ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

A Review on Genetic algorithm to deal with Optimization of Parameters of Constructive Cost Model

Sonam Bhatia[1] , Aanchal Bawa[2] , Varinder Kaur Attri [3]

Master of Technology, Department of Computer Science and Engineering, GNDU RC, Jalandhar, India[1,2]

Assistance Professor, Department of Computer Science and Engineering, GNDU RC, Jalandhar, India[3]

Abstract: Software cost estimation is one of the most challenging issues in software project management. To produce the accurate estimation, many models have been developed, but no model has efficient with the uncertainty of the project development. Most of these models are based on the size measure, such as Lines of Code (LOC) and Function Point (FP). Size estimation accuracy directly effect on cost estimation accuracy. The COCOMO model is the most important model for Software Cost Estimation. Today's effort estimation models are based on soft computing techniques as, genetic algorithm, the fuzzy logic, neural network etc for finding the accurate predictive software development effort and time estimation. The aim of this paper is to optimize the parameters of COCOMO Model with genetic algorithms.

Keywords: Software cost estimation, COCOMO model, genetic algorithm, Variance Account For

I. INTRODUCTION

An important part of software development is Software cost estimation. [4] In the software development processes, Software cost and effort estimation is one of the most critical and complex, but also it is an inevitable activity. [1] Accurate cost estimates are critical to both customers and developers. To predict effort required for software development, many software cost estimation methods available for software developers like expert based methods, complex algorithmic modeling methods and the analogy-based methods [2]. Machine learning method is used to group together a set of techniques that represent some of the facts of human mind for example regression trees, fuzzy systems, genetic algorithms, Bayesian networks, artificial neural networks. The approaches are classified as soft computing group. [5]

The cost estimation relies on the accurate prediction of the size.[3] The reasons that software cost estimation is difficult and error prone include human bias and Experience is required for developing estimates, especially for large projects. In 1981, COCOMO was introduced by Boehm. This model gives mathematical equations that identify the developed time, the effort and also the maintenance effort. Based on 63 software project, the model was developed for cost estimation. The estimation accuracy is improved when adopting models such as the Intermediate and Complex COCOMO models. [6] Therefore, the aim of this research is to optimize the model coefficients [7]

In evolutionary computing techniques, Genetic algorithms are optimization algorithms that proposed in 1975 by a scientist Holland. Genetic algorithm is a natural heuristic algorithm that is used to search the exact and approximate solutions. This is based on the iterative improvement of the current solution, but a solution set is used instead of one solution. Most genetic algorithm applications are linked to the development of prediction models and large-scale information processing. [8]

The remainder of this paper is organized as follows. Section II, describes the COCOMO Model. And Genetic Algorithm .COCOMO Model parameters optimization using genetic algorithms is presented in Section III. In Section IV, overviews related work. In Section V, Open issues and future scope are presented. In the end the conclusion is made in Section VI.

II. BACKGROUND

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

A. COCOMO MODEL

It is an algorithmic software cost estimation model developed by Barry W. Boehm. This model uses a basic regression formula with metrics that are derived from historical project data and current as well as future project traits. Boehm has given three levels of the model: basic, intermediate, detailed. [9] It describe relationship between the size of the project, development cost (pm), and the EAF. Effort and development time is given by

$E = a.KLOC^b$

a, b = COCOMO model coefficients; KLOC = kilo-lines of code;E = t effort (man-months).

$TDEV = c.E^d$

c, d = COCOMO model coefficient

These coefficients a, b and c depend on the mode of the development.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

TABLE1:THREE MODES OF DEVELOPMENT [9].

Development Mode	Basic Effort Equation	Time Duration (D)
Organic	Effort=2.4KLOC$^{1.0}$ ^{5}PM	D=2.5*(Effort) $^{0.38}$M
Semi-detached	Effort=3.0KLOC$^{1.1}$ ^{2}PM	D=2.5*(Effort) $^{0.35}$M
Embedded	Effort=3.6KLOC$^{1.2}$ ^{0}PM	D=2.5*(Effort) $^{0.32}$M

B. GENETIC ALGORITHM

These algorithms are adaptive heuristic search algorithms based on the Darwin theory of natural selection .They came into existence by John Holland and extensively studied by Goldberg De Jong [12]. Algorithm is based on the iterative improvement of the current solution, but a solution set is used instead of one solution. It is an optimization technique used in researching the optimal solution randomly [11]. Genetic Algorithm is one of the evolutionary techniques for the effort estimation. [13].Genetic algorithm depends on the representation of individuals, objective function, and genetic operators. It includes a population of encoded solutions to a given problem through generations using genetic operations. The individuals at each step, called a generation, are calculated and a score is given to each one, called its fitness value that show how good the solution it represents is. [16]Genetic algorithm consists four main components:

Chromosome: It is an individual representing one of task solutions. And the line of numbers that could be encoded using the binary encoding, integer number encoding etc. Each position in chromosome is called a bit or gene. [11][14]

Initial population: The first population is a set of task solutions that is generated randomly. The main condition of the generation process of the first population is to get a variety of sets of solutions.

