Parameters Estimation of the COCOMO Model Using Hybrid Algorithm of Genetic Algorithm and Cuckoo Search Algorithm

Alaa Abu-Srhan(1), **Azzam Sleit**(2), **Ahmad Sharieh**(2)
Hashemite University(1),University of Jordan(2)
alaaa@hu.edu.jo, azzam.sleit@ju.edu.jo, sharieh@ju.edu.jo

Abstract

The Constructive Cost Model (COCOMO) is one of the famous software cost estimation model developed by Barry W. Boehm. This model is used to estimate software costs, duration and maintenance efforts early in the development life cycle. COCOMO model has a simple function with two parameters to be estimated. Many techniques were used to estimate those parameters such as fine tuning and better prediction that can be achieved. In this paper, hybrid algorithm of cuckoo search algorithm and genetic algorithm (CSGA) have been used to solve the parameter estimation problem that leads to what we call COCOMO-CSGA. A data set from NASA software projects has been used in the experiments. The experiments shows that CSGA improve the accuracy of effort estimation.

Keywords: COCOMO, CSGA, Effort Estimation

1. INTRODUCTION

Software development is an important process for many organizations. The productivity of people is different, so it is difficult at the beginning of any project to estimate the software, because we don’t know lot about the software being developed. It is important to estimate the software effort during the early stage of software development, and it becomes less important as the project progresses [1].This lead us to think about developing model to predicate the software effort.

There are many models used in computing the effort for software project and it gave us accurate results but there is still a challenge. This is because of many reasons:

1) The collected measurements are uncertainty.

2) There are many drawbacks in the current estimated methods.

3) The cost drivers which have many characteristics deepened on the methodology of development [2].

One of the most powerful mathematical models for effort estimation is COCMO model, this model used to estimate the effort of any project on the early stage of project and this leads to minimize the cost of overall project.

In this paper, review on the algorithmic software effort estimation models is presented. The proposed hybrid algorithm of cuckoo search algorithm and genetic algorithm to build COCOMO-CSGA for the software effort is explained. The hybrid algorithm was tested on 18 software projects based on NASA data set.

2. LITERATURE REVIEW

The definition of soft computing techniques has been presented in 1994 by Zedeh [3]. He said that soft computing is not homogeneous body with regard to concepts and techniques. The techniques used in soft computing domain at that time were fuzzy logic, neurocomputing, and probabilistic reasoning. These techniques were explored to build efficient effort estimation models structures [4], [5]. This domain was expanded to cover other techniques such as Swarm Intelligence (SI), Genetic Algorithms (GAs), Differential Evolution (DE), and other techniques. The use of Neural Networks (NNs), GAs and Genetic Programming (GP) for cost estimation of software has been explored by author in [6]. Later on, a detailed study on using Genetic Programming (GP), Neural Network (NN) and Linear Regression (LR) in solving the software project estimation was provided by author in [7], [8] and [9]. Many datasets have been explored with acceptable results. Artificial neural networks has been used on the cost estimation models. A survey on this is in

[10]. A fuzzy COCOMO model has been developed in [4].

Many software engineering problems can be solved by using Soft Computing and Machine Learning Techniques. These problems include the effort and cost estimation problems. Set of modified COCOMO models has been provided by the author in [11]. These models gave us an interesting results. Many authors made some modification to COCOMO model [12]–[14] and made a comparison to the work presented in [11]. Takagi Sugeno Fuzzy Logic has been used to see how the software effort estimation problem can be solved using rule based system, it is presented in [15]. An extended work on using Soft Computing Techniques has been presented by the authors in [16]. It's important to say here that NASA software projects [17] have been used to test the developed model. Particle Swarm Optimization (PSO) has been used by the authors to estimate the parameters of the COCOMO model. COCOMO-PSO, Fuzzy Logic (FL), Halstead, Walston-Felix, Bailey-Basili and Doty models with excellent performance results have been used by the authors to made a comparison between them.

3. CONSTRUCTIVE COST MODEL

There are many software effort estimation models used to help project manager to take best decisions for a project. These models provide a high quality software estimation. One of the most famous model is COCOMO, this model was introduced in 1981 by Boehm [18], [19]. COCOMO model used to identify the effort, the developed time and the maintenance effort by the mathematical equations that this model used. It was developed based on 63 software projects. The estimation accuracy is suggestively improved when adopting models such as the Intermediate and Complex COCOMO models [2]. Equation 1 shows the basic COCOMO model.

$$E = a(KLOC)^b \quad (1)$$

Where E is the software effort, KLOC stands for Kilo Line Of Code. α and β are the values of the parameters depend mainly on the kind of software project. Based on the complexity of the project, Software projects were classified into three categories. They are: Organic, Semidetached and Embedded models [2]. Although, there are extensions of COCOMO such as COMCOMO II, however, for the purpose of research reported in this paper, the basic COMCOMO model is used. The three models are given in Table 1 . According to the type of software projects these models are give different results.

