Neural Network Prediction Model for Construction Project Duration

Silvana Petruseva 1, Vahida Zujo 2, Valentina Zileska-Pancovska 3

1 Assis. Prof. PhD, University “St. Cyril and Methodius”, Faculty of Civil Engineering, (Mathematics Department) Skopje, Macedonia,
2 Prof. PhD, C.E., University “Džemal Bijedić” in Mostar, Faculty of Civil Engineering, Bosnia and Herzegovina,
3 Prof. PhD, C.E., University “St. Cyril and Methodius”, Skopje, Faculty of Civil Engineering, Macedonia, (Department for Management of Construction)

Abstract

This paper presents neural network model for predicting construction project duration.

Key data of the total of 75 buildings constructed in the Federation of Bosnia and Herzegovina have been collected through field studies. The collected data contain information for the contracted and real time of construction, the contracted and real price of construction and there are also data for the use of these 75 objects and for the year of construction.

First, a linear regression using “time-cost” model was applied to these data for forecasting the construction time. Then, a multilayer perceptron neural network (MLP - NN) predictive model to the same data was applied and significant improvement of the accuracy of the prediction was obtained.

Keywords: artificial neural network, construction time, construction cost, linear regression, multilayer perceptron

1. Introduction

Project deadline is an essential element of each construction contract, but in numerous construction projects there is non-compliance of the contracted time (that is included in the building contract and certified by the contractor and client) and planned time (that is determined by the technical preparation department with one of the planning methods) of construction. This problem is confirmed by many authors worldwide [12], [14], [4], [30], [31].

Numerous scientific studies have been addressed to the investigation of contracted construction time and contracted price at various construction markets.

Bromilow investigated the relation between construction price and construction time. From his research stems the well known “time – cost” model. Afterwards the model has been confirmed in many countries [7], [4], [30], [15], [8], [24], [13], [19].

Yakubu Adisa Olawale and Ming Sun identified causes of cost and time overruns. They studied factors inhibiting the ability of practitioners to effectively control their projects [18].

Jieh-Haur Chen and Wei-Hsiang Chen derived a mathematical way of defining the contractor's costs for factoring account receivables [6].

Yanshuai Zhang and S. Thomas Ng applied an evolutionary-based optimization algorithm known as an ant colony system to solve the multi-objective time-cost optimization problems [29].

Abbas Afshar and Habib Fathi presented a multi-objective model to search the non-dominated solutions considering total duration, required credit, and financing cost as three objectives. Fuzzy-sets theory is used to account for uncertainties in direct cost of each activity for determining the required credit and financing cost [2].

Ehsan Eshtehardian, Abbas Afshara and Reza Abbasnia presented an approach to investigate stochastic time-cost trade-off problems employing fuzzy logic theory [9].
This paper presents linear regression and neural network (NN) prediction models applied to the data of 75 objects built in the Federation of Bosnia and Herzegovina in the period from 1999 to 2012. Comparison of these two models shows that the application of NN prediction model significantly improves the accuracy of the prediction.

Artificial neural networks (ANN) have been found to be powerful and versatile computational tools for many different problems in engineering over the past 2 decades. Many applications of ANN have proven that they can be applied successfully to many engineering’s problems and that in many cases they perform better than conventional methods. They have proved useful for solving certain types of problems, which are too complex, or too resource-intensive nonlinear problems to tackle using more traditional computational methods, such as the finite element method. ANNs are intelligent tools, which have gained strong popularity in a large array of engineering applications such as pattern recognition, function approximation, optimization, forecasting, data retrieval, automatic control or classification, where conventional analytical methods are difficult to pursue, or show inferior performance.

The first journal article on neural network application in civil engineering was published in 1989. Since then, many articles on NN application in different areas of civil engineering like structural engineering, management [3], [11], environmental and water resources engineering, traffic, geotechnical and geomechanical engineering have been published [1],[20].

