№ 3 (4), 2022 ISSN: 2682 – 9517 (print) ISSN: 2683 – 0930 (online)

INSTITUTE OF INFORMATICS AND INNOVATIVE TECHNOLOGIES

Construction Projects Risk Prediction Using Artificial Intelligence Model

Phuong Thao Cao
Faculty of Construction Management
University of Transport and Communications
Hanoi, Vietnam
thaocp@utc.edu.vn

Thi Hau Nguyen
Faculty of Construction Management
University of Transport and Communications
Hanoi, Vietnam
haunt@utc.edu.vn

***Abstract*. Rick delay in construction projects is one of the major problems that affect construction quality and implementation costs. However, the identification of factors as well as the prediction of risk delay is a major challenge. Currently, there are many methods of risk identification using statistics, analysis, and expert opinion consultation. This paper presents a method that detects and predicts risk using an intelligence model. The criteria to assess risks will serve as inputs in the neural network, and the output will be the ranking of risk delay in three levels. The proposed neural network was tested on 101 construction projects in Vietnam and the result shows the high accuracy level of this method in construction risk delay detection.**

Keywords: project management, risk delay prediction, neural network

I. Introduction

Delay in traffic construction projects is a prominent problem in construction projects in general and traffic construction in particular. There are many reasons for the delay, from subjective to objective, although beyond the control of both the investors and the contractors. The delay in projects will affect the efficiency of exploitation, capital pool, capital accumulation and costs increased. Therefore, predicting the risk delay in construction projects is receiving much attention from researchers and project managers. Currently, with the development of computational systems, especially data sources about many projects, prediction models based on old project data have developed. It is urgent to build an artificial intelligence model to predict the risk delay in traffic construction projects based on the collected data.

There have been many studies focusing on the causes of delay, even predicting the possibility of delay in projects. Linear regression model has been used to estimate the relationship between time and cost for construction projects in Malaysia [1]. This model is used by project managers and owners to estimate the average time to complete a project. The multivariate regression model is also used to analyze data related to the effectiveness of projects over time and gives positive results in estimating time in construction projects [2]. With objective factors, the Monte Carlo analysis method is used to predict the effects on the progress of the project by analyzing the relationship between the objective factors in construction projects [3]. In Jordan, the time and cost of construction projects are predicted through statistical models and hypothesis testing also gives reliable results [4]. To estimate the preventive time of construction projects, the AHP method has been used to evaluate on many different criteria [5]. The stochastic Monte Carlo method is also used to manage the risk of delay in BOT projects and has predicted risks with many objective factors [6].

Although there are many research works that apply artificial intelligence in risk delay progress, these methods focus on evaluating the cause of risk delay or the efficiency of the projects. In this paper, we present a method of risk delay prediction in construction projects using a neural network. Criteria and factors to the delay in the construction project will be used as the inputs for a multi-layer perceptron neural network and the output are the thlevelsevel of risk delay including low, medium, and high risk delay progress. We have tested the method using data from 200 construction projects in Vietnam. The experimental results show the efficiency of the method. In the first section, we describe the progress of the project and how the neural network is applied to risk delay prediction. In the next section, real data will be used to illustrate the performance of the method. Finally, we a draw conclusion about the study and its implication for future work.

II. MAIN CONTENTS

A. Factors affection to risk delay in construction project

The most important factors that affect progress in construction projects were identified from a survey document, and these factors were classified into different sources. These sources include investors, design consultants, contractors, projects, materials, equipment, labor and external factors. To obtain information on delay issues and factors in construction projects, surveys were conducted with 101 experts in the construction field. Through this survey, sources and factors are identified related to the construction industry. Based on the survey and survey documents, the most important sources and factors of delay are identified in Table 1.

Table 1. List of factors affected to delay in construction projects.

