Hubert Anysz (h.anysz@il.pw.edu.pl)
Nabi Ibadov

The Team of Production Engineering and Construction Management, Civil Engineering Department, Warsaw University of Technology

Neuro-fuzzy predictions of construction site completion dates

Neuro-rozmyte prognozowanie terminów zakończenia przedsięwzięć budowlanych

Abstract

The results from two types of multi-layer perceptron artificial neural networks (Matlab R2015a was used) were compared. The first one, with only one neuron in an output layer having the value of delay in completion date of building site. The output layer of the second artificial neural network is created by three neurons. These three values represent the same delay, but in a form of three values of membership functions to fuzzy sets. In order to evaluate the accuracy of predictions, the mean squared error was used. It was necessary to find the best method of defuzzyfication predicted delays to compare the results from these two, aforementioned artificial neural networks. The level of prediction accuracy measured by mean squared error was discussed, too.

Keywords: MLP artificial neural networks, fuzzy sets, delays in completion dates

Streszczenie

W artykule porównano wyniki prognoz uzyskane wielowarstwowymi sztucznymi sieciami neuronowymi (obliczenia wykonano pakietem Matlab R2015a). Pierwsza z nich miała jeden neuron w warstwie wyjściowej, któremu przypisano wartość opóźnienia w terminie zakończenia danej budowy. Warstwę wyjściową drugiej sieci stanowiły trzy neurony, zawierające wartości funkcji przynależności tego samego opóźnienia do trzech zborów rozmytych. Do porównania dokładności prognoz uzyskanych z dwóch ww. sztucznych sieci neuronowych zastosowano średni błąd kwadratowy. Wymagało to znalezienia najlepszej metody defuzzyfikacji prognoz otrzymanych w postaci liczb rozmytych. Dokładność prognoz została porównana i przedyskutowana.

Słowa kluczowe: sztuczne sieci neuronowe MLP, zbiory rozmyte, opóźnienia terminu zakończenia budowy

1. Introduction

The delays in completion dates of construction sites have a negative impact on operations' efficiency of contractors as well as the clients. Moreover, taking into account public building investments like roads, hospitals, schools, the public suffers from the aforementioned delays.

Before the completion date of building works is set in the agreement between the client and the contractor, it is a subject of the client's analysis: what is the nearest possible, reasonable time of finishing an already designed building object. Officers can rely on knowledge obtained from past investments. They can analyze the scope of works necessary to complete already erected building objects, in order to estimate the time needed to build a new one. The other approach proposed in this paper is to verify planned time by predicting possible delays of the completion date.

Predicted values of the completion dates or delays can vary due to different capabilities of bidders. This is an opportunity to utilize delays of completion date predictions as one of the factors for multi-criteria evaluations of bidders. According to Tadeusiewicz [1], artificial neural networks (ANN) can be successfully applied for complex problems, where we suspect the existence of rules between input factors and the final effect of the process, but we are not able to fully recognize them. The complexity of public building investment is undisputed. The moment of prediction, i.e. before any building works started, makes liaison between parameters describing bidders, planned building object, and delay in completion date very weak. It was decided to check if Zadeh's theory of fuzzy sets [2, 3], applied for number of days as a unit of delay of a building site completion date, can improve the prediction accuracy of ANN.

2. ANN trained with the set of real numbers as an output, representing delays

The predictions of delays in completion dates of building sites were limited to the construction of express roads and highways completed in 2009-2013. The following data were chosen as an input:

- the value of construction works (taken as the best offer),
- the length of the section of highway to be built,
- the scope of works ("build" or "design and build"),
- the type of works (building a new road or modernization, upgrading),
- the total number of employees at the disposal of the bidder,
- the planned time of work execution (number of days) up to completion.

The delays in completion date represented by number of days were chosen as an output. The accuracy of predictions was measured by Mean Squared Error (MSE) [4].

$$MSE = \frac{\sum_{i=1}^{N}(c_i - r_i)^2}{N} \qquad (1)$$

where:

c_i – the delay calculated by ANN for i set of test input data,

r_i – the real value of delay observed for i building site belonging to test set,

N – the number of test data set.

The total number of 156 sections of highways and express roads in Poland completed between 2009 and 2013 was limited to 119 data sets for calculations. This was caused by the necessity of using complete input data (number of employees mainly was not available). The building sites with extremely high delays in completion dates (exceeding 370 days) were excluded from the data set. Extremely high delays in completion dates were usually caused by very unusual reasons (e.g. protests of ecologists). It was assumed that completing the section of highway before the planned date causes delay equal to zero. The linear-maximum method [10] of the data standardization method was applied for input and output of ANN.

$$a_i = \frac{a_{0i}}{\max a_{0i}} \tag{2}$$

where:

a_i – the value of type a of input set, from i set of input data after standardization,

a_{01} – the value of type of input observed in process (before standardization),

i – consecutive number of the process being observed.

Multi-Layer Perceptron (MLP) type of ANN was used, as it was found that MLP networks have higher abilities to generalize input-output relations than ANN of Radial Basis Function (RBF) [5, 6]. It was checked empirically that for this case, i.e. predicting delays of completion dates, basing on the aforementioned inputs and [9], having 119 cases of input datasets, the following ANN gives the lowest MSE of predictions:

▸ 8 input neurons (two input parameters had to be represented by vectors [0, 1] or [1, 0], so it has risen the number inputs from 6 to 8),

▸ one hidden layer with 14 neurons,

▸ single output neuron – the delay in completion date,

▸ activation function in the hidden layer – linear with limits <–1, 1> (called *satlins* in Matlab),

▸ activation function in the output layer – logistic (called *logsig* in Matlab),

▸ teaching algorithm – conjugate gradient backpropagation (the order *trainscg* in Matlab).

It was set that first 79 cases create a teaching set of data. The rest, i.e. 40 sets, were kept for testing purposes (34%). Then, the complete set of data was randomized and calculations were proceeded. As the result of ANN calculation, a set of 40 predictions of delays was obtained (for every set of data from test data). Then, the predictions could be compared to the real values, by calculating MSE. The lowest MSE achieved (for test data) was 11 709.

3. ANN trained with the set of delays expressed by the values of membership functions to three fuzzy sets as an output

The delays in completion dates of building sites can be described by values of membership functions to fuzzy sets [8]. Three fuzzy sets were created for short, medium and long delays of completion dates. Membership function to the fuzzy set "short delays" μ_S was described by following formula:
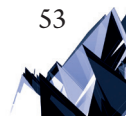
$$(x) = \begin{cases} 1 \ \text{or} \ x = 0 \\ \dfrac{185 - x}{185 - 0} \ \text{for} \ 0 \le x \le 185 \\ 0 \ \text{for} \ x \ge 185 \end{cases} \tag{3}$$

where:

x – value of delay in days.

The membership function μ_M to the fuzzy set "medium delays" was defined as:

$$\mu_M(x) = \begin{cases} 0 \ \text{for} \ x = 0 \ \text{or} \ x \ge 370 \\ \dfrac{x}{185} \ \text{for} \ 0 \le x \le 185 \\ \dfrac{370 - x}{370 - 185} \ \text{for} \ 185 \le x \le 370 \end{cases} \tag{4}$$

where:

x – value of delay in days.

The membership function μ_L to the fuzzy set "long delays" was created too as:

$$\mu_L(x) = \begin{cases} 0 \ \text{for} \ x \le 185 \\ \dfrac{x - 185}{370 - 185} \ \text{for} \ 185 \le x \le 370 \\ 1 \ \text{for} \ x \ge 370 \end{cases} \tag{5}$$

where:

x – value of delay in days.

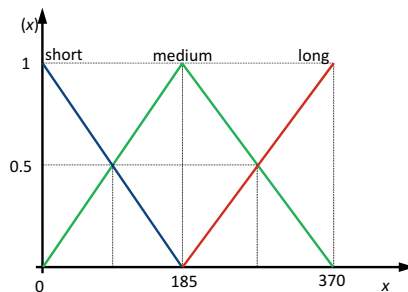All three fuzzy sets (short, medium sized and long) can be shown graphically in Fig. 1.

Fig. 1. Membership functions for short, medium and long delays fuzzy sets

The names of the fuzzy sets described above are partly inconsistent with their literal meaning. Nobody would say that the delay in completion date is short, when it is equal to half a year, even when the planned duration of work execution was three years. However, for the machine learning, it does not have any meaning. Much more important for the calculation is that proposed three membership functions provide that:

- for each x from 0 to 370 there is only one fuzzy set with the membership value not lower than 0.5,
- there is no x for which every membership value is lower than 0.5,
- every delay expressed in days (from 0 to 370 days) can be expressed by three membership values μ_S, μ_M, μ_S to the three fuzzy sets; one equal to 0, one higher than 0.5 and one smaller than 0.5 or one equal to 1 and the rest equal to 0.

Some examples of the fuzzyfication process concerning the set of 119 observed delays are shown in Tab. 1.

Table 1. Examples of delays shown as real numbers and values of membership functions to fuzzy sets

Real numbers	Values of membership functions to fuzzy sets		
Delay in days	μ_S	μ_M	μ_L
0	1	0	0
46	0.7514	0.2486	0
102	0.4486	0.5514	0
144	0.2216	0.7784	0
185	0	1	0
242	0	0.6919	0.3081
280	0	0.4865	0.5135
314	0	0.3027	0.6973
370	0	0	1

In order to apply the fuzzy sets theory to an output layer, the topology of artificial neural network had to be modified. Now output (the delay) is represented by three values of membership functions to three fuzzy sets, so the number of neurons in the output layer was increased to 3. All the other parameters remained the same (as described in chapter 2). As the values of membership functions are from 0 to 1, there was no need to standardize them. The result of calculations produced by ANN was a set of 40 predictions of delays (for test data). Each one was expressed by three predicted values of membership functions to the three fuzzy sets. MSE was not calculated for these predicted "fuzzy delays" (for each membership function value separately or as a sum of MSE), because it could not be compared to MSE calculated for predictions of delays given in days. Therefore, the defuzzyfication of the calculated predictions was required.

That is the main reason of applying the fuzzy set theory to an output of artificial neural network: by choosing the defuzzyfication method, the researcher can adjust the predictions coming with them closer to the original (not fuzzyficated) values from test data. As a result of this process, MSE can be lowered, which means that predictions will be more accurate.

In order to obtain real, crisp numbers of predicted delays (for test data), the following four methods of defuzzyfication were applied:

▸ the center of gravity method [7] (labeled as A):

$$x^* = \frac{\int x\mu(x)dx}{\int \mu(x)dx}$$ (6)

where:

    x^* – real (crisp) number,

    x – the value of the output variable,

    $\mu(x)$ – the membership function of the output variable,

▸ the weighted average method [11] (labeled as B);

$$x^* = \frac{\sum_{i=1}^{n}\mu(x_i)x_i}{\sum_{i=1}^{n}\mu(x_i)}$$ (7)

    x^* – real (crisp) number,

    x_i – the value of the output variable for i fuzzy set,

    $\mu(x_i)$ – the membership function of the output variable for i fuzzy set,

    i – number of fuzzy set of n fuzzy sets,

▸ the maximum of membership function method [11] (labeled as C):

$$x^* = \max_{i=1}^{n}\left(\mu(x_i)\right)$$ (8)

    x^* – real (crisp) number,

    x_i – the value of the output variable for i fuzzy set,

    $\mu(x_i)$ – the membership function of the output variable for i fuzzy set,

    i – number of fuzzy set of n fuzzy sets,

▸ the weighted average method modified (labeled as D) where only two fuzzy sets were taken into account: medium always, and the other set with higher predicted μ (μ_S or μ_L), then formula (7) is applied.

Table 2. MSE values calculated after different methods of defuzzyfication

defuzzyfication method	MSE value
A	22 623
B	**9526**
C	10 534
D	28 699

The modification of method B was based on the fact that delays calculated for teaching ANN set of data are represented by three numbers (as shown in tab. 1) and at least one is equal to zero. Two of the three values of membership functions are equal to zero only for

0 days of delay, 185 and 370 days. Moreover, for all domains of medium delays (except 0 and 370 days), membership function μ_M has different than zero values.

All A, B, C and D methods of defuzzyfication were applied to 40 predictions received from ANN and predicted delays (in days) for the test data could be calculated. By comparing them to original delays from test data, MSE could be calculated for each method of defuzzyfication. The results are shown in Tab. 2. The lowest MSE is for the weighted average method of defuzzyfication, and it is equal to 9526.