2. Linear regression and neural network prediction models for construction project duration

Data were collected for a total of 75 structures built in the Federation of Bosnia and Herzegovina in the period from 1999 to 2012 using questionnaires and interviews. Key data were: structure type; construction year; contracted and realized construction time; contracted and realized construction price; reasons for non-compliance of deadline. From a total of 75 structures, disregard of the contracted deadline was registered at 55 of them (73%), disregard of the contracted price was registered at 40 structures (53%), while simultaneously the contracted deadline and the contracted price overrun were registered at 36 structures (48%). The maximum contracted deadline overrun was 100%, and the price was 68.75% while the average contracted deadline has been exceeded for 11.55% and 2.77% for the price. Contracted deadline reduction was registered at 11 structures (14.67%), while simultaneously the contracted deadline and the contracted price reduction was registered at 2 structures (2.67%).

Total of five different reasons for non-compliance appear: approvals and permits; climate; incomplete and inaccurate technical documentation; material delivery and terms of financing [5],[31].

2.1. Linear regression prediction model for construction project duration

Linear regression is an approach to model the relationship between a scalar dependant variable y and one or more explanatory variables denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, it is called multiple linear regression [26], [25].

In linear regression, data are modeled using linear predictor functions, and unknown model parameters are estimated from the data.

Since linear regression is restricted to fitting linear (straight line/plane) functions to data, it rarely works as well on real-world data as more general techniques such as neural networks which can model non-linear functions. However, linear regression has many practical uses and a number of strengths, for example linear regression analysis can be applied to quantify the strength of the relationship between y and the xj, to assess which xj may have no relationship with Y at all, and to identify which subsets of the xj contain redundant information about Y, also linear regression models are simple and require minimum memory to implement, so they work well on embedded controllers that have limited memory space [22].

For construction time prediction Bromilow’s “time cost” model given in Eq. (1) is used.

\[ T = K \cdot C^B \]  

(1)

where: T - contracted time; C - contracted price; 
K - model parameter that is a specific way to measure productivity because it shows the average time needed for the construction of a monetary value;
B - model parameters that shows time dependence of costs change.

For the requirements of linear regression model, we shall write this model in linear form, using logarithmic transformation [4], as shown in Eq. (2):

\[ \ln{T} = \ln{K} + B \ln{C} \]  

The linear form of the equation allows usage of the simple statistical procedure, i.e., a single linear regression. We shall determine the values of parameters K and B in this model.

For creating the linear regression model for predicting of the construction time for the 75 objects, DTREG software package was used [23].

DTREG analysis results of the model are presented in Table 1 to Table 3. Considering eq. (2) variable \( \ln(\text{real time}) \) is used as target variable, and \( \ln(\text{real price}) \) as predicted variable.

From Table 1 the coefficients of the linear regression model can be read: \( B = 0.550208 \) (which is multiplied with variable \( \ln(\text{real price}) \)), and the constant \( \ln{K} = -2.37546 \) and from here \( K = e^{-2.37546} \).

The linear regression model for prediction of construction time will be:

\[ T = e^{-2.37546} \cdot 0.550208 \]

A part of the data was used for training the model (training data, Table 2) and a part of the data was used for validation of the model (Table 3).

We can estimate the accuracy of the model from the statistics of validation data (Table 3):

Most often used estimators of a predictive model are \( R^2 \) and MAPE. In statistics, the coefficient of determination denoted \( R^2 \) indicates how well data points fit a line or curve. It is a measure of global fit of the model. In linear regression \( R^2 \) equals the square of Pearson correlation coefficient between observed and modeled (predicted) data values of the dependent variable. \( R^2 \) is an element of [0,1] and is often interpreted as the proportion of the response variation “explained” by the regressors in the model. So, the value \( R^2 = 0.73341 \) from our model may be interpreted: around 73% of the variation in the response can be explained by the explanatory variables. The remaining 27% can be attributed to unknown, lurking variables or inherent variability [27].

MAPE (Mean Absolute Percentage Error) is a measure of accuracy of a method for constructing fitted times series values in statistics. It usually expresses accuracy as a percentage [28]. For this model \( \text{MAPE} = 10.355481 \) which means that the error of the model is around 10%.