Construction investment procedures	
1. Delayed preparation, appraisal and approval of projects	*2. Difficulty in project adjustment and design*
3. Delay in design verification and appraisal	*4. Changes in state policies on construction activities*
5. Violation of construction investment procedures	
Ground clearance	
6. Compensation and support for site clearance and resettlement	
Investor/Project Management Board	
7. The investor's financial ability does not meet the project requirements	*8. Weakness of the PMU/insufficient experience in project management*
9. The investor is late in paying for the completed work	*10. Choosing an incompetent design consultant*
11. Investor changes decision/changes investment objective	*12. Selecting an inexperienced construction contractor*
13. Contract is not tight, lack of binding	*14. Investors are slow to make decisions/The process of solving projects related to the project is not good*
15. Lack of information exchange among project stakeholders	*16. Delay in the acceptance of completed works*
Consulting contractor	
17. Inexperienced survey/design consultant	*18. Wrong in design and ambiguous in design drawings*
19. Change design	*20. Estimates are not accurate*
21. Inexperienced supervision consultant	*22. Poor project planning*
Construction contractor	
23. The contractor's finances are in trouble	*24. Poor organization of site management and supervision*
25. The difference between the actual conditions and the survey, design	*26. Rework due to wrong implementation of items / not in accordance with the design*
27. Weakness of subcontractors or having to change subcontractors too much	*28. Outdated/inappropriate construction methods and technology*
29. Lack of information exchange between consultants and construction contractors	*30. An occupational accident occurred during construction*

Materials	
31. Market shortage of materials/scarce resources	*32. Slow delivery of materials*
33. Substandard material	*34. Material prices soar*
35. Delay in importing materials separately from abroad	
Labor	
36. Insufficient human resources	*37. Lack of skilled workers*
38. Low labor productivity	
Construction machine	
39. Terrorism and political instability	*40. Inflation*
41. Changing the legal bases of the State	*42. Ground problems arise*
External factors	
43. Unstable economic/political situation	*44. Inflation/price volatility in the market*
45. Changing the legal bases of the State	*46. Issues with the ground*
47. The problem of the surrounding population	*48. Influenced by cultural and social factors*
49. Bad weather condition	*50. Affected by disease*

These factors will be quantification by coding in range from 0 to 1. These values will be the inputs of neural network. Outputs of neural network are risks, measured on three levels of audit risk as low, medium and high.

B. *Neural network*

An Artificial Neural Network (ANN) is a computational model that simulates biological neurons and functions in the brain. Typically, an ANN has layers of interconnected nodes. The nodes and their inter-connections are similar to the network of neurons in the brain. Any basic ANN will always have multiple layers of nodes, specific connection patterns and links between the layers, connection weights and activation functions for the nodes that convert weighted inputs to outputs. The learning process for the network typically involves a cost function and the objective is to optimize the cost function (typically minimize the cost). The weights keep getting updated in the process of learning.

For risk delay detection, we have considered them as a classification problem. In this paper, we use a multiple outputs three-layer structure of a multilayer perceptron (MLP) neural network. Although this classifier needs quite a large training time but it is able to process data and classification fast [7]. Figure 1 presents an example of MLP structure that consists of one input layer, one hidden layer and one output layer [8].

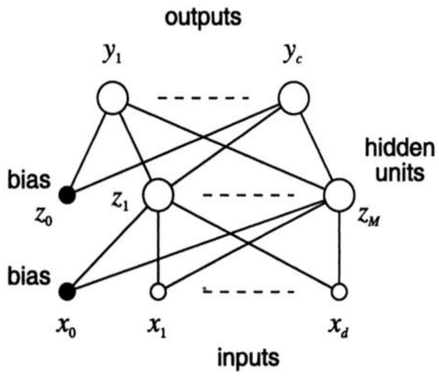

Fig. 1. Two layer feed-forward neural network

Let x_i, i = 1..d is the input value to the network, the output forms M linear combinations of these inputs to $a_j^{(1)}$ as:

$$a_j^{(1)} = \sum_{i=1}^{d} w_{ji}^{(1)} x_i + b_j^{(1)}, j = 1, \ldots, M \qquad (1)$$

where w_{ji} are element of the weight matrix and b_j are the bias parameters associated with the hidden unit. Also, each variable aj was associated with each hidden unit and then transformed by the non-linear activation functions of the hidden layer. The output of the hidden units are then given by:

$$z_j = \tanh(a_j^{(1)}), j = 1, \ldots, M \qquad (2)$$

№ 3 (4), 2022 ISSN: 2682 – 9517 (print) ISSN: 2683 – 0930 (online)
INSTITUTE OF INFORMATICS AND INNOVATIVE TECHNOLOGIES

The z_j are then combined with weights and biases of the next layer to produce values $a_k^{(2)}$

$$a_k^{(2)} = \sum_{j=1}^{M} w_{kj}^{(2)} z_j + b_k^{(2)}, k = 1,\ldots,c \quad (3)$$

where *c* is the number of outputs.

