

PREDICTING HIGHWAY PROJECTS' ACTUAL DURATION USING NEURAL NETWORKS

Euthimios Glymis¹, Athanasios Kanelakis², Georgios Aretoulis³, and Theodoros Mastoras⁴

Abstract: The current study predicts the actual duration of highway projects, based on the initial planned schedule. Highway projects suffer from delays and deviations, which in Greece are more often associated with law disputes, project financing, archaeological findings, environmental issues and private land acquisition procedures. In this research, data were obtained from 37 road projects in Greece and the purpose was to estimate the construction duration, using an artificial neural network. The Fast Artificial Neural Network (FANN) Tool program was used. FANN based on the available data, identifies the optimal training algorithm. The training algorithm and the activation function with the lowest mean square error (MSE) are selected. In order to achieve the best possible solution, numerous trials were made, applying different input data combinations, different variations in the architecture of the network, and different data values were used. This paper presents the three more reliable and effective networks produced from the current study. The results indicated that Artificial Neural Networks, employing the appropriate parameters do provide a relatively high accuracy in predicting actual construction time and more specifically appear as one of the most optimal methods for actual highway construction time prediction.

Keywords: Highway projects, actual project duration, estimated project duration, artificial neural networks.

1 INTRODUCTION

The procurement process of any major infrastructure project, like highway projects, includes the design phase, tender phase and construction phase. In the design phase, the owner determines the need for the realization of the project, carries out a feasibility study and completes the required designs. In the construction phase, the project is realized according to the contractual documents (Antonioni et al. 2013a,b). Prediction of required time for the realization of a construction project is of great importance for most professionals in the construction industry, and is traditionally being identified as a key success factor in a construction project (Chan and Kumaraswamy 1996). But despite the many advances in the science of project management, in recent decades construction projects exhibit poor performance regarding the planned duration and cost (Chan and Kumaraswamy 2002; Ng et al. 2001; Aretoulis et al. 2006; Aretoulis et al. 2015). Accurate prediction of the project duration in the early stages of planning is critical for success, as

¹ Graduate Civil Engineer, School of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, glymis@hotmail.gr

² Graduate Civil Engineer, School of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, kanelaka@civil.auth.gr

³ Assistant Professor, School of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, garet@civil.auth.gr

⁴ Doctor's degree in Applied Informatics, School of Information Sciences, University of Macedonia, Thessaloniki, Greece, mastoras@uom.gr

reported by Guerrero et al. (2013). Delays are a common problem in the construction industry from one task to another. They can be attributed to various reasons, causing serious financial consequences for the various project stakeholders involved in the construction process. From a basic preliminary analysis, the delay of a building project may be due to poor performance, such as failure of the contractor, changes in project design or bad weather (Guerrero et al. 2013). It is important to know these parameters, the relationship between them and their degree of influence on the construction time in order to increase the accuracy of the initial estimates (Chan 1998) and reduce the magnitude of the delays. All this led to the attempt to estimate the actual time of construction projects, and delays by means of expert / intelligent systems, based on artificial intelligence, such as neural networks. Neural networks can be used to predict the actual cost and time of a project, two parameters which are perhaps the most important for a project realization. According to Skitmore and Ng (2003) construction duration is the most influential parameter that affects the cost of construction.

Attal (2010); Elhag and Boussabaine (1998); Gunaydin and Dogan (2003); Cheng, Tsai, and Sudjono (2009); Berlin, Raz, Glezer and Zviran (2008); Zhang, Liu and Coble (2002); Tatari and Kucukvar (2010); ElSawy, Hosny and Razek (2001), made some of the first attempts to apply neural network tools to the problem of predicting the final project construction schedule. The following sections describe the research methodology, available data, and the conclusions of the current study. FANN Tool was used to develop an artificial neural network as one of the best possible approaches for the assessment of the duration of construction and more specifically, for a series of selected highway construction projects.

2 RESEARCH METHODOLOGY

The main goal of this study, is to create a simple predictive model for the duration of road construction projects, based on the development of an artificial neural network.

