An Empirical Validation of Software Cost Estimation Model Using Fuzzy Technique

Ritu Agarwal 1 , Isha Awasthi 2, Saoud Sarwar 3

1 ASSIST. PROFF. VISHVESHWARYA GROUP OF INSTITUTIONS, GREATER NOIDA, AFFILIATED TO GBTU
Rituagarwal8585@Gmail.Com
2 ASSIST. PROFF. VISHVESHWARYA GROUP OF INSTITUTIONS, GREATER NOIDA, AFFILIATED TO GBTU
3 HOD, DEPARTMENT OF CSE, AFSET, DHUJ, FARIDABAD, AFFILIATED TO M.D.U., ROHTAK, HARYANA,

Abstract: Software effort estimation is the greatest challenge for software developers. Software effort estimation is the process of predicting most realistic use of effort required to develop or maintain software based on incomplete and uncertain input. There are various models suggested by researchers for calculating effort. But no one can give accurate result. The best result are achieved by using soft computing technique. The various methodology used in soft computing for effort estimation are based on fuzzy logic, neural network and genetic algorithm. In this paper we have used the fuzzy logic approach. MATLAB is used for tuning the parameters of famous various cost estimation models. On published software projects data, the performance of the model is evaluated. Comparison of results from our model with existing ubiquitous models is done.

Keywords: LOC based model, fuzzy logic, triangular fuzzy number, membership function and fuzziness

1. INTRODUCTION

The ability to estimate accurately the size, effort and duration of a software development project is crucial to project success. Inaccurate software estimate causes trouble in business process related to software development. As the demand for software application increases the software companies need accurate estimation of project under development. Cost Estimation is achieved in terms of person-months (PM), which can be translated into actual dollar cost. The concept of software cost estimation has been growing rapidly due to practicality and demand for it. Today the peoples are expecting high quality software with a low cost, the main objective of software engineering. So many popular cost estimation models like COCOMO, FP, Delphi, Halsted Equation, Bailey- Basili, Doty, barry boehm and Anish Mittal Model had came into existence. These models are created as a result of regression analysis and power regression analysis methods applied to historical data. Today most of the software companies follow COCOMOII for estimating the cost of products; we found some variations in this model [2], [6-11]. Newer computation techniques to cost estimation that are non-algorithmic were desired in the 1990’s. Researchers particularly have turned their attention to a set of approaches that are soft computing based. These include fuzzy logic models artificial neural networks, and genetic algorithms.

Fuzzy logic with its features of a powerful linguistic representation can signify imprecision in inputs and outputs, while providing a more expert knowledge based approach to model building. The fuzzy logic model uses the fuzzy logic concepts introduced by Lofti A. Zadeh [4], [5], [14].

Fuzzy reasoning consists of following three primary components: fuzzification process, inference from fuzzy rules and defuzzification process. Fuzzification process is where the objective term is transformed into a fuzzy concept. The membership functions are applied to the actual values of variables to determine the confidence factor or membership value (MV for short). Fuzzification allows the input and output to be expressed in linguistic terms. Inferencing involves defuzzification of the conditions of the rules and propagation of the confidence factors of the conditions to the conclusion of the rules. A number of rules will be fired and the inference engine assigned the particular outcome with the maximum MV from all the fired rules. Defuzzification process refers to the translation of fuzzy output into objective terms. Fuzzy numbers are one of the ways to describe data vagueness, obscurity and imprecision. A fuzzy number is an extension of a regular number in the sense that it does not refer to one single value but rather to a connected set of possible values, where
An Empirical Validation of Software Cost Estimation Model Using Fuzzy Technique

...each possible value has its own weight between ‘0’ and ‘1’. This weight is called the membership function. The membership function is increasing towards the mean and decreasing away from it. The Fuzzy number can be of three types 1) Triangular fuzzy Number 2) Trapezoidal fuzzy number 3) Bell shaped fuzzy number. The following figure shows the three curves [3]

(a) Fuzziness
Fuzziness results from imprecise boundary of fuzzy set. Fuzziness of any set is measured by metric distance between its membership grade function and its nearest crisp set. A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The graphical representations may include different shape. The “shape” of the membership function is an important criterion that has to be considered. There are different methods to form membership functions. Zadeh proposed a series of membership functions that could be classified into two groups: those made up of straight lines, or “linear,” and Gaussian forms, or “curved.” [4]

(b) Fuzzy Logic
Fuzzy logic is a methodology, to solve problems which are too complex to be understood quantitatively, based on fuzzy set theory [4,5]. Use of fuzzy sets in logical expression is known as fuzzy logic. A fuzzy set is characterized by a membership function, which associates with each point in the fuzzy set a real number in the interval [0,1], called degree or grade of membership. The membership function may be triangular, trapezoidal, parabolic etc. A triangular fuzzy number (TFN) is described by a triplet \((\alpha, m, \beta)\), where \(m\) is the model value, \(\alpha\) and \(\beta\) are the right and left boundary respectively. Size of the project, especially in the beginning of the project, cannot be taken precisely. It can be taken as a fuzzy number. The effort is estimated in terms of distribution described by membership function of the effort. A single fuzzy estimate of effort is calculated by defuzzification technique given in defuzzification section of the paper.