These values are then passed through the output layer to produce output values $y_k, k = 1..c$. There are several forms of activation functions. For the classification purpose, we consider the logistic sigmoidal activation functions as follows:

$$y_k = \frac{1}{1+\exp(-a_k^{(2)})} \quad (4)$$

The network needs to train to model the data in order to make a best predictions of new input data. In this paper, we consider the back propagation algorithm [8]. Assume we have the target vector t for input data *x*, the error of the network, *E*, is defined as:

$$E = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{c}(y_k^n - t_k^n) \quad (5)$$

Where y_k^n is the actual value of k[th] output unit for the n[th] input pattern, t_k^n is desired value of the k[th] output unit for the *n*[th] input pattern.

The derivative of E with respect to the second layer weights are given by:

$$\frac{\partial E}{\partial w_{kj}^{(2)}} = \sum_{n=1}^{N} \delta_k^{(2)n} z_j^n, \quad (6)$$

where $\delta_k^{(2)n} = y_k^n - t_k^n$ is the error for the output unit on pattern *n*.

Also, the derivatives for the output unit biases are given by:

$$\frac{\partial E}{\partial b_k^{(2)}} = \sum_{n=1}^{N} \delta_k^{(2)n}. \quad (7)$$

The back propagation can be calculate as below

$$\delta_j^{(1)n} = g'(a_j^{(1)n})\sum_{k=1}^{c} w_{kj}^{(2)}\delta_k^{(2)n} = (1-(z_j^n)^2)\sum_{k=1}^{c} w_{kj}^{(2)}\delta_k^{(2)n}, \quad (8)$$

The derivatives of the first layer weights are then given by

$$\frac{\partial E}{\partial w_{ji}^{(1)}} = \sum_{n=1}^{N} \delta_j^{(1)n} x_j^n, \quad (9)$$

And the derivatives for the hidden unit biases are given by

$$\frac{\partial E}{\partial b_j^{(1)}} = \sum_{n=1}^{N} \delta_j^{(1)n}. \quad (10)$$

The difference between the calculated output and the desired output is back-propagated to the previous layers, usually modified by the derivative of the activate function, and the connection weights are normally adjusted using the Delta Rule. This process proceeds for the previous layers until the input layer is reached.

C. *Risk delay prediction model*

In this paper, a neural network model has been used to predict the risk delay in the construction project. The neural network is trained using the factors described in the table 1. The framework of the method is shown in figure 2.

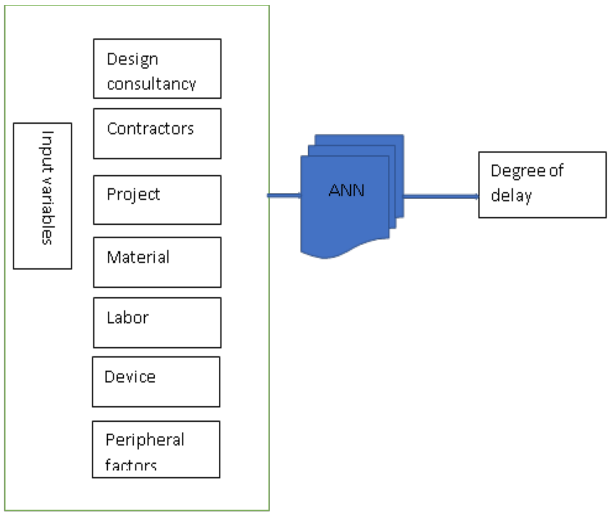

Fig.2. The delay prediction framework

The neural network structure here is designated with four layers, one input layer, two hidden layers and an output layer. The number of nodes in each layer is selected by experiment, we used twelve nodes for the first hidden layer, 32 nodes for second hidden layer, 16 nodes for the third layer and three nodes for the output layer, corresponding three levels of audit risk, low, medium and high, respectively. Activation function for each hidden layer was rectified linear unit (ReLU), and sigmoid function for the output layer. The network architecture is shown in table 2.

