Software Effort Estimation For Size Metric Framework Modelling Using Loc And Fp Oriented Estimation Models And Fuzzy Logic Approach

Shama Kousar Jabeen.A¹, Mrs.B.Arthi², Mrs.V.Vani³

PG Student [SOFTWARE ENGINEERING], Dept. of IT, Easwari Engineering College, Chennai, Tamilnadu, India¹
Assistant professor, Dept. of IT, Easwari Engineering College, Chennai, Tamilnadu, India²
Assistant professor  Senior Grade, Dept. of IT, Easwari Engineering College, Chennai, Tamilnadu, India³

ABSTRACT— As software grew in size and requirements it also successively grew in complexity and cost. Evaluating size estimates accurately at an initial stage in the software conglomeration is of high priority. Conventional techniques have the problems of uncertainty and precision during the evaluation of size estimates. Software engineering cost models and estimation techniques are used for a number of purposes. In our work we have compared the results using three function point based effort estimation models. We have also compared MMRE, MMER, MRE, MER values by training the dataset using fuzzy logic approach which overcomes the problems of traditional methods. In this paper effort estimate has been obtained by modeling the size metric framework.

KEYWORDS— size metric, fuzzy logic software effort ,software engineering, cost estimation models, MMRE, MER,MRE,MMER.

I.INTRODUCTION

Delivering the software on time and within budget is a critical concern for many organizations Cost estimations refers to the prediction in terms of time, staff, and effort. In many papers cost estimation is referred to as the effort prediction and hence it is used interchangeably. Effort prediction is usually made at an early stage of software development. Difficulty prevails in estimation because the estimates are very often uncertain and little knowledge is known. The size proxy metric framework considers the mean and variance of effort which is not considered in traditional metrics. In effort prediction laziness and ignorance must be taken in to account as the software designer does not consider all the factors during the prediction process. We consider uncertainty in our work which is considerable across specific project size metrics. The work has been implemented with a framework where the size metric is considered for better effort prediction. The entire project in this paper has been implemented in two major parts. The first part consists of extraction of function points from requirements document. In the second part the extracted function points are used to calculate effort estimates by training the dataset. This training of the dataset has been done using fuzzy logic approach.

This paper is organized as follows. Section 2 discusses about the related works. Architecture implementation and the concepts carried out in our work has been discussed in Section 3. Section 4 shows the experimental results and the implementation of the system. Section 5 draws conclusions and the future work.

II.RELATED WORKS

Boehm’s[1] describes that software engineering practices in the industry determines the cost and the quality of the software product .Thus a large and increasingly costly item also makes a large and increasing impact on human welfare.

The work done by Moataz A.Ahamed, Irfan Ahmad and Jarallah S.Alghamdi[9] explains the size proxy framework which considers normal distribution in combination with the regression model .They have considered LOC as inputs for early software effort prediction. There are basically two areas of research dealt in effort prediction: (1) developing prediction techniques and models, (2) developing size metrics that could be used as a proxy for effort estimation.
B. Boehm, C. Abts, S. Chulani [2] have illustrated in their work about the first the area of research. In this paper, different methodologies of effort prediction like expert judgment, analogy based prediction, algorithmic models, non algorithmic approaches have been detailed. Expert judgment is a time consuming approach. Prediction using analogy requires a predetermined effort estimate which is then used to compare similar or analogous projects. Algorithmic models are many which involves COCOMO, SLIM, SEER-SEM models. Non algorithmic modeling methods are based on machine learning and soft computing techniques. Some of them are Bayesian belief networks, fuzzy logic approach, artificial neural networks, and evolutionary computation. There are also researches done using soft computing approaches in combinations, such as neuro-fuzzy, neuro-genetic approaches.

Moataz A.Ahamed and Zeeshan Muzaffar[10] have suggested that traditional approaches for software projects effort prediction such as the use of the mathematical formulae are derived from the historical data, or the use of expert judgment are plagued with issues pertaining to effectiveness and robustness in their results. Thus type-II fuzzy logic systems must be allowed to handle imprecision and uncertainty. They have considered COCOMO model in their work.

