Software Cost and Effort Estimation using a New Optimization Algorithm Inspired by Strawberry Plant

Muhammad Sufyan Khan, CH Anwar ul Hassan, Munam Ali Shah, Azra Shamim
Department of Computer Science
COMSATS University, Islamabad 44000, Pakistan
muhammadsufyan012@gmail.com

***Abstract*—Nowadays, accurate software effort and time estimation are one of the main challenges in the software engineering community. Correct and precise estimation plays an important role in successful software development and organization productivity. Constructive Cost Model (COCOMO) is an algorithmic model commonly used in estimating time and effort having four coefficients. From the last few decades, many researchers work on optimization of the COCOMO model by using naturally inspired algorithms. Such optimization algorithms help the software industry in predicting accurate and genuine values of cost and effort used for software project development and maintenance. In this paper, we are using a new meta-heuristic algorithm inspired by the strawberry plant for optimization of COCOMO effort estimation method. NASA 93 data set is used in the proposed approach. The Magnitude of Relative Error (MRE) and Mean Magnitude of Relative Error (MMRE) is evaluated. Experimental results of the proposed method with the COCOMO model shows a decline in MMRE to 23.8%**

I. INTRODUCTION

Softwares play an essential role in every field of life. Software costing and estimation is a very hot topic form last decades in the software community. To estimate the efforts and development time not only increase the success ability of software also enriches the company performance by managing human resources, project schedules, cost estimation and more benefits. These benefits control the project delay and reduce the failure probability of the software [1]. Software time and effort estimation are one of the most critical parts, before the software development. So, the people involved in software project management is responsible to estimate cost and efforts.

$$Effort = a(KLOC)b \tag{1}$$

Researcher introduces many models and methods for costing and estimation from last decades. The most common model which is used for software costing and estimation is COCOMO. In 1981 COCOMO introduce firstly that is used for software effort estimation, with growing software industry COCOMO cannot fulfill the requirements. In 2000 COCOMO II is introduced that is used to estimate development time and efforts. It contains coefficients which affect the estimation accuracy [2]. To adjust the COCOMO II parameters, there are several methods to estimate the optimum value, such as metaheuristic algorithms, nature heuristic algorithms, and machine learning algorithms. Now, in modeling this effort in software development makes heuristics algorithms and machine learning algorithms attain much more attention. The aim of these algorithms in estimating time and effort in development to find the best solution. Some of these algorithms are, Differential Evolution (DE), Ant Colony Optimization (ACO), Genetic Algorithm (GA), Firefly Algorithm (FA), Fuzzy Logic Model (FLM), Harmony Search Algorithm (HSA), Neural Network (NN), Cuckoo Search Algorithm (CSA) etc. [3,4,5,6] Moreover, the efficiency is improved through the hybridization of and deterministic optimization methods and algorithms. In this paper, a new nature-inspired heuristic algorithm is used: Strawberry Algorithm (SBA), to estimate the software effort, time and cost. Also, to calculate the Mean Magnitudes Relative Error (MMRE) to show the effectiveness of SBA. In figure 1 show the growth of strawberry roots.

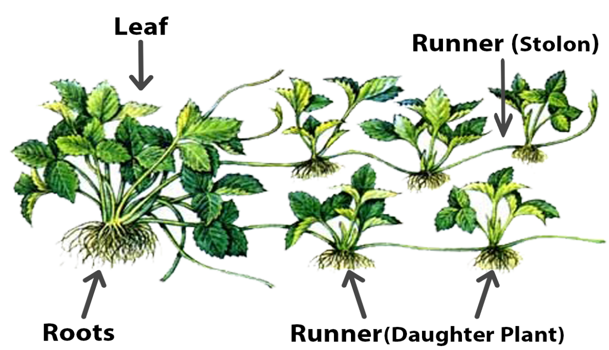

Fig. 1: Strawberry roots working

.

Rest of the paper is organized like Section II contains Literature review, Section III Proposed Method, Section IV Experiments and Results and in last conclusions to summarize the whole work.

II. RELATED WORK

The state of the art review shows software projects development often neither completed within allocated budget nor in

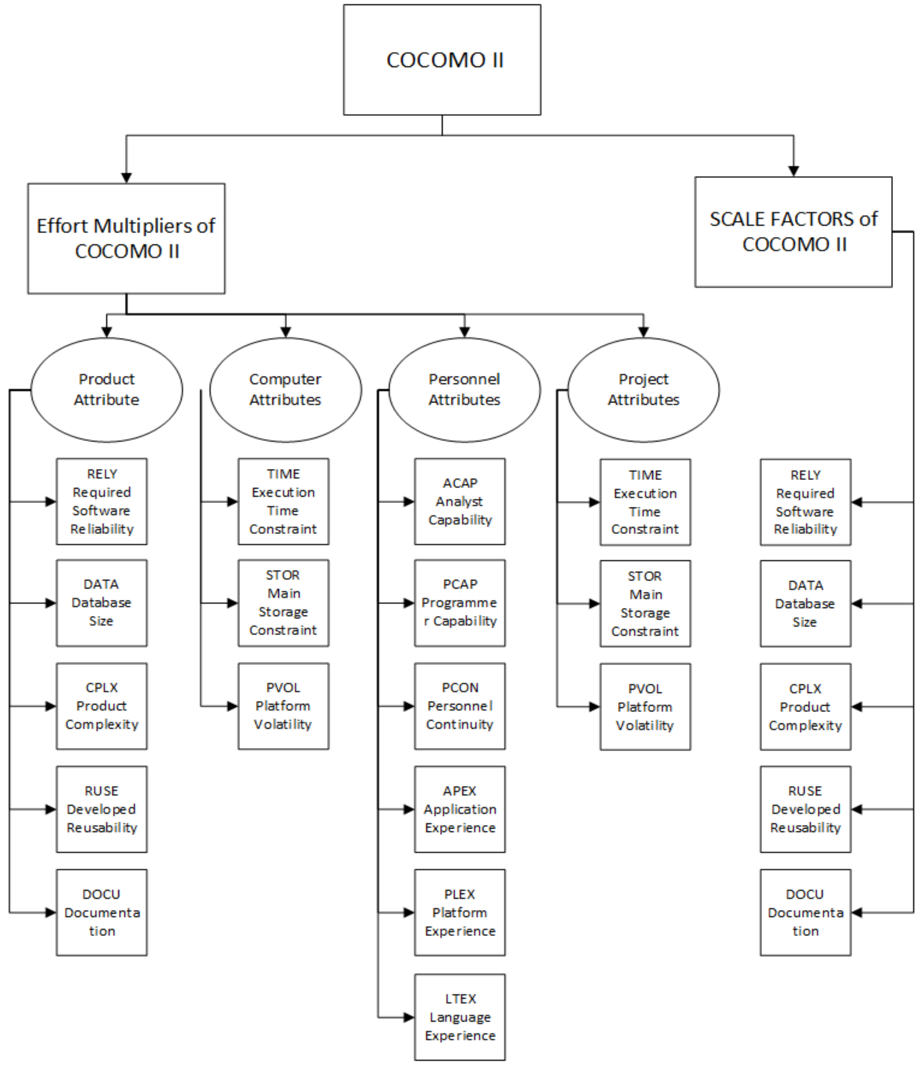

Fig. 2: COCOMO II

time. Metaheuristic technique has been dominating over the last few decades. For software cost estimation many meta-heuristic algorithms have already been implemented [1]. GA, PSO, and HSA are a few examples of algorithms used in cost estimation methods. In [2] GA has been used for optimization of estimation techniques. In another study, parameters of the basic COCOMO model is optimized by using the simplified GA. NASA software project dataset is used. The comparison shows better truthful estimation over the basic COCOMO model [3]. GA and analogy-based algorithms are bringing closer to proposed a model for estimation with high efficiency. MMRE is decreased in comparison with COCOMO [4]. PSO is used to optimized analogy-based estimation technique. The final model gave better results than all above as it consists of stepwise and multiple regression, artificial neural network, regression and classification tree [5]. In [6] COCOMO model and PSO algorithm are combined for effort estimation based on calculative intelligence techniques. The final results showed that by combining PSO with clustering algorithms give more precise results in the evaluation of the COCOMO model in contrast with the primary COCOMO estimation model. A research by [7] the best values for a and b in effort estimation equation of the COCOMO model is achieved by the combination of basic COCOMO with PSO. With other estimation models: Halstead, Bailey-Basili, Doty and basic COCOMO, this model is compared and results show that the proposed model gave smaller MMRE. In [8] meta-heuristic HSA is used to optimized COCOMO effort estimation model. To test the results, the NASA dataset was used. The software effort estimation in the propose optimization methods shows a decrease in MMRE to almost 21%. Optimization of COCOMO II coefficients is adjusted using Cuckoo optimization algorithm (COA), Turkish software industry Dataset was used. The proposed system is not being trapped in obtaining the best solution by the local optimum. [9]. By using BCO technique on an interactive voice response software project dataset of a company, tuning of COCOMO model parameter is done [11]. This approach generates various partial solutions and the finest solution is selected which has the least MMRE. BCO based model, improve the accuracy of cost estimation. In [12] FA is used for optimization. Three COCOMO-based models: basic COCOMO model, a proposed model using GA and estimation model using PSO is evaluated. Comparison results show that high accuracy is achieved by using FA over other techniques [14].

III. Proposed Method

This research is carried out on NASA dataset by using Strawberry Algorithm (SBA) for optimization of basic COCOMO effort estimation method using MATLAB software.

A. Dataset used for optimization:

NASA dataset is used for execution of the proposed technique. The dataset was divided into three main categories, on the basis of project design: organic, semidetached and embedded. In the proposed method, NASA dataset is used because it has more data than other datasets. There are many options for implementation of SBA on basic COCOMO as it has organic, semidetached and embedded status. There are two stages in the planned technique. In the first phase, all the 93 NASA projects data is used at once for the optimization. The optimization is done on basic COCOMO formula. In the second phase, the projects are to split into three categories having different folds of organic, semidetached and embedded. The experiments are carried out on each of the categories with separate basic COCOMO formula.

B. Optimization by the strawberry algorithm:

Strawberry algorithm is inspired by the movement of roots of strawberry plant for resolving ceaseless multi-variable issues [15]. Strawberry plant has to be spread through the so-called runner (or stolen). From the mother (parent) plant leaf axil the runner is a crawling stalk which is produced and grows out. At the endpoint of the runner, there comes a new plant, known as daughter plant, which grew large and then again, a new root arises and also runner comes out on the daughter plant that time it becomes mother (parent) plant and to produce another new daughter plant and so on. Underneath equation is used to produce random root and runner.

$$ZProp(i) = [Zroot(i)Zrunner(i)] = [Z(i)Z(i)] + [xrootr1xrunnerr2] \quad (2)$$

Following steps are performed during COCOMO optimization.

1) Import data from the dataset.
2) Specify parent root and number of daughter root, moves and stopping criteria.
3) Initialize parameters of the COCOMO model.
4) Execute basic COCOMO module.
5) Optimize using the strawberry algorithm.
6) Calculate effort values for each root in a forward pass by using formula (2).
7) Generate a partial solution for each root by optimizing parameters with different values according to changes required in a forward pass by using formula (3).

$$MRE = \frac{actualeffort - predictedeffort}{actualeffort} \quad (3)$$

8) Evaluate the basic COCOMO and SBA for MRE and MMRE to choose the best solution to prove the effectiveness of strawberry algorithm by using fitness function in formula (4).

$$MMRE = \frac{1}{N}\sum \frac{abs(actualeffort - predictedeffort)}{actualeffort} \quad (4)$$

9) Discard the results which are not reliable to their solutions.
10) Acquire a global best solution from the best root. Results having minimum MMRE is the global solution.

In this research, we are implementing both options. In first option we are using 93 NASA projects (datasets), in the second option, Nasa projects are classified into three groups; organic projects (3 datasets), semidetached projects (69 datasets) and embedded projects (21 datasets) according to the design modes of the NASA projects. The datasets in organic mode are less (3 datasets) so, no optimization performed on that mode. Therefore, basic COCOMO optimization is carried out on the remaining two modes of Nasa projects.

C. Strategy for trebling the datasets

Training datasets and test datasets are essential to perform optimization, for that reason each project (NASA datasets) are isolated into three folds using MATLAB cvpartition (k-fold cross validation partition) function. Each fold contains 2 datasets of training and test. This cvpartition function splits the NASA datasets into three clusters according to the number of datasets. In case of using all the projects datasets are divided into the chunks of 31 projects (31+31+31 =93). In semidetached case, chunks of 23 datasets (23+23+23=69) are created and same as in embedded case (7+7+7 = 21)

Datasets of each fold are put into training datasets. Test datasets as training dataset members, two times then training dataset number. As an outcome, training and test datasets are different in each fold

D. Fitness function

Many researcher and practitioner performed optimization on effort estimation method to assist accuracy under different estimation criteria. In this article, we are using most common criteria Mean Magnitude of Relative Error. MMRE is formed by using equation 5 and 6. RE is the size difference between the estimated and actual value as shown in equation 5. In equation 7 combined all MRE values, as in equation 8 summarized MMRE, Fitness function is defined.

$$RE.i = \frac{Estimate.i - Actual.i}{Actual.i} \quad (5)$$

$$MRE.i = abs(RE.i) \tag{6}$$

$$MMRE.i = \frac{1}{N}(MRE.1 + MRE.2 + ... + MRE.N) \tag{7}$$

$$MMRE = \frac{1}{N}\sum_{i=1}^{N}\frac{|Estimate - Actual|}{Actual} \tag{8}$$

In these equations, estimation refers to the value of predicted efforts and actual refer to actual effort required for the project and N defines the number of projects.

TABLE I: SBA Parameters and Values

Parameters	Values
Number of Mother plants	25
Number of variables	10
Dimensions	[-3, +3]
Length of Runners	200
Length of Roots	20
Maximum iterations	100

IV. Experiments and Results

As we have mentioned earlier in the proposed method section that COCOMO has three modes and optimization was carried out on all these modes. In the first mode, all 93 NASA projects from the data set are taken and optimization of basic COCOMO was carried out. In the second, semidetached projects are taken from the Nasa and optimization was done using the proposed method. Embedded NASA projects were optimized in the third mode. The proposed method results in all the fold were compared to those obtained from basic COCOMO formula.The average outcome on all the folds by 100 times simulation of results was compared.

A. Basic COCOMO model Optimization:

During optimization of the basic COCOMO model, we optimized coefficients a and b of general basic COCOMO formula by using all 93 projects of the dataset of NASA. We have obtained the result of coefficients as a = 5.06 and b = 0.89. As shown in figure 3, MMRE was being minimized in all the three folds as compared to the general basic COCOMO model. After doing the average of three folds, we have obtained a diagram shown in figure 4. As we can see in the diagram, MMRE managed from the planed process shows a diminution as compared to the COCOMO method and effects 23.8% enhancement in MMRE.

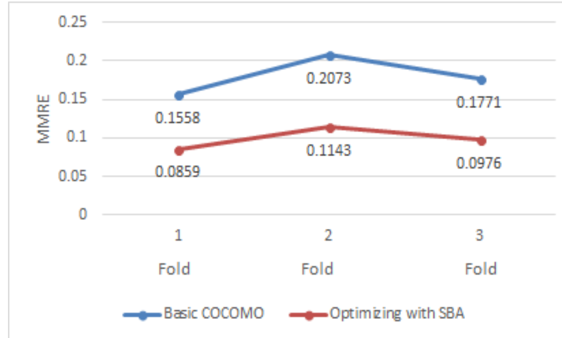

Fig. 3: Basic COCOMO, Optimization with SBA

.

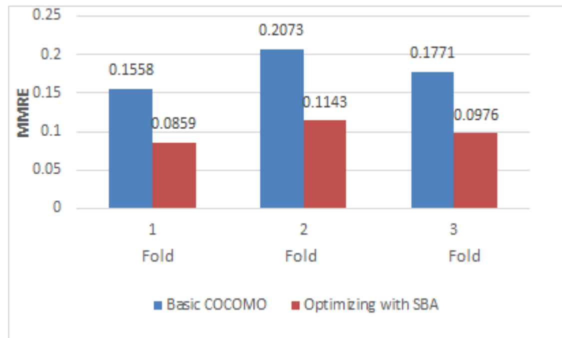

Fig. 4: Average Comparison of Basic COCOMO

.

B. Semidetached- Basic COCOMO Optimization:

We carried out the simulation and optimization of coefficients of the semidetached COCOMO formula. The values of a = 3.0 and b = 1.12 is used and simulation is done using all 63 semidetached projects of NASA datasets. The resultant coefficients are a = 4.6 and b = 0.94. As illustrated in figure 5, MMRE in all the fold shows a diminution as compared to semidetached basic COCOMO model. After averaging all the three folds, figure 6 is obtained. As illustrated in the graph, the MMRE generated by the anticipated methods is diminution as associated with the current COCOMO model and leads 16% enhancement in MMRE.

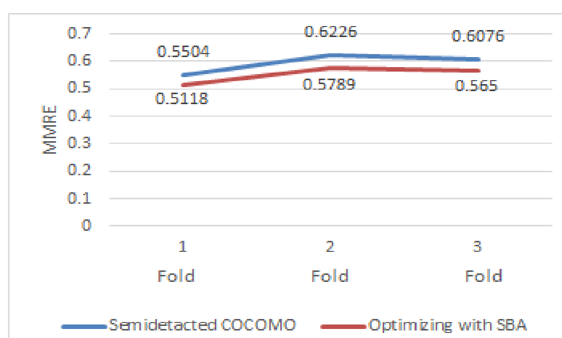

Fig. 5: Semidetached-Basic COCOMO, Optimization with SBA

.

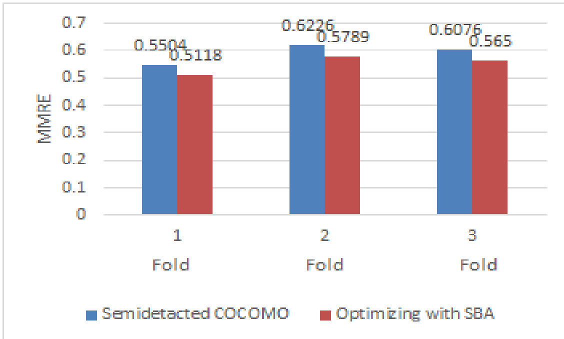

Fig. 6: Average Comparison of Semidetached-Basic COCOMO

.

C. Embedded- Basic COCOMO Optimization:

NASA dataset has 21 embedded projects, optimization of its coefficients a = 3.6 and b = 1.20 is done using embedded COCOMO formula. As results shown in figure 7, MMRE in all the three-fold shows a diminution when optimization is done using the strawberry algorithm as a blend with the embedded basic COCOMO model. The average of the three-fold is shown in figure 8. As shown in the graph, the MMRE obtained from COCOMO model is high as a comparison with the graph obtained from the method proposed. The results obtained from the optimization done by strawberry is 23.8% of the current COCOMO model

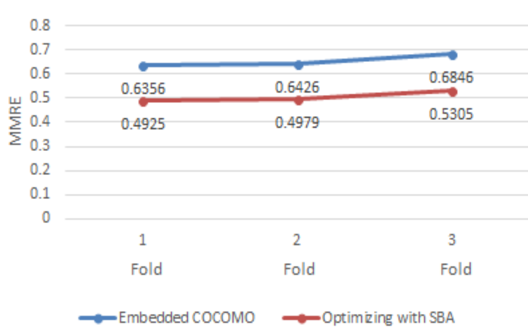

Fig. 7: Embedded-Basic COCOMO, Optimization with SBA

.

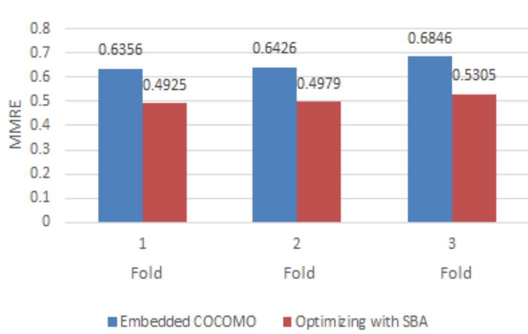

Fig. 8: Average Comparison of Embedded-Basic COCOMO

.