<table>
<thead>
<tr>
<th>Table 1. Linear regression model for predicting real time of construction using DTREG package</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>===============</th>
<th>Linear Regression Parameters ===============</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Coefficient (Beta) Values</td>
</tr>
<tr>
<td>( \ln(\text{real price}) )</td>
<td>0.550208</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.37546</td>
</tr>
</tbody>
</table>

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 11, November - 2013
2. Statistics for training data for the linear regression model (DTREG)

--- Training Data ---

Mean target value for input data = 4.7038902
Mean target value for predicted values = 4.7038902
Variance in input data = 0.8306976
Residual (unexplained) variance after model fit = 0.2753121
Proportion of variance explained by model (R^2) = 0.66858 (66.858%)
Coefficient of variation (CV) = 0.111546
Normalized mean square error (NMSE) = 0.331423
Correlation between actual and predicted = 0.817666
Maximum error = 1.7033044
RMSE (Root Mean Squared Error) = 0.5247019
MSE (Mean Squared Error) = 0.2753121
MAE (Mean Absolute Error) = 0.4108599
MAPE (Mean Absolute Percentage Error) = 9.2084657

Table 3. Statistics for validation data for the linear regression model (DTREG)

--- Validation Data ---

Mean target value for input data = 4.6575062
Mean target value for predicted values = 4.6797136
Variance in input data = 1.0179999
Residual (unexplained) variance after model fit = 0.2713927
Proportion of variance explained by model (R^2) = 0.73341 (73.341%)
Coefficient of variation (CV) = 0.111852
Normalized mean square error (NMSE) = 0.266594
Correlation between actual and predicted = 0.858757
Maximum error = 0.8396821
RMSE (Root Mean Squared Error) = 0.5209536
MSE (Mean Squared Error) = 0.2713927
MAE (Mean Absolute Error) = 0.462848
MAPE (Mean Absolute Percentage Error) = 10.355481

2.2 Neural network prediction model for construction project duration

Neural networks are intelligent systems that are based on simplified computing models of the biological structure of the human brain, whereas the systems based on traditional computer logic require comprehensive programming in order to perform a given task.

There are several aspects that make ANNs attractive, such as: relation with biological neural networks; relation with the concept of parallel distributed processing and relation with the concept of learning and self-organization[10].

ANNs are suitable for multivariable applications where they can easily identify interactions and patterns between inputs and outputs. ANN models do not require complicated and time consuming finite element input file preparation for routine design applications. They are able to infer important information for the task, which is being solved by them, if data that is representative of the underlying process to be implemented, is provided.

Neural networks have a self-learning ability, which is particularly useful where comprehensive models that are required for conventional computing methods are either too large or too complex to represent accurately, or simply don’t exist at all. The highly connected, distributed nature of neural networks also provides a high degree of generalization capability and noise immunity. In the last 20 years there have been several pointers toward the issue of feelings and emotions as needed features for development of artificial neural networks, especially in the speeding up the process of their learning [21].

Software package DTREG, which is used in this paper implements the most widely used types of neural networks: Multilayer Perceptron Networks (MLP), Probabilistic Neural Networks (PNN) and General Regression Neural Networks (GRNN), Radial Basic Function (RBF) networks, Polynomial Neural Networks (GMDH), and Cascade Correlation networks [22].

In this paper, Multilayer Perceptron Neural network is used as a predictive model.

A multilayer perceptron (MLP) is a feed forward artificial neural network that maps sets of input data onto a set of appropriate outputs. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. For training the network MLP utilizes a supervised learning technique called back propagation [17].

What makes a multilayer perceptron different is that each neuron uses a nonlinear activation function which was developed to model the frequency of action potentials. This function is modeled in several ways, but always must be normalizable and differentiable.
There are two main activation functions used in current applications which are both sigmoid:

\[ \phi(v_i) = \tanh(v_i) \quad \text{and} \quad \phi(v_i) = (1 + e^{-v_i})^{-1} \]

where: the former function is a hyperbolic tangent which ranges from -1 to 1, and the latter, the logistic function, is similar in shape but ranges from 0 to 1.

MLPs were a popular machine learning solution, finding applications in diverse fields such as speech recognition, image recognition, forecasting and machine learning software.

MLP network (Figure1.) has an input layer (on the left), one hidden layer (in the middle) and an output layer (on the right) [22]. There is one neuron in the input layer for each predictor variable \((x_1, \ldots, x_p)\).

A vector of predictor variable values \((x_1, \ldots, x_p)\) is presented to the input layer. The input layer (or processing before the input layer) standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer.

Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight \((w_{ij})\), and the resulting weighted values are added together producing a combined value \(u_i\). The weighted sum \((u_i)\) is fed into a transfer function, \(\sigma\), which outputs a value \(y_i\). The outputs from the hidden layer are distributed to the output layer.

Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight \((w_{kj})\), and the resulting weighted values are added together producing a combined value \(v_j\). The weighted sum \((v_j)\) is fed into a transfer function, \(\sigma\), which outputs a value \(y_j\). The \(y\) values are the outputs of the network.

The goal of the training process is to find the set of weight values that will cause the output from the neural network to match the actual target values as closely as possible.

One of the most important characteristics of a multilayer perceptrons network is the number of neurons in the hidden layer(s). If an inadequate number of neurons are used, the network will be unable to model complex data, and the resulting fit will be poor.

If too many neurons are used, the training time may become excessively long, and, worse, the network may over fit the data. When over fitting occurs, the network will begin to model random noise in the data. The result is that the model fits the training data extremely well, but it generalizes poorly to new, unseen data. Validation must be used to test for this. Software DTREG includes an automated feature to find the optimal number of neurons in the hidden layer.

In this model three layer MLP with one hidden layer is used.

A typical neural network might have a couple of hundred weighs whose values must be found to produce an optimal solution. If neural networks were linear models like linear regression, it would be easy to find the optimal set of weights. But the output of a neural network as a function of the weights is often highly nonlinear and this makes the optimization process complex. DTREG uses the Nguyen-Widrow algorithm to select the initial range of starting weight values. It then uses the conjugate gradient algorithm [16] to optimize the weights.

The training algorithms follows this cycle to refine the weight values:

1. Run the predictor values for a case through the network using a tentative set of weights.
2. Compute the difference between the predicted target value and the actual target value for the case. This is the error of the prediction.
3. Average the error information over the entire set of training cases.
4. Propagate the error backward through the network and compute the gradient (vector of derivatives) of the change in error with respect to changes in weight values.
5. Make adjustments to the weights to reduce the error.

Each cycle is called an epoch. Because the error information is propagated backward through the network, this type of training method is called backward propagation or “backprop”.

Table 4, Table 5 and Table 6 present the result of the modeling MLP using DTREG package.

Considering the eq. (2) for dependence of the time of construction from the price of construction, here the target variable is \(\ln(\text{real time})\) and predictor variables are chosen to be \(\ln(\text{contracted time})\), \(\ln(\text{contracted price})\) and \(\ln(\text{real price})\).

Table 4 presents general statistics and parameters of the MLP model. Table 5 presents the statistics for training data, and Table 6 for validation data.
Figure 1. Multilayer perceptron (with permission of the author Phill Sherrod of DTREG package)

Table 4. General statistics of the MLP predictive model (using DTREG package)

<table>
<thead>
<tr>
<th>Number</th>
<th>Variable</th>
<th>Class</th>
<th>Type</th>
<th>Missing rows</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of project</td>
<td>Unused</td>
<td>Continuous</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>use of the object</td>
<td>Unused</td>
<td>Categorical</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>year of construction</td>
<td>Unused</td>
<td>Continuous</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>contracted time (days)</td>
<td>Unused</td>
<td>Continuous</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>real time of constr. (days)</td>
<td>Unused</td>
<td>Continuous</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>difference (days)</td>
<td>Unused</td>
<td>Continuous</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>price contracted [KM]</td>
<td>Unused</td>
<td>Continuous</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>real price</td>
<td>Unused</td>
<td>Continuous</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>difference of prices</td>
<td>Unused</td>
<td>Continuous</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ln(treal time)</td>
<td>Target</td>
<td>Continuous</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>11</td>
<td>ln(real price)</td>
<td>Predictor</td>
<td>Continuous</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>12</td>
<td>ln(contracted time)</td>
<td>Predictor</td>
<td>Continuous</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>13</td>
<td>ln(contracted price)</td>
<td>Predictor</td>
<td>Continuous</td>
<td>0</td>
<td>75</td>
</tr>
</tbody>
</table>