The second area of research in effort prediction is to develop early and better size metrics for a good software effort prediction. Some of them which involves usecases, class diagrams, source code. The most frequently and commonly used size metrics in the early phases of software development lifecycle are LOC (lines of code) and FP (Function points), project features and use cases.

The function point metric was proposed by Albrecht as a method for measuring software size and productivity. Function point metric sizes software from the end user perspective by measuring the functionality delivered to the end user. A function point is defined as one end user business function. It employs functional and logical entities such as inputs, outputs, files and inquiries that are believed to relate more closely to the functions performed by the software as compared to other measures such as lines of code.

The counting of function points is based on IFPUG[3] as suggested in their standards. There are many other derivatives of function point metric proposed in the literature trying to address and overcome issues related to using function point metric measures size. Mark II function point metric is one of the widely used in the industry.

Justin Wong, Danny Ho, Luiz Fernando, Capretz[4] have suggested that the neuro fuzzy function point backfiring (NFFPB) makes use of neural network which is used for tuning the fuzzy logic membership functions where backfiring approach was used. In this method function points are converted into SLOC estimates and programming languages where grouped based on the fuzzy levels.

K.K.Shula[6] has discussed in his work on the substantial improvement in the prediction accuracy by the neuro genetic approach as compared to both a regression tree based conventional approach. They have made use of COCOMO data set comprising of 63 projects and kemerer data set comprising 15 projects which were merged randomly into a single database of 18 projects.

III. PROPOSED METHODOLOGY

The work in this paper has been done for size estimation in order to predict the overall effort of a software project. When a project's effort estimate is obtained, it is equivalent to the size estimate made by the user. Here, sizing represents the project planner's first major challenge to be accomplished.

The framework which has been cited in the work of M.A. Ahamed, Irfan Ahmad, and Jarallah S. Alghamdi[9] is used as a proxy to train the size estimates. We have considered this paper as a major source for our work. Since there work has the problem of uncertainty, we overcome this issue with the help of soft computing approach like fuzzy logic which is highly suitable to train the size proxy framework.

Moreover it is proposed in there work that training of the dataset will also lead to better and accurate estimates. Henceforth, we have carried out our work using fuzzy logic toolbox and function points which results in precision and accuracy of results in the developed framework to illuminate suitable size proxy metric for effort prediction.

Roger S. Pressman[11] has suggested the overall structure of software models as follows:

\[ E = A + B \cdot (ev)^c \]  

(1)
In equation (1), A, B, C are empirical constants. E is effort in person months and ev is the estimation variable (either LOC or FP). In addition to the above equation he has also suggested the effort equations (2), (3), (4) to estimate effort from function points using various FP oriented models as given below:

\[
E = -91.4 + 0.355 \text{FP} \\
E = -37 + 0.96 \text{FP} \\
E = -12.88 + 0.405 \text{FP}
\]

Here, equation (2) is the Albrecht and Gaffney model. Equation (3) is the Kemerer model. Equation (4) is the small project regression model. A quick examination of these models indicates that each will yield a different result for the same values of LOC or FP. The implications is clear that the estimation models must be calibrated for local needs of our estimation.

Roger S. Pressman [11] has also suggested about the LOC (Lines Of Code) oriented estimation models proposed in the literature as follows:

\[
E = 5.2 \times (K\text{LOC})^{0.91} \\
E = 5.5 + 0.73 \times (K\text{LOC})^{1.16} \\
E = 3.2 \times (K\text{LOC})^{1.05} \\
E = 5.288 \times (K\text{LOC})^{1.047}
\]

Here, equation (5) represents the Walston Felix model. Equation (6) represents the Bailey–Basili model. Equation (7) represents the Boehm–simple model and equation (8) represents the Doty model for KLOC greater than 9.

L. Putnam and W. Myers [7] have suggested four methods to overcome the sizing problem as follows:

1) Fuzzy Logic Sizing approach: In this approach, the project manager must identify the application nature and type to subsequently establish its magnitude within the original range.

2) Function point Sizing approach: The requirements engineer develops estimates of the information domain features.

