Software Effort Estimation Using Particle Swarm Optimization with Inertia Weight

CH.V.M.K.Hari[1] and Dr. Prasad Reddy[2]

[1] GITAM University

Received: 24 August 2011 Accepted: 19 September 2011 Published: 1 October 2011

Abstract

Software is the most expensive element of virtually all computer based systems. For complex custom systems, a large effort estimation error can make the difference between profit and loss. Cost (Effort) Overruns can be disastrous for the developer. The basic input for the effort estimation is size of project. A number of models have been proposed to construct a relation between software size and Effort; however we still have problems for effort estimation because of uncertainty existing in the input information. Accurate software effort estimation is a challenge in Industry. In this paper we are proposing three software effort estimation models by using soft computing techniques: Particle Swarm Optimization with inertia weight for tuning effort parameters. The performance of the developed models was tested by NASA software project dataset. The developed models were able to provide good estimation capabilities.

Index terms — PM- Person Months, KDLOC-Thousands of Delivered Lines of Code, PSO - Particle Swarm Optimization, Software Cost Estimation

1 INTRODUCTION

he modern day software industry is all about efficiency. With the increase in the expanse and impact of modern day software projects, the need for accurate requirement analysis early in the software development phase has become pivotal. The provident allocation of the available resources and the judicious estimation of the essentials form the basis of any planning and scheduling activity. For a given set of requirements, it is desirable to cognize the amount of time and money required to deliver the project prolifically. The chief aim of software cost estimation is to enable the client and the developer to perform a costbenefit analysis. The software, the hardware and the human resources involved add up to the cost of a project. The cost / effort estimates are determined in terms of person-months (pm) which can be easily interchanged to actual currency cost.

The basic input parameters for software cost estimation is size, measured in KDLOC (Kilo Delivered Lines Of Code). A number of models have been evolved to establish the relation between Size and Effort **??**13]. The parameters of the algorithms are tuned using Genetic Algorithms [5], Fuzzy models [6] A common approach to the estimation of the software effort is by expressing it is as a single variable function of the project size. The equation of effort in terms of size is considered as follows:Effort= a * (Size) b (1)

Where a, b are constants. The constants are usually determined by regression analysis applied to historical data.

2 b) Standard PSO with Inertia Weights

In order to meet the needs of modern day problems several optimization techniques have come been introduced. When the search space is too large to search exhaustively, population based searches may be a good alternative, however, population based search techniques cannot guarantee you the optimal (best) solution. We will discuss a population based search technique, Particle Swarm Optimization (PSO) with Inertia Weights [Shi and **??**berhart

1998]. Particle Swarm has two primary operators: Velocity update and Position update. During each generation each particle is accelerated toward the particles previous best position and the global best position. At each iteration a new velocity value for each particle is calculated based on its current velocity, the distance from its previous best position, and the distance from the global best position. The new velocity value is then used to calculate the next position of the particle in the search space. The inertia weight is multiplied by the previous velocity in the standard velocity equation and is linearly decreased throughout the run. This process is then iterated a set number of times or until a minimum error is achieved.

The basic concept of PSO lies in accelerating each particle towards its Pbest and Gbest locations with regard to a random weighted acceleration at each time. The modifications of the particle's positions can be mathematically modeled by making use of the following equations:V i k+1 = w * V i k + c 1 * rand() 1 * (Pbest -S i k) + c 2 * rand() 2 * (Gbest -S i k)**(2)**S i k+1 = S i k + V i k**(3)**

Where, S i k is current search point, S i

3 THE STANDARD PSO WITH INERTIA WEIGHT FOR SOFTWARE EFFORT ESTIMATION

The software effort is expressed as a function of a single variable of effort in terms of the project size as shown in equation-1. The parameters a, b are measured by using regression analysis applied to historical data. In order to tune these parameters we use the standard PSO with inertia weights. A nonzero inertia weight introduces a preference for the particle to continue moving in the same direction it was going on the previous iteration. Decreasing the inertia over time introduces a shift from the exploratory (global search) to the exploitative (local search) mode. The updating of weighting function is done with the following formula.W new = [(T mi -T ci) * (W iv -W fv)] / T mi + W fv**(4)**

Where W new is new weight factor, T mi is the maxium numer of iteration specified, T ci is the current iteration number, W iv is the initial value of the weight, W fv is the final value of the weight. Empirical experiments have been performed with an inertia weight set to decrease linearly from 0.9 to 0.4 during the course of simulation. In the first experiment we keep the parameters c1 and c2 (weighting factors) fixed, while for the following experiment we change c1 and c2 (weighting factors) during subsequent iterations by employing the following equations [Rotnaweera, A. Halgamog S.K. and Watson H.C, 2004].