The network will be built using 7 neurons for hidden layer 1.
Table 5. Statistic for the training data for the MLP predictive model (DTREG)

<table>
<thead>
<tr>
<th>--- Training Data ---</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean target value for input data = 4.6946134</td>
</tr>
<tr>
<td>Mean target value for predicted values = 4.7049379</td>
</tr>
<tr>
<td>Variance in input data = 0.8685023</td>
</tr>
<tr>
<td>Residual (unexplained) variance after model fit = 0.021151</td>
</tr>
<tr>
<td>Proportion of variance explained by model (R^2) = 0.97565 (97.565%)</td>
</tr>
<tr>
<td>Coefficient of variation (CV) = 0.030979</td>
</tr>
<tr>
<td>Normalized mean square error (NMSE) = 0.024353</td>
</tr>
<tr>
<td>Correlation between actual and predicted = 0.987832</td>
</tr>
<tr>
<td>Maximum error = 0.594998</td>
</tr>
<tr>
<td>RMSE (Root Mean Squared Error) = 0.1454337</td>
</tr>
<tr>
<td>MSE (Mean Squared Error) = 0.021151</td>
</tr>
<tr>
<td>MAE (Mean Absolute Error) = 0.1033324</td>
</tr>
<tr>
<td>MAPE (Mean Absolute Percentage Error) = 2.2472572</td>
</tr>
</tbody>
</table>

Table 6. Statistic for the validation data for the MLP predictive model (DTREG)

<table>
<thead>
<tr>
<th>--- Validation Data ---</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean target value for input data = 4.6946134</td>
</tr>
<tr>
<td>Mean target value for predicted values = 4.6900164</td>
</tr>
<tr>
<td>Variance in input data = 0.8685023</td>
</tr>
<tr>
<td>Residual (unexplained) variance after model fit = 0.0261502</td>
</tr>
<tr>
<td>Proportion of variance explained by model (R^2) = 0.96989 (96.989%)</td>
</tr>
<tr>
<td>Coefficient of variation (CV) = 0.034446</td>
</tr>
<tr>
<td>Normalized mean square error (NMSE) = 0.030110</td>
</tr>
<tr>
<td>Correlation between actual and predicted = 0.984851</td>
</tr>
<tr>
<td>Maximum error = 0.6874137</td>
</tr>
<tr>
<td>RMSE (Root Mean Squared Error) = 0.1617104</td>
</tr>
<tr>
<td>MSE (Mean Squared Error) = 0.0261502</td>
</tr>
<tr>
<td>MAE (Mean Absolute Error) = 0.1127981</td>
</tr>
<tr>
<td>MAPE (Mean Absolute Percentage Error) = 2.4984055</td>
</tr>
</tbody>
</table>

We can see from table 6 that estimators R^2 and MAPE for the validation data of the model are: $R^2 = 0.96989$, and $MAPE = 2.4984055$, indicating significant improvement of the accuracy of the prediction in comparison with the linear regression model.

CONCLUSION

The paper presents two models for predicting the time construction of an object: linear regression model and neural network model (multilayer perceptron).

75 objects structured in the period of 1999 to 2012 in the federation of Bosnia and Herzegovina have been analyzed. First, the conventional linear regression model was built using the “time cost” model. Then, the neural network multilayer perceptron predictive model was applied to the same data.

- The linear regression model for predicting the real time of construction was estimated by $R^2 = 0.73341$, and $MAPE=10.355481$ (for the validation data).

- MLP model for predicting the real time of construction was estimated by: $R^2 = 0.96989$, and $MAPE= 2.4984055$ (for the validation data).

Models’ comparison results show that application of the MLP-NN model indicates significant improvement of the accuracy of the prediction.

The authors hope that the presented model will prove very useful for improving planning in construction industry in general.

REFERENCES


2004, Faculty of Civil Engineering, University of Zagreb.