3) Component Sizing approach: The software application is composed of various types of components that are unique to a particular application area.

4) Change sizing approach: This method is used when a project circumscribes the usage of software that must be modified as a subdivision of a software project.

The system architecture implemented in our work is as shown below in Figure 1. It is split up into two larger modules. The first module deals with generation of size estimates from the requirements document. The second module deals with training of the dataset using the fuzzy logic editor in our fuzzy logic toolbox in MATLAB R2009a. The second module has four submodules which are fuzzifier, inference, defuzzifier and rules modules. In function point estimation process, decomposition work is canonical. The subsequent estimation process does not account only for functionality, but rather it focuses on the domain characteristics and complexity issues as well. The resultant estimates can then be used to obtain a FP value that can be correlated to past data and used to achieve a general estimate. The work has involved calculation of function points by extracting use cases from the requirements document. This work has been implemented using Visual Use case design tool [16] where the use cases are imported from the requirements document. The use case diagrams are rated as simple, average or complex. This obtains the unadjusted function point count. Then the value adjustment factors (VAF) are also calculated using function point calculator. On calculation of VAF and UFCs they are displayed to the client for acceptance or rejection of the project. This display to the client is done using Extensible Mark up Language (XML) tags. This function point (FP) dataset is then trained using the fuzzy logic toolbox in MATLAB and it’s corresponding MMRE, MMER, MRE, MER and other parameters are calculated in order to identify the accuracy of the size estimates. Stephen H. Kan [12] refers to the calculation of value adjustment factors (VAF) which forms the second part of function point computation. This is done using 14 general system characteristics (GSC’S). These are rated in a scale of 0 to 5 in order to assess their impact. The 14 GSC’s are as follows:
1) Data Communication functions.
2) Distributed functions in system.
3) Performance of the system.
4) Heavily used configuration of the system.
5) Transaction Rate of the system.
6) Online data entry of the system.
7) End user efficiency of the system.
8) Online Update in system.
9) Complex Processing of the system.
10) Reusability of the system.
11) Installation Ease of the system.
12) Operational Ease of the system.
13) Multiple sites usage.
14) Facilitation of Change of the system.

IFPUG[3] has cited Table1 as shown below, which represents the complexity factors and ratings that helps the customer in the estimation of different unadjusted function points. The unadjusted function point depends upon the complexity judgment of the software application in terms of five components which are:

1) External Input.
2) External Output.
3) Logical internal file.
4) External Interface file.
5) External inquiry.

Stephen H.Kan[12] has cited equation(5) to calculate VAF as shown below. This has been used in this paper to rate the complexities of the usecase diagrams used in our study dataset.
Stephen H. Kan[12] has referred to equation (6) to calculate function points as shown below. This equation has been implemented in our work to estimate function points for the study dataset.

\[ FP = FC \times VAF \]  

Equation (6) has been used in this work which is a simple derivation used in the calculation of function points. A project manager or a software analyst must understand the full documented methods such as International Function Point User’s Group (IFPUG)[3] standard for a complete and accurate implementation. Thus the size estimated, dataset is then applied to Mamdani systems which is a typical type-2 fuzzy logic system as detailed in the work of Moutaz A. Ahamed and Zeeshan Muzaffar.

M.Wasif Nasar, Yong-Ji Wang and Manzoor Elahi[8] in their work have expressed about fuzzy logic as a mathematical tool for dealing with uncertainty and imprecision. It is a theory of unsharp boundaries and is used to solve problems that are too complex to be understood qualitatively. The concept of fuzzy sets must be viewed as a generalization of the concept of a classical crisp set.