C 1 (t) = 2.5 -2 * (t / max_iter), which is the cognitive learning factor.

()**5**

C 2 (t) = 0.5 + 2* (t / max_iter), which is the social coefficient.

The particles are initialized with random position and velocity vectors the fitness function is evaluated and the Pbest and Gbest of all particles is found out. The particles adjust their velocity according to their Pbest and Gbest values. This process is repeated until the particles exhaust or some specified number of iterations takes place. The Gbest particle parameters at the end of the process are the resultant parameters.

4 III.

5 MODEL DESCRIPTION

In this model we have considered "The standard PSO with inertia weights" with /without changing the weighting factors (c1, c2). PSO is a robust stochastic optimization technique based on the movement of swarms. This swarm behavior is used for tuning the parameters of the Cost/Effort estimation. As the PSO is a random weighted probabilistic model the previous benchmark data is required to tune the parameters, based on that data, swarms develop their intelligence and empower themselves to move towards the solution.

The following is the methodology employed to tune the parameters in each proposed models following it.

6 a) METHODOLOGY (ALOGORITHM)

Input: Size of Software Projects, Measured Efforts, Methodology (Effort Adjustment factor-EAF). Output: Optimized Parameters for Estimating Effort.

The following is the methodology used to tune the parameters in the proposed models for Software Effort Estimation.

Step 1: Initialize "n" particles with random positions P i and velocity vectors V i of tuning parameters .We also need the range of velocity between [-V max ,V max]. The Initial positions of each particle are Personally Best for each Particle.

Step 2: Initialize the weight function value w with 0.5 and weightening parameters cognitive learning factor c1, social coefficient c2 with 2.0.

Step 3: Repeat the following steps 4 to 9 until number of iterations specified by the user or Particles Exhaust.

Step 4: for i = 1,2, ???, n do // For all the Particles For each particle position with values of tuning parameters, evaluate the fitness function. The fitness function here is Mean Absolute Relative Error (MARE). The objective in this method is to minimize the MARE by selecting appropriate values from the ranges specified in step 1.

Step 5: Here the Pbest is determined for each particle by evaluating and comparing measured effort and estimated effort values of the current and previous parameters values. If fitness (p) better than fitness (Pbest) then: Pbest = p.

Step 6: Set the best of 'Pbests' as global best -Gbest. The particle value for which the variation between the estimated and measured effort is the least is chosen as the Gbest particle.

Step 7: Update the weightening function is done by the following formulaW new = [(T mi -T ci) * (W iv -W fv)] / T mi + W fv**(7)**

Step 8: Update the weightening factors is done with the following equations for faster convergence.

Step 9: Update the velocity and positions of the tuning parameters with the following equations for j = 1, 2, ????m do // For number of Parameters, our case m is 2or 3 or 4 beginV ji k+1 = w * V ji k + c 1 * rand() 1 * (Pbest -S ji k) + c 2 * rand() 2 * (Gbest -S ji k)**(10)**S ji k+1 = S ji k + V ji k+1**(11)**

end;

Step 10: Give the Gbest values as the optimal solution.

Step 11: Stop b) PROPOSED MODELS i. MODEL 1:

A prefatory approach to estimating effort is to make it a function of a single variable , often this variable is project size measure in KDLOC (kilo delivered lines of code) and the equation is given as , Effort = a (size) b Now in our model the parameters are tuned using above PSO methodology.The Update of velocity and positions of Parameter "a" isV ai k+1 = w * V ai k + c 1 * rand() 1 * (Pbest -S ai k) + c 2 * rand() 2 * (Gbest -S ai k)**(12)**

S ai k+1 = S ai k + V ai k+1

The Update of velocity and positions of Parameter "b" isV bi k+1 = w * V bi k + c 1 * rand() 1 * (Pbest -S bi k) + c 2 * rand() 2 * (Gbest -S bi k) S bi k+1 = S bi k + V bi k+1 Table 1 : Effort Multipliers ii. MODEL 2:

Instead of having resources estimates as a function of one variable, resources estimates can depend on many different factors, giving rise to multivariable models. Such models are useful as they take into account the subtle aspects of each project such as their complexity or other such factors which usually create a non linearity. The cost factors considered are shown below. The product of all the above cost factors is the Effort Adjustment Factor (EAF).A model of this category starts with an initial estimate determined by using the strategic single variable model equations and adjusting the estimates based on other variable which is methodology. The equation is,Effort = a *(size) b + c* (ME).

Where ME is the methodology used in the project. The parameters a, b, c are tuned by using above PSO methodology. The Update of velocity and positions of Parameter "a", "b" are shown in Model 1 and Parameter "c" isV ci k+1 = w * V ci k + c 1 * rand() 1 * (Pbest -S ci k) + c 2 * rand() 2 * (Gbest -S ci k) S ci k+1 = S ci k + V ci k+1 iii. MODEL 3

There are a lot of factors causing uncertainty and non linearity in the input parameters. In some projects the size is low while the methodology is high and the complexity is high, for other projects size is huge but the complexity is low. As per the above two models size and effort are directly proportional. But such a condition is not always satisfied giving rise to eccentric inputs. This can be accounted for by introducing a biasing factor (d). So the effort estimation equation is: The Update of velocity and positions of Parameter "a", "b", "c" are shown in Model 1,2 and Parameter "d" isV di k+1 = w * V di k + c 1 * rand() 1 * (Pbest -S di k) + c 2 * rand() 2 * (Gbest -S di k) S di k+1 = S di k + V di k+1

IV.

7 MODEL ANALYSIS a) Implementation

We have implemented the above methodology for tuning parameters a,b,c and d in "C" language. For the parameter' a 'the velocities and positions of the Where ME represents Measured Effort, EE represents Estimated Effort.

V.

8 MODEL EXPERIMENTATION

9 EXPERIMENT -1:

For the study of these models we have taken data of 10 NASA **??**13] Table **??** : NASA software projects data By running the "C" implementation of the above methodology we obtain the following parameters for the proposed models. The following are the results obtained by running the above PSO algorithm implemented in "C" with changing weighting factors on each iteration. VI.

10 RESULTS AND DISCUSSIONS

The following table shows estimated effort of our proposed model:

EXPERIMENT -1:

11 CONCLUSION

Software cost estimation is based on a probabilistic model and hence it does not generate exact values. However if good historical data is provided and a systematic technique is employed we can generate better results. Accuracy of the model is measured in terms of its error rate and it is desirable to be as close to the actual values as possible. In this study we have proposed new models to estimate the software effort. In order to tune the parameters we use particle swarm optimization methodology algorithm. It is observed that PSO gives more accurate results when juxtaposed with its other counterparts. On testing the performance of the model in terms of the MARE, VARE and VAF the results were found to be futile. These techniques can be applied to other software effort models.

12 Tad Gonsalves

Figure 1:

[1] [2]

[1] Global Journal of Computer Science and Technology Volume XI Issue XVIII Version I 16 2011 October Software Effort Estimation Using Particle Swarm Optimization with Inertia Weight

[2] . Software Effort Estimation Using Particle Swarm Optimization with Inertia Weight

COST FACTORS	DESCRIPTION	RATING				
		VERY LOW	LOW	NOMINAL	HIGH	VERY HIGH
	Product					
RELY	Required software reliability	0.75	0.88	1	1.15	1.4
DATA	Database size	-	0.94	1	1.08	1.16
CPLX	Product complexity	0.7	0.85	1	1.15	1.3
	Computer					
TIME	Execution time constraint	-	-	1	1.11	1.3
STOR	Main storage constraint	-	-	1	1.06	1.21
VIRT	Virtual machine volatility	-	0.87	1	1.15	1.3
TURN	Computer turnaround time	-	0.87	1	1.07	1.15
	Personnel					
ACAP	Analyst capability	1.46	1.19	1	0.86	0.71
AEXP	Application experience	1.29	1.13	1	0.91	0.82
PCAP	Programmer capability	1.42	1.17	1	0.86	0.7
VEXP	Virtual machine volatility	1.21	1.1	1	0.9	-
LEXP	Language experience	1.14	1.07	1	0.95	-
	Project					
MODP	Modern programming practice	1.24	1.1	1	0.91	0.82
TOOL	Software tools	1.24	1.1	1	0.91	0.83
SCED	Development schedule	1.23	1.08	1	1.04	1.1

Figure 2:

SIZE	MEASURED EFFORT	METHODOLOGY	ESTIMATED EFFORT OF OUR MODELS C1,C2 ARE CONSTANT DURING THE ITERATION (CASE-I)			ESTIMATED EFFORTOF OUR MODELS C1,C2 ARE CHANGED DURING THE ITERATION(CASE-II)		
			MODEL-I	MODEL-II	MODEL-III	MODEL-I	MODEL-II	MODEL-III
2.1	5	28	5.000002	4.998887	5.000007	5.000002	5.502722	5.000001
3.1	7	26	6.982786	7.047925	7.07543	6.982786	7.071439	6.975912
4.2	9	19	9.060186	9.222874	8.999259	9.060186	8.47359	9.154642
12.5	23.9	27	23.08629	23.40447	24.05549	23.08629	21.65101	22.82118
46.5	79	19	71.2293	71.75396	71.84614	71.2293	68.24138	71.03909
54.5	90.8	20	81.61792	82.10557	82.04368	81.61792	78.82941	81.44935
67.5	98.4	29	98.05368	98.39988	98.39998	98.05368	96.18965	97.79541
78.6	98.7	35	111.7296	111.9449	111.8526	111.7296	110.7037	111.4518
90.2	115.8	30	125.7302	125.8721	125.048	125.7302	125.0572	125.6834
100.8	138.3	34	138.3002	138.3003	137.2231	138.3002	138.523	138.2999

Figure 3: ©

Model	VAF (%)	Mean Absolute Relative Error (%)	Variance Absolute Relative Error (%)
Bailey –Basili Estimate	93.147	17.325	1.21
Alaa F. Sheta G.E. Model I Estimate	98.41	26.488	6.079
Alaa F. Sheta Model II Estimate	98.929	44.745	23.804
Harish model1	98.5	12.17	80.859
Harish model2	99.15	10.803	2.25
CASE-I MODEL -I	98.92	4.6397	0.271
CASE-I MODEL-II	98.92	4.6122	0.255
CASE-I MODEL-III	98.9	4.4373	0.282
CASE-II MODEL -I	98.92	4.6397	0.271
CASE-II MODEL-II	98.89	7.5	0.253
CASE-II MODEL-III	98.95	4.9	0.257

1

Figure 4: Model 1 :

3

Fig 1 : Measured Effort Vs Estimated Efforts of
Proposed Models
COMPARISON WITH OTHER MODELS
Refer Table 4 for the comparison with other models.

Figure 5: Table 3 :

2011
October
18

Project No	Size In KDLOC	Methodology (ME)	Measured Effort
13	2.1	28	5
10	3.1	26	7
11	4.2	19	9
17	12.5	27	23.9
3	46.5	19	79
4	54.5	20	90.8
6	67.5	29	98.4
15	78.6	35	98.7
1	90.2	30	115.8
18	100.8	34	138.3

Figure 6:

5

Figure 7: Table 5 :

Techniques", 2008 IEEE Congress on Evolutionary
, Atsushi Ito, Ryo Kawabata and Kiyoshi Itoh , Swarm Intelligence in the Optimization of Software Development Project Schedule, 0730-
Computation (CEC 2008), 978-1-4244-1823-7/08
3157/08 , 2008 IEEE. 8. J.S.Pahariya ,V. Ravi, M. Carr, Software Cost Estimation using Computational Int

Biologically Inspired Computing (NaBIC 2009).

9. 20 2011 October	International Journal of Software Engineering, IJSE Vol.3 Table 4 : Measured Efforts of Various Models

[Note: © 2011 Global Journals Inc. (US)]

Figure 8:

[Sheta et al.] *Development of Software Effort and Schedule Estimation Models Using Soft Computing*, Alaa Sheta , David Rine , Aladdin Ayesh .

[Robert et al. ()] *Elitist Multi evolutionary algorithm for environmental/economic dispatch*, T F Robert , Ah King , C S Harry , Rughooputh . 2003. IEEE.

[Sheta] 'Estimation of the COCOMO Model Parameters Using Genetic Algorithms for NASA Software Projects'. Alaa F Sheta . *Journal of Computer Science* 1549-36362006. 2 (2) p. .

[Coello Coello ()] *Evolutionary Algorithms for Solving Multi-Objective Problems*, C A Coello Coello . 2002. Kluwer.

[Goldberg ()] *Genetic Algorithms in Search, Optimization and Machine Learning*, D E Goldberg . 1989. Addison-Wesley.

[Deb ()] *Multi-Objective Optimization Using Evolutionary Algorithms*, K Deb . 2002. John Wiley and Sons.