

Project Time and Cost Forecasting Using Monte Carlo Simulation and Artificial Neural Networks

Mahdieh Akhbari

Mahdieh Akhbari, Department of Industrial Engineering, Electronic Branch, Islamic Azad University, Tehran, Iran

KEYWORDS

Project management;
Earned value management;
Estimate at completion
(EAC);
Monte carlo simulation;
Artificial neural network

ABSTRACT

The aim of this study is to present a new method to predict project time and cost under uncertainty. Assuming that what happens in projects implementation which is expressed in the form of Earned Value Management (EVM) indicators is primarily related to the nature of randomness or unreliability, in this study, by using Monte Carlo simulation and assuming a specific distribution for the time and cost of project activities, a significant number of predicting scenarios will be simulated. According to the data, an artificial neural network is used as an efficient data mining method to estimate the project time and cost at completion.

© 2018 IUST Publication, IJIEPR. Vol. 29, No. 2, All Rights Reserved

1. Introduction

Predictions of project performance can guide project manager and project team to engineer critical issues scientifically, allowing for proactive project performance management. Several researchers have proposed models and approaches to estimate project performance. In recent years, many studies have been conducted to evaluate and predict the project cost and time based on Earned Value Management (EVM).

EVM was presented for the first time by the United State Department of Defense (DOD) in 1967. It was proposed as an acceptable method for managing the cost and time of large projects. This method is able to evaluate and control the project by defining three main indices: (EV) Earned Value, or (BCWP) Budgeted Cost of Work Performed, (PV) Planned Value, or (BCWS) Budget Cost of

Work Schedule, and (AC) Actual Cost, or (ACWP) Actual Cost of Work performed.

As the primary indicators of projects, previous studies have been carried out on the basis of the cost of projects. Meanwhile, all performance criteria, such as cost variance ($CV=EV-AC$), schedule variance ($SV=EV-PV$), cost performance index ($CPI=EV/AC$), and schedule performance index ($SPI=EV/AC$), were estimated according to the three main cost-based EVM indices (Figure1). The estimate at completion (EAC) was also based on the above criteria [1].

Many studies have been conducted to forecast the project performance based on EVM criteria investigated in a review article by Christensen et al. (1995). They divided methods into three main categories: Index-based methods, Regression-based methods, and other heuristic-based methods. It is shown that, in the case of having sufficient data, regression-based methods are much more accurate than others [2].

In 2003, in his study on Earned Value Management (EVM), Anbari presented

* Corresponding author: Mahdieh Akhbari

Email: m_akhbari@iauec.ac.ir

Received 22 October 2017; revised 26 December 2017; accepted 15 April 2018

evaluation indices in the form of various practical charts and compared six conventional methods to predict the Cost Estimate at Completion (CEAC) as well as Time Estimate at Completion (TEAC) according to the EVM [3]. In the same year, Lipke presented the new index of Earned Schedule (ES). In fact, (ES) represented the specific time to which Earned Value based on baseline must have been achieved. In accordance with this new index, time criteria such as Schedule Variance ($SV(t)=ES-AT$) and Schedule Performance Index ($SPI(t)=ES/AT$) are introduced as a suitable alternative to the previous indices of Schedule Variance (SV) and Schedule Performance Index (SPI) where AC is in real-time or the time at which the earned value (EV) is obtained (Figure 1). Then, in addition, to compute performance measures based on cost, project managers could calculate the project progress based on the time schedule [4].

Henderson (2003), in an article, investigated the evaluation indices based on the Earned Schedule (ES) and Earned Value indices (EV). He tested a portfolio of six projects. Cost at Completion (CAC) was calculated using criteria based on both methods [5]. In 2004, he developed an article to predict the project time duration and the date of completion on the basis of Earned Schedule (ES) [6].

Vanhoucke & Vandevoorde (2007) evaluated three different methods by using simulation based on (EV) to predict project completion time including earned schedule method (ES), developed by Lipke (2003), Planned Value (PV), presented by Anbari (2003), and Earned Duration method (ED), provided by Jacob (2003) [7]. The results determined that Earned Schedule (ES) was more efficient than other predicting methods [8].