Sandeep Kad and Vinay Chopra[13] have described about the type-2 fuzzy systems in their work with COCOMO approach. Here fuzzification and defuzzification of cost drivers of COCOMO is dealt in their work. Type-2 Mamdani fuzzy logic systems consist of four components:

1) Fuzzifier: It converts the crisp inputs in to a fuzzy set. Membership functions used to graphically describe a situation.
2) Fuzzy Rule Base: It uses if-then rules.
3) Fuzzy Inference Engine: A collection of if-then rules stored in fuzzy rule base is known as inference engine. It performs two operations i.e. aggregation and composition.
4) Defuzzification: It is the process that refers to the translation of fuzzy output in to crisp output.
IV. EXPERIMENTAL RESULTS AND DISCUSSION

In our work, we have proposed fuzzy based software effort estimation model rules which contain linguistic variables related to the project. The rule base for fuzzy inference system (FIS) makes use of OR, AND logical operation with unadjusted function point, value adjustment factor and function points as input variables to form a large number of rules in our work. We have implemented our work with 3 inputs and 1 output resulting in a combination of 27 rules. The rules are implemented individually for both FP and LOC oriented estimation models. After estimating the effort values other parameters such as MMRE, MMER, MRE and MER are found out and they are being compared in order to conclude the best model suiting our local dataset. The fuzzy logic rules for OR operation are as follows:

Pseudo code:

Rule 1: If (UFC is low) or (VAF is low) or (FP is low) then (EFFORT is low) (1).
Rule 2: If (UFC is low) or (VAF is low) or (FP is medium) then (EFFORT is medium) (1).
Rule 3: If (UFC is low) or (VAF is low) or (FP is high) then (EFFORT is high) (1).
Rule 4: If (UFC is low) or (VAF is high) or (FP is low) then (EFFORT is high) (1).
Rule 5: If (UFC is low) or (VAF is high) or (FP is medium) then (EFFORT is high) (1).
Rule 6: If (UFC is low) or (VAF is high) or (FP is high) then (EFFORT is high) (1).
Rule 7: If (UFC is low) or (VAF is medium) or (FP is low) then (EFFORT is medium) (1).
Rule 8: If (UFC is low) or (VAF is medium) or (FP is medium) then (EFFORT is medium) (1).
Rule 9: If (UFC is medium) or (VAF is low) or (FP is low) then (EFFORT is medium) (1).
Rule 10: If (UFC is medium) or (VAF is low) or (FP is medium) then (EFFORT is medium) (1).
Rule 11: If (UFC is medium) or (VAF is low) or (FP is high) then (EFFORT is high) (1).
Rule 12: If (UFC is medium) or (VAF is medium) or (FP is low) then (EFFORT is medium) (1).
Rule 13: If (UFC is medium) or (VAF is medium) or (FP is medium) then (EFFORT is medium) (1).
Rule 14: If (UFC is medium) or (VAF is medium) or (FP is high) then (EFFORT is high) (1).
Rule 15: If (UFC is medium) or (VAF is high) or (FP is low) then (EFFORT is high) (1).
Rule 16: If (UFC is medium) or (VAF is high) or (FP is medium) then (EFFORT is high) (1).
Rule 17: If (UFC is medium) or (VAF is high) or (FP is high) then (EFFORT is high) (1).
Rule 18: If (UFC is high) or (VAF is low) or (FP is low) then (EFFORT is high) (1).
Rule 19: If (UFC is high) or (VAF is low) or (FP is medium) then (EFFORT is high) (1).
Rule 20: If (UFC is high) or (VAF is low) or (FP is high) then (EFFORT is high) (1).
Rule 21: If (UFC is high) or (VAF is medium) or (FP is low) then (EFFORT is high) (1).
Rule 22: If (UFC is high) or (VAF is medium) or (FP is medium) then (EFFORT is high) (1).
Rule 23: If (UFC is high) or (VAF is medium) or (FP is high) then (EFFORT is high) (1).
Rule 24: If (UFC is high) or (VAF is high) or (FP is low) then (EFFORT is high) (1).
Rule 25: If (UFC is high) or (VAF is high) or (FP is medium) then (EFFORT is high) (1).
Rule 26: If (UFC is high) or (VAF is high) or (FP is high) then (EFFORT is high) (1).
Rule 27: If (UFC is high) or (VAF is high) or (FP is high) then (EFFORT is high) (1).

Similarly the fuzzy logic rules for AND operation are as follows:

Pseudo code:

Rule 1: If (UFC is low) and (VAF is low) and (FP is low) then (EFFORT is low) (1).
Rule 2: If (UFC is low) and (VAF is low) and (FP is medium) then (EFFORT is low) (1).
Rule 3: If (UFC is low) and (VAF is low) and (FP is high) then (EFFORT is low) (1).
Rule 4: If (UFC is low) and (VAF is high) and (FP is low) then (EFFORT is low) (1).
Rule 5: If (UFC is low) and (VAF is high) and (FP is medium) then (EFFORT is low) (1).
Rule 6: If (UFC is low) and (VAF is high) and (FP is high) then (EFFORT is low) (1).
Rule 7: If (UFC is low) and (VAF is medium) and (FP is low) then (EFFORT is low) (1).
Rule 8: If (UFC is low) and (VAF is medium) and (FP is medium) then (EFFORT is low) (1).
Rule 9: If (UFC is low) and (VAF is medium) and (FP is high) then (EFFORT is low) (1).
Rule 10: If (UFC is medium) and (VAF is low) and (FP is low) then (EFFORT is low) (1).
Rule 11: If (UFC is medium) and (VAF is low) and (FP is medium) then (EFFORT is low) (1).
Rule 12: If (UFC is medium) and (VAF is low) and (FP is high) then (EFFORT is low) (1).
Rule 13: If (UFC is medium) and (VAF is medium) and (FP is low) then (EFFORT is low) (1).
Rule 14: If (UFC is medium) and (VAF is medium) and (FP is medium) then (EFFORT is low) (1).
Rule 15: If (UFC is medium) and (VAF is medium) and (FP is high) then (EFFORT is low) (1).
Rule 16: If (UFC is medium) and (VAF is high) and (FP is low) then (EFFORT is low) (1).
Rule 17: If (UFC is medium) and (VAF is high) and (FP is medium) then (EFFORT is low) (1).
Rule 18: If (UFC is medium) and (VAF is high) and (FP is high) then (EFFORT is low) (1).
Rule 19: If (UFC is high) and (VAF is low) and (FP is low) then (EFFORT is low) (1).
Rule 20: If (UFC is high) and (VAF is low) and (FP is medium) then (EFFORT is low) (1).
Rule 21: If (UFC is high) and (VAF is low) and (FP is high) then (EFFORT is low) (1).
Rule 22: If (UFC is high) and (VAF is medium) and (FP is low) then (EFFORT is low) (1).
Rule 23: If (UFC is high) and (VAF is medium) and (FP is medium) then (EFFORT is low) (1).
Rule 24: If (UFC is high) and (VAF is medium) and (FP is high) then (EFFORT is low) (1).
Rule 25: If (UFC is high) and (VAF is high) and (FP is low) then (EFFORT is low) (1).
Rule 26: If (UFC is high) and (VAF is high) and (FP is medium) then (EFFORT is low) (1).
Rule 27: If (UFC is high) and (VAF is high) and (FP is high) then (EFFORT is low) (1).
Copyright to IJAREEIE www.ijareeie.com 463

Rule 11: If (UFC is medium) and (VAF is low) and (FP is medium) then (EFFORT is low) (1).
Rule 12: If (UFC is medium) and (VAF is low) and (FP is high) then (EFFORT is low) (1).
Rule 13: If (UFC is medium) and (VAF is medium) and (FP is low) then (EFFORT is low) (1).
Rule 14: If (UFC is medium) and (VAF is medium) and (FP is high) then (EFFORT is medium) (1).
Rule 15: If (UFC is medium) and (VAF is medium) and (FP is high) then (EFFORT is medium) (1).
Rule 16: If (UFC is medium) and (VAF is high) and (FP is low) then (EFFORT is low) (1).
Rule 17: If (UFC is medium) and (VAF is high) and (FP is high) then (EFFORT is medium) (1).
Rule 18: If (UFC is medium) and (VAF is high) and (FP is high) then (EFFORT is medium) (1).
Rule 19: If (UFC is high) and (VAF is low) and (FP is low) then (EFFORT is low) (1).
Rule 20: If (UFC is high) and (VAF is low) and (FP is medium) then (EFFORT is low) (1).
Rule 21: If (UFC is high) and (VAF is low) and (FP is high) then (EFFORT is low) (1).
Rule 22: If (UFC is high) and (VAF is medium) and (FP is low) then (EFFORT is low) (1).
Rule 23: If (UFC is high) and (VAF is medium) and (FP is medium) then (EFFORT is low) (1).
Rule 24: If (UFC is high) and (VAF is medium) and (FP is high) then (EFFORT is medium) (1).
Rule 25: If (UFC is high) and (VAF is high) and (FP is low) then (EFFORT is low) (1).
Rule 26: If (UFC is high) and (VAF is high) and (FP is medium) then (EFFORT is medium) (1).
Rule 27: If (UFC is high) and (VAF is high) and (FP is high) then (EFFORT is medium) (1).

The Effort estimated for Albrecht FP oriented model AND, OR operation are 9.6 and 8.63 Man/Hours. The Effort Estimated for Kemerer FP oriented model using AND, OR operation are 371 and 406Man/Hours. The effort estimated for Small regression FP oriented model using AND,OR operation are 159 Man/Hours respectively. The figure2,figure3,figure4,figure5, figure6 shows the rule viewer implementations done in Fuzzy logic toolbox in MATLAB R2009a. The models have been simulated and the respective MRE,MER,MMRE,MMER values have been calculated as shown in Table 2.

Figure 2: Rule Viewer For Albrecht AND Fuzzy System
Figure 3: Rule Viewer For Albrecht OR Fuzzy System

Figure 5: Rule Viewer For Kemerer OR Fuzzy System
Shama Kousar Jabeen.A and Mrs. B. Arthi[15] have shown in their work for effort estimation using function points with training the dataset using fuzzy logic approach. Shama Kousar Jabeen.A and Mrs. B. Arthi[14] have shown in their another work for size estimation with neuro fuzzy logic approach for Albrect dataset.

V. CONCLUSION

The MMRE and MMER values for Bailey and Kemerer Models are found to be effective. In our work further training of the dataset must be done using other soft computing techniques like neural networks, Support Vector Machines, genetic algorithms and other soft computing techniques as such. Also other parameters must be analyzed for effective data analysis in the future work. Apart from this work which has been implemented using triangular membership function other membership functions and other logical operations can be implemented for the same dataset to analyze it’s effect on the estimation parameters.

Table 2: Comparison of FP and LOC Oriented Estimation Models for Various Parameters

<table>
<thead>
<tr>
<th>S.NO</th>
<th>ESTIMATION ORIENTED MODELS</th>
<th>MRE</th>
<th>MER</th>
<th>MMRE</th>
<th>MMER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ALBRECHT MODEL-OR OPERATION</td>
<td>0.1598</td>
<td>0.1378</td>
<td>0.0145</td>
<td>0.0125</td>
</tr>
<tr>
<td>2</td>
<td>ALBRECHT AND OPERATION</td>
<td>0.0450</td>
<td>0.0431</td>
<td>0.0041</td>
<td>0.0039</td>
</tr>
<tr>
<td>3</td>
<td>KEMERER OR OPERATION</td>
<td>0.0614</td>
<td>0.0655</td>
<td>0.0056</td>
<td>0.0060</td>
</tr>
<tr>
<td>4</td>
<td>KEMERER AND OPERATION</td>
<td>0.0271</td>
<td>0.0264</td>
<td>0.0025</td>
<td>0.0024</td>
</tr>
<tr>
<td>S.NO</td>
<td>ESTIMATION ORIENTED MODELS</td>
<td>MRE</td>
<td>MER</td>
<td>MMRE</td>
<td>MMER</td>
</tr>
<tr>
<td>------</td>
<td>---------------------------</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>14</td>
<td>WALSTON FELIX MODEL-OR OPERATION</td>
<td>0.0215</td>
<td>0.0210</td>
<td>0.0020</td>
<td>0.0019</td>
</tr>
</tbody>
</table>

### Acknowledgement

I sincerely thank all those who have offered their support and valuable guidelines in accomplishing this work. I also thank my family for their constant support in all my endeavours.
REFERENCES