II. PROPOSED METHOD

We Use Gaussian bell shape Fuzzy number \(G(n)\) which is defined as follows:

\[
G(N) = \begin{cases} 
 e^{-k(N-m)} & \text{if } m < n < m \\
 e^{-k(N-\beta)} & \text{if } m < n < m
\end{cases}
\]

Where \(n\) is the size as input, \(E\) is the effort as output, \(\alpha, \beta, m, n\) are the parameter of membership function \(G(n)\). \(m\) is the mid value. \(\alpha\) and \(\beta\) are the left and right boundaries respectively.

From S Function

\[
m = (\alpha + \beta)/2 \tag{1}
\]

\[
\alpha + \beta = 2m \tag{2}
\]

as by the definition of fuzziness
\[
F = \frac{\beta - \alpha}{2m} \quad (3)
\]
\[
\beta - \alpha = 2mf \quad (4)
\]
by solving 2 and 4
\[
\alpha = (1-F)\times m \quad (5)
\]
\[
\beta = (1+F)\times m \quad (6)
\]
Similarly, the GFN \( \mu(E) \) is defined as

\[
\mu(E) = \begin{cases} 
\frac{e^{-(b+c)x}}{e^{-(b+c)x}+e^{-(a+b)x}} & x < (c-a)/(b+a) \\
\frac{e^{-(a+b)x}}{e^{-(a+b)x}+e^{-(a+b)x}} & x > (c-a)/(b+a) 
\end{cases}
\]

Table 1 gives the values for \( \alpha \) and \( \beta \) for \( F=0.1, 0.2, 0.3 \) for various values using equation 5 and 6, where \( m \) is the size estimate in KLOC.

\[
\begin{array}{|c|c|c|}
\hline
F & \alpha & \beta \\
\hline
0.1 & 0.9m & 1.1m \\
0.2 & 0.8m & 1.2m \\
0.3 & 0.7m & 1.3m \\
\hline
\end{array}
\]

\((a)\) Defuzzification

Fuzzy effort estimate (Em) is given as
An Empirical Validation of Software Cost Estimation Model Using Fuzzy Technique

\[
E_a = \frac{1}{2a(2m^2 + 1)} + \frac{1}{2b(2m^2 + 1)} + \frac{1}{2c(2m^2 + 1)} \sqrt[3]{(w_1 + w_2 + w_3)^2}
\]

Where \( a = 0.01179 \), \( b = -0.00002 \) obtained by using MATLAB, \( m \) represents the size in KLOC

\[
\alpha = (1 - F) \times m
\]

\[
\beta = (1 + F) \times m
\]

Here \( F \), \( w_1 \), \( w_2 \) and \( w_3 \) are arbitrary constants. The effort is estimated in man months (MM).

### III. EXPERIMENT RESULT ANALYSIS

The data is taken from [2] and given in table 2. Let \( F = 0.3 \), then from Table 1, \( \alpha = 0.7m \), \( \beta = 1.3m \). We have taken \( W_1 = 1 \), \( W_2 = 50 \), \( W_3 = 1 \) for our Model. Table 3 below gives the experimental results.

<table>
<thead>
<tr>
<th>s. no</th>
<th>Project no</th>
<th>kloc</th>
<th>Actual effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>39</td>
<td>72</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>40.5</td>
<td>82.5</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>50</td>
<td>84</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>128.6</td>
<td>230.7</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>161.4</td>
<td>157</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>164.8</td>
<td>246.9</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>200</td>
<td>130.3</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>214.4</td>
<td>86.9</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>253.6</td>
<td>287</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>254.2</td>
<td>258.7</td>
</tr>
</tbody>
</table>

**TABLE 2**
IV. COMPARISON BETWEEN VARIOUS MODELS

The mean magnitude of relative error (MMRE) as the main performance measure. Using MMRE we evaluate the impact of estimation accuracy using evaluation criteria, for each model. Model with minimum MMRE value gives better estimation result. The mean magnitude of relative error (MMRE), defined as:

\[
MMRE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{E_{\text{actual}} - E_{\text{estimated}}}{E_{\text{actual}}} \right|
\]