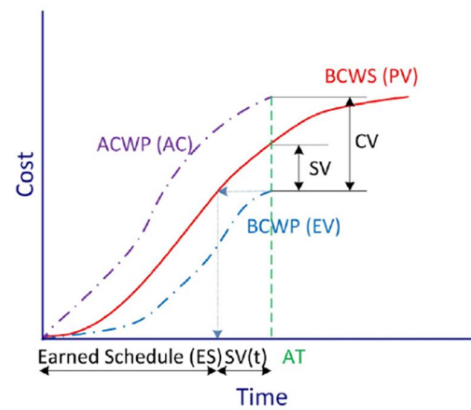


Fig. 1. EVM main criteria

Lipke (2006) examined the use of statistical methods in Earned Value Management (EVM). In his study, by assuming the mean and variance for CPI, Confidence Interval (CI) was determined; based on it, Independent Estimate at Completion (IEAC) was defined [9]. Later, Lipke et al. (2009) offered an article aimed at improving prediction method of project cost and completion time with the same statistical approach [10]. Tseng (2011) implemented the use of Weibull analysis on earned value metrics for comparison of overall performances of projects [11]. Narbaev & Marco (2014) proposed a new CEAC methodology based on a modified index-based formula to predict the expected cost for the remaining work with the Gompertz growth model using nonlinear regression curve fitting [12].

One area of development in the field of project management is the combination of Earned Value Management and risk or uncertainty in project information. In this area, Naeni & Salehipour (2011) proposed an approach to evaluating and controlling a project by providing new indices to measure both project performance and project progress under fuzzy circumstances [13]. In another article, Naeni et al. (2014) proposed a new fuzzy-based earned value model with the advantage of developing and analyzing the earned value indices, and the time and the cost estimates at completion under unreliability [14]. Wauters & Vanhoucke (2014), in an article, used the combination of Monte Carlo simulation and Support Vector Machine (SVM) approaches to predict the time and cost at completion [15]. In another paper, they also proposed the use of five other Artificial Intelligence (AI) methods including

Decision Tree Learning (DTL), Support Vector Machine (SVM), Boosting, Bagging, Random Forest, Principal Component Analysis (PCA), Cross-Validation, and Grid Search. Finally, strengths and weakness of the proposed method were determined by sensitivity analysis [16]. Azaron and Fatemi Ghomi (2007) applied the stochastic dynamic programming to estimate the mean project completion time in dynamic Markov PERT networks. The activity durations were assumed to be independent random variables with exponential distributions; however, some social and economic problems affect the mean of activity durations [17].

Baghiabad and KhademiZare (2015) developed a three-stage algorithm for software cost and time estimation in fuzzy environment [18].

Acebes et al. (2014) offered a new approach by using Monte Carlo simulation and Earned Value Method (EVM) for controlling projects under uncertainty [19]. Then, in 2015, another article with statistical learning techniques combined with the previous methods has proposed an algorithm to predict the time and the final cost of the project [20].

This paper also presents a method to predict the time and cost at completion of the project under unreliability based on the Earned Value Method using a hybrid algorithm including Monte Carlo simulation and an artificial neural network.

The rest of the paper is organized as follows. Section 2 presents the research methodology. In Section 3, for clarification purposes, a case study is studied. The paper ends with the conclusion.

2. Methodology

2-1. Traid definition (x_t, t, AC_t)

According to Acebes et al. (2015), each point of a project can be shown by three characteristics: time (t), the percentage of completion at the time of t (x_t), and Actual Cost at time t (AC_t). This point can be revealed in the form of a triple (x_t, t, AC_t) called the traid of project. For example, (50%, 3, 150\$) indicates that, on the third day of the project, the percentage of completion is 50% and the actual cost of the project is 150\$.

