

Improving the Accuracy of Effort Estimation through Fuzzy Set Representation of Size

Ch. Satyananda Reddy and KSVSN Raju

Department of Computer Science and Systems Engineering, College of Engineering,
Andhra University, Visakhapatnam, India

Abstract: Problem statement: The precision and reliability of the effort estimation is very important for the competitiveness of software companies. The uncertainty at the input level of the Constructive Cost Model (COCOMO) yields uncertainty at the output, which leads to gross estimation error in the effort estimation. Fuzzy logic-based cost estimation models are more appropriate when vague and imprecise information was to be accounted for and was used in this research to improve the effort estimation accuracy. This study proposed to extend the COCOMO by incorporating the concept of fuzziness into the measurements of size. The main objective of this research was to investigate the role of size in improving the effort estimation accuracy by characterizing the size of the project using trapezoidal function which gave superior transition from one interval to another. **Approach:** The methodology adopted in this study was use of fuzzy sets rather than classical intervals in the COCOMO. Using fuzzy sets, size of a software project can be specified by distribution of its possible values and these fuzzy sets were represented by membership functions. Though, Triangular membership functions (TAMF) was used in the literature to represent the size, but it was not appropriate to clear the vagueness in the project size. Therefore, to get a smoother transition in the membership function, the size of the project, its associated linguistic values were represented by trapezoidal shaped MF and rules. **Results:** After analyzing the results attained by means of applying COCOMO, triangular and trapezoidal MF models to the COCOMO dataset, it had been found that proposed model was performing better than ordinal COCOMO and trapezoidal function was performing better than triangular function, as it demonstrated a smoother transition in its intervals and the achieved results were closer to the actual effort. The relative error for COCOMO using trapezoidal function is lower than that of the error obtained using TAMF. **Conclusion:** From the experimental results, it was concluded that, by fuzzifying the project size using TPMF, the accuracy of effort estimation can be improved and the estimated effort can be very close to the actual effort.

Key words: Constructive cost model, fuzzy based effort estimation, trapezoidal membership function, software cost estimation, software effort estimation and project management

INTRODUCTION

Software development involves a number of interrelated factors which affect development effort and productivity. Since many of these relationships are not well understood, accurate estimation of software development time and effort is a difficult problem. The precision and reliability of the effort estimation is very important for software industry because both overestimates and underestimates of the software effort are harmful to software companies. Nevertheless, accurate estimation of software development effort in reality has major implications for the management of software development. If a manager's estimate is too

low, then the software development team will be under considerable pressure to finish the product quickly. On the other hand, if a manager's estimate is too high, then too many resources will be committed to the project. In reality, estimating software development effort remains a complex problem attracting considerable research attention. It is very important to investigate novel methods for improving the accuracy of such estimates. As a result, many models for estimating software development effort have been proposed and are in use.

Fuzzy logic-based cost estimation models are more appropriate when vague and imprecise information is to be accounted for. This study proposed to extend the Constructive Cost Model (COCOMO)^[4] by incorporating

the concept of fuzziness into the measurements of size. The size of the project in COCOMO is represented by fixed numerical values. In fuzzy logic based cost estimation models, this size is represented with fuzzy interval values. The advantages of this over quantization are that they are more natural and they mimic the way in which humans interpret linguistic values.

Though, many membership functions were used in the literature^[10] to represent the size, but it is not appropriate to clear the vagueness in the project size. The TAMF was being used in COCOMO to replace the conventional quantization by using fuzzy interval values. So, the transition from one interval to an adjacent interval is abrupt rather than gradual. Therefore, to get a smoother transition in the Membership Function (MF), this study attempts to achieve a fuzzy based effort by using Trapezoidal Membership Function (TAMF). Hence, in this study it has been proposed and validated empirically, that the size of the software project can be specified by distribution of its possible values and the uses of TAMF to represent the size in the COCOMO. It has been found that TAMF is performing better than the triangular function, as it demonstrates a smoother transition in its intervals and the achieved results were closer to the actual effort.

Related work: Papers were reviewed regarding aspects related to research on software development effort estimation based on a fuzzy logic model. The fuzzy logic model uses the fuzzy logic concepts introduced by Zadeh^[13]. Study showed that fuzzy logic model has a place in software effort estimation. Attempts have been made to fuzzify some of the existing models in order to handle uncertainties and imprecision problems. Using real project data, Gray and MacDonell^[8] compared Function Point Analysis, Regression techniques, feed forward neural network and fuzzy logic in software effort estimation. Their results showed that fuzzy logic model achieved good performance, being outperformed in terms of accuracy only by neural network model with considerably more input variables. In their fuzzy logic model, triangular membership functions were defined for the small, medium, large intervals of size.

Fuzzy logic had also been applied to algorithmic models to cater for the need of fuzziness in the input. The first realization of the fuzziness of several aspects of COCOMO was that of Fei and Liu^[7]. The researchers observed that an accurate estimate of delivered source instruction (KDSI) cannot be made before starting a project and it is unreasonable to assign a determinate number for it. Ryder^[11] researched on the application of fuzzy logic to COCOMO and Function

Points models. Musflele *et al.*^[10] worked on fuzzifying basic COCOMO model without considering the adjustment factor. On the other hand, Idri *et al.*^[2] proposed fuzzy intermediate COCOMO with the fuzzification of cost drivers. The effort multiplier for each cost driver is obtained from fuzzy set, enabling its gradual transition from one interval to a contiguous interval. Validation results showed that the fuzzy intermediate COCOMO can tolerate imprecision in its input (cost drivers) and generate more gradual outputs.

Ahmed and Saliu^[1] geared up further by fuzzifying the two different portions of the COCOMO model i.e., nominal effort estimation and the adjustment factor. They proposed a fuzzy logic framework for effort prediction by integrating the fuzzified nominal effort and the fuzzified effort multipliers of the intermediate COCOMO model. So far, the mainstream of the work is concentrated on fuzzifying cost drivers with the representation of triangular membership functions. Hence, in this study, it is proposed to use fuzzy set interval values using TPMF for the size of the project in the effort estimation of Constructive Cost Model.

MATERIALS AND METHODS

Problem-formulation: In COCOMO effort is expressed as Person Months (PM). It determines the efforts required for a project based on software project's size in Kilo Source Line of Code (KSLOC) as well as other cost drivers known as scale factors and effort multipliers as shown in Eq. 1:

$$PM = A \cdot (\text{Size})^{1.01 + \sum_{i=1}^5 SF_i} \cdot \prod_{i=1}^{17} EM_i \quad (1)$$

where, A is a multiplicative constant and the set of Scale Factors (SF) and Effort Multipliers (EM) are defined the model^[5]. It contains 17 effort multipliers and 5 scale factors. The standard numeric values of the cost drivers are given in Table 1.

Traditionally, the problem of software cost estimation relies on a single (numeric) value of size of given software project to predict the effort. However, the size of the project is, based on some previously completed projects that resemble the current one (especially at the beginning of the project). Obviously, correctness and precision of such estimates are limited. 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,

Table 1: COCOMO cost drivers

| Cost drivers | Range | Description |
|--------------|-----------|---|
| RELY | 0.82-1.26 | Required software reliability |
| DATA | 0.90-1.28 | Database size |
| CPLX | 0.73-1.74 | Product complexity |
| RUSE | 0.95-1.24 | Developed for reusability |
| DOCU | 0.81-1.23 | Documentation match to life-cycle needs |
| TIME | 1.00-1.63 | Execution time constraint |
| STOR | 1.00-1.46 | Main storage constraint |
| PVOL | 0.87-1.30 | Platform volatility |
| ACAP | 1.42-0.71 | Analyst capability |
| PCAP | 1.34-0.76 | Programmer capability |
| PCON | 1.29-0.81 | Personnel continuity |
| APEX | 1.22-0.81 | Applications experience |
| PLEX | 1.19-0.85 | Platform experience |
| LTEX | 1.20-0.84 | Language and tool experience |
| TOOL | 1.17-0.78 | Use of software tools |
| SITE | 1.22-0.80 | Multi site development |
| SCED | 1.43-1.00 | Required development schedule |

size 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.

It is important to stress that uncertainty at the input level of the COCOMO model yields uncertainty at the output^[10]. This becomes obvious and, more importantly, bears a substantial significance in any practical endeavor. By changing the size using fuzzy set, we can model the effort that impacts the estimation accuracy. Obviously, a certain monotonicity property holds, which is less precise estimates of size gives rise to less detailed effort estimates. Overlapped symmetrical triangles reduce fuzzy systems to precise linear systems^[3]. Furthermore there is a possibility when using a triangular 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 superior function i.e., trapezoidal membership function for representing size of the project.

Proposed research method: In this investigation it is projected to characterize the size of the project using TPMF which gives superior transition from one interval to another. For example, a small software project can be described by a fuzzy set K in the form shown in Fig. 1. The grades of membership capture a notion of partial membership of an element to the concept (fuzzy set). In general, a fuzzy set K is described by its membership function $K(x)$ which expresses the degree of membership of x to the fuzzy set K describing a certain concept (say, small project, high reliability).

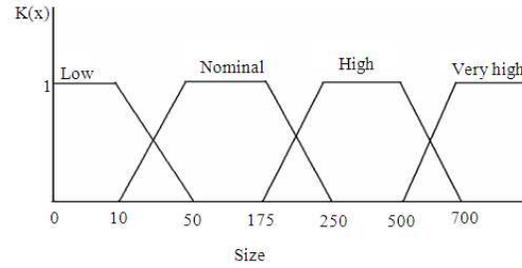


Fig. 1: Representation of size using a trapezoidal membership function

TPMF gives more continuous transition from one interval to another^[9]. A typical representation of project size using TPMF is shown in Fig. 1 and its function is represented by Eq. 2:

$$\mu_A(x) = \text{Trapezoidal}(x, a, b, c, d) \quad (2)$$

In this research, a new fuzzy effort estimation model is proposed by using trapezoidal function to deal with the size and to generate fuzzy MF and rules. In the next step, we evaluate the COCOMO model using Eq. 1 and size obtained from fuzzy set (F_Size) rather than from the classical size. F size is calculated from Eq. 4, the classical size and the membership functions μ defined for the size:

$$F_Size = F(\mu_A(P), \text{Size}) \quad (3)$$

For ease, F is taken as a linear function, where the μ_A is the MF of the fuzzy set A is shown in Eq. 3:

$$F_Size = \mu_A(P) \times \text{Size} \quad (4)$$

Experimental design: 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 size using ‘addmf’ command. The MF editor ‘mfedit’ that allows us to inspect and modify all the membership functions. For each MF we can change the name, type and parameters. The size of the project is defined and customized to the TPMF using the command ‘trapmf’ (x, [a b c d]). In designing the above model, we have used COCOMO dataset. The assignment of linguistic

values to the size uses conventional quantification where the values are intervals. For example, in the case of the size attribute, we have defined a fuzzy set for each linguistic value with a trapezoidal shaped MF shown in Fig. 1. We note that the fuzzy set associated with the size satisfies the normal condition. The evaluation consists in comparing the accuracy of the estimated effort with the actual effort. A common criterion for the evaluation of cost estimation models is the Magnitude of Relative Error (MRE)^[6], which is defined in the following Eq. 5:

$$MRE = \frac{|ActualEffort - Predicted Effort|}{Actual Effort} \times 100 \quad (5)$$

The TPMF that has been proposed in this work gives accurate effort than by using TAMF. When it uses triangular function the peak value is linear but in trapezoidal function it touches the peak at only one point. Hence, trapezoidal function is performing better than triangular 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.

RESULTS AND DISCUSSION

Experiments were done by taking original data from COCOMO dataset^[12]. The software development efforts obtained when using COCOMO and other membership functions were observed. After analyzing the results attained by means of applying COCOMO, triangular and trapezoidal MF models, it is observed that the effort estimation of the proposed model is giving more precise results than the other models. The effort estimated by means of fuzzifying size using TPMF is yielding better estimate which is very nearer to the actual effort. Therefore, using fuzzy sets, size of a software project can be specified by distribution of its possible values, by means of which we can evaluate the associated imprecision residing within the final results of cost estimation.

Table 2 shows the sample results obtained for some of the data sets taken from COCOMO dataset, which includes the effort estimated using Constructive Cost Model and the effort obtained using TAMF for the size and the effort achieved using TPMF for size i.e., the proposed fuzzified model. It has been found that proposed model is performing better than ordinal COCOMO and trapezoidal function is performing better than triangular function, as it demonstrates a smoother transition in its intervals and the achieved results were closer to the actual effort.

Table 2: Results and comparison of effort estimation in person months

| Project ID | Actual effort | Effort in Person Months (PM) | | |
|------------|---------------|------------------------------|---------------------|----------------------|
| | | COCOMO | Using triangular MF | Using trapezoidal MF |
| 1 | 61 | 45.63 | 52.53 | 55.39 |
| 2 | 237 | 214.10 | 249.16 | 234.52 |
| 3 | 599 | 539.60 | 575.44 | 580.80 |
| 4 | 603 | 553.43 | 578.12 | 589.36 |
| 5 | 702 | 1335.10 | 1253.90 | 1146.20 |
| 6 | 523 | 278.86 | 314.04 | 365.57 |
| 7 | 1075 | 661.30 | 739.63 | 806.34 |
| 8 | 2455 | 1945.40 | 2016.90 | 2096.30 |
| 9 | 958 | 408.33 | 476.60 | 563.65 |
| 10 | 1063 | 1275.90 | 1209.60 | 1164.40 |

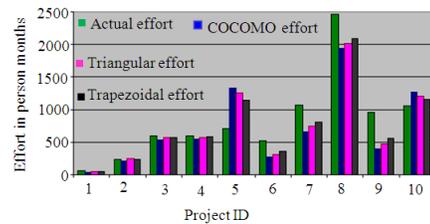


Fig. 2: Chart representing the comparisons of effort estimation

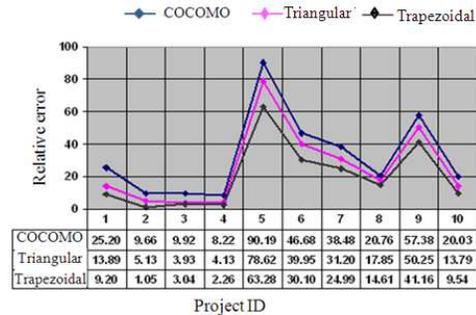


Fig. 3: Assessments of magnitude of relative errors

Figure 2 shows the bar chart representing comparative analysis of actual effort with that of the effort estimated using COCOMO, triangular and trapezoidal membership functions. Effort in person months is scaled along with y-axis. Actual effort, COCOMO effort and effort obtained using TAMF for size and effort obtained using TPMF for size, were represented for each sample projects, which were taken along with x-axis.

The magnitude of relative errors was calculated using (5). For example, the relative error calculated for project 1 for COCOMO, triangular and for the proposed model is 25.20, 13.89 and 9.20 respectively. In the case of second project it is 9.66, 5.13 and 1.05. The Mean Magnitude of Relative Error (MMRE) is 32.65, 25.87 and 19.92 respectively. Figure 3 shows the chart

representing relative errors which are represented along with y-axis against each project, which is taken along with x-axis. This clearly shows that there is a decrement in the relative error, so that the proposed model is more suitable for effort estimation.

CONCLUSION

In this study it has been proposed and examined the use of fuzzy sets rather than classical intervals in the COCOMO. Using fuzzy sets, size of a software project can be specified by distribution of its possible values and these fuzzy sets were represented by membership functions. For the size of the project, its associated linguistic values are represented by trapezoidal shaped MF. The relative error for COCOMO using trapezoidal function is lower than that of the error obtained using TAMF.

From the experimental results, it is concluded that, by fuzzifying the size of the project using TPMF, it can be proved that the resulting estimate impacts the effort. The effort generated using the proposed model gives better result than that of using ordinal COCOMO. This illustrates that by fuzzifying size using TPMF, the accuracy of effort estimation can be improved and the estimated effort can be very close to the actual effort. Moreover, by capturing the uncertainty of the initial data (estimates), one can monitor the behavior (quality) of the cost estimates over the course of the software project. This facet adds up a new conceptual dimension to the models of software cost estimation by raising awareness of the decision making with regard to the quality of the initial data needed by the model. This study can be extended by integrating with neural networks. By using this extended approach with the standard COCOMO models, we can take advantage of the features of neural network, such as learning ability and good interpretability. Therefore, a promising line of future work is to extend to the neuro-fuzzy approach.

REFERENCES

1. Ahmed, M.A., M.O. Saliu and J. Alghamdi, 2005. Adaptive fuzzy logic-based framework for software development effort prediction. Elsevier Sci., Inform. Software Technol., 47: 31-48. <https://eprints.kfupm.edu.sa/8688/>
2. Ali Idri and Alain Abran, 2000. COCOMO cost model using fuzzy logic. Proceedings of the 7th International Conference on Fuzzy Theory and Technology, Feb. 27-Mar 03, Atlantic City, New Jersey, pp: 1-4. <http://www.lrgl.uqam.ca/publications/pdf/465.pdf>
3. Idri, A. and A. Abran, 2004. La logique floue appliquee aus modeles d'Estimation d'Effort de developement de logiciels cas du modele COCOMO'81. <http://www.gelog.etsmtl.ca/publications/pdf/560.pdf>
4. Boehm, B.W., 1981. Software Engineering Economics. Prentice-Hall, Englewood Cliffs, New Jersey, ISBN: 10: 0138221227, pp: 767.
5. Boehm, B.W. *et al.*, 1995. Cost models for future software life cycle processes: COCOMO2.0. Ann. Software Eng., 1: 57-94. DOI: 10.1007/BF02249046
6. Briand, L.C., K.E. Emam, D. Surmann and I. Wiczorek, 1999. An assessment and comparison of common software cost estimation modeling techniques. Proceedings of the International Conference on Software Engineering, May 22-22, IEEE Xplore Press, Los Angeles, CA., USA., pp: 313-323. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=841022
7. Fei, Z. and X. Liu, 1992. F-COCOMO, fuzzy constructive cost model in software engineering. Proceedings of IEEE International Conference on Fuzzy System, Mar. 8-12, IEEE Xplore Press, San Diego, CA., USA., pp: 331-337. DOI: 10.1109/FUZZY.1992.258637
8. Gray, A.R. and S.G. MacDonell, 1997. Applications of fuzzy logic to software metric models for development effort estimation. Proceedings of the Fuzzy Information Processing Society Annual Meeting of the North American, Sept. 21-24, IEEE Xplore Press, Syracuse, USA., pp: 394-399. DOI: 10.1109/NAFIPS.1997.624073
9. Azzeh, M., D. Neagu and P. Cowling, 2008. Improving analogy software effort estimation using fuzzy feature subset selection algorithm. Proceedings of the 4th International Workshop on Predictor Models in Software Engineering, May 12-13, Leipzig, Germany, pp: 71-78. <http://portal.acm.org/citation.cfm?id=1370805>
10. Musflek, P., W. Pedrycz, G. Succi and M. Reformat, 2000. Software cost estimation with fuzzy models. Applied Comput. Rev., 8: 24-29. <http://portal.acm.org/citation.cfm?id=373975.373984>
11. Ryder, J., 1998. Fuzzy modeling of software effort prediction. Proceedings of IEEE Information Technology Conference, Sept. 1-3, IEEE Xplore Press, Syracuse, New York, USA., pp: 53-56. DOI: 10.1109/IT.1998.713380
12. Menzies, T., 2005. Promise software engineering repository. <http://promise.site.uottawa.ca/SERpository/>
13. Zadeh, L.A., 1965. Fuzzy sets. Inform. Control, 8: pp: 338-353. <http://www-bisc.cs.berkeley.edu/Zadeh-1965.pdf>