<table>
<thead>
<tr>
<th>Project → Method</th>
<th>actual effort</th>
<th>cocomo basic</th>
<th>cocomo semidetached</th>
<th>cocomo embedded</th>
<th>barry bohm</th>
<th>Halsted model</th>
<th>walston felix</th>
<th>fuzzy model</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>72</td>
<td>112.41</td>
<td>181.59</td>
<td>292.13</td>
<td>149.88</td>
<td>1266.48</td>
<td>145.83</td>
<td>39.53</td>
</tr>
<tr>
<td>2</td>
<td>82.5</td>
<td>116.96</td>
<td>189.43</td>
<td>305.66</td>
<td>155.94</td>
<td>1340.25</td>
<td>150.93</td>
<td>26.5</td>
</tr>
<tr>
<td>6</td>
<td>84</td>
<td>145.92</td>
<td>239.86</td>
<td>393.61</td>
<td>194.56</td>
<td>1838.47</td>
<td>182.83</td>
<td>19.34</td>
</tr>
<tr>
<td>11</td>
<td>230.7</td>
<td>393.47</td>
<td>690.98</td>
<td>1222.9</td>
<td>524.62</td>
<td>7583.4</td>
<td>431.92</td>
<td>5.93</td>
</tr>
<tr>
<td>12</td>
<td>157</td>
<td>499.47</td>
<td>891.19</td>
<td>1606.15</td>
<td>665.96</td>
<td>10662.4</td>
<td>531.12</td>
<td>0.98</td>
</tr>
<tr>
<td>13</td>
<td>246.9</td>
<td>510.52</td>
<td>912.24</td>
<td>1646.83</td>
<td>680.69</td>
<td>11001.1</td>
<td>541.29</td>
<td>0.714</td>
</tr>
<tr>
<td>7</td>
<td>130.3</td>
<td>625.59</td>
<td>1133.11</td>
<td>2077.48</td>
<td>834.12</td>
<td>14707.2</td>
<td>645.56</td>
<td>0.419</td>
</tr>
<tr>
<td>4</td>
<td>86.9</td>
<td>672.97</td>
<td>1224.87</td>
<td>2258.25</td>
<td>897.29</td>
<td>16324.5</td>
<td>687.72</td>
<td>0.245</td>
</tr>
<tr>
<td>1</td>
<td>287</td>
<td>802.72</td>
<td>1478.31</td>
<td>2762.36</td>
<td>1070.3</td>
<td>21000.38</td>
<td>801.27</td>
<td>0.11</td>
</tr>
<tr>
<td>10</td>
<td>258.7</td>
<td>804.72</td>
<td>1482.23</td>
<td>2770.2</td>
<td>1072.9</td>
<td>21074.9</td>
<td>802.995</td>
<td>0.015</td>
</tr>
</tbody>
</table>

TABLE 3

TABLE 4
An Empirical Validation of Software Cost Estimation Model Using Fuzzy Technique

<table>
<thead>
<tr>
<th>Model</th>
<th>Barry-Boehm Model</th>
<th>Halstead Model</th>
<th>Waslton Felix Model</th>
<th>Fuzzy Model</th>
<th>Cocomo Basic (Organic)</th>
<th>Cocomo Basic (Semidetached)</th>
<th>Cocomo Basic (Embedded)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMRE</td>
<td>0.2818</td>
<td>6.428</td>
<td>0.2008</td>
<td>0.09426</td>
<td>0.1863</td>
<td>0.4149</td>
<td>0.8373</td>
</tr>
</tbody>
</table>

The above table 4 gives the comparison among various models and it is observed that fuzzy based model gives the best result as it has minimum MMRE value

COMPARISON OF EFFORT BETWEEN VARIOUS MODELS

The following figure 2 shows the comparison of effort calculated among various models.

V. CONCLUSION

In this paper we proposed a model that performs better than other models in achieving the accuracy of effort estimation. We proposed fuzzy software cost estimation model that handles ambiguity, obscurity and then compared with other popular software cost estimation models. From the experiments evaluation we concluded that: proposed fuzzy logic model showed better software effort estimate in view of the MMRE evaluation criteria as compared to the traditional estimation models. Furthermore, the fuzzy logic model presents better estimation accuracy as compared to the other models. The utilization of fuzzy logic for other applications in the software engineering field can also be explored in the future. The future work may also be done by using hybrid model neuro fuzzy model that combine fuzzy logic with neural network for more accurate estimation result.
REFERENCES