Originally, thirty-seven road projects were used as a sample. The project contracts which were examined extend over approximately the mid-1990s until the year 2010. The first step towards this task is to identify the factors which decisively influence the actual completion time of highway projects through comprehensive investigation of the literature and consideration of the local existing project conditions. According to Attal (2010), the parameters used internationally for forming reliable roadwork assessment models include but are not limited to the following: the location of the project, the size of the project, the duration of the contract, the design method, the type of project, the availability of equipment, the cost of labour and materials, the number of lanes, temperature, soil composition and the population of the construction area. Additional factors that may be used according to Attal (2010) in highway project's time forecasting models, include: the geometric design standards, the annual daily traffic, the length of the project, the type of bridge, length of bridges, the bridge width and the projects for procuring proper road signs.

As far as the current research is concerned, the selected input variables include the following: length of the highway project, number of lanes, number of associated technical projects, existence / number of bridges, existence / number of tunnels, height of slopes, the region where the project is being constructed, contract signing date, tender budget, initial contractual physical object, initial project schedule and actual project schedule.

As mentioned above, the FANN Tool application, was implemented to produce the neural network, which is a free-to-use software. The user can select one of the following

learning algorithms: FANN Train Incremental (gradually increasing), FANN Train Batch (clustering), FANN Train Rprop (Resilient backpropagation), FANN Train Quickprop.

The basic functions of the software are:

Neural Network → Detect → Optimum Training Algorithm: Each possible Training Algorithm is used for several epochs. All other parameters are fixed and the weight initialization is identical. The Training Algorithm showing the lowest MSE is picked (Fann Tool Users Guide 2016).

Neural Network → Detect → Optimum Activation Functions: Each possible Activation Function is used for several epochs. All other parameters are fixed and the weight initialization is identical. The Activation Function yielding the lowest MSE is picked (Fann Tool Users Guide 2016).

Neural Network → Train → Normal: Fixed topology training. The size and topology of the ANN is determined in advance and the training alters the weights in order to minimize the difference between the desired output values and the actual output values (Fann Tool Users Guide 2016).

Neural Network → Train → Cascade: Evolving topology training. The training starts with an empty ANN, only consisting of input and output neurons. Hidden neurons and connections are added during training, in order to reach the same goal as for fixed topology training (Fann Tool Users Guide 2016).

The Fann Tool software environment is presented in the following figure 1:

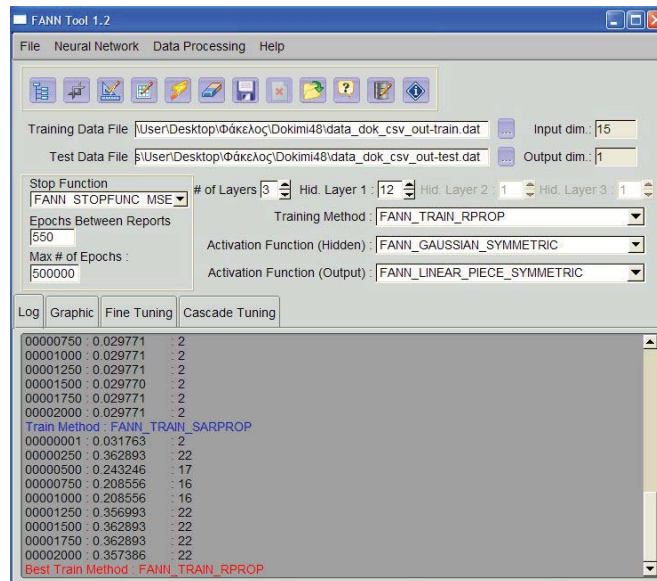


Figure 1. FANN Tool Software Environment

3 RESEARCH RESULTS

The data sets that were selected both for network training and control are 37. Part of these data were used for training the network and the rest for testing. In the context of the current research numerous trials have taken place. These trials focused on variations of the structure of the neural network, on different number of training and testing data,

alternative approaches for detecting optimum activation function, for training the network. The three most successful trials are presented in the following:

In the first test attempt, a network with the following structure 1-7-15 was selected. This structure means that the input level includes fifteen (15) neurons, the output level contains one (1) neuron and the hidden layer consists of seven (7) neurons. Twenty-five sets of data for network training were used and twelve sets for testing. Initially, the program processes the data, all thirty-seven (37) projects in total, and provides a training data file (Training Data File) and a test data file (Test Data File). The optimum options for producing an effective neural network are the following:

Optimal training algorithm: FANN Train Rprop

Optimal activation function of hidden level: FANN Elliot

Optimal output level activation function: FANN Sigmoid Symmetric Stepwise

The application of the neural network provided the results which are presented in the Table 1 below. The values are given on a scale from 0 to 1.

