

# International Journal of Advance Research in Computer Science and Management Studies

Research Article / Paper / Case Study

Available online at: [www.ijarcsms.com](http://www.ijarcsms.com)

## *An Efficient Method for the Estimation of Effort in Software Cost*

**Dr. S. Malathi<sup>1</sup>**

Professor

Department of CSE

Panimalar Engineering College

Chennai – India

**B. S. Lijin<sup>2</sup>**

Final Year Student

Department of CSE

Panimalar Engineering College

Chennai – India

*Abstract: Several useful models have been developed by the software engineering community to elucidate the periodic growth of life cycle and calculate the effort of cost estimation in a precise manner. One of the commonly used machine learning technique is the analogy method that cannot handle the categorical variables efficiently. In general project attributes of cost estimation are often measured in terms of linguistic values. These imprecise values leads to analogous while explaining the process. The proposed method is an integrated approach of combining analogy with the fuzzy, based on reasoning by analogy for handling both numerical and categorical variables where the uncertainty and imprecision solution is also identified by the behaviour of linguistic values utilized in the software projects. The performance of this method validates the results based on historical NASA93 dataset. The outcome of the proposed method is analyzed and indicates that better results have achieved by comparing with the other machine learning methods.*

*Keywords: Cost estimation; categorical variables; fuzzy, analogy; linguistic; dataset.*

### I. INTRODUCTION

The software environment has evolved significantly in the last 30 years. To estimate software development effort, the use of the neural networks has been viewed with skepticism by majority of the cost estimation community. Even though neural networks have exposed their strengths in solving multifarious problems, their limitation of being 'black boxes' has limited their usage as a common practice for cost estimation [1]. Some models carry a few advantageous features of the neuro-fuzzy approach, such as learning capability and excellent interpretability, while maintaining the qualities of the COCOMO model [2].

Estimation by analogy is simple and flexible, compared to algorithmic models. Analogy technique is applied effectively even for local data which is not supported by algorithmic models [3], [4]. It can be used for both qualitative and quantitative data, reflecting closer types of datasets found in real life. Analogy based estimation has the potential to mitigate the effect of outliers in a historical data set, since estimation by analogy does not rely on calibrating a single model to suit all the projects. Unfortunately, it is difficult to assess the preliminary estimation as the available information about the historic project data during early stages is not sufficient [5]. The proposed method effectively estimates the software effort using an incorporated approach.

The research paper is organized as follows: Section 2 deals with some of the recent research works related to the proposed technique. Section 3 describes the proposed technique and Section 4 discusses about the experimentation and comparative results with necessary tables and graphs and Section 5 concludes the paper.

## II. RELATED WORK

The proposed technique elucidates the effective estimation of effort. Several researchers have carried out researches in the field of effort estimation for the software projects using various techniques [6]. A few of the significant researches have been highlighted here for iris recognition.

Fuzzy logic has been applied to the COCOMO using membership functions such as Symmetrical Triangles and Trapezoidal Membership Function (TMF) to signify the cost drivers. The limitation of the latter function is that a few attributes were assigned the maximum degree of compatibility instead of lower degree. To overcome this drawback, Ch. Satyananda Reddy et al. [7] proposed the usage of Gaussian Membership Function (GMF) for the cost drivers by studying the behaviour of COCOMO cost drivers. Ahmeda and Muzaffar [8] dealt with the imprecision and uncertainty in the inputs of effort prediction. M.Kazemifard et al. [9] uses a multi agent system for handling the characteristics of the team members in fuzzy system. There are many studies that utilized the fuzzy systems to deal with the ambiguous and linguistic inputs of software cost estimation [10].

Wei Lin Du et al. [11] proposed an approach combining the neuro-fuzzy technique and the SEER-SEM effort estimation algorithm. The continuous rating values and linguistic values are the inputs of the proposed model for avoiding the deviation in estimation among similar projects. The performance of the proposed integrated method has been optimized by designing and evaluated with the data published in the historical projects. The evaluated results indicate that the estimation with the proposed fuzzy model containing analogy reasoning produce better results in comparison with the existing techniques[12] that uses feature selection algorithm.

## III. PROPOSED WORK

### A. Analogy

Analogy based effort estimation method belongs to machine learning category. The basic idea of analogy prediction is “Projects that are similar with respect to project and product features such as size and complexity will be similar with respect to project effort”. The strong point of this method is that the estimate is based on actual project experience. However, it is not clear to what extent the previous project is actually representative of the limitation, situation and role to be performed by the new system

### B. Fuzzy Logic

Fuzzy logic is based on human behaviour and reasoning. It has an affinity with fuzzy set theory and applied in situations where decision making is difficult. A Fuzzy set can be defined as an extension of classical set theory by assigning a value for an individual in the universe between the two boundaries that is represented by a membership function.

$$A = \int \mu_A(x) / x \quad (1)$$

Where  $x$  is an element in  $X$  and  $\mu_A(x)$  is a membership function. A Fuzzy set is characterized by a membership function that has grades between the interval  $[0, 1]$  called grade membership function. There are different types of membership function, namely, triangular, trapezoidal, Gaussian etc.

### C. Fuzzy Analogy

Fuzzification of classical analogy procedure is Fuzzy analogy. It comprises the following procedures, viz., 1) Identification of cases, 2) Retrieval of similar cases and 3) Case adaptation. Each step is the fuzzification of its equivalent classical analogy procedure. The framework is shown in Fig.1

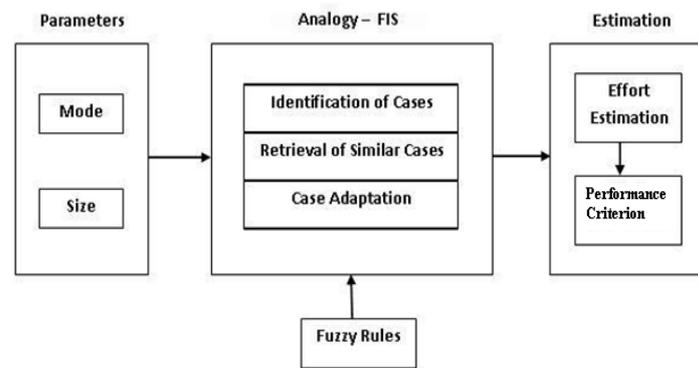


Fig.1 Framework of the proposed method

1. **Identification of Cases:** The goal of this step is the characterization of all software projects by a set of attributes. Selecting attributes, which will describe software projects, is a complex task in the analogy procedure. Consequently, the attributes must be relevant for the effort estimation task. The objective of the proposed Fuzzy Analogy approach is to deal with categorical data. So, in the identification step, each software project is described by a set of selected attributes which can be measured by numerical or categorical values. These values will be represented by fuzzy sets.

In the case of numerical value  $x_0$ , its fuzzification will be done by the membership function which takes the value of 1 when  $x$  is equal to  $x_0$  and 0 otherwise. For categorical values,  $M$  attributes are considered and for each attribute  $M_j$ , a measure with linguistic values is defined ( $A_k^j$ ). Each linguistic value  $A_k^j$  is represented by a fuzzy set with a membership function ( $\mu_{A_k^j}$ ).

It is preferable that these fuzzy sets satisfy the normal condition. The use of fuzzy sets to represent categorical data, such as 'very low' and 'low', is similar to how humans interpret these values and consequently it allows dealing with imprecision and uncertainty in the case identification step.

2. **Retrieval of Similar Cases:** This step is based on the choice of software project similarity measure. These measures assess the overall similarity of two projects  $P_1$  and  $P_2$ ,  $d(P_1, P_2)$  by combining all the individual similarities of  $P_1$  and  $P_2$  associated with the various linguistic variables  $V_j$  describing the project  $P_1$  and  $P_2$ ,  $d_{V_j}(P_1, P_2)$ . After an axiomatic validation of some proposed candidate measures for the individual distances  $d_{V_j}(P_1, P_2)$ , two measures have been retained [13].

$$d_{V_j}(P_1, P_2) = \begin{cases} \max_k \min(\mu_{A_k^j}(P_1), \mu_{A_k^j}(P_2)) & \text{max-min aggregation} \\ \sum_k \mu_{A_k^j}(P_1) \times \mu_{A_k^j}(P_2) & \text{sum-product aggregation} \end{cases} \quad (2)$$

Where  $A_k^j$  are the fuzzy sets associated with  $V_j$  and  $\mu_{A_k^j}$  are the membership functions representing fuzzy sets  $A_k^j$ .

The effort is estimated using the formula:

$$Effort = A * (SIZE)^{B+0.01 * \sum_{i=1}^N d_i} * \prod_{i=1}^N EM_i \quad (3)$$

Where  $A$  and  $B$  are constants,  $d$  is the distance and EM is effort multipliers. By using the above formula the effort is estimated. Rules are developed with cost driver in the antecedent part and corresponding effort multiplier in the consequent part.

The defuzzified value for each of the effort multiplier is obtained from individual Fuzzy Inference Systems after matching, inference aggregation and subsequent Defuzzification. Total Effort is obtained after multiplying them together. The high values for the cost drivers lead an effort estimate that is more than three times the initial estimate, whereas low values reduce the estimate to about one third of the original. This highlights the vast differences between different types of projects and the difficulties of transferring experience from one application domain to another

3. **Case Adaptation:** The objective of this step is to derive an estimate for the new project by using the know effort values of similar projects. We are not convinced in fixing the number of analogies in this step. In our proposed method, all the projects in a dataset are used to derive the new project estimate. There are two issues that have to be addressed, (i) the choice of how many similar projects should be used in the adaptation, and (ii) how to adapt the chosen analogies in order to generate an estimate for the new project. In the available literature, it can be clearly noticed that there is no definite rule to guide the choice of the number of analogies. Fixing the number of analogies for the case adaptation step is considered here neither as a requirement nor as a constraint.

#### IV. EXPERIMENTAL RESULTS

This section explains the accuracy of effort estimation by the proposed work as well as the performance against other methods.

The standard datasets are chosen from the available software engineering public domain as follows. In this method, NASA 93 [14] was selected consisting of 93 projects in various programming languages. The implementation is done using the default packages of JAVA net beans. The estimated values of the proposed method for NASA 93 dataset is compared with the fuzzy method and tabulated in Table 1.

TABLE 1  
Comparison of Proposed Effort with Existing and Actual Effort

Pjt .ID	Actual Effort	Estimated Effort	
		Fuzzy Method	Proposed Method
1	36	45.31	34.14
2	42	32.37	36.48
3	42	43.83	36.60
4	50	60.74	43.625
5	60	80.99	52.125
6	120	113.77	103.122
7	72	94.04	61.794
8	192	172.59	163.201
9	239	268.04	203.150
10	300	354.73	254.512
11	352.8	354.73	324.538
12	420	483.07	355.993

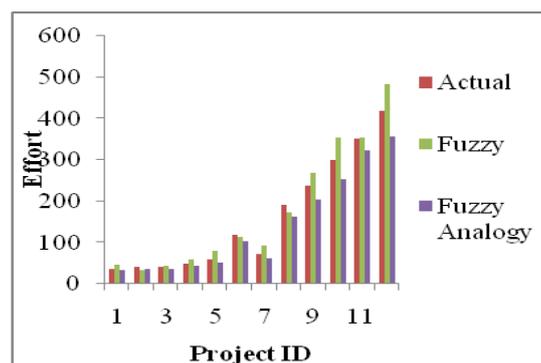


Fig.2 Comparison of proposed effort with Actual and Fuzzy method

In Fig.2, the performance of the actual effort in the proposed method is compared with the existing method [15] and found that the fuzzy analogy method is very efficient.

#### A. Performance Criterion

To assess the accuracy of the Fuzzy Analogy method, the common evaluation criteria is Prediction (PRED) [10] which represents the percentage of MRE that is less than or equal to the value  $p$  among all projects. The definition of PRED ( $p$ ) is given as follows:

$$PRED(p) = \frac{K}{N} \quad (4)$$

Where  $N$  is the total number of observations and  $k$  is the number of observations and a common value for  $p$  is 25. The overall comparative results of the PRED are tabulated for the NASA93 dataset in Table 2.

Table 2  
Comparative Results of Prediction Accuracy

Methods	Pred (0.25)
Proposed Method	0.86
Fuzzy Method	0.81

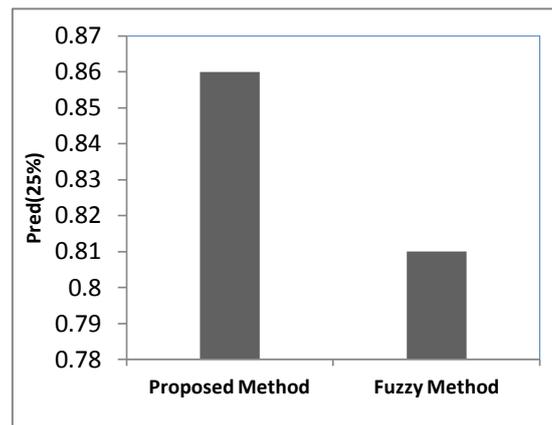


Fig.3 Comparison of Prediction for NASA93 dataset

From the comparative results given in Figure 3, it is clearly observed that the proposed Fuzzy Analogy outperforms the accuracy achieved with the existing Fuzzy method.

#### V. Conclusion

An improved incorporated method of Fuzzy Analogy approach has been proposed in this paper to estimate the effort for the historical dataset. This approach is based on reasoning by analogy, fuzzy logic and linguistic variables. This method can be applicable both for numerical and categorical datasets and can handle the imprecision and uncertainty in an effective manner. From the experimental results, it is observed that the fuzzy analogy outperforms for the software effort estimation and in future it can be optimized further by dealing many datasets in consideration.

#### References

1. Pichai Jodpimai, Peraphon Sophatsathit and Chidchanok Lursinsap, "Estimating Software Effort with Minimum Features using Neural Functional Approximation", ICCSA.2010.
2. M.A. Ahmeda, Z. Muzaffar, "Handling imprecision and uncertainty in software development effort prediction: a type-2 fuzzy logic based framework", Information and Software Technology 51 (2009) 640–654.
3. Ekrem Kocaguneli, Tim Menzies, Ayse Bener, Jacky W. Keung, "Exploiting the Essential Assumptions of Analogy-based Effort Estimation", Journal of IEEE Transactions on Software Engineering, Vol.34, No.4, pp. 471–484, July 2010.

4. J.Keung, "Empirical evaluation of analogy-x for software cost estimation", in ESEM '08: Proceedings of the second ACM-IEEE international symposium on Empirical engineering and measurement. New York, NY, USA: ACM, 2008, pp.294-296.
5. Hasan Al-Sakran, "Software Cost Estimation Model Based on Integration of Multi-agent and Case-Based Reasoning", Journal of Computer Science, Vol. 2, No. 3, Pp. 276-282, 2006.
6. M.Jorgensen and M.Shepperd, "A systematic review of software development cost estimation studies", IEEE Trans.Softw.Eng., vol.33, no.1, pp.33 -53, 2007.
7. Ch. Satyananda Reddy and KSVN Raju, "An Improved Fuzzy Approach for COCOMO's Effort Estimation using Gaussian Membership Function", Journal Of Software, Vol. 4, No. 5, Pp. 452-459, July 2009.
8. M.A. Ahmeda, Z. Muzaffar, "Handling imprecision and uncertainty in software development effort prediction: a type-2 fuzzy logic based framework", Information and Software Technology 51 (2009) 640-654.
9. M.Kazemifard,A.Zaeri,N.ghasem-ghaee, M.A.Nematbakhsh, F.Mardukhi, "Fuzzy Emotional COCOMO II Software Cost Estimation (FECSC) using Multi-Agent Systems", Applied Soft Computing, Elsevier, pg.2260-2270, 2011.
10. Iman Attarzadeh and Siew Hock Ow, "Improving the Accuracy of Software Cost Estimation Model Based on a new Fuzzy Logic Model", World applied sciences Journal, Vol. 8, No. 2, pp. 177-184, 2010.
11. Wei Lin Du, Danny Ho and Luiz Fernando Capretz, "Improving Software Effort Estimation Using Neuro-Fuzzy Model with SEER-SEM", Global Journal of Computer Science and Technology, Vol. 10, No. 12, Pp. 52-64, Oct 2010.
12. Pichai Jodpimai, Peraphon Sophatsathit and Chidchanok Lursinsap, "Estimating Software Effort with Minimum Features using Neural Functional Approximation", ICCSA.2010.
13. A.Idri and A. Abran, "Towards A Fuzzy Logic Based Measures For Software Project similarity", In Proc. of the 7th International Symposium on Software Metrics, England, pp.85-96, 2001.
14. Sayyad Shirabad, J. and Menzies, T.J. 2005. The PROMISE Repository of Software Engineering Databases. School of Information Technology and Engineering University of Ottawa, Canada. Available: <http://promise.site.uottawa.ca/SERepository>.
15. Prasad Reddy P.V.G.D, Sudha K.R,Rama Sree P, 2011. " Application of Fuzzy Logic Approach to Software Effort Estimation," International Journal of Advanced Computer Science and Applications,Vol. 2, Issue 5.

#### AUTHOR(S) PROFILE



**Dr.S.Malathi** graduated from Coimbatore Institute of Technology, Bharathiyar University. Then she obtained her Masters Degree in Computer Science and Engineering and Doctorate from Sathyabama University, Chennai, Tamilnadu in the field of Software Engineering. She has more than 18 years of teaching experience and currently working as Professor, Dept. of CSE, Panimalar Engineering College, Chennai. She has published/presented several research papers in peer-reviewed International/National Journals and Conferences, respectively. She has recently received the award from CSI for "Best Paper Presenter in International Conferences". Her future interests lay in the field of Software Engineering, Image Processing and Networks.



**B.S.Lijin** doing his final year UG degree (B.E-C.S.E) in Panimalar Engineering College affiliated to Anna University, Chennai. He has presented papers in various international and national Conferences. His areas of Interest are cloud computing, image processing and automobile.