Table 1: BASIC COCOMO MODELS [2]

Model Name	Effort (E)	Time (T)
Organic	$E = 2.4(KLOC)^{1.05}$	$T = 2.5(E)^{0.38}$
Semi-Detached	$E = 3.0(KLOC)^{1.12}$	$T = 2.5(E)^{0.35}$
Embedded	$E = 3.6(KLOC)^{1.20}$	$T = 2.5(E)^{0.32}$

The definition of organic, semidetached, and embedded systems are reported in [18]. the condition to consider the software project of organic type is when the project deals with developing a well understood application program, the size of the development team is reasonably small, and the team members are experienced in developing similar types of projects.

The condition to consider the software project of semidetached type is when the development consists of a mixture of experienced and inexperienced staff. Team members may have limited experience on related systems but may be unfamiliar with some aspects of the system being developed.

The condition to consider the software project of embedded type is when the software being developed is strongly coupled to complex hardware, or when the stringent regulations on the operational procedures exist.

4. CSGA ALGORITHM

CSGA combined the advantages of Genetic Algorithm(GA) and Cuckoo Search (CS) together. The main disadvantage of the GA is that it easily trapped within local minimum [20]. In order to overcome the above drawback, CS is used. Through using CS the local search will be performed faster than the GA [20]. Moreover, in the Cuckoo search there is only a single parameter apart from the population size. CSGA has the following operations: The Search New Nest Operator, the Abandon Operator, and Genetic Algorithm Operation.

CSGA starts with CS operations and then proceed to genetic operations in order to estimate COCOMO parameters, so we can say that the initial population of GA is not generated randomly but it uses the results of CS, after genetic operations finished the algorithm start over. Flowchart in Figure 1 shows the steps of parameter estimation of COCOMO-CSGA.

4.1 Cuckoo search operations

Cuckoo Search (CS) algorithm [20] is based on the following three principles:

1) each cuckoo lays its egg in a randomly chosen nest, one at a time.

2) the nests with high quality of eggs are considered best nests for the next generation to follow.

3) the host bird can discover the egg laid by cuckoo with a probability pa $\in$ (0, 1), and the number of available nests is fixed.

The search pattern used in CS is Lévy flight, which mimic animals and birds when they search for foods randomly, searching of the next step based on the current location, and it is based also on the probability of going to the next location.

Our goal of parameters estimation is to find the optimal values of a and b that minimize MMRE (Mean Magnitude of Relative Error) and this considered as the fitness function, The MMRE shown in equation 2. The objective function is the equation of COCOMO II, the input of this model are the actual effort for the project and Kilo line of code (KLOC).

$$MMRE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{AEffort - EEffort}{AEffort}\right| \qquad (2)$$

Where N is the number of projects, AEffort is actual effort taken from NASA dataset , EEffort is the estimated effort which is the effort we gain from our proposed algorithm.

The algorithm starts with randomly specifies the initial population of nests, we use lower and upper boundaries to define the search space. A lévy flight is used to get the new nest for a cuckoo *i*, the equation of lévy flight shown in equation 3.

$$x_i^{t+1} = x_i^t \oplus \alpha \text{levy}(\lambda) \qquad (3)$$

Where α >0 is the step size and t is the current generation , $\oplus$ is an entry-wise multiplication, The random step length is drawn from a levy distribution which has an infinite variance with an infinite mean.

The worst nests will be discarded and the best nests will be kept for next generation, new nest will generate using Lévy flight. In each generation we calculate the fitness function for the new nests and rank these nests according to their fitness, and find the best nest from the current nests, we will repeat the previous procedures until the stop condition satisfied. It is important to say here that before checking the stop criteria we have to proceed to GA in order to enhance the result.

4.2 Genetic Algorithm operations

the basic operations of GA[21] are crossover (also called recombination), and mutation . In order to produce other new solutions, a pair of "parent" solutions is selected for breeding from the pool that has been previously selected. By producing a "child" solution by the methods of crossover and mutation, a new solution is created and this solution typically shares many of the parents characteristics. The process of producing new parents continues until generating the new population of solutions with the appropriate size. It is suggested that more than two "parents" generate higher quality chromosomes in spite of that the reproduction methods basing on the use of two parents are more "biology inspired".