Table 1: Results of the first test.

| Cases / Projects | Estimated | Actual |
|------------------|-----------|--------|
| 1 | 0,6721 | 1 |
| 2 | 0,677 | 0,4966 |
| 3 | 0,5281 | 0,6706 |
| 4 | 0,3035 | 0,3309 |
| 5 | 0,5118 | 0,4149 |
| 6 | 0,1834 | 0 |
| 7 | 0,228 | 0,2827 |
| 8 | 0,4738 | 0,3446 |
| 9 | 0,5149 | 0,3933 |
| 10 | 0,5227 | 0,3736 |
| 11 | 0,3583 | 0,5447 |
| 12 | 0,3498 | 0,4248 |

Two more test trials followed. In the second test, the structure of the network was kept unchanged, and there was just a change in the number of training data and test data. The chosen ratio was 8:2, for training and testing, respectively, which translates into twenty-nine (29) training data projects and eight (8) control data projects. The optimum options for producing an effective neural network are the following:

Optimal training algorithm: FANN Train Rprop

Optimal activation function of hidden level: FANN Elliot

Optimal output level activation function: FANN Sigmoid Symmetric

The application of the neural network provided the results which are presented in the Table 2 below. The values are given in project days.

Table 2: Results of the second test.

| Cases / Projects | Estimated | Actual |
|------------------|-----------|--------|
| 1 | 882 | 865 |
| 2 | 299 | 218 |
| 3 | 861 | 793 |
| 4 | 1235 | 1081 |
| 5 | 937 | 720 |
| 6 | 927 | 978 |
| 7 | 1167 | 1175 |
| 8 | 874 | 1082 |

Considering the satisfactory results of the first two tests, it was decided to decrease the number of the input variables and look at how the network behaves with the new available data. More, specifically the third test which took place, used six (6) input variables: tender budget, length of highway project, number of lanes, number of technical projects, number of bridges, and tunnels. Proportion of training data to control data was defined as 7: 3, so there were created 25 data sets for training and 12 sets of data for control. The network which was selected had a structure: 1-2-6, and as a result the input level has six (6) neurons, the output level one (1) and the hidden layer consists of two (2) neurons. The optimum options for producing an effective neural network are the following:

Optimal training algorithm: FANN Train Incremental

Optimal activation function of hidden level: FANN Elliot Symmetric

Optimal output level activation function: FANN Linear Piece Symmetric

The application of the neural network provided the results which are presented in the Table 3 below. The values are given in project days.

Table 3: Results of the third test.

| Cases / Projects | Estimated | Actual |
|------------------|-----------|--------|
| 1 | 1063 | 2252 |
| 2 | 1189 | 1565 |
| 3 | 110 | 353 |
| 4 | 701 | 891 |
| 5 | 1276 | 1076 |
| 6 | 804 | 689 |
| 7 | 784 | 793 |
| 8 | 18 | 919 |
| 9 | 984 | 1018 |
| 10 | 1030 | 978 |
| 11 | 925 | 1175 |
| 12 | 899 | 1082 |

By reverting the data to days, the mean squared error for the first test was calculated as $RSME = 322$ days and mean absolute percentage error $MAPE = 34,53\%$. Correspondingly, $RSME = 127$ days and $MAPE = 14,63\%$, for the second test and $RSME = 467$ days and $MAPE = 29,02$, for the third test.

It should be mentioned that several of the projects have major deviations between the calculated and the actual values. A number of successive tests have taken place, in which either alter the structure of the network or completely exclude projects which are identified as outliers, from the original data. The test results have slightly improved and deviations continued to appear on individual projects.

