as follows.

3.1. NN development

Proofing the good fitting tools for the practical functions, neural network should simply fit with (nftool) through the dataset and describe the problem in NN toolbox. So, this process passed through using with MATLAB R2018b NN toolbox until models result come out are as follows.

Stage 1 Database Analysis
Stage 2 Neural Network fitting and training
Stage 3 Case Study for Validation
Stage 1 Database Analysis
In this study, having a database of 51 projects; it is decided to utilize about 70% of the database for training, 15% for testing and 15% for validation.

Stage 2 Neural Network fitting and training
Once neural networks fitting, the data set firstly imported as a numeric matrix then classified input and output factors. During this process, the neural network architect came out and filled the acceptable hidden layers which is 25 almost four times greater than input numbers. The below figure 1, is NN toolbox in MATLAB with sigmoid function applied and describes with the data numberings in each segment.

![Figure 1. NN Toolbox in MATLAB](image)

In this study, the default Levenberg-Marquardt used as the training test and showed the validation error failed to decrease for twenty-one iterations until its stop.

Stage 3 Case Study for Validation
The MSE display the network outputs with respect to targets for training, validation, and test sets. For a perfect fit, the result data performed to be the best validation at epoch 15.

![Figure 2. Trained result](image)  ![Figure 3. Validated result](image)

Both figure 2 and 3, all input data in training results showed a significance outcome as the form of $R=1$ and the result for validation also found on the way of line graph around confidential level at 0.05. For the final stage of the model simulation, the MATLAB function applied to analyze in matrix and cell array arguments for the output results. So, the below equation is showing about the cost model processes referred from MATLAB function.

Input weight layer,
\[ H = \sigma (W \cdot X^T + b) \]  

(1)

Non-linear cost model equation,

\[ Y = W_2 \cdot [\sigma (W_1 \cdot X^T + b_1)]^T + b_2 \]  

(2)

where, \( \sigma \) is sigmoid function, \( X \) is input vector, \( W \) is controller weight of input as in matrix, and \( b \) is the bias.

3.2. Regression analysis

The reason to find the model from using regression analysis based on regression learner, is to test as another way for comparing NN toolbox. The model is developed through the following phases using regression learner in MATLAB summarized as follows:

Stage 1 Database Analysis
Stage 2 Apply in regression learner
Stage 3 Case Study for Validation

Stage 1 Database Analysis

The first stage is used for importing data and selected for all dataset in the column because inputs (modeling factors) are in the first six columns and output (project cost) is the last.

Stage 2 Apply in regression learner

To predict the cost models, the regression learner can also be used to train regression models in MATLAB. Automated regression training was used to search for the best regression model type from the linear, interaction, robust, and stepwise.

Table 2. Best linear regressions

| Linear Regression  | RMSE  |
|--------------------|-------|
| Linear             | 5.24e-10 |
| Interactions       | 5.62e-10 |
| Robust             | 5.24e-10 |
| Stepwise           | 5.35e-10 |

As shown in table 2, linear one is suitable as the best regression model of the projects and the above results were implemented only with automated regression model training because the training of this option is fast to fit the multiple models simultaneously.

Stage 3 Case Study for Validation

The training set assesses its performance with holdout validation set. Therefore, this validation is implemented with 15% of the database for all results. From extracting the results, the coefficients of cost model would be printed out from the exported model in the workplace as shown in table 3.

Table 3. Linear model coefficients

| Factors   | Estimate  |
|-----------|-----------|
| (Intercept) | -2.36e-10 |
| X1        | -1.02e-08 |
| X2        | 1.01e-08  |
| X3        | 4.71e-09  |
| X4        | 0.7621    |
| X5        | 0.1456    |
| X6        | 0.0923    |
Therefore, the equation can be simulated with these estimations from the coefficients. The general equation is

\[ Y = C^T X + b \]  

(3)

where, \( Y \) is output cost, \( C \) is the coefficients, \( X \) is the factors and \( b \) is intercept.

So, the cost model of this project described as below.

\[ Y = -1.02e - 08X_1 + 1.01e - 08X_2 + 4.71e - 09X_3 + 0.7621X_4 + 0.1456X_5 + 0.0923X_6 - 2.36e - 10 \]  

(4)

where, \( X_1, X_2, X_3, X_4, X_5 \) and \( X_6 \) are length, width, lanes, pavement types, earthwork, and miscellaneous.

