



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY
Volume 11 Issue 22 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Fuzzy and Swarm Intelligence for Software Cost Estimation

By Srinivasa Rao.T, Prasad Reddy P.V.G.D, Hari Ch.V.M.K

GITAM University

Abstract - Software cost estimation is the process of predicting the amount of time, effort and resources required to complete the project successfully. Software development is a collection of activities includes feasibility study, analysis, design, coding, testing, implementation and maintenance. Each phase requires resources- people, time, software and hardware which should be predicted well before the software development. The prediction means lot of uncertainty. So far many models are proposed by using Fuzzy Logic, Neural Networks, Machine Learning, Regression analysis and Soft Computing techniques. In this paper we are proposed a new model structure basing on Alaa F. Sheta using Fuzzy logic for controlling prediction uncertainty and the parameters of the cost model tuned by using swarm intelligence-Particle Swarm Optimization. The proposed model results are verified with NASA software dataset and results are compared with the existing models. The Results show that the value of MARE (Mean Absolute Relative Error) applying fuzzy-swarm intelligence was substantially lower than MARE of other models exists in the literature.

Keywords : *Software Cost Estimation, Swarm Intelligence, Fuzzy Logic, COCOMO, Particle Swarm Optimization.*

GJCST Classification : *D.2.9*



FUZZY AND SWARM INTELLIGENCE FOR SOFTWARE COST ESTIMATION

Strictly as per the compliance and regulations of:



Fuzzy and Swarm Intelligence for Software Cost Estimation

Srinivasa Rao.T^α, Prasad Reddy P.V.G.D^α, Hari Ch.V.M.K^β

Abstract - Software cost estimation is the process of predicting the amount of time, effort and resources required to complete the project successfully. Software development is a collection of activities includes feasibility study, analysis, design, coding, testing, implementation and maintenance. Each phase requires resources- people, time, software and hardware which should be predicted well before the software development. The prediction means lot of uncertainty. So far many models are proposed by using Fuzzy Logic, Neural Networks, Machine Learning, Regression analysis and Soft Computing techniques. In this paper we are proposed a new model structure basing on Alaa F. Sheta using Fuzzy logic for controlling prediction uncertainty and the parameters of the cost model tuned by using swarm intelligence-Particle Swarm Optimization. The proposed model results are verified with NASA software dataset and results are compared with the existing models. The Results show that the value of MARE (Mean Absolute Relative Error) applying fuzzy-swarm intelligence was substantially lower than MARE of other models exists in the literature.

Keywords : Software Cost Estimation, Swarm Intelligence, Fuzzy Logic, COCOMO, Particle Swarm Optimization.

I. INTRODUCTION

Software project management is collection of two activities: Project Planning and Project Monitoring and control. Planning is predicting the activities that must be done before starting development work. Once project work is started it is the responsibility of project manager to monitor the work and see the goal-high quality of software must be produced with low cost and within a time and budget. The input for the planning is SRS-Software Requirement Specification Document and output is project plan mainly includes Cost estimation and Schedule estimation.

Software cost estimation is the process of predicting the amount of time required to build a software system. The time is measured in terms of Person-Months (PM's) which is later on converted into dollar cost. The basic input for the cost model is size measured in terms of KDLOC (Kilo Delivered Lines Of Code) and set of Cost parameters. The advantage of cost estimation is Cost benefit analysis, proper resource utilization (software, hardware and people), staffing plans, functionality trade-offs, risks and modify budget.

Author ^{αβ} : GITAM Institute Of Technology, GITAM University.
E-mail : tsr.etl@gmail.com,

Author ^α : Department Of CS&SE, Andhra University.
E-mail : kurmahari@gmail.com

The software cost estimation problem deserves a special attention because of development of product is unique under taking results in uncertainty, with increased size of software projects estimation mistakes could cost lot in terms of resources allocated to the project.

II. BACKGROUND

In this section we briefly discuss the COCOMO (Constructive Cost Model), Fuzzy Logic and Swarm Intelligence-Particle Swarm Intelligence.

a) COCOMO

[Boehm, 1981][4] described COCOMO as a collection of three variants, they are Basic model, Intermediate model, and Detailed model. Boehm described three development modes and Organic is for relatively simple projects, Semidetached is for relatively intermediate projects, Embedded for a project developed under tight constraints.

The Basic COCOMO Model computes effort E as function of program size, and it is same as single variable method. The Effort calculated using the following equation

$$\text{Effort} = a * (\text{size})^b \quad (1)$$

Where a and b are the set of values depending on the complexity of software (for organic projects a=2.4, b=1.05, for semi-detached a=3.0, b=1.1.2 and for embedded a=3.6, b=1.2).

An Intermediate COCOMO model effort is E is function of program size and set of cost drivers or effort multipliers. The Effort calculated using the following equation

$$\text{Effort} = a * (\text{size})^b * \text{EAF} \quad (2)$$

where a and b are the set of values depending on the complexity of software (for organic projects a=3.2, b=1.05, for semi-detached a=3.0, b=1.1.2 and for embedded a=2.8, b=1.2) and EAF (Effort Adjustment Factor) which is calculated using 15 cost drivers. Each cost driver is rated from ordinal scale ranging from low to high.

In Detailed COCOMO the effort E is function of program size and a set of cost drivers given according to each phase of software life cycle. The phases used in detailed COCOMO are requirements planning and product design, detailed design, code and unit test, and

integration testing. The weights defined accordingly. The Effort calculated using the following equation

$$\text{Effort} = a * (\text{size})^b * \text{EAF} * \sum(W_i) \quad (3)$$

Boehm and his colleagues have refined and updated COCOMO called as COCOMO II. It is a collection of three variants, Application composition model, early design model, and Post architecture model.

b) Fuzzy Logic

A fuzzy set is a set with a smooth boundary. Fuzzy set theory generalizes classical set theory to allow partial membership[5,6]. The best way to introduce fuzzy sets is to start with a limitation of classical sets. A set in classical set theory always has a sharp boundary because membership in a set is a black-and-white concept, i.e. an object either completely belongs to the set or does not belongs to the set at all. The degree of membership in a set is expressed by a number between 0 and 1; 0 means entirely not in the set, 1 means completely in the set, and a number in between means partially in the set. This way a smooth and gradual transition from the region outside the set to those in the set can be described. A fuzzy set is thus defined by a function that maps objects in a domain of concern to their membership value in the set. Such a function is called the Membership Function and usually denoted by the Greek symbol μ . The membership function of a fuzzy set A is denoted by μ_A , and the membership value of x in A is denoted by $\mu_A(x)$. The domain of membership function, which is the domain of concern from which elements of the set are drawn, is called the Universe Of Discourse. We may identify meaningful lower and upper bounds of the membership functions. Membership functions of this type are known as interval values fuzzy sets. The intervals of the membership functions are also fuzzy then it is known as interval Type-2 fuzzy sets.

c) Swarm Intelligence-Particle Swarm Optimization

Swarm Intelligence (SI) is an innovative distributed intelligent paradigm for solving optimization problems that originally took its inspiration from the biological examples by swarming, flocking and herding phenomena in vertebrates. Particle Swarm Optimization (PSO) incorporates swarming behaviors observed in flocks of birds, schools of fish, or swarms of bees, and even human social behavior, from which the idea is emerged. PSO is a population-based optimization tool, which could be implemented and applied easily to solve various function optimization problems, or the problems that can be transformed to function optimization problems. Particle Swarm Optimization was first introduced by Dr. Russell C. Eberhart and Dr. James Kennedy in 1995. As described by Eberhart and Kennedy, the PSO algorithm is an adaptive algorithm based on a social-psychological metaphor; a population of individuals (referred to as particles) adapts by

returning stochastically toward previously successful regions. The basic concept of PSO lies in accelerating each particle towards its Pbest and Gbest locations with a random weighted acceleration at each time. The modification mbns of the particles positions can be mathematically modeled according to the following equations:

$$V_i^{k+1} = w * V_i^k + c_1 * \text{rand}() * (V_{pbest} - S_i^k) + c_2 * \text{rand}() * (V_{gbest} - S_i^k) \quad (4)$$

$$S_i^{k+1} = S_i^k + V_i^{k+1} \quad (5)$$

Where, S_i^k is current search point, S_i^{k+1} is modified search point, V_i^k is the current velocity, V_i^{k+1} is the modified velocity, V_{pbest} is the velocity based on Pbest, V_{gbest} = velocity based on Gbest, w is the inertia weight, c_i is the weighting factors, $\text{rand}()$ are uniformly distributed random numbers between 0 and 1. In order to guide the particle effectively in the search space, the maximum moving distance during each iteration must be changed in between the maximum velocity $[-V_{max}, V_{max}]$.

III. LITERATURE REVIEW

In this section we discuss the some previous models proposed using Genetic Algorithms[8], Fuzzy models[9], Soft-Computing Techniques[10], Computational Intelligence Techniques[3], Heuristic Algorithms, Neural Networks[7], Radial Basis[11], and Regression[1,2,4].

The Cost of product is a function of many parameters which are Size (coding size), Cost Drivers and Methodology used in the project. The Walston Felix uses 36 cost drivers, 16 by Boheam and 30 other factors considered by the Bailey-Basili for the cost estimation. The parameters are estimated by using regression analysis and the effort equation is[1]

$$E = 5.5 + 0.73(KLOC)^{1.16} \quad (6)$$

Where E is effort and KLOC is kilo lines of code-coding size

The Alaa F. Sheta proposed two new model structures by using genetic algorithms for tuning parameters. The Model 1 and 2 equations is

$$\text{Effort} = 3.1938(DLOC)0.8209 - 0.1918(ME) \text{ for model 1} \quad (7)$$

$$\text{Effort} = 3.3602(DLOC)0.8116 - 0.4524(ME) + 17.8025 \text{ for model 2} \quad (8)$$

Where ME is the methodology used in the project.

The Harish proposed two model structures based on triangular fuzzy sets [13]. Interval Type-2 fuzzy logic, Particle Swarm Optimization for is proposed by Prasad Reddy. [12].

IV. PROPOSED METHODOLOGY AND ALGORITHM

a) Methodology

The uncertainty about cost estimation is usually quite high, because of prediction of basic element size, cost drivers and other parameters. By introducing some modifications in the interval type-2 fuzzy logic we can control the uncertainty. In the present work fuzzy sets are used for modeling uncertainty and imprecision in an efficient way. The inputs of the standard cost model include an estimation of project size and evaluation of the parameters, rather than a single number, the software size can be regarded as a fuzzy set yielding the cost estimate also in the form of a fuzzy set. We emphasize a way of propagation of uncertainty and ensuring violation of the resulting effort. Fuzzy sets create a more flexible, high versatile development environment. They generate a feedback also the resulting uncertainty of the results. The decision-maker is no longer left with a single variable estimate which could be highly misleading in many cases and lead to the belief as a to the relevance of the obtained results.

In the present work on the proposed models the parameter tuning is done by using Particle Swarm Optimization. For each particle position with values of tuning parameters, fitness function is evaluated with an objective to minimize the fitness function. The objective is to minimize or maximize fitness function. The particles moving towards optimal parameters by doing several iterations until particles exhaust or derivative of velocity becomes nearly zero then we get the optimal parameters which are later used for effort estimation.

b) Proposed Model

The Proposed model consists of three major components. First component is fuzzification process which identifies the suitable firing intervals for the input parameter size. Second component is parameter tuning using particle swarm optimization. Finally effort estimation done through weighted average defuzzification method using the results obtained in first and second steps.

i. Fuzzification process

The input size is fuzzified by using two triangular fuzzy sets. The Triangular member function is shown below.

$$\mu_p(\text{size}) = \begin{cases} 0 & \text{size} \leq L \\ \frac{\text{size} + L}{2L} & -L \leq \text{size} \leq L \\ 1 & \text{size} > L \end{cases} \quad (9)$$

Where L is the Mean of input sizes

After fuzzification of the data set find the shaded regions (overlap) of the left hand side and right hand side those are called meaningful lower and upper bounds of the data set- Foot Print of Uncertainty. The

means of the Foot Print of Uncertainty as firing intervals.

ii. Parameter tuning using Particle Swarm Optimization

The effort equation we considered is Alaa F. Sheta model-2

$$\text{Effort} = a * (\text{Size})^b + c * (\text{ME}) + d \quad (10)$$

The parameters a, b, and c of the equation 10 are tuned by using particle swarm optimization with inertia weight and MARE as the fitness function (minimize).

iii. Defuzzification

The defuzzification is done through weighted average method is as shown below

$$E = \{w_1 * [(a * \alpha^b) + c * (\text{ME}) + d] + w_2 * [(a * m^b) + c * (\text{ME}) + d] + w_3 * [(a * \beta^b) + c * (\text{ME}) + d]\} / w_1 + w_2 + w_3 \quad (11)$$

Where w_i is the weighting factor and α , m , and β are the fuzzified sizes obtained from triangular member function.

V. RESULTS AND DISCUSSIONS

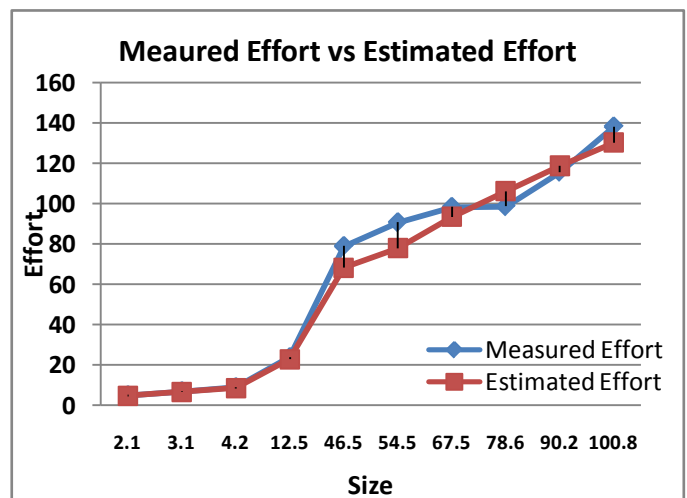
NASA dataset is considered for experimentation. The firing intervals obtained after the fuzzification are [0.7362, 0.8998]. The parameters obtained after tuning PSO methodology $a=3.131606$, $b=0.820175$, $c=0.045208$ and $d=-2.020790$. While performing defuzzification $w_1=1, w_2=10$ and $w_3=10$. The following table 1 shows the efforts the proposed model. The estimated efforts are very close to the measured efforts.

The proposed model results are compared with the existing models in the literature and the results are shown in the following table 2.

The performance measure considered here is Mean Absolute Relative Error (MARE)

$$\% \text{ MARE} = \text{mean}$$

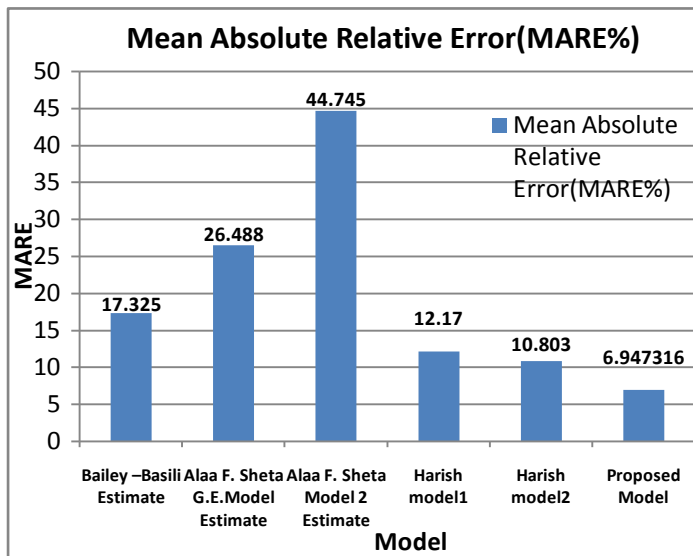
$$\left[\frac{\text{abs}(\text{Measured Effort} - \text{Estimated Effort})}{(\text{measured effort})} \right] \times 100$$



The MARE of various models is shown below.

Model	Mean Absolute Relative Error(MARE%)
Bailey –Basili Estimate	17.325
Alaa F. Sheta G.E.Model Estimate	26.488
Alaa F. Sheta Model 2 Estimate	44.745
Harish model1	12.17
Harish model2	10.803
Proposed Model	6.947316

The Results show that the value of MARE (Mean Absolute Relative Error) applying fuzzy-swarm intelligence was substantially lower than MARE of other models exists in the literature.



VI. CONCLUSION

Software cost estimation is based on a probabilistic model and hence it does not generate exact values. However availability of good historical data coupled with a systematic technique can generate better results. In this paper we proposed new model structure to estimate the software cost (Effort) estimation. Fuzzy sets is used for modeling uncertainty and impression to better the effort estimation and particle swarm optimization for tuning parameters. It is observed from the results that Fuzzy-Swarm intelligence gives accurate results when juxtaposed with its other counterparts. On testing the performance of the model in terms of the MARE the results were found to be useful.

REFERENCES REFERENCES REFERENCIAS

1. John w. Bailey and Victor R.Basili,(1981) "A meta model for software development resource expenditures", proceedings of the Fifth International Conference on Software Engineering, CH-1627-9/81/0000/0107500.75@IEEE, pp:107-129,1981.C

2. Chris F. Kemerer, (1987) "An Empirical Validation of Software Cost Estimation Models", Management Of Computing-Communications of ACM, Vol:30 No. 5, pp:416-429, May 1987.
3. KrishnaMurthy.S and Douglas Fisha (1995), "Machine Learning Approaches to estimating software development effort", IEEE Transactions On Software Engineering, Vol. 21, No. 2, pp: 126-137, February 1995.
4. Barry Boehm, Chris Abts, Sunita Chulani(1998), Software Development Cost Estimation Approaches – A Survey1,IBM, pp: 1-46 ,1998.
5. Hao Ying,(1998), The Takagi-Sugeno fuzzy controllers using the simplified linear control rules are nonlinear variable gain controllers, ELSEVIER: Automatica, Vol:34,No.2,pp.157-167, 1998.
6. Magne Jørgensen,(2005), Evidence-Based Guidelines for Assessment of Software Development Cost Uncertainty, IEEE Transactions On Software Engineering, Vol. 31, No. 11, pp: 942-954, November 2005.
7. Xishi Huang, Danny Ho, Jing Ren, Luiz F. Capretz (2005), Improving the COCOMO model using a neuro-fuzzy approach, DOI:10.1016 /j.asoc. 2005.06.007, Applied Soft Computing 7 (2007) pp: 29–40, @2005 Elsevier.
8. Alaa F. Sheta, (2006), Estimation of the COCOMO Model Parameters Using Genetic Algorithms for NASA Software Projects, Journal of Computer Science 2, (2) pp:118-123, 2006.
9. Alaa Sheta (2006),Software Effort Estimation and Stock Market Prediction Using Takagi-Sugeno Fuzzy Models, Proceedings of 2006 IEEE International Conference on Fuzzy Systems, 0-7803-9489-5/06/IEEE 2006,pp:171-178,July 16-21, 2006.
10. Alaa Sheta, David Rine and Aladdin Ayesh (2008), Development of Software Effort and Schedule Estimation Models Using Soft Computing Techniques, IEEE Transaction, 978-1-4244-1823-7/08/pp: 1283-1289@2008 IEEE.
11. Prasad Reddy P.V.G.D, Sudha, K.R., Rama Sree .P, Ramesh .S.N.S.V.S.C. (2010), Fuzzy Based Approach for Predicting Software Development Effort, International Journal of Software Engineering (IJSE), Vol.1 No.1,pp:1-11,2010.
12. Prasad Reddy P.V.G.D (2010), Particle Swarm Optimization In The Fine Tuning Of Fuzzy Software Cost Estimation Models, International Journal of Software Engineering (IJSE), Vol.1 No.2, pp: 12-23,2010.
13. Harish Mittal and Pradeep Bhatia, (2007), Optimization Criteria for Effort Estimation using Fuzzy Technique, CLEI Electronic Journal, Vol. 10, Number 1, Paper 2,pp: 1-11, June 2007.

Table 1 : Estimated effort of the proposed model.

Size (m)	Methodology	Measured Effort	$\alpha = 0.7362m$	m	$\beta = 0.8998m$	$a*\alpha^b + c*ME + d$	$a*m^b + c*ME + d$	$a*\beta^b + c*ME + d$	Estimated Effort
2.1	28	5	1.546	2.100	1.890	3.722	5.000	4.523	4.712
3.1	26	7	2.282	3.100	2.789	5.316	7.075	6.418	6.679
4.2	19	9	3.092	4.200	3.779	6.742	8.999	8.156	8.490
12.5	27	23.9	9.203	12.500	11.248	18.535	24.055	21.994	22.811
46.5	19	79	34.233	46.500	41.841	55.630	71.846	65.790	68.190
54.5	20	90.8	40.123	54.500	49.039	63.572	82.044	75.145	77.879
67.5	29	98.4	49.694	67.500	60.737	76.386	98.400	90.179	93.437
78.6	35	98.7	57.865	78.600	70.724	86.911	111.853	102.538	106.229
90.2	30	115.8	66.405	90.200	81.162	97.125	125.048	114.620	118.753
100.8	34	138.3	74.209	100.800	90.700	106.636	137.223	125.800	130.327

Table 2 : Estimated effort of various models & proposed model.

Size(m)	Methodology	Measured Effort	Estimated Effort-Proposed Model	Bailey – Basili Estimate	Alaa F. Sheta G.E.model Estimate	Alaa F. ShetaModel 2 Estimate	Harish model1	Harish model2
2.1	28	5	4.712	7.226	8.44	11.271	6.357	4.257
3.1	26	7	6.679	8.212	11.22	14.457	8.664	7.664
4.2	19	9	8.490	9.357	14.01	19.976	11.03	13.88
12.5	27	23.9	22.811	19.16	31.098	31.686	26.252	24.702
46.5	19	79	68.190	68.243	81.257	85.007	74.602	77.452
54.5	20	90.8	77.879	80.929	91.257	94.977	84.638	86.938
67.5	29	98.4	93.437	102.175	106.707	107.254	100.329	97.679
78.6	35	98.7	106.229	120.848	119.27	118.03	113.237	107.288
90.2	30	115.8	118.753	140.82	131.898	134.011	126.334	123.134
100.8	34	138.3	130.327	159.434	143.0604	144.448	138.001	132.601