TABLE 2. NETWORK ARCHITECHTURE

Layer	Node
Input (factors)	50
Dense	32
Dense1 (hidden layer)	16
Output (three level of risk delay)	3

D. *Results*

Data used in this paper have been collected from 101 construction projects in Vietnam. The survey forms for the risk delay includes 50 factors as described in table 1. From the data collected, we quantify these factors to form the matrix with score from 0 to 1. In this data set, each row presents data from one project and each column presents a factor. These factors have brought to the input of neural network for train the model.

The data set was divided into two parts, 70% for training and 30% for testing. The program was written using Python with keras backend tensorflow support GPU run on the computer Core i7, RAM 8GB. The training result is shown in figure 3.

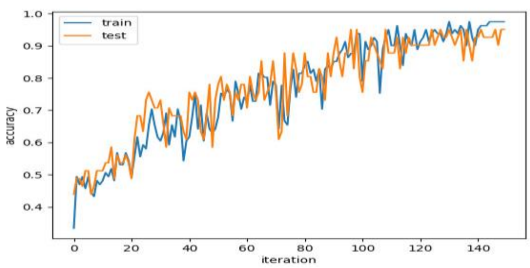

Fig. 3. Performance of training process

Figure 3 indicates that the values of accuracy in both testing set and training set are equal at the starting point. In both training and testing phases, we see that the accuracy increases with each iteration, from the 120th iteration, the accuracy has been achieved close to the optimal level. In the testing phase, the accuracy curve of the test data also increases linearly with the training accuracy curve, which shows that the network model is suitable for the collected data set. Table 3 shows the confusion matrix for the test result.

TABLE 3. CONFUSION MATRIX FOR THE RISK DELAY PREDICTION

	Predicted label		
True label	10	0	0
	2	17	0
	0	3	9
	Low	**Medium**	**High**

As in the confusion matrix, the model can predict all 10 projects in the low risk delay. In the medium risk delay, the model predicts correct 17 projects and miss classification of two projects. In the high risk delay project, the model predict 9 projects correctly and miss 3 projects. Through the matrix table, we can see that the accuracy in detecting the possibility of delay of the project is 88%. The prediction accuracy will be higher as more data of the project is available.

III. CONCLUSION

This paper proposed the neural network structure to detect the risk delay in construction projects. By quantifying the factor survey, these variables can be used as inputs to the neural network to train the model which can be used to detect the audit risk in any new project. The experimental results show the efficiency of the method. This method can be applied to information system to quickly detect the risk delay. This method can be applied to detect risk delay and can serve as a framework to identify risk in a comprehensive manner for construction projects.

ACKNOWLEDGMENT *(Heading 5)*

This research was supported by University of Transport and Communications with the grant number T2022-QLXD-003.

REFERENCES

[1] P. C. Albert Chan, "Time–cost relationship of public sector projects in Malaysia", International Journal of Project Management, Vol. 19(4), pp. 223-229, 2001.

[2] A. P. C. Chan, D. W. Chan, "Developing a benchmark model for project construction time performance in Hong Kong", Build. Environ. Vol. 39, pp. 339–349, 2004.

[3] K. Rezaie, M. Amalnik, A. Gereie, B. Ostadi, M. Shakhseniaee, “Using extended Monte Carlo simulation method for the improvement of risk management: Consideration of relationships between uncertainties”, Appl. Math. Comput. Vol. 190, pp. 1492–1501, 2007

[4] A. A. A. Hammad, S. M. A. Ali, G. J. Sweis, A. Bashir, “Prediction model for construction cost and duration in Jordan”, Jordan J. Civ. Eng. Vol. 2, pp. 250–266, 2008.

[5] D. Mohamed, F. Srour, W. Tabra, T. A. Zayed, “Prediction model for construction project time contingency”, In Construction Research Congress 2009: Building a Sustainable Future; ASCE: Reston, VA, USA, pp. 736–745, 2009..

[6] O. Dursun, C. Stoy, “Time–cost relationship of building projects: statistical adequacy of categorization with respect to project location”, Constr. Manag. Econ. Vol. 29, pp-97–106, 2011.

[7] 7. B. D. Ripley, “Pattern Recognition and Neural Networks”, Cambridge University Press, UK, 1996.

[8] Ian Nabney, 'Netlab: Algorithms for Pattern Recognition', Advances in Pattern Recognition, Springer, 2004.