By considering Budget at completion (BAC), the percentage of completion at the time of t (x_t) can be defined as $x_t = EV_t / BAC$. In the

previous example, assuming the total budget of 240\$, the earned value (EV) on the third day is 120\$.

2-2. Monte carlo simulation

A project under uncertainty is assumed in which the time and cost of the project activities are randomly determined by a statistical distribution. Moreover, it is supposed that the time schedule of the project is planned based on a specific scenario including the mean of time and cost of activities. Therefore, using the Monte Carlo simulation, a considerable number of conditions that may occur up to the completion of the project can be generated based on the statistical distribution of the time and cost of the activities.

In each of Monte Carlo simulation run, such as in the j th run, $j \in \{1, 2, \dots, n\}$, Earned Value (EV_{ij}), and Actual Cost (AC_{ij}) can be obtained in each point of project. In addition, the j th project percentage of completion in each t (x_{ij}) can be calculated based on the Budget at Completion (BAC). In the j th run of Monte Carlo simulation, the point of time is assumed in which the percentage of completion is received to a certain amount of progress (x), or Earned Value takes a specific amount of $EV = x * BAC$, called AT_j with the cost of AC_{AT_j} ; in addition, the cost and time at completion in this run are considered as CAC_j and TAC_j , respectively (Figure 2).

Therefore, assuming that the delay (rush) and deficit (surplus) budget for the project occur on the basis of the random nature of the project.

Based on the results of Monte Carlo simulations, by assuming a certain amount of x , data consisting of AT_j , AC_{AT_j} , CAC_j , and TAC_j , $j \in \{1, 2, \dots, n\}$ can be gathered.

By considering AT_j , AC_{AT_j} as inputs and CAC_j and TAC_j as outputs, the aim is to develop a model for mapping inputs to outputs.

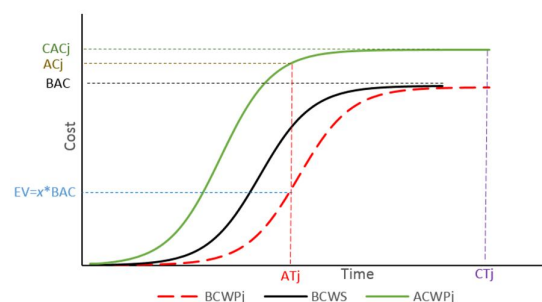


Fig. 2. Traid in the j th Monte Carlo simulation run

The method used in this study for prediction is Artificial Neural Network (ANN) that will be discussed briefly below.

2-3. Artificial neural network (ANN)

Artificial Neural Network (ANN) is one of the powerful techniques among the existing prediction techniques [21]. In this study, an artificial neural network (ANN) with Back-Propagation Algorithm was used to predict the time and cost of completing the project. ANN comprised two or more non-linear functions in a multi-layer structure. The proposed network consists of three layers: an input layer, middle (hidden) layer, and output layer (Figure 3).

The output of each layer, O , is a function of net or $O(net)$. (Equation 1).

$$= w_0x_0 + w_1x_1 + \dots + w_nx_n = w_0 + \sum_{i=1}^n w_i x_i \quad (1)$$

In Equation (1), w_i and x_i are model weights (parameters) and input variables of the layer, respectively. w_0 determines bias and x_0 is a vector of ones.

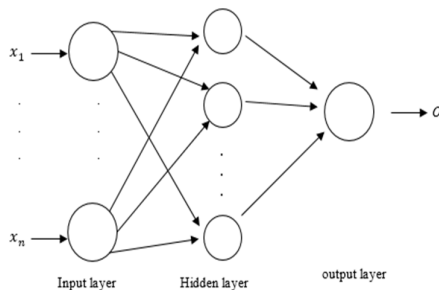


Fig. 3. ANN structure

In the three-layer model, two output functions can be demonstrated as follows:

Equation (2) is related to the middle layer and Equation (3) is related to the output layer.

$$h_{j,k} = \psi \left[w_{j0} + \sum_{i=1}^q w_{ji} x_{i,k} \right] \quad j=1, \dots, n \quad (2)$$

$$O_k = \phi \left[w_o + \sum_{j=1}^n w_j h_{j,k} \right] \quad (3)$$

Or, in short, the output will be equal to:

$$O_k = \phi \left[w_o + \sum_{j=1}^n w_j \psi \left[w_{j0} + \sum_{i=1}^q w_{ji} x_{i,k} \right] \right] \quad (4)$$

Transfer functions in both layers are often the same ($\Psi = \Phi$).