4. Comparing between linear and nonlinear cost model’s performance

All training models from dialog box can be compared the accuracy of the results in regression model process and neural network performance. These test results of two models are shown in figure 4 and 5.

![Figure 4. Trained result](image1)

![Figure 5. Validated result](image2)

Above, linear is rather accurate result than non-linear, so that the below table 4 clearly compared between linear and non-linear of their performance of cost models results.

| Table 4. Cost model performance between linear and non-linear |
|-------------------------------------------------------------|
| Results | Linear | Non-linear |
| MSE     | 2.57e-19 | 2.21e-6 |
| R       | 1       | 9.96e-1 |
| RMSE    | 5.24e-10 | 1.49e-3 |

As cost model results, linear one is the most accurate in this paper. However, the models were developed based on 51 sets of data collected in Thailand and any model's performance is going to be data dependent. For small data sets with less variance, linear model works well, and neural network-based models fail to find meaningful relationships. However, as the complexity and scale of the datasets increase, non-linear models with bigger learning capacity will become more relevant.

5. Conclusion

This research work aimed at formulating a reliable and practical model for conceptual cost estimate that can be used by actual utilizations of the rural road projects. In neural network, the model was trained using Levenberg-Marquardt algorithm. According to [8, 10], the output results were acceptable as mentioned in table 4. As a practical issue, the model was adapted in GUI as MATLAB function to classify the input parameters for the projects and, output results for predicting the cost.
To develop a linear regression models, it is very useful, especially in its simplicity and ability to be handled by simple program as regression learner. It has also a good benefit in estimating project cost at early stages of the project [10].

Accordingly, the relationship between the independent and dependent variables of the developed models are acceptable because of extracting as the best regression results in two methods. Also, the predicted values from a forecast model fit with the actual data, but the root mean squared error of linear model got the less result error than the non-linear one so that the cost model from linear is the best performance in this paper. However, it is also seemed that neural network method based on NN toolbox could be a reasonable tool to solve problems with uncertainties as well [9].

The study is to simulate the cost model with acceptable stage for future rural road projects which are supporting for owners, consultants and contractors under this kind of projects. In the future, if there is a certain amount of distance to estimate the final cost, the cost model will be predicted overall cost. Another recommendation of future work will be developed as cost estimation model by analyzing with other methods such as fuzzy logic, case based reasoning, and other up to date techniques [8, 9].

Acknowledgments
The authors thank professor, colleagues, technical staff and financial support from faculty of graduate studies and department of civil engineering at Mahidol university. Lastly, the authors appreciate for the valuable comments by ICCEMS committee members and staffs.

6. References
[1] Meeampol S and Ogunlan S O 2006 Factors affecting cost and time performance on highway construction projects: evidence from Thailand J. Financial Management Property Construction 111 3-20
[2] Pomlaktong N, Jongwilaiwan R, Theerawattanakul P and Pholpanich R 2014 Road transport in Thailand In priorities and pathways in services reform: Part II—political economy studies pp 227-243
[3] Doniges C, Edmonds G and Johannessen B 2007 Rural road maintenance: Sustaining the benefits of improved access Bangkok International Labour Office
[4] Hollar D A 2011 Predicting preliminary engineering costs for highway projects (Doctoral dissertation, North Carolina State University)
[5] Turochy R E, Hoel L A and Doty R S 2001 Highway project cost estimating methods used in the planning stage of project development Virginia Transportation Research Council (No. VTRC 02-TAR3)
[6] Skitmore R M and Ng S T 2003 Forecast models for actual construction time and cost Building and environment 38 8 1075-1083
[7] Wilmot C G and Mei B 2005 Neural network modeling of highway construction costs J. Construction Engineering Management 131 7 765-771
[8] Adel K, Elyamany A, Belal A M and Kotb A S 2016 Developing Parametric Model for Conceptual Cost Estimate of Highway Projects Int. J. Engineering Science 6 7 1728-1734
[9] Attal A 2010 Development of neural network models for prediction of highway construction cost and project duration (Doctoral dissertation, Ohio University)
[10] Peurifoy R L and Oberlender G D 2002 Estimating construction costs (New York: McGraw-Hill)