4 CONCLUSION, DISCUSSION AND FURTHER WORK

The aim of the current research was to predict the actual duration of highway projects constructed in Greece between mid-1990s and 2010, applying artificial neural network application. From the tests, described in detail above, it is noted that any change in the network structure or the input variables, could provide different results, while important remains the fact that removing outliers, restores uniformity of results and leads to the belief, that in these cases / projects there were some delay factors which were unforeseen in the initial scheduling process. Finally, the tests performed and the results obtained showed that artificial neural networks could predict the course of road construction as far as time is concerned. The second neural network model was the most successful one, based on the $RSME$ and $MAPE$ values. Within the future work, research could focus in the form of neural networks, because the combinations that can be made are numerous. Developing the artificial neural network allows for further research, which in relation to the extent and the reliability of the available data, could provide even more accurate forecasts. These approaches could be applied both in the case of highway projects and in any other construction project. Finally, the human factor, involved in the project implementation, should also be taken into account in the neural network model. Especially, the performance

of the involved project managers, may affect significantly the project performance and outcome.

5 REFERENCES

- Antoniou, F., Aretoulis, G.N., Konstantinidis, D. and Kalfakakou, G.P. (2013a). An empirical study of researchers' and practitioners' views on compensating major highway project contractors. *Inderscience IJMDM*, 12(4), pp. 351-375.
- Antoniou, F., Aretoulis, G.N., Konstantinidis, D. and Kalfakakou, G.P. (2013b). Engineers' Perception and Assessment of Contract Types for Highway Construction Projects. *IGI-Global IJABE*, 2(4), pp. 1-24.
- Aretoulis, G.N., Angelides, D.C., Kalfakakou, G.P., Fotiadis, G.S. and Anastasiadis, K.I. (2006). A Prototype System for the Prediction of Final Cost in Construction Projects, *Springer Operational Research, An International Journal*, Special Issue, 6(3), pp. 423 – 432.
- Aretoulis, G.N., Kalfakakou, G.P. and Seridou, A.A. (2015). Project Managers' Profile Influence on Design and Implementation of Cost Monitoring and Control Systems for Construction Projects. *IGI-Global IJITPM*, 6(3), pp. 1-25.
- Attal, A. (2010). *Development of neural network models for prediction of highway construction cost and project duration*. MSc. Diss., Civil Engineering, Ohio University, Athens.
- Berlin S., Raz T., Glezer C. and Zviran M. (2009). Comparison of estimation methods of cost and duration in IT projects. *Elsevier Inform. Software Tech.*, 51(4), pp. 738-748.
- Chan D. W. M. and Kumaraswamy M. M. (1996). An evaluation of construction time performance in the building industry. *Elsevier Build. Environ.*, 31(6), pp. 569-578.
- Cheng M.Y., Tsai H.C. and Sudjono E. (2012). Evolutionary fuzzy hybrid neural network for dynamic project success assessment in construction industry. *Elsevier Automat. Constr.*, 21, pp. 46-51.
- Elhag, T. M. S. and Boussabaine, A. H. (1998). An artificial neural system for cost estimation of construction projects. In Proceedings of the 14th ARCOM Annual Conference.
- ElSawy I., Hosny H. and Razeq M. A. (2011). A Neural Network Model for Construction Projects Site Overhead Cost Estimating in Egypt. *IJCSI*, 8(3), pp. 273-283.
- Guerrero, M. A. Villacampa, Y. and Montoyo, A. (2013). Modeling construction time in Spanish building projects. *Elsevier Int. J. Proj. Manag.*, 32(5), pp. 861-873.
- Gunaydin H. M. and Dogan S. Z. (2004). A neural network approach for early cost estimation of structural systems of buildings. *Elsevier Int. J. Proj. Manag.*, 22(7), pp. 595-602.
- Skitmore R. M. and Ng S. T. (2003). Forecast models for actual construction time and cost. *Elsevier Build. Environ.*, 38(8), pp. 1075-1083.
- Tatari O. and Kucukvar M. (2011). Cost premium prediction of certified green buildings: A neural network approach. *Elsevier Build. Environ.*, 46(5), pp. 1081-1086.
- Zhang J. P., Liu L. H. and Coble R. J. (2002). Hybrid intelligence utilization for construction site layout. *Elsevier Automat. Constr.*, 11(5), pp. 511-519.