Operator set: Operator set allows producing new solutions on the base of current population. Operator set includes selection, crossover and mutation.

a) Selection: In selection, individuals are selected in the intermediate population. Different types of selection are known: Roulette wheel selection and Tournament selection. [10][11]

b) Crossover: It involves a mating of two randomly selected strings. This process result is the generation of new individuals. The simplest crossover is n type crossover and can change by another crossover type. [11]

c) Mutation: The use of the mutation chromosome gene with defined probability interchanges its value. The new value of gene is also evaluated with defined probability. The uniform single bit mutation function is used, in which a bit has a probability to change it states from 0 to 1 or from 1 to 0[10].

Fitness function: It is the individual estimation attribute. An objective function is used that decide the fitness of an individual. The individual having large fitness value is supposed to be more stable and close to the solution. [10] [11][12] .

C. Evolutionary Process

The evolutionary process starts by the computation of the fitness of each individual in the initial population [15] In Fitness function the evaluation criterion to measure the performance of the developed GA based models is selected to be the Variance-Accounted-For (VAF).The VAF is calculated as:

[1 - var (Effort –Estimated Effort)/ var (Effort)] × 100%

The variance is calculated as:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i)^2 \quad - \quad \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2 \qquad (1)$$

Here, x is the variable and n is the number of values of that variable.

MMRE and PRED are evaluated from the relative error, which is the relative size of the difference between the actual and estimated value of individual effort i :

REi = (Estimated Effort$_i$ – Actual Effort$_i$) /Actual Effort$_i$)

The magnitude of relative error can be calculated by taking the absolute value of that relative error that is,

MREi = abs (RE$_i$)

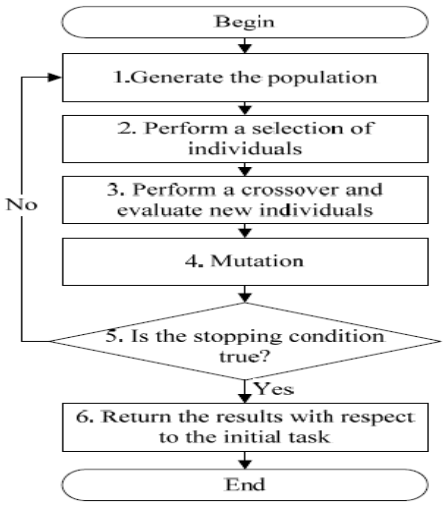

Fig. 1.The steps of Basic genetic algorithm [11]

III. COCOMO MODEL PARAMETERS OPTIMIZATION USING GENETIC ALGORITHM

It is used to estimate the COCOMO model parameters. The main steps of algorithm are described following:

1) Individuals are created randomly and a specific chromosome structure is chosen.

E.g. x x x x | x x x x | x x x x | x x x x

a b c d

Each of above four parameters is represented by 4 genes.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

2) Evaluate predicted development Effort for each Project j in the dataset using coefficients from an Individual i
3) Evaluate Individual i fitness for the Project j, the equation used is:

$$\text{Fitness}_{ij} = \frac{\text{TDEV}^{\,j}_{R} - \text{TDEV}^{\,ij}_{P}}{\text{TDEV}^{\,j}_{R}} \quad (2)$$

Where i = the individual number; j =the project number

4) Individual fitness is calculated as the average value of all Project specific fitness values of an Individual achieved during steps 2 , 3. The fitness value depends on the difference between real development effort and predicted development effort.

$$\text{Fitness}_i = \frac{1}{n} \cdot \sum_{j=1}^{n} \text{Fitness}_{ij} \quad (3)$$

5) The stopping condition describes when the algorithm must be finished. The iteration number is describes as a stopping condition.
6) The selection method, roulette wheel selection is used to form a set of individuals.
7) Using individuals, selected in a previous step, new individuals are produced.
8) None or a small number of Individuals is randomly selected in mutation. The probability to be selected the individual should be low and random scheme of mutating genes is defined for each Individual.
9) The new population is generated and using similar selection method applied to step 6, the best Individuals among parents and children are chosen.

Genetic algorithms versus conventional search algorithms:
One of the major advantages of Genetic Algorithms is that it operates on a population of solutions not only a single point as compared to conventional search algorithms .It gives more accurate result.. The solution give by this algorithm is more optimal and global in nature. This algorithm does not involve derivative information about the fitness criterion .Therefore it is perfect for continuous and discrete optimization problems. There are some traits that make genetic algorithms different from conventional search algorithms. Goldberg stated that:

- Genetic Algorithms implement the search using a coded solution not the solutions themselves.
- Genetic Algorithms is not just a single solution but based on a population of candidate solutions.
- This algorithm evaluate individual based on their fitness function not the derivative of the Function
- Probabilistic operators are used by Genetic Algorithms.