4. Comparing the obtained results and conclusions

As it was suspected, the application of fuzzy sets theory as the output of artificial neural network (instead of one crisp number) and then searching for the best method of defuzzyfication, has brought a lower mean squared error for predictions. The obtained results are shown in Tab. 3.

Table 3. MSE calculated for ANN predictions with different outputs

Type of ANN output	MSE value
delay in days – one neuron in the output layer	11 709
delay as three values of membership function to fuzzy sets – three neurons in the output layer	9526

Applying ANN with different outputs (real, crisp value of delay – one neuron in the output layer or three values of membership functions for the same delay – three neurons in output layer) with the other parameters of ANN topology unchanged, allows one to state that the net with fuzzy set theory applied to output layer has given MSE lower by 18,6% then the net with real, crisp output.

This kind of result was expected. The fuzzyfication of delay in the completion date of a construction site, in a fact, adds some more information to the single number of delays. Moreover, the defuzzyfication of achieved predictions for test data set is a kind of post-process, made on rough predictions. Analyzing which method of defuzzyfication makes MSE the lowest resulted in it lowering, even below MSE calculated by ANN with a single, real number in an output layer. For the net with a single, real number as an output, the predictions were not post-processed. MSE equal to 9526 makes square root from it (RMSE) equal to 97 days of possible error in predicting delay in the completion date of a given construction site. It is not a precise prediction. However, taking into account the fact that the prediction can be done before any work starts on the building site, the meaning of 97 days of RMSE changes. On the other hand, it has to be emphasized that post-processing is independent from the input data, which means that the application of both tools (artificial neural networks and fuzzy sets block) can be worked out separately in order to achieve lower prediction errors.

Accurate predictions of delays in completion dates of construction sites can help in choosing the contractor providing the highest probability of completing works on time. When there
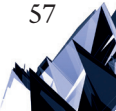
is no such contractor, it may mean that the planned time for work execution was assumed too short. The usefulness of this kind of predictions is also a subject of their accuracy. In order to achieve it, the fuzzy sets theory was applied to the output layer of MLP artificial neural networks. The block of defuzzyfication gives the chance to refine predictions, resulting in lowering the prediction error. It was proved in analysed case that post-processing of results received from ANN – when the fuzzy sets theory is applied for an output layer – can significantly lower the mean squared error of predictions.

References

[1] Tadeusiewicz R., Nałęcz M., *Biocybernetyka i inżynieria medyczna 2000. T. 6 Sieci Neuronowe*, Akademicka Oficyna Wydawnicza EXIT, Warsaw 2000.Zadeh L.A., *Fuzzy Sets*, Information and Control, 8, 1965, 338–353.

[2] Zadeh L.A., *Knowledge representation in Fuzzy Logic,* IEEE Transaction on Knowledge and Data Engineering, Vol. 1, No. 1, March 1989.

[3] Bartkiewicz J. (red. Zieliński J.S.), *Inteligentne systemy w zarządzaniu*, PWN, Warszawa 2000.

[4] Osowski S., *Sieci neuronowe w ujęciu algorytmicznym*, WNT, Warszawa 1996.

[5] Statsoft Polska Sp. z o.o., *Wprowadzenie do sieci neuronowych*, Statsoft, Kraków 2001.

[6] Ibadov N., *Determination of the Risk Factors Impact on the Construction Projects Implementation Using Fuzzy Sets Theory*, Acta Physica Polonica A, Vol. 130, No. 1, 2016, p. 107–111, DOI: 10.12693/APhysPolA.130.107.

[7] Ibadov N., *Fuzzy estimation of activities duration in construction projects*, Archives of Civil Engineering, Vol. 61, Issue 2, 2015), 23–34. DOI: 10.1515/ace-2015-0012.

[8] Anysz H. *Porównanie trafności prognoz sztucznych sieci neuronowych MLP z jedną i dwoma warstwami ukrytymi na przykładzie prognoz wyników finansowych przedsiębiorstw budowlanych*, Logistyka 3/2014, 87–94.

[9] Anysz H., Zbiciak A., Ibadov N., *The influence of input data standardization method on prediction accuracy of artificial neural networks*, Procedia Engineering 153, 2016, 66–70, DOI: 10.1016/j.proeng.2016.08.081.

[10] Rutkowski L., *Metody i techniki sztucznej inteligencji*, PWN, Warszawa 2012.