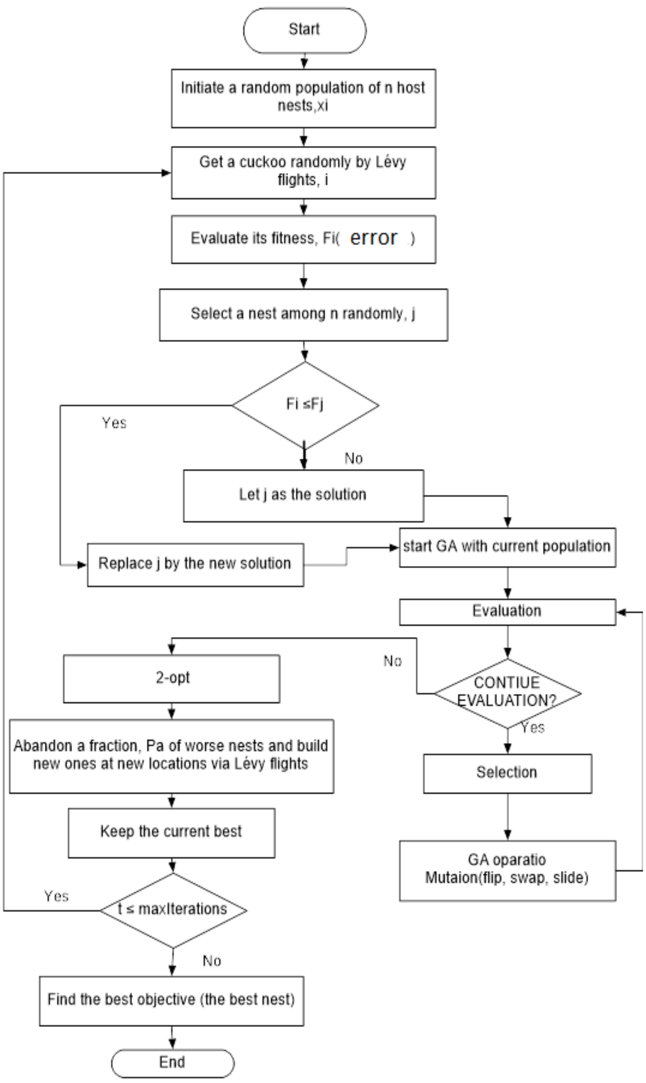

Fig. 1: The Proposed CSGA Algorithm

5. COMPUTATION CRITERIA

The performance of the developed model shall be evaluated using number of evaluation criteria which are:

- Variance-Accounted-For (VAF), as in equation (4).

$$VAF = [1 - \frac{\mathrm{var}(y - \hat{y})}{\mathrm{var}(y)}] \times 100\% \qquad (4)$$

- Euclidian distance (ED), as in equation (5).

$$ED = \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad (5)$$

- Manhattan distance (MD), as in equation (6).

$$MD = \sum_{i=1}^{n} | y_i - \hat{y}_i | \qquad (6)$$

- Mean Magnitude of Relative Error (MMRE), as in equation (7).

$$MMRE = \frac{1}{N} \sum_{i=1}^{N} \frac{| y_i - \hat{y}_i |}{y_i} \qquad (7)$$

where y and $\hat{y}$ are the actual effort and the estimated effort , respectively, based on the developed CSGA- COCOMO model, and N is the number of measurements used in the experiments.

6. EXPERIMENTAL RESULTS

6.1 *NASA data set*

Experiments have been conducted on a data set presented by Bailey and Basili to explore strengthen of the developed CSGA based model. This data set [2] consists of Kilo Line of Code (KLOC),Methodology (ME), and Effort (E) variables. KLOC and E are used in this paper, ME will be used in next work. NASA data set is given in Table 2.

Table 2: NASA DATA SET[2]

Project No.	KLOC	ME	Measured Effort
1	90.2	30	115.8
2	46.2	20	96.0
3	46.5	19	79.0
4	54.5	20	90.8
5	31.1	35	39.6
6	67.5	29	98.4
7	12.8	26	18.9
8	10.5	34	10.3
9	21.5	31	28.5
10	3.1	26	7.0
11	4.2	19	9.0
12	7.8	31	7.3
13	2.1	28	5.0
14	5.0	29	8.4
15	78.6	35	98.7
16	9.7	27	15.6
17	12.5	27	23.9
18	100.8	34	138.3

6.2 COCOMO-CSGA model

We developed a COCOMO-CSGA model for the effort estimation, CSGA is used to estimate the optimal parameters a and b of equation 1, the values of a and b chosen to minimize the error between actual effort and estimated effort which we refer to it as MMRE, from table 4 we can notice that we gain good MMRE value. Actual effort and estimated effort based on the COCOMO-CSGA model are given in Table 3. Figure 2 shows the actual and estimated effort using COCOMO-CSGA model. Figure 3 shows the convergence of the estimated model parameters a and b. Figure 4 shows minimum error at each generation. The developed model's performance were computed using five different criterias as reported in Table IV. A very high VAF has been received which reflects a good performance modelling. It can be seen that the performance of the developed COCOMO-CSGA model based historical data were able to achieve significant modelling results.

Table 3: ACTUAL AND ESTIMATED EFFORT

Project NO.	Actual Effort	CSGA Estimated Effort
1	115.8	145.8570
2	96.0	76.9340
3	79.0	77.4116
4	90.8	90.0996
5	39.6	52.6966
6	98.4	110.5482
7	18.9	22.5506
8	10.3	18.6601
9	28.5	37.0253
10	7.0	5.8123
11	9.0	7.7704
12	7.3	14.0438
13	5.0	4.0052
14	8.4	9.1799
15	98.7	127.8698
16	15.6	17.2984
17	23.9	22.0450
18	138.3	162.2045

The developed model parameters are estimated as equation (6).

$$E = 0.9704(kLOC)^{0.9561} \qquad (6)$$

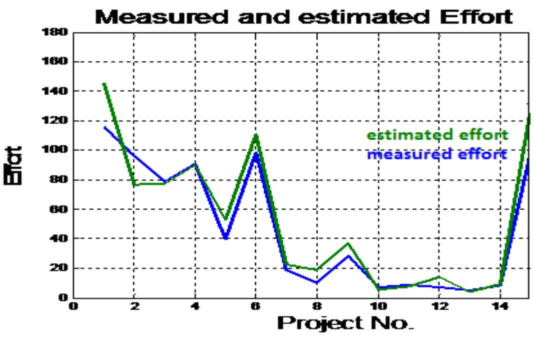

Fig.2 Measured and Estimated effort for COCOMO-CSGA Model

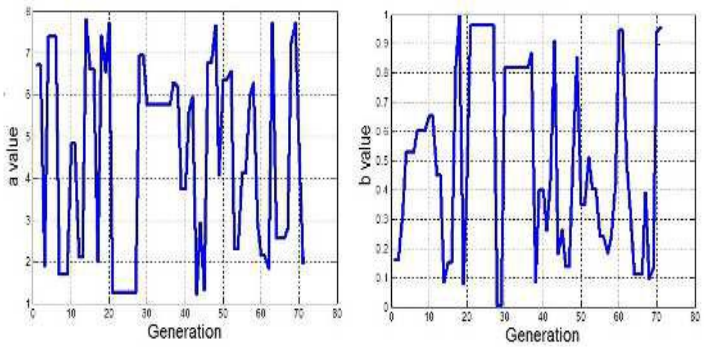

Fig. 3: Convergence of the estimated model parameters a and b

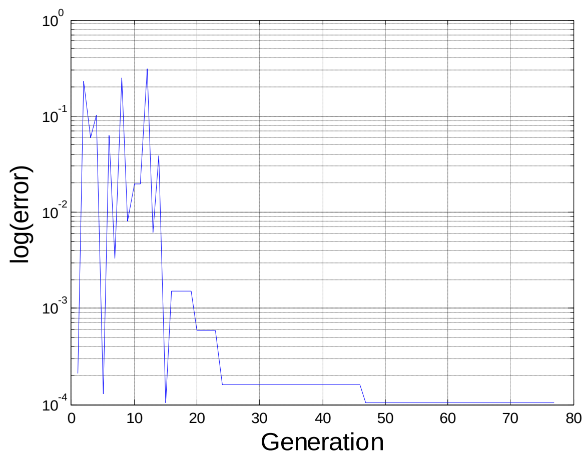

Fig. 4: minimum error at each generation

Table 4: COMPUTATION CRITERIA FOR THE COCOMO-CSGA MODEL

VAF	ED	MS	MD	MMRE
94.5635%	56.777	179.09	9.1531	0.18971

7. CONCLUSION

The problem of estimation effort for software project is a challenging problem for software project manager. In this paper CSGA is modified to estimate the parameter of COCOMO model. The developed COCOMO-CSGA model has been tested using NASA software project dataset. This model was capable of providing good effort estimation of compared to other known model in the literature.