$x_{i,k}$ independent variables (where i represents the corresponding variable to K index) send some signals by using transfer function Ψ with various weights w_{ji} (where i and j indices represent middle node and independent variables, respectively). The activation of each intermediate unit (h_{jk}) will activate output layer by using a transfer function with variable w_j weights. Thus, the result of such a network output is O_t .

w weights are in the range of 0 and 1, and changing layers are performed on the basis of corresponding signals in middle layers. The outputs obtained from the network, O_k , are compared with the (t_k) target, and the error rate ($t_k - o_k$) is obtained to observe k value and the sum of the squared errors for all observations. Then, using the error backpropagation algorithm, these errors are propagated from the output layer to the middle one; moreover, in a duplicate process, the error (total squared error) is minimized by adjusting the weights. The whole process can be described in two steps: A forward step where the weights and outputs of the network are determined, and a backward step in which the errors are propagated reversely and, then, minimized in an iterative process. On the other hand, error index of

$$Err = 1/2 \sum_{k=1}^N (t_k - o_k)^2$$

can also be used as a

stopping point. In fact, the error functions in this method and the traditional regression methods are the same. As mentioned, this error rate is propagated to previous layers and, in order to minimize it, the weights of the transfer function are adjusted; thus,

$$\frac{\partial Err}{\partial w_i} = \frac{\partial \left[\frac{1}{2} (t - o)^2 \right]}{\partial w_i} \quad (5)$$

And

$$w_i = w_i - \mu \cdot \frac{\partial Err}{\partial w_i} \quad (6)$$

In Equation (6), μ is the learning rate that defines the lengths of the step of each repetition in a negative direction of the gradient. This process continues until the error indicator is less than the predetermined

($Err < \xi$) ξ value, or if a predetermined number of repetitions occurs.

Although neural networks may provide good results in the range of data used in training, they are very susceptible to encountering many errors when using new data and, thus, providing poor results for out-of-sample predictions. In fact, this problem occurs due to overfitting of the networks with model data. To solve this problem, an early stopping method is one of the simplest and most widely used methods. In this method, the sum of model data is divided into two groups: a group of data used to estimate network weights is called training data. In addition, the second group is called validation set. During the process of teaching the model using training data, the error is controlled in validation data. When the network encounters an overfitting problem, the error in the validation data is increased. When the error shows the increment in several repetitions, teaching process will stop and estimated weights in a repetition with minimal error in validation data are considered as the final weights.

In order to increase the speed of learning, several algorithms are presented: a Levenberg-Marquardt algorithm has increased the network

weighting speed by using Jacobin matrix. This algorithm is used for network training.

3. Case Study

In this section, in order to illustrate the proposed method, the example of previous researches [20, 22] with activity on node network (AON) is used (Fig. 4).

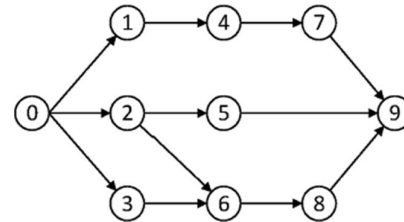


Fig. 4. AoN of case study

Duration of each activity is assumed to be a random variable with a normal distribution and mean and variance which is shown in Table 1. Furthermore, the variable cost of each activity, which represents the cost of each activity per time unit, is also presented in Table 1. In fact, in this example, the cost of each activity is linearly dependent on the time of activity.

This analysis is based on the assumption that the percentage of project progression is 50%.