IV. RELATED WORK

Alaa F. Sheta[2006] described two new model structures to estimate the software effort for projects. Modified versions of the famous COCOMO model were provided to consider the effect of methodology in effort estimation. This model structure defined the effect of ME as linearly related to the effort and extending the model by adding a new bias parameter to the above model and re-estimate the new model parameters, model 2, by Genetic algorithm.[15]

Anna Galinina, Olga Burceva, Sergei Parshutin [2012] presented a paper and explained current research proposes a genetic algorithm based method for optimization of the COCOMO model coefficients for organic and semi-detached modes. By compairing COCOMO model modes, it can be defined that use of the coefficients optimized by the Gentic algorithm in the organic mode produces better results in comparison with the results achieved using the current COCOMO model coefficients and at the same time, coefficients for the semi-detached mode produced by the proposed algorithm do not result in the high accuracy.[11]

Brajesh Kumar Singh ,A. K. Misra[2012] presented a paper and described model structure is proposed to estimate the software effort for projects sponsored by NASA using binary genetic algorithm and it can be examined that taking into consideration the effect of ME and adding new bias d help to improve the computed VAF. Proposed model provides around 40% improvement in performance.[12]

Astha Dhiman1, Chander Diwaker[2013] presented a Paper concluded that having the appropriate statistical data describing the software development projects, Gentic algorithm based coefficients can be used to generate better results in comparison with the results achieved using the current COCOMO II PA model coefficients and also shows that the results achieved using the coefficients optimized with the propose algorithm are less than the real effort values[13]

Pushpendra K Rajput, Geeta Sikka, and Aarti [2014] explained clustered based Genetic Algorithm model that estimates the cost of software project. The previous developed projects are classified into cluster and the optimized parameters are obtained using GA algorithm. Neural network is used to classify the new project that is supposed to be estimated and also explained that the hybrid technique has a significant improvement over the COCOMO model.[10]

Vishali Anshu Sharma Suchika Malik[2014] explained that genetic algorithms and ant colony optimization can be used to optimize the COCOMO model coefficients. The goal of this research was to optimize the COCOMO model coefficients with the help of the genetic algorithms and ant colony optimization[14]

Efi Papatheocharous, Harris Papadopoulos and Andreas S. Andreou presented a paper and explained the use of the dual variables Ridge Regression (RR) technique with a Genetic Algorithm (GA) to evolve the selection of cost drivers used as inputs for the estimation of software effort. The results achieved from the evaluation phase of this experiments show that the proposed algorithm did not only reduce the number of cost drivers used to less than half,

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

but it provides improvement to the performance of the dual variables RR technique. [16]

V. OPEN ISSUES & FUTURE SCOPE

- Other optimization algorithms may be more efficient or effective than genetic algorithms in terms of speed of convergence for specific optimization problems.
- Genetic algorithms do not scale well with complexity. That is, where the number of elements which are exposed to mutation is large there is often an exponential increase in search space size. The second problem of complexity is the issue of how to protect parts that have evolved to represent good solutions from further destructive mutation, particularly when their fitness assessment requires them to combine well with other parts.

This survey indicates directions for further research. Trying to enhance the performance of existing methods and familiarize the new methods for estimation based on today's software project requirements can be future works in this area. So the research is on the way to integrate different techniques or methods for evaluating the best estimate

VI. CONCLUSION

It has been a assert task to conveniently estimate the cost and effort of projects in its early phases of software development. For finding the accurate predictive software development effort and time estimation ,today's Software development effort estimation models are based on soft computing techniques as neural network, genetic algorithm, the fuzzy logic modeling ,ACO ,ridge regression etc. Genetic Algorithm can provide some significant enhancement in accuracy and has the potential to be a valid additional tool for software effort estimation. Genetic algorithm has been used for difficult numerical optimization problems. They have been successfully used to solve system identification, signal processing and path searching problems.3

ACKNOWLEDGMENT

We would like to thank acknowledge almighty for his constant blessings. Then we like to thank our family and friends for helping and supporting us throughout the making of this paper.

REFERENCES

[1] Saleem Basha, Dhavachelvan P "Analysis of Empirical Software Effort Estimation Models" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 3, 2010.
[2] Jyoti G. Borade " Software Project Effort and Cost Estimation Techniques" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 8, August 2013.
[3] Vahid Khatibi, Dayang N. A. Jawawi "Software Cost Estimation Methods: A Review Journal of Emerging Trends in Computing and Information Sciences" Volume 2 No. 1 ISSN 2079-8407 2011.
[4] Stein Grimstad Magne ,Jørgensen "A Framework for the Analysis of Software Cost Estimation Accuracy".
[5] Iman Attarzadeh, Siew Hock Ow "Improving Estimation Accuracy of the COCOMO II Using an Adaptive Fuzzy Logic Model"IEEE International Conference on Fuzzy Systems 2011.
[6] Alaa F. Sheta and Sultan Aljahdali "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, 2013
[7] X. Huang, D. Ho, J. Ren, L.F. Capretz, Improving the COCOMO model using a neuro-fuzzy approach. Vol. 7, Applied Soft Computing, 2007
[8] J. Horn, N. Nafpliotis and E. Goldberg, "A niched Pareto genetic algorithm for multiobjective optimization." Orlando, USA: IEEE; 1994,pp. 82-87.
[9] Nancy Merlo – Schett "COCOMO (Constