

# A Bee Colony Optimization Algorithm Approach for Software Cost Estimation

Zahra Ashegi Dizaji  
Department of Computer  
Engineering, Urmia  
Branch, Islamic Azad  
University, Urmia, Iran

Reza Ahmadi  
Department of Computer  
Engineering, Urmia  
Branch, Islamic Azad  
University, Urmia, Iran

Hojjat Gholizadeh  
Department of Computer  
Engineering and  
Information Technology,  
Amirkabir University

Farhad Soleimanian  
Gharehchopogh  
Department of Computer  
Engineering, Hacettepe  
University, Ankara, Turkey

## ABSTRACT

The Software Cost Estimation (SCE) is one of the most important stages of the production cycle of a system. Therefore, the managers need to accurately determine the requirements of the project to have accurate estimations. But given the fact that the SCE is done at the beginning of the system development, it can be problematic due to lack of accurate information about the project. To solve this problem, researchers have proposed several methods to estimate the cost of software projects but haven't been successful in estimating the costs with 100% accuracy.

In this paper, we intend to improve the accuracy of the cost estimation by using Bee Colony Optimization algorithm. It should be mentioned that the proposed method is compared with the intermediate COCOMO. The results indicate that the proposed method have reduced the mean absolute relative error to 0.1619.

## Keywords

Software Cost Estimation, bee colony optimization algorithm, intermediate COCOMO.

## 1. INTRODUCTION

Nowadays, given the great importance of the software systems, production and development of software systems have been the focus of many companies and organization in a way that they have started to produce and develop them. But it should be carefully attended that the produced system should be affordable for the customers and the producer companies and they should also comply with the needs of the users and be able to meet their needs. If software manufacturing companies aren't able to estimate the true costs they would have trouble in the upcoming stages of system production, so that if the estimated cost is less than the actual cost of the project this will cause time and financial loss and project failure. And if the estimated cost is more than the actual cost it will end up in customer dissatisfaction [1]. Accordingly, SCE is of utmost importance and the process is done before starting the software project [2]. But given the fact that there isn't enough information about system requirements at the beginning of software development and the project is not fully understood, SCE faces some difficulties [3].

Several factors affect the cost estimation of a software project and the most important among these factors is the size of the project which can be measured based on Source Lines of Code (SLOC), Kilo SLOC (KSLOC) and Function Point (FP) [4]. Researchers have proposed several methods to solve this problem. The most primitive of these solutions are mathematical and algorithmic methods. These algorithms started to appear in the late 1970s [5]. The most important

documented of these methods is COCOMO which was developed by Boehm in 1980 in three types of: Basic COCOMO [6], Intermediate COCOMO [7] and Detailed COCOMO [8]. Among these models, the Intermediate COCOMO has been considered by many researchers. But with the passage of time and increasing complexity of the software projects and with the formation of new ways of programming, researchers needed to offer new ways to solve these problems. After some time machine learning methods including data mining methods and meta-heuristic algorithms have been proposed to solve this problem. However, it should be noted that researchers are still exploring to find a new solution with the accuracy of 100 percent in cost estimation.

We have organized this paper into these sections: section 2: an overview to previous works. Section 3: the basic concepts (intermediate COCOMO, BCO). Section 4: the proposed method. Section 5: evaluating the performance of the proposed model. Section 6: conclusion and future works.

## 2. PREVIOUS WORKS

With the increasing importance of software systems, nowadays, software is among the most expensive parts of computer systems. Therefore the importance of SCE has increased in the recent years. Thus, every year researchers try to obviate this problem by presenting new methods. Also, considering the changing nature of modern software systems, researchers are attempting to use modern methods to increase the accuracy of the cost estimates. A number of proposed methods based on machine learning algorithms can be outlined as follows:

A.F.Sheta offered two new models for SCE using a combination of Genetic Algorithm (GA) and COCOMO (intermediate COCOMO and detailed COCOMO) [9]. In their paper, to estimate the cost of software projects, COCOMO formula is used in a way that in the first model a, b and c parameters are obtained by GA. In the second model, in addition to the parameters a, b and c, parameter d is also obtained by GA. Based on the results and the statements of the researcher it can be said that the presented combined models cause improvement in the accuracy of cost estimation. In [10] to increase the accuracy of the estimated costs researchers used a combination of GA and Functional Link Neural Network. The combined method consists of two separate stages. The first stage is training stage. At this stage FLANN algorithm is used to classify the NASA dataset and GA is used to initialize the effective parameters in SCE based on the COCOMO formula. At the second stage which is the testing stage, FLANN algorithm is used in the prediction process. These researchers have evaluated the results according to criterion like MMRE, MdMRE and RED (0.25)

and compared them with various methods such as SWR, OCRWFLANN, OFWFLANN and CART and the results show the success of the presented method.

Another meta-heuristic algorithm that is used by many researchers to solve various problems is the Particle Swarm Optimization algorithm which is usually used in combination with other algorithms. Due to the good results obtained by this algorithm for solving various problems, many researchers have used it to solve SCE problems. For instance in [11] PSO algorithm and combination of it with Tent Mapping is used as PSO & Chaos theory for SCE. In this paper, the best result at every stage is selected from among two proposed algorithms and considered as the output of that stage. According to the claims of these researchers, the combined method has performed better than PSO and PSO has performed better than COCOMO. Or in [12] combination of PSO algorithm and Fuzzy C-means Clustering as one model and combination of PSO and Learning Automata model as another model have been used for SCE. These researchers have evaluated the results obtained from the presented new models with COCOMO according to MMRE, MRE and PRED and they have seen improved accuracy of estimated cost compared to COCOMO.

### 3. BASIC CONCEPTS

In this paper, Bee Colony Optimization algorithm is used to increase the accuracy in estimating the cost of software projects. Therefore, in this section, a brief explanation of Intermediate COCOMO and Bee Colony Optimization will be given.

#### 3-1 Intermediate COCOMO

Intermediate COCOMO as the most popular mathematical and logarithmic method was presented by Boehm in 1981 [6]. In this model, the parameters  $a$  and  $b$  are suggested and the values for these parameters are initialized according to the type of the project (1. Organic projects 2. Semidetached projects 3. Embedded projects). The calculation of cost estimation is shown in formula 1.

$$\text{Formula 1: } PM = a * (\text{size})^b * \prod_{i=1}^{15} EM_i$$

According to formula 1 in addition to the mentioned parameters and the project size, 15 other features also affect the cost calculation [13, 14].

#### 3-2 Bee Colony Optimization Algorithm

Bee Colony Optimization algorithm was introduced by Teodorovic. In this algorithm artificial bees get their food resources in two stages. At the first stage “moving forward” and at the second stage “moving backwards” is performed. At the first stage bees find a number of new solutions as basic solutions and at the second stage they participate in a decision making meeting and present their solutions. Then according to the provided solutions and food source quality they produced a factor called “loyalty” in themselves and if the value of loyalty is bigger than the randomly generated number the bee continues exploring the food source, otherwise it leaves its food source and moves on to another food source [15].

The probability of selecting a solution by bees is calculated by formula 2 [16].

$$\text{formula2: } V_j = \frac{\text{Max}(F) - F_j}{\text{Max}(F) - \text{Min}(F)}, J = 1, 2, \dots, M$$

In formula 2  $M$  denotes the number of solutions,  $F$  denotes all solutions and  $F_j$  denotes the current solution. Formula 3 and

formula 4 are used to determine the loyalty of each bee for the food source [16].

$$\text{Formula 3: } O_b = \frac{C_{\text{max}} - C_b}{C_{\text{max}} - C_{\text{min}}}, b = 1, 2, \dots, B$$

Parameter  $B$  is the number of bees and  $C$  is the popularity of the solutions.

$$\text{Formula4: } P_b^{u+1} = e^{\frac{O_{\text{max}} - O_b}{U}}, b = 1, 2, \dots, B$$

Parameter  $B$  is the number of bees and  $U$  is the number of steps forward.

## 4. PROPOSED METHOD

SCE is considered as one of the important aspects of software project management. Therefore, researchers attempt to solve this problem and reduce the error rate in cost estimation. This paper as well tries to reduce the error rate using the BCO algorithm. As was mentioned in the second part, SCE is affected by many factors including fixed parameters like  $a$  and  $b$  for same type projects. We have sought to calibrate these values in this paper. Actually the proposed method performs in this way that it divides NASA data set completely randomly into training and testing sets. Then the projects within each set are classified according to the project types presented in intermediate COCOMO. In the next step values of the parameters  $a$  and  $b$  for same type projects are initialized by BCO algorithm according to training and learning. Then the specified values for parameters  $a$  and  $b$  are used to estimate the cost of projects in testing data set. The procedure of the presented algorithm is shown in Table 1 and Figure 1.

Table 1: purpose Method Pseudo code

Step 1: Read data from NASA data set
Step 2: Separate data set to training and testing data set
Step 3: Categorize training and testing data set
Step 4: Do the forward pass for every bees
Step 4-1: Develop many solution based on constructive moves
Step 4-2: Select solution with roulette wheel
Step 5: All bees back the hive and start backward pass
Step 6: Sort bees by fitness value
Step 7: Develop random number and compare with fitness value
Step 8: If fitness value is better than random number continue own exploration else select other solution by roulette wheel
Step 9: If the stopping condition is not met go step 3
Step 10: Give value of $a, b$ parameters
Step 11: Calculate cost for test data set projects by $a, b$ parameters
Step 12: Show result

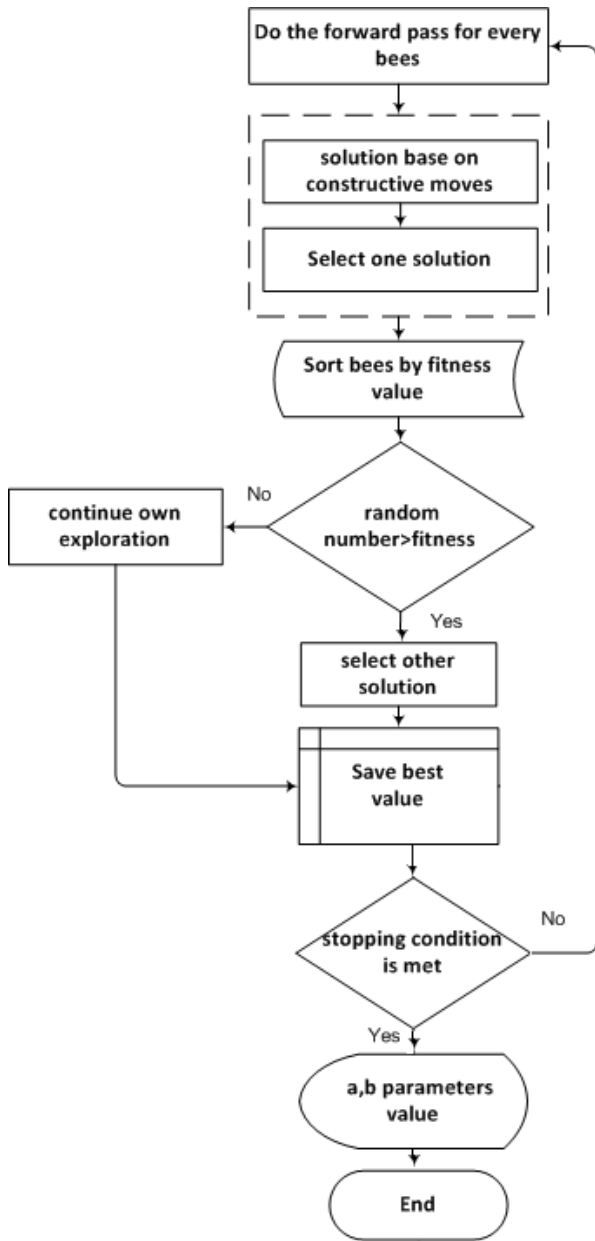


Figure 1: purpose Method Flowchart

## 5. DISCUSSION

SCE nowadays is expressed as an important issue in software engineering domain. The reason can be stated as if the estimated cost has a high precision and if it roughly matches the actual end cost the project is completed successfully. But if the estimates are less accurate the project will certainly fail. Therefore, in this paper we have tried to increase the accuracy of estimated cost using BCO algorithm on NASA data set based on intermediate COCOMO. The obtained results show an improvement in accuracy of estimated cost. It should be mentioned that in this paper, evaluation is based on mean absolute relative error, which is calculated by formula 5 and formula 6.

$$\text{Formula 5: } MARE_i = \frac{|Actual_i - Estimate_i|}{Actual_i}$$

$$\text{Formula 6: } MMARE = \frac{1}{N} \sum_{i=1}^N MARE_i$$

The procedure of the presented algorithm is shown in Figure 2, Figure 3 and Table 2. Based on these results, it can be said that using BCO algorithm improves the accuracy of the estimated cost of the projects.

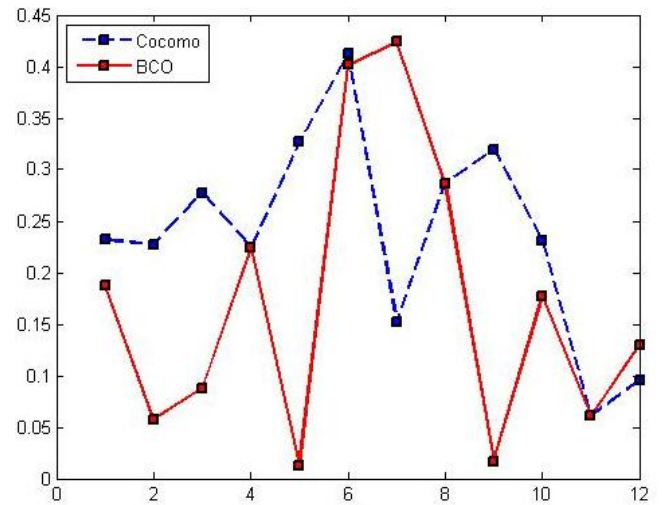


Figure 2: compare purpose method with COCOMO model on test dataset

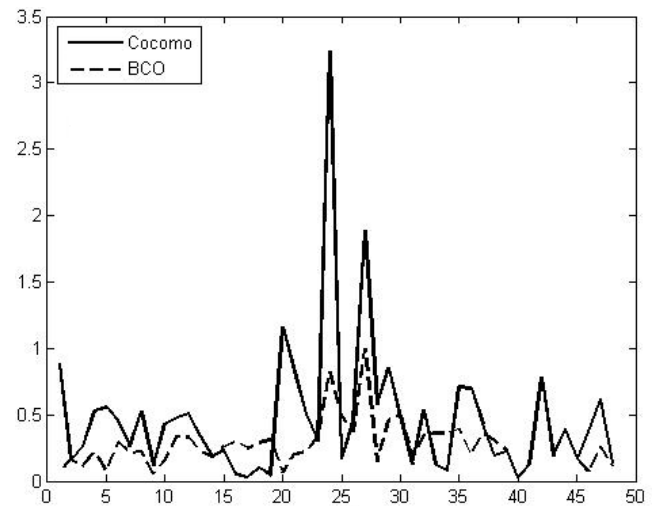


Figure 3: compare purpose method with COCOMO model on training dataset

Figure 2 shows the comparison between the proposed method and COCOMO. This comparison is done for the testing data and in Figure 3 comparison is done for the training data. The results show improved accuracy of the estimated cost in most of the training and testing data sets.

Table 2: compare purpose method with COCOMO model on MMARE

Model	Mean Absolute Relative Error
COCOMO	0.2371
Purpose Method	0.1619

Table 2 shows the MMARE of COCOMO and the proposed method. Based on these results it can be said that with the new method MMARE has decreased 0.0752 percent.

## **6. CONCLUSION AND FUTURE WORKS**

Today with the increasing complexity of software systems and following it the increasing cost of these systems, accurate estimation of software systems in the early stages of system production is considered essential. But a method capable of predicting the costs with 100 percent accuracy has not yet been provided. This problem is still a challenge for the researchers in the field of software engineering. Accordingly, this paper seeks to increase the accuracy of the estimated cost using bee colony optimization algorithm on the basis of intermediate COCOMO.

In this paper NASA data set projects have been classified according to the project types defined in COCOMO and then the values of fix parameters has been calibrated by BCO algorithm. The results indicate an increase in the accuracy of estimated costs.

The calculated MMARE for COCOMO is 0.2371 percent and it is 0.1619 percent for the proposed method. These numbers show a 0.0762 percent decline in MMARE. Given the favorable results, the proposed method can be used for solving other problems like process scheduling, automatic text summarization and so on.

## **7. REFERENCES**

- [1] J.Caper, "Estimating Software Costs (English) 2nd Edition", Tata McGraw - Hill Education, 2007.
- [2] Z.A. Khalifelu, F.S. Gharehchopogh, "A New Approach in Software Cost Estimation Using Regression Based Classifier ", *AWER Procedia Information Technology & Computer Science*, Vol:2, pp. 252-256, December 2012.
- [3] H. Park, S. Baek, "An Empirical Validation of a Neural Network Model for Software Effort Estimation", *Expert Systems with Applications*, Vol. 35, pp. 929-937, 2008.
- [4] Z.A. Khalifelu, F.S. Gharehchopogh, "Comparison and Evaluation Data Mining Techniques with Algorithmic Models in Software Cost Estimation", *Elsevier Press, Procedia-Technology Journal*, ISSN: 2212-0173, Vol: 1, pp. 65-71, 2012.
- [5] L. H. Putnam, "A general empirical solution to the macro software sizing and estimating problem", *IEEE Trans. Soft. Eng.*, pp. 345-361, 1978.
- [6] B. W. Boehm, "Software engineering economics", Englewood Cliffs, NJ: Prentice-Hall, 1981.
- [7] B.Boehm, W. Royce , "Ada COCOMO and the Ada Process Model", *Proceedings, Fifth COCOMO Users Group Meeting, Software Engineering Institute, Pittsburgh, , 1989*
- [8] B.Clark, E.Horowitz, R.Madachy, R.Shelby, Ch.Westland, "The COCOMO® 2.0 Software Cost Estimation Model", *International Society of Parametric Analysts*, 1995.
- [9] A.F.Sheta, "Estimation of the COCOMO Model Parameters Using Genetic Algorithms for NASA Software Projects", *Journal of Computer Science*, vol.2, pp. 118-123, 2006
- [10] T.R.Benala, S.Dehuri, S.Ch.Satapathy, S. Madhurakshara, "Genetic Algorithm for Optimizing Functional Link Artificial Neural Network Based Software Cost Estimation", *Springer-Verlag Berlin Heidelberg* , pp: 75–82,2012.
- [11] Z.A.Dizaji , K.Khalilpour, "Particle Swarm Optimization Ana Chaos Theory Based Approach for Software Cost Estimation", *International Journal of Academic Research*, Vol. 6, No. 3,pp.130-135, 2014
- [12] F.S.Gharehchopogh, L.Ebrahimi, I.Maleki and S.J.Gourabi, "A Novel PSO based Approach with Hybrid of Fuzzy C-Means and Learning Automata in Software Cost Estimation", *Indian Journal of Science and Technology*, Vol 7(6), pp.795–803, 2014
- [13] F.S. Gharehchopogh, "Neural Networks Application in Software Cost Estimation: A Case Study", 2011 *International Symposium on Innovations in Intelligent Systems and Applications (INISTA 2011)*, pp. 69-73, IEEE, Istanbul, Turkey, 15-18 June 2011.
- [14] T.Menzies, D.Port, Z.Chen, J.Hihn, Sh.Stukes, "Validation Methods for Calibrating Software Effort Models", *International Conference on Software Engineering*, St Louis Missouri, USA, 15-21 May, 2005.
- [15] P.Lucic, D .Teodorovic, "Computing with bees: attacking complex transportation engineering problems", *International Journal on Artificial Intelligence Tools* , vol:12, pp:375–94 , 2003
- [16] M .Nikolic, D.Teodorovic, "Empirical study of the Bee Colony Optimization (BCO) algorithm", *Expert Systems with Applications* 40, pp: 4609–4620, , 2013