REFRANCE

[1] A. K. Geeta Nagpal, Moin Uddin, "Ahybrid technique using gray relational analysis and regression for software effort estimation using feature selection," International Journal of Soft Computing and Engineering, vol. 1, pp. 20–27, Jan. 2012.

[2] S. A. Alaa F. Sheta, "Software effort estimation inspired by cocomo and fp models: A fuzzy logic approach," International Journal of Advanced Computer Science and Applications, vol. 4, no. 11, pp. 192–197, 2013.

[3] L. A. Zadeh, "Soft computing and fuzzy logic," IEEE Softw., vol. 11,pp. 48–56, Nov. 1994.

[4] J. Ryder, Fuzzy COCOMO: Software Cost Estimation. PhD thesis, Binghamton University, 1995.

[5] A. C. Hodgkinson and P. W. Garratt, "A neuro-fuzzy cost estimator,"in Proceedings of the Third Conference on Software Engineering and Applications, pp. 401–406, 1999.

[6] M. A. Kelly, "A methodlolgy for software cost estimation using machine learning techniques," Master's thesis, Naval Postgratuate School,Monterey, California, 1993.

[7] J. J. Dolado and L. F. andez, "Genetic programming, neural network and linear regression in software project estimation," in Proceedings of the INSPIRE III, Process Improvement through training and education, pp. 157–171, British Company Society, 1998.

[8] A. J. Albrecht and J. R. Gaffney, "Software function, source lines of code, and development effort prediction: A software science validation,"IEEE Trans. Software Engineering, vol. 9, no. 6, pp. 630–648, 1983.

[9] J. E. Matson, B. E. Barret, and J. M. Mellinchamp, "Software developemnet cost estimation using function points," IEEE Trans. Software Engineering, vol. 20, no. 4, pp. 275–287, 1994.

[10] M. Shepper and C. Schofield, "Estimating software project effort using analogies," IEEE Tran. Software Engineering, vol. 23, pp. 736–743,1997.

[11] A. F. Sheta, "Estimation of the COCOMO model parameters using genetic algorithms for NASA software projects," Journal of Computer Science, vol. 2, no. 2, pp. 118–123, 2006.

[12] H. Mittal and P. Bhatia, "A comparative study of conventional effort estimation and fuzzy effort estimation based on triangular fuzzy numbers," International Journal of Computer Science and Security, vol. 1,no. 4, pp. 36–47, 2007.

[13] M. Uysal, "Estimation of the effort component of the software projects using simulated annealing algorithm," in World Academy of Science, Engineering and Technology, vol. 41, pp. 258–261, 2008.

[14] P. S. Sandhu, M. Prashar, P. Bassi, and A. Bisht, "A model for estimation of efforts in development of software systems," in World Academy of Science, Engineering and Technology, vol. 56, pp. 148–152, 2009.

[15] A. Sheta, "Software effort estimation and stock market prediction using takagi-sugeno fuzzy models," in Proceedings of the 2006 IEEE Fuzzy Logic Conference, Sheraton, Vancouver Wall Centre, Vancouver, BC, Canada, July 16-21, pp. 579–586, 2006.

[16] A. Sheta, D. Rine, and A. Ayesh, "Development of software effort and schedule estimation models using soft computing techniques," in Proceedings of the 2008 IEEE Congress on Evolutionary Computation (IEEE CEC 2008) within the 2008 IEEE World Congress on Computational Intelligence (WCCI 2008), Hong Kong, 1-6 June, pp. 1283–1289,2008.

[17] J. W. Bailey and V. R. Basili, "A meta model for software development resource expenditure," in Proceedings of the International Conference on Software Engineering, pp. 107–115, 1981.

[18] B. Boehm, Software Engineering Economics. Englewood Cliffs, NJ, Prentice-Hall, 1981.

[19] B. Boehm, Cost Models for Future Software Life Cycle Process: COCOMO2. Annals of Software Engineering, 1995.

[20] Ala'a Abu-Srhan , Essam Al Daoud, " A Hybrid Algorithm Using a Genetic Algorithm and Cuckoo

Search Algorithm to Solve the Traveling Salesman Problem and its Application to Multiple Sequence Alignment", International Journal of Advanced Science and Technology Vol.61, pp.29-38, 2013.

[21] Maram Alajlan , Nejmeddine Tagoug,"Optimization of COCOMO-II Model for Effort and Development Time Estimation using Genetic Algorithms", Proc. Of The International Conference on Communications, Computer Science and Information Technology,2016.