

A FUZZY APPROACH FOR SOFTWARE EFFORT ESTIMATION

Geetika Batra and Mahima Trivedi

Department of Computer Science and Engineering, L.N.C.T., Indore, India.

batra.geetika@yahoo.com

ABSTRACT

The most significant activity in software project management is Software development effort prediction. Ubiquitous availability of COCOMO model revealed many possibilities with a perspective of optimization of cost. Cost drivers have significant influence on the COCOMO and this research investigates the role of cost drivers in improving the precision of effort estimation Fuzzy logic has been applied to the COCOMO using membership functions to represent the cost drivers. Using Trapezoidal Membership Function (TMF), a few attributes are assigned the maximum degree of compatibility when they should be assigned lower degrees. To overcome the above limitation, in this paper, it is proposed to use Gaussian Membership Function (GMF) for the cost drivers by studying the behavior of COCOMO cost drivers. It has been found that Gaussian function is performing better than the trapezoidal function, as it demonstrates a smoother transition in its intervals, and the achieved results were closer to the actual effort.

KEYWORDS

COCOMO, Fuzzy based effort estimation, Gaussian membership function, Software cost estimation, Software effort estimation and Project management.

1. INTRODUCTION

Cost estimation is one of the most crucial tasks in software project management. Sometimes, there is huge deviation between the actual cost and an estimated cost and hence accurate cost estimate is highly desired. The accuracy in the effort estimation is very essential because both underestimation and overestimation are harmful to the companies. If the estimate is too low then there will be pressure to complete the project as soon as possible. On other side if the estimate is too high, then more number of resources will be required. It implies that effort estimation is complex problem and it is very essential to investigate appropriate method for improvement of estimation accuracy. The cost estimation is classified as: Algorithmic and Non Algorithmic Cost methods. The most popular techniques used for the software cost estimation is algorithmic cost models [2] such as COCOMO, Function Point Analysis, Putnam's SLIM etc. The algorithmic methods differ in two aspects: selection of cost factors and the form of functions. This can also be classified on the basis of empirical and analytical. From the above models the Cost Constructive Model (COCOMO) [1] is widely used for estimating the cost which is developed by the Barry Boehm in 1981, which uses a basic regression formula with parameters that are derived from historical project data and current project characteristics. Sometimes due the lack of information regarding the project this model fails to work in an appropriate work, to overcome this disadvantage of COCOMO, fuzzy approach is used. A fuzzy logic based cost estimation models are more appropriate as it is suitable when the information is vague, uncertain, and imprecise. The proposal to extend the COCOMO by including the concept of fuzziness into the measurement of cost drivers; fuzzy set rather than classical set to represent the linguistic values are more advantageous as it is much more close to the real world. Though many membership functions like trapezoidal, triangular, Gaussian etc were used to represent the cost drivers as many of them are not appropriate to clear the vagueness. Some of the membership functions like triangular, trapezoidal are used for replacing the conventional method of using the fixed numeric values by

the fuzzy interval [8]. As the transition from one interval to another is not fine so as to attain the smoother transition in the membership functions the use of Gaussian membership function can be attempted to represent the cost drivers in the COCOMO. And it is found that the Gaussian is performing better than trapezoidal as the transition is smoother and the achieved result is closer to the actual cost. The introduction summarized that estimation is the critical activity to be performed in the software project management and to solve the problems regarding estimation novel method is to be used. It is suggested to be the fuzzy approach using Gaussian membership function as it is better and appropriate option to solve the ongoing challenge of effort estimation. While the next chapter i.e literature survey gives the general idea of the methods used.

2. LITERATURE REVIEW

2.1. Software Effort Estimation

The software effort estimation is one of the critical tasks in the software project management and is not possible every time to expect accurate estimations in the development of software. As due to inconsistency and uncertainty the estimations should not be systematic and appropriate so that some improvement made in the effort estimation should be a great achievement and it is valuable to the software companies. Some of the factors that depend upon the estimation should be forecasted by the project planners they are: software price determination, resource allocation, schedule arrangement and process monitoring. Some of the techniques that are included in the estimation are expert, analogy based, algorithmic method, fuzzy logic [6] etc.

2.2. COCOMO model

A number of algorithmic models [2] have been studied for estimating the effort, cost and schedule of the projects. This model uses a mathematical formula to predict the cost, project size, number of engineers and many other process and product factors. The survey resemble that most widely used model for the cost estimation is COCOMO Model [1], an algorithmic software cost estimation model that uses the regression formula with parameters that are derived from historical project data and current project characteristics. These data were analyzed such that to discover formula which is the best. The formula links the size of the system and product, project and team factors to the effort which helps in developing the system. In this model effort is described in Person Months (PM), which determines the effort required that is based on the project size and other cost factors like scale factors and effort multipliers [10]. It contains 17 effort multipliers and 5 scale factors. The formula is shown in (1).

$$Effort = A \times [Size]^{1.01 + \sum_{i=1}^5 SF_i} \times \prod_{i=1}^{17} EM_i \quad (1)$$

Whenever algorithmic cost models are used the cost drivers should be measured in order to drive the effort estimate. The cost drivers at the initial stage are uncertain so it is difficult to calculate the accurate cost. As the vagueness and uncertainty of cost drivers should be avoided by using the fuzzy approach which is advantageous and enhancing the model capabilities.

2.3. Fuzzy logic

Fuzzy logic is successful methodology to solve the realistic problems which include the simplicity and flexibility. It handles the problems with imprecise and incomplete data which is based on the fuzzy set theory which was formalized by Prof. Lofti Zadeh in 1965 [4]. It is used to represent linguistic values like low, old and complex. The fuzzy set is represented as where x is an element in X and $\mu_A(x)$ is a membership function of set A which defines the membership of fuzzy set A in the universe of discourse, X.

2.4. Fuzzy membership functions

The notion of membership in fuzzy sets thus becomes a matter of degree, which is number between 0 and 1. A membership degree of 0 represents complete non membership, while a membership degree of 1 represents a complete membership. It is characterized by a mapping from its universe of discourse into interval [0,1]. This mapping is known as membership function which is denoted by μ . The membership function may be trapezoidal, triangular [6], Gaussian etc. The trapezoidal function is described as a function of a vector, x , and depends on four scalar parameters a , b , c , and d , as given by

$$f(x: a, b, c, d) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & b \leq x < c \\ \frac{d-x}{d-c} & c \leq x < d \\ 0 & x \geq d \end{cases} \quad (2)$$

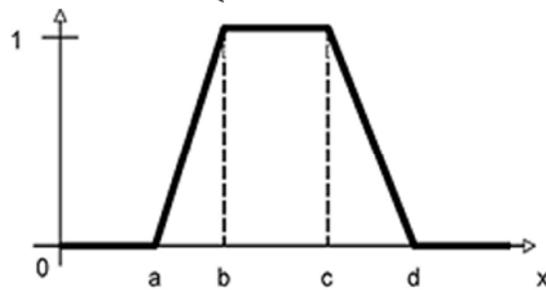


Figure 1. Trapezoidal membership function

And Gaussian membership function can be defined as function depends on two parameters σ and c as given by:

$$\text{Gaussian}(x; c, \sigma) = \exp\left(-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2\right) \quad (3)$$

The figure 3 shows the gaussian membership function which helps us to understand better about the results which are to be driven.

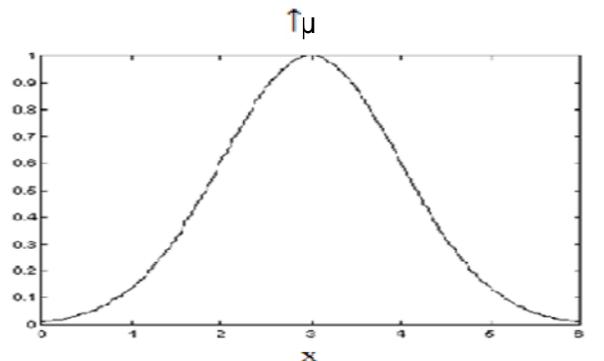


Figure 2. Gaussian Membership function

The literature survey shows the better technologies used to solve the complexity of cost estimation. The studied also shows, fuzzy sets can be represented by other membership functions

also, here Gaussian membership is used for the better result. The different datasets like NASA 93, Iris can used; currently COCOMO dataset are used and different costing parameters like complexity, coupling, development time etc can be used; here estimates is calculated on the basis of person-month.

3. RESEARCH METHODOLOGY

3.1. Problem Formulation

The cost estimation is the most essential activity and the widely used model is COCOMO, yields uncertainty at the output to improve the accuracy fuzzy based cost model [7]. The use of cost drivers produced the significant effect to improve the accuracy as the incorporating fuzziness into the measurement of cost drivers which give the positive effect on the result. As it is characterized by the fix numeric values, the interval values are used which is represented by the membership functions. The participation of the mathematical formulation of fuzzy values of COCOMO attributes, by using the trapezoidal function as some attributes are assigned the maximum degree instead of assigning the lower degree and to overcome this problem Gaussian membership function is used for the cost drivers. Cost drivers are often expressed through an unclear category which needs subjective assessment. The effort multipliers and scale factors of the COCOMO were described in natural language as very low, low, nominal, high, very high and extra high and these were represented by fixed numerical values. More conventionally, the problem of software cost estimation using COCOMO relies on a single (numeric) value of cost driver of a given software project to predict the effort. But it is not an appropriate way to fix numerical number to each of these scales. It is of principal importance to recognize this situation and come up with a technology using which we can evaluate the associated imprecision residing within the final results of cost estimation. The technology endorsed here deals with fuzzy sets. Using fuzzy sets, cost drivers of a software project can be specified by distribution of its possible values. Commonly, this form of distribution is represented in the form of a fuzzy set. By changing the cost drivers using fuzzy set, we can model the effort that impacts the estimation accuracy. Instead of using fixed numbers to characterize the cost drivers, interval values were used and these were represented using various membership functions triangular, trapezoidal etc. However still there was some linearity by using these functions. Overlapped symmetrical triangles or trapezoids reduce fuzzy systems to precise linear systems.

Furthermore there is a possibility when using a trapezoidal function that some attributes are assigned the maximum degree of compatibility when they should be assigned lower degrees. In order to avoid this linearity it is proposed to use more continuous Gaussian function to represent the cost drivers.

3.1. Research method

By studying the behavior of COCOMO cost drivers, in this investigation it is proposed to characterize the use GMF for cost drivers to represent the linguistic values. GMF gives more continuous transition from one interval to another. Gaussian Bell curve sets give richer fuzzy system with simple learning laws that tune the bell curve variance. The Gaussian Function is represented by:

$$\mu_{A_i}(x) = \text{Gaussian}(x, c_i, \sigma_i) = e^{-\frac{(x-c_i)^2}{2\sigma_i^2}}$$

(4)

Where c_i is the center of the i th fuzzy set and σ_i is the width of the i th fuzzy set.

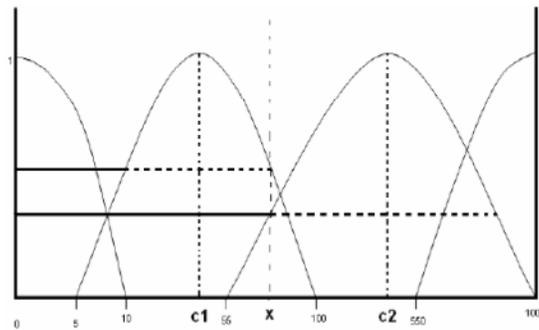


Figure 3.Representation of DATA Cost driver using Gaussian Membership Function

For example, in the case of DATA cost driver, we define a fuzzy set for each linguistic value with a Gaussian shaped membership function μ is shown in “Fig 3.” We have defined the fuzzy sets corresponding to the various associated linguistic values for each cost driver. In the research, a new fuzzy effort estimation model is proposed by using Gaussian function to deal with linguistic data, and to generate fuzzy membership functions and rules for cost drivers obtained from “(5)”. In the next step, we evaluate the COCOMO model using the “(1),” and cost drivers obtained from fuzzy sets (F_EMij) rather than from the classical EMij. F_EMij is calculated from “(6),” the classical EMij and the membership functions μ defined for the various fuzzy sets associated with the cost drivers.

$$F_EM_{ij} = F \left[\mu_{A_1}^{V_i(p)} \dots \mu_{A_i}^{V_i(p)}, EM_{i1} \dots EM_{ij} \right] \quad (5)$$

For ease, F is taken as a linear function, where the $\mu_{A_j}^{V_i}$ is the membership function of the fuzzy set A_j associated with the cost driver V_i is shown in “(5)”.

$$F_EM_{ij} = \sum_{j=1}^{k_i} \mu_{A_i}^{V_i(p)} \times EM_{ij} \quad (6)$$

4. DESIGN METHODOLOGY

The proposed cost estimation model was implemented using fuzzy logic tool box of MATLAB software. The fuzzy inference system (FIS) is used in order to implement the various processing steps. Options were provided for creating and editing FIS with fuzzy logic tool box software using graphical tools or command line functions.

This GUI tool allows us to edit the higher level features such as number of input and output variables of the FIS. Using FIS editor, membership functions can be added for each cost driver using ‘addmf’ command. Each cost driver in fuzzy COCOMO can be defined with membership function. The membership function editor ‘mfedit’ that allows us to inspect and modify all the membership functions. For each membership function we can change the name, type and parameters. All the cost drivers are defined and customized to the Gaussian membership function using the command ‘gaussmf’ (x, [sig c])[9]. The software development effort is recorded in terms of unit of person-month. The evaluation consists in comparing the accuracy of the estimated effort with the actual effort. There are many evaluation criteria for software effort estimation introduced in the literature, among them we applied the most frequent evaluation

criteria such as: Magnitude of Relative Error (MRE), Mean Magnitude of Relative Error (MMRE)[6] which is defined in (7)

$$MRE = \frac{|ActualEffort - EstimatedEffort|}{ActualEffort} \times 100$$

(7)

The Gaussian membership function that has been proposed in this work gives accurate effort than by using any other membership functions. When it uses trapezoidal function the peak value is linear but in Gaussian function it touches the peak at only one point. Hence, Gaussian function is better than trapezoidal function, as it demonstrates a smoother transition between its intervals. The results clearly indicate that such fuzzy set modeling approach affects significantly the estimation outcomes. This methodology or existing approach helps to solve the critical problem of software project management. This helps to reduce the huge deviation between estimated and actual cost. The proposed approach is the use of iris dataset on the different costing parameters like development time, complexity, person month etc. this will be apply in the formula s\and statistical computation is performed. On the basis of the above approach the results and conclusion is drawn The above methodology helps to complete the work with the help of the facilities required.

5. CONCLUSIONS

A critical issue for project managers is the accurate and reliable estimates of the required software development effort, especially at the initial stages of the software development life cycle. Software effort drivers usually have properties of uncertainty and vagueness when they are measured by human judgment. Cost drivers in algorithmic software cost estimation are often expressed through linguistic assessments and they usually represent concepts for which a single, precise measurement scale is not available. This desire the use of fuzzy techniques to estimate with accuracy. As fuzzy logic modeling techniques is used to handle imprecise data. It is projected an improved approach to estimate the software project effort by the use of fuzzy sets rather than classical intervals in the COCOMO model. These fuzzy sets are represented by Gaussian-shaped membership functions, yields a better estimate which is very nearer to the actual effort.

FUTURE WORK

This work can be extended by integrating with neural networks to take the advantage of its features, such as learning ability and good interpretability. Thus, a promising line of future work is to extend to the neuro-fuzzy approach that allows the integration of numerical data and expert knowledge.

REFERENCES

- [1] Boehmand B.W., NJ(1981),Software Engineering Economics, Prentice-Hall, Englewood Cliffs.
- [2] RamilJ.F, Algorithmic cost estimation for software evolution,701-703 <http://www.mendeley.com/research/algorithmic-cost-estimation-for-software-evolution/page-1>
- [3] Yen Langari, fuzzy logic, intelligence, Control and information,Pearson edition
- [4] Zadeh. L. A.,1965,Fuzzy Sets, Information and Control, Volume 8, pp. 338-353.

- [5] Ali Idri and Alain Abran, La Logique Floue Appliquee Aus Modeles d'Estimation d'Effort de Development deLogicielsCas Du Modele COCOMO'81 <http://www.gelog.etsmtl.ca/publications/pdf/560.pdf>.
- [6] Prasad Reddy P.G.V.D, Su.,et.al.,2011,Application of fuzzy logic approach to software effort estimation Vol. 2 NO 5.
- [7] Fei. Z, X. and Liu., "f-COCOMO - Fuzzy constructive cost model in software engineering, Proceedings of IEEE International Conference on Fuzzy System, pp. 331-337
- [8] Leonard J.,et.al., *Estimation of f-COCOMO model parameters using optimization techniques*, [http://sunset.usc.edu/events/2006/CIIForum/pages/presentations/2006SEWorld Jowers-Buckley-Reilly-c.pdf](http://sunset.usc.edu/events/2006/CIIForum/pages/presentations/2006SEWorld%20Jowers-Buckley-Reilly-c.pdf).
- [9] <http://www.mathworks.in/help/toolbox/fuzzy/trapmf.html>
Ch. Satyananda Reddy and KSVSN Raju ,JULY 2009,An ImprovedFuzzyApproach for COCOMO's Effort Estimation using Gaussian Membership Function JOURNAL OF SOFTWARE, VOL. 4, NO. 5.

Authors

Geetika Batra, Department of Computer Science and Engineering, L.N.C.T.,Indore, India.
(e-mail id:- batra.geetika@yahoo.com)