V. Conclusion

A new technique for estimating software effort is proposed in the current article. This method used a blend of COCOMO cost and effort estimation with SBA to achieve optimized coefficients. NASA dataset is used for simulation of the results in MATLAB software. MMRE is reduced in the majority of the optimization methods. Analysis of results shows a decrease in MMRE in all the three modes, general basic COCOMO, semidetached basic COCOMO, and embedded basic COCOMO. This decline in results shows that a combination of the SBA and COCOMO effort estimation technique increase an accuracy in software estimation technique. In the future, we use other optimization algorithms to improve the accuracy of our results.

References

[1] Puri, Ruchi, and I. Kaur. "Novel Meta-Heuristic Algorithmic Approach for Software Cost Estimation." International Journal of Innovations in Engineering and Technology (IJIET) 5.2 (2015): 456-463.

[2] Rijwani, Poonam, and Sonal Jain. "Enhanced software effort estimation using multi layered feed forward artificial neural network technique." Procedia Computer Science 89 (2016): 307-312.

[3] Choudhary, Kavita. "GA based Optimization of Software Development effort estimation." IJCST, September (2010).

[4] Bardsiri, V. Khatibi, A. Khatibi, and E. Khatibi. "An optimization-based method to increase the accuracy of software development effort estimation." Journal of Basic and Applied Scientific Research (2013): 159-166.

[5] Bardsiri, Vahid Khatibi, et al. "A PSO-based model to increase the accuracy of software development effort estimation." Software Quality Journal 21.3 (2013): 501-526.

[6] Puspaningrum, Alifia, and Riyanarto Sarno. "A Hybrid Cuckoo Optimization and Harmony Search Algorithm for Software Cost Estimation." Procedia Computer Science 124 (2017): 461-469.

[7] Gharehchopogh, Farhad Soleimanian, Isa Maleki, and Seyyed Reza Khaze. "A NOVEL PARTICLE SWARM OPTIMIZATION APPROACH FOR SOFTWARE EFFORT ESTIMATION." International Journal of Academic Research 6.2 (2014).

[8] Jafari, SM Sabbagh, and F. Ziaaddini. "Optimization of software cost estimation using harmony search algorithm." Swarm Intelligence and Evolutionary Computation (CSIEC), 2016 1st Conference on. IEEE, 2016.

[9] Parwita, I. Made Mika, Riyanarto Sarno, and Alifia Puspaningrum. "Optimization of COCOMO II coefficients using Cuckoo optimization algorithm to improve the accuracy of effort estimation." Information and Communication Technology and System (ICTS), 2017 11th International Conference on. IEEE, 2017.

[10] Maleki, Isa, Ali Ghaffari, and Mohammad Masdari. "A new approach for software cost estimation with hybrid genetic algorithm and ant colony optimization." International Journal of Innovation and Applied Studies 5.1 (2014): 72.

[11] Jafari, SM Sabbagh, and F. Ziaaddini. "Optimization of software cost estimation using harmony search algorithm." Swarm Intelligence and Evolutionary Computation (CSIEC), 2016 1st Conference on. IEEE, 2016.

[12] Wadhwa, Ankita, Somya Jain, and Chetna Gupta. "An Effective Precision Enhancement Approach to Estimate Software Development Cost: Nature Inspired Way." Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 9.3 (2017): 85-91.

[13] Sharma, Ashish, and Dharmender Singh Kushwaha. "Estimation of software development effort from requirements based complexity." Procedia Technology 4 (2012): 716-722.

[14] ul Hassan, CH Anwar, et al. "Energy Optimization in Smart Grid Using Grey Wolf Optimization Algorithm and Bacterial Foraging Algorithm." International Conference on Intelligent Networking and Collaborative Systems. Springer, Cham, 2017.

[15] Khan, Muhammad Sufyan, et al. "A new meta-heuristic optimization algorithm inspired from strawberry plant for demand side management in smart grid." International Conference on Intelligent Networking and Collaborative Systems. Springer, Cham, 2017.