Software Cost Estimation Using Artificial Neural Network

Devbrat Sahu
M.Tech Research Scholar, CSE, Chattrapati Shivaji Institute of Technology
Durg, India

Mrs. Deepty Dubey
Associate Professor, CSE, Chattrapati Shivaji Institute of Technology
Durg, India

Abstract— The software project cost and effort estimation is to scientifically estimate the required workload and its corresponding costs in the life cycle of software system. Software cost estimation is a complex activity that requires knowledge of a number of key attributes that affect the outcomes of software projects, both individually and in concert. The most critical problem is the lot of data is needed, which is often impossible to get in needed quantities. Artificial Neural Network model is one of building design that is utilized all the more as a part of software cost estimation. In this paper Artificial Neural Network (ANN) has been used using dataset for prediction of software effort. The performance indices such as Sum-Square-Error (SSE), Mean-Square-Error (MSE), Root-Mean-Square-Error (RMSE) and Mean-Absolute-Error (MAE) are calculated from ANN technique and the results obtained from these parameters are discussed for software cost estimation.

Keywords—Software effort estimation, Artificial Neural Network, Source lines of code.

I. INTRODUCTION

Software has played an increasingly important role in systems acquisition, engineering and development, particularly for large, complex systems. For such systems, accurate estimates of the software costs are a critical part of effective program management [4]. The bulk of the cost of software development is due to the human effort and most cost estimation methods focus on cost and effort aspect and give estimates in terms of person-months. Accurate cost estimates are critical to both developers and customers. They can be used for request for proposal, contract negotiations, scheduling, monitoring and control. Underestimating the costs may result in management approving proposed systems that then exceed their budgets, with underdeveloped functions and poor quality, and failure to complete on time. Overestimating may result in too many resources committed to the projects or during contract bidding result is not winning the contract, which can lead to loss of jobs. Software project failures have been an important subject in the last decade. Software projects usually don’t fail during the implementation and most project fails are related to the planning and estimation steps. Hence accurate cost estimation is become a challenge for IT industries.

Cost estimation is usually measured in terms of effort [2]. The most common metric used is person months or years (or man months or years). The effort is the amount of time for one person to work for a certain period of time.

An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks.

II. RESEARCH METHODOLOGY

A. ARTIFICIAL NEURAL NETWORK

A neural network is a simplified model of the biological neuron system. It is a massively parallel distributed processing system made up of highly interconnected neural computing elements that have ability to learn and thereby acquire knowledge and make it available for use. McCulloch and Pitts (1943) proposed the first computational model of a neuron, namely the binary threshold unit, whose output was either 0 or 1 depending on whether [3] its net input exceeded a given threshold. The most common algorithm for training or learning is known as error back-propagation algorithm. The error back propagation learning consists of two phases: a forward pass and a backward pass, an input is presented to the neural network, and its effect is propagated through the network layer by layer. This is also called Testing Phase. During the forward pass the weights of the network are all fixed. During the backward pass or “Training Phase”, the weights are all updated and adjusted according to the error computed. An error is composed from the difference between the desired response and the system output. This error information is feedback to the system and adjusts the system parameters in a learning rule. The process is repeated until the performance is acceptable.

B. EVALUATING THE PERFORMANCE OF THE MODELS

The main measures used for evaluating the performance of machine learning techniques for predicting the software effort are as follows[3]:-
1. **Sum Squared Error (SSE)**: The sum squared error is defined as.

\[
E = \left( \sum_{i=1}^{n} (P_i - A_i)^2 \right)
\]

Where \( P_i \) = Estimated value for data point \( i \);
\( A_i \) = Actual value for the data point \( i \);
\( n \) = Total number of data points.

2. **Mean Squared Error (MSE)**: The mean squared error is defined as.

\[
E = \left( \frac{1}{n} \sum_{i=1}^{n} (P_i - A_i)^2 \right)
\]

Where \( P_i \) = Estimated value for data point \( i \);
\( A_i \) = Actual value for the data point \( i \);
\( n \) = Total number of data points.

3. **Root Mean Squared Error (RMSE)**: The root mean squared error is defined as.

\[
E = \left( \frac{1}{n} \sum_{i=1}^{n} (P_i - A_i)^2 \right)^{1/2}
\]

Where \( P_i \) = Estimated value for data point \( i \);
\( A_i \) = Actual value for the data point \( i \);
\( n \) = Total number of data points.

4. **Mean Absolute Error (MAE)**: The mean absolute error measures of how far the estimates [3] are from actual values. It could be applied to any two pairs of numbers, where one set is “actual” and the other is an estimate prediction.

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |P_i - A_i|
\]

Where \( P_i \) = Estimated value for data point \( i \);
\( A_i \) = Actual value for the data point \( i \);
\( n \) = Total number of data points.

### III. RESULT ANALYSIS

The Artificial Neural Network (ANN) has been used for predicting the software efforts using dataset. The performance indices and the results obtained from these models. These indices are Sum-Square-Error (SSE), Mean-Square-Error (MSE), Root-Mean-Square-Error (RMSE), Mean-Absolute-Error (MAE).

MATLAB programs were developed for training and testing of various models and also for computation of performance indices. The results are tabulated in Table 2.

### Table 1 Data Set of Software Cost Estimation

<table>
<thead>
<tr>
<th>KLOC</th>
<th>ME</th>
<th>a</th>
<th>b</th>
<th>AE</th>
<th>EE</th>
</tr>
</thead>
<tbody>
<tr>
<td>902000</td>
<td>30</td>
<td>3.3602</td>
<td>0.8116</td>
<td>115.8000</td>
<td>115.4923</td>
</tr>
<tr>
<td>462000</td>
<td>20</td>
<td>3.3820</td>
<td>0.7920</td>
<td>96</td>
<td>95.6978</td>
</tr>
<tr>
<td>465000</td>
<td>19</td>
<td>3.4520</td>
<td>0.7890</td>
<td>90.8000</td>
<td>82.5803</td>
</tr>
<tr>
<td>311000</td>
<td>35</td>
<td>3.2400</td>
<td>0.8120</td>
<td>36.6000</td>
<td>18.3678</td>
</tr>
<tr>
<td>675000</td>
<td>29</td>
<td>3.2480</td>
<td>0.8116</td>
<td>98.4000</td>
<td>88.9900</td>
</tr>
<tr>
<td>128000</td>
<td>26</td>
<td>3.1580</td>
<td>0.8470</td>
<td>82.5803</td>
<td>10.1283</td>
</tr>
<tr>
<td>105000</td>
<td>34</td>
<td>3.3240</td>
<td>0.8321</td>
<td>10.3000</td>
<td>7.1042</td>
</tr>
<tr>
<td>215000</td>
<td>31</td>
<td>3.3167</td>
<td>0.8321</td>
<td>28.5000</td>
<td>20.4246</td>
</tr>
<tr>
<td>310000</td>
<td>26</td>
<td>3.2980</td>
<td>0.8120</td>
<td>7</td>
<td>5.8316</td>
</tr>
</tbody>
</table>

### Table 2 Performance indices with Artificial Neural Network

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Performance Measures</th>
<th>Artificial Neural Network (ANN) Ten Hidden Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Sum Squared Error (SSE)</td>
<td>3404.9442</td>
</tr>
<tr>
<td>2.</td>
<td>Mean Squared Error (MSE)</td>
<td>340.4944</td>
</tr>
<tr>
<td>3.</td>
<td>Root Mean Squared Error (RMSE)</td>
<td>18.4524</td>
</tr>
<tr>
<td>4.</td>
<td>Mean Absolute Error (MAE)</td>
<td>3404.9</td>
</tr>
</tbody>
</table>

### IV. CONCLUSION

The Artificial Neural Network (ANN) has been used to analyze the results using dataset for software development effort. A similar study can be carried out to predict software effort using prediction models based on other machine learning algorithms such as Genetic Algorithms. Cost benefit
analysis of models may be carried out to determine whether a given effort prediction model would be economically viable.

References