Tab 1. Duration and cost of activities of the case study. Duration activities are modelled as normal distributions and costs depend linearly on duration

| Id. activity | Duration | Variance | Variable cost |
|--------------|----------|----------|---------------|
| 1 | 2 | 0.15 | 755 |
| 2 | 4 | 0.83 | 1750 |
| 3 | 7 | 1.53 | 93 |
| 4 | 3 | 0.56 | 916 |
| 5 | 6 | 1.72 | 34 |
| 6 | 4 | 0.28 | 1250 |
| 7 | 8 | 2.82 | 875 |
| 8 | 2 | 0.14 | 250 |

For scheduling the base line of the project, the mean of duration of activities is considered.

Thus, it is possible to make calculations for the planned value (PV). Table 2 shows the obtained values during the project base line.

Tab. 2. Planned value (PV)

| Id. activity | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|--------------|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|
| 1 | 755 | 0 | 0 | 1510 | 1510 | 1510 | 1510 | 1510 | 1510 | 1510 | 1510 | 1510 | 1510 |
| 2 | 175 | 350 | 525 | 7000 | 7000 | 7000 | 7000 | 7000 | 7000 | 7000 | 7000 | 7000 | 7000 |

| | | | | | | | | | | | | | |
|---|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|
| | 0 | 0 | 0 | | | | | | | | | | |
| 3 | 93 | 186 | 279 | 372 | 465 | 558 | 651 | 651 | 651 | 651 | 651 | 651 | 651 |
| 4 | 0 | 0 | 916 | 1832 | 2748 | 2748 | 2748 | 2748 | 2748 | 2748 | 2748 | 2748 | 2748 |
| 5 | 0 | 0 | 0 | 0 | 34 | 68 | 102 | 136 | 170 | 204 | 204 | 204 | 204 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1250 | 2500 | 3750 | 5000 | 5000 | 5000 |
| 7 | 0 | 0 | 0 | 0 | 0 | 875 | 1750 | 2625 | 3500 | 4375 | 5250 | 6125 | 7000 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 250 | 500 |
| P | 259 | 519 | 795 | 1071 | 1175 | 1275 | 1376 | 1592 | 1807 | 2023 | 2236 | 2348 | 2461 |
| V | 8 | 6 | 5 | 4 | 7 | 9 | 1 | 0 | 9 | 8 | 3 | 8 | 3 |

To analyze the above analysis and based on the data in Table 1, 10000 samples were simulated by using Crystal Ball software. In order to display the process of analysis, one of the simulated samples is presented in Table 3.

Tab. 3. Duration and cost of the simulated sample

| Id. activity | Duration | Cost |
|--------------|----------|--------|
| 1 | 2.5 | 1904.9 |
| 2 | 3.5 | 6050.3 |
| 3 | 7.2 | 669.6 |
| 4 | 3.2 | 2931.2 |
| 5 | 5.7 | 192.7 |
| 6 | 3.8 | 4723.7 |
| 7 | 8.5 | 7437.5 |
| 8 | 2.0 | 498.8 |

According to Table 3, the actual cost of activities (AC) can be calculated. The results of these calculations are presented in Table 4. As can be seen, in the sample, the total completion time of project (CT) is 15 units with a cost at completion (CAC) of 24409 currency unit. Moreover, regarding the data in Tables 3 and 4, the earned value (EV) can be calculated for this

sample which is provided in Table 5. In the last row of Table 5, the percentage of progress (x) at the end of each time unit is observed.

Since x is analyzed in 50 percent according to the assumptions of projects, actual time (AT) and actual cost (AC) should be obtained in the simulated project. As shown in Table 5, the actual time is between 6- and 7-time units. It can be estimated using linear interpolation; accordingly, the actual time (AT) will be 6.192 units. On the other hand, considering data in Table 4, actual cost (AC) equal to 11969.81 is obtained by using a linear interpolation. In other words, it can be said that the project at 6,192-time units (Actual Time(AT)) with a cost of 11969.81 currency unit (Actual Cost(AC)) has had 50 percent(x) progress. Moreover, the project ended up at a time unit of 15 (CT) and at a cost of 24,630 (CAC). These values are calculated for all 10,000 generated samples. Finally, a dataset of 10,000 records with two input fields, containing actual time (AC) and actual cost (AC), and two output fields, including completion time (CT) and cost at completion (CAC), are obtained to enter the next stage.

Tab. 4. Actual cost of the simulated sample (AC)

| Id. activity | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|--------------|-----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 755 | 15 | 19 | 19 | 190 | 190 | 190 | 190 | 190 | 190 | 190 | 190 | 190 | 190 | 190 |
| 2 | 175 | 35 | 52 | 60 | 605 | 605 | 605 | 605 | 605 | 605 | 605 | 605 | 605 | 605 | 605 |
| 3 | 93 | 18 | 27 | 37 | 465 | 558 | 651 | 670 | 670 | 670 | 670 | 670 | 670 | 670 | 670 |
| 4 | 0 | 0 | 43 | 13 | 226 | 293 | 293 | 293 | 293 | 293 | 293 | 293 | 293 | 293 | 293 |
| 5 | 0 | 0 | 0 | 18 | 52 | 86 | 120 | 154 | 188 | 193 | 193 | 193 | 193 | 193 | 193 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 225 | 350 | 472 | 472 | 472 | 472 | 472 |
| 7 | 0 | 0 | 0 | 0 | 0 | 242 | 111 | 199 | 286 | 374 | 461 | 549 | 636 | 724 | 743 |

| | | | | | | | | | | | | | | | |
|---|-----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | | | | | | 7 | 2 | 7 | 2 | 7 | 2 | 7 | 2 | 8 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 255 | 499 | 499 | 499 |
| A | 259 | 51 | 78 | 96 | 107 | 117 | 127 | 147 | 168 | 189 | 210 | 222 | 233 | 242 | 244 |
| C | 8 | 96 | 71 | 99 | 42 | 73 | 75 | 03 | 62 | 91 | 95 | 20 | 38 | 13 | 09 |

Tab. 5. Earned value of the simulated sample (EV) and the percentage of progress (x)

| Id. activity | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|--------------|-----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 598 | 11 | 15 | 151 | 151 | 151 | 151 | 151 | 151 | 151 | 151 | 151 | 151 | 151 | 151 |
| 2 | 202 | 97 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 5 | 40 | 60 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 |
| 4 | 90 | 49 | 74 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 18 | 27 | 362 | 452 | 543 | 633 | 651 | 651 | 651 | 651 | 651 | 651 | 651 | 651 |
| 6 | 0 | 1 | 1 | 41 | 126 | 212 | 274 | 274 | 274 | 274 | 274 | 274 | 274 | 274 | 274 |
| 7 | 0 | 0 | 0 | 0 | 8 | 7 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 8 | 0 | 0 | 0 | 20 | 56 | 92 | 128 | 164 | 200 | 204 | 204 | 204 | 204 | 204 | 204 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 105 | 238 | 370 | 500 | 500 | 500 | 500 | 500 |
| 10 | 0 | 0 | 0 | 0 | 0 | 228 | 105 | 8 | 2 | 5 | 0 | 0 | 0 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 187 | 269 | 352 | 434 | 516 | 599 | 681 | 700 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 9 | 2 | 6 | 9 | 3 | 6 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 256 | 500 | 500 | 500 |
| E | 271 | 54 | 82 | 101 | 111 | 121 | 130 | 150 | 171 | 193 | 214 | 225 | 236 | 244 | 246 |
| V | 4 | 27 | 65 | 60 | 45 | 20 | 70 | 06 | 89 | 40 | 64 | 38 | 06 | 29 | 13 |
| x | 11 | 22 | 34 | 41 | 45 | 49 | 53 | 61 | 70 | 79 | 87 | 92 | 96 | 99 | 100 |
| | % | % | % | % | % | % | % | % | % | % | % | % | % | % | % |

Considering the differences between units of measurement of these two outputs and to increase the accuracy of prediction, two separate neural networks were used to predict each of completion time (CT) and cost at completion (CAC). Then, nntool of MATLAB software was utilized to build, train, and test the above networks. The percentages of training data, validation, and test are 70, 15, and 15, respectively. In addition, the learning algorithm of Leonberg-Marquard algorithm was used to train the network.

Prior to entering data for network training, preprocessing was done to normalize them; to do so, the premmx function of MATLAB software was used.

In order to set the two main parameters consisting of the number of middle layer nodes and the learning rate which affect the network accuracy (mean square error), the Response Surface Methodology (RSM) was used for both networks.

For example, for the network of estimating CT, the middle layer at levels of (3,7,9,11,15) and learning rates in the range of 0.01 to 0.2% were

tested. Then, 37 experiments consisted of 32 non-central and 5 central points (with the number of repetitions of non-central points equal to 4 and the number of axes equal to 4) were designed through Design-Expert software. After analyzing the ANOVA table and estimating the coefficients, function of the MSE is fitted:

$$MSE = 0.039992 + 0.098085\mu + 1.24382 * 10^{-3}n - 0.010467\mu n \quad (7)$$

The optimal MSE value was obtained based on this function. This value is 0.0505 at 0.073 levels for the learning rate and 7 for the number of nodes. The trained network with this setting will be named "netct".

Furthermore, in order to adjust the number of nodes in the middle layer of the network of CAC, a similar process is followed. The number of middle nodes in this network equals 11 and the learning rate is set to 0.105. At the end, the trained network is saved as "netcac".

After training and testing these two networks, both the actual cost and actual time can be estimated at 50%. For instance, whereas, in the reality, the percentage of progress is 50% in the time unit of 5 with an actual cost of 11,000 currency units; first, these input data should be normalized by using `trannmx` command and enter into two networks; then, the obtained outputs must be extracted. To estimate the cost of (CEAC) and the actual time of (ECT) by using `Postmnmx` command, postprocessing is done on the obtained results. In this example, with the above steps, it is expected that the project will be completed at a time unit of 14 with a cost of 23283.

In this example, the assumption of the percentage of progress is equal to 50%. Obviously, for any amount of x , this process will be implemented and, at any time of the project, the time and cost of completion of the project can be predicted by knowing the percentage of project progress and its actual cost.

4. Conclusion

Since predicting the cost and completion time of project represents the significant demands of project stakeholders, in this research, the project activities are unreliable and the deviations occurring during the project are all due to this unreliability. Therefore, the project's behavior was obtained from the beginning to the end by using Monte Carlo simulation with the assumption of normal distribution for the time and cost of activities in a significant number of runs. In each run, certain amounts of the project progress percentage (x), the actual time (AT), and the actual cost (AC) were obtained correspondingly. In addition, by specifying the project completion time (CT) and cost at completion (CAC) in each run, a database of performances was formed. Finally, to predict the time and cost of the project and by knowing the time and cost of the project with the percentage of progression x , an artificial network was trained and tested.

References

- [1] Q. W. Fleming and J. M. Koppelman, "Earned value project management," 2016: Project Management Institute.
- [2] D. S. Christensen, R. C. Antolini, and J. W. McKinney, "A review of estimate at completion research," in *The Journal of Cost Analysis*, Vol. 12, No. 1, (1995), pp. 41-62.
- [3] F. T. Anbari, "Earned value project management method and extensions," *Project management journal*, Vol. 34, No. 4, (2003), pp. 12-23.
- [4] W. Lipke, "Schedule is different," *The Measurable News*, Vol. 31, No. 4, (2003), pp. 31-34.
- [5] K. Henderson, "Earned schedule: A breakthrough extension to earned value theory? A retrospective analysis of real project data," *The Measurable News*, Vol. 1, No. 2, (2003), pp. 13-23.
- [6] K. Henderson, "Further developments in earned schedule," *The measurable news*, Vol. 1, No. 1, (2004), pp. 15-22.
- [7] D. Jacob, "Forecasting project schedule completion with earned value metrics," *The Measurable News*, Vol. 1, No. 11, (2003), pp. 7-9.
- [8] M. Vanhoucke, & Vandevoorde, S. "Measuring the accuracy of earned value/earned schedule forecasting predictors." *Measurable News. WINTER 2007-2008*, (2007), pp. 26-30.
- [9] K. Henderson and W. Lipke, "Earned schedule: An emerging enhancement to earned value management," *Cross Talk, Journal of Defense Software Engineering*, (2006), pp. 26-30.
- [10] W. H. Lipke, *Earned Schedule: An Extension to Earned Value Management--for Managing Schedule Performance*. Lulu Pub., (2009).
- [11] C.-C. Tseng, "Statistical analysis for comparison of overall performance of projects using Weibull analysis on earned value metrics," *Journal of the Chinese Institute of Industrial Engineers*, Vol. 28, No. 3, (2011), pp. 165-178.

- [12] T. Narbaev and A. De Marco, "An earned schedule-based regression model to improve cost estimate at completion," *International Journal of Project Management*, Vol. 32, No. 6, (2014), pp. 1007-1018.
- [13] M. Naeni and A. Salehipour, "Evaluating fuzzy earned value indices and estimates by applying alpha cuts," *Expert systems with Applications*, Vol. 38, No. 7, (2011), pp. 8193-8198.
- [14] L. M. Naeni, S. Shadrokh, and A. Salehipour, "A fuzzy approach for the earned value management," *International Journal of Project Management*, Vol. 32, No. 4, (2014), pp. 709-716.
- [15] M. Wauters and M. Vanhoucke, "Support vector machine regression for project control forecasting," *Automation in Construction*, Vol. 47, (2014), pp. 92-106.
- [16] M. Wauters and M. Vanhoucke, "A comparative study of Artificial Intelligence methods for project duration forecasting," *Expert Systems with Applications*, Vol. 46, (2016), pp. 249-261.
- [17] A. Azaron , S.M. Fatemi Ghomi , " Expected Duration of Dynamic Markov PERT Networks, " *International Journal of Industrial Engineering & Production Research*, Vol 18, No. 3, (2007), pp. 1-5.
- [18] F. ZareBaghiabad , H. KhademiZare, " A Three- Stage Algorithm for Software Cost and Time Estimation in Fuzzy Environment," *International Journal of Industrial Engineering & Production Research*, September 2015, Vol. 26, No. 3, (2015), pp.193-211.
- [19] F. Acebes, J. Pajares, J. M. Galán, and A. López-Paredes, "A new approach for project control under uncertainty. Going back to the basics," *International Journal of Project Management*, Vol. 32, No. 3, (2014), pp. 423-434.
- [20] F. Acebes, M. Pereda, D. Poza, J. Pajares, and J. M. Galán, "Stochastic earned value analysis using Monte Carlo simulation and statistical learning techniques," *International Journal of Project Management*, Vol. 33, No. 7, (2015), pp. 1597-1609.
- [21] H. Farughi, A. Hakimi, and R. Kamranrad, "Availability Prediction Of The Repairable Equipment Using Artificial Neural Network And Time Series Models," *International Journal of Industrial Engineering & Production Research*, Vol. 29, No.1, (2018), pp. 79-90.
- [22] O. Lambrechts, E. Demeulemeester, and W. Herroelen, "Proactive and reactive strategies for resource-constrained project scheduling with uncertain resource availabilities," *Journal of scheduling*, Vol. 11, No. 2, (2008), pp. 121-136.

Follow This Article at The Following Site

Akhbari M. Project Time and Cost Forecasting using Monte Carlo simulation and Artificial Neural Networks. *IJIEPR*. 2018; 29 (2) :231-239
URL: <http://ijiepr.iust.ac.ir/article-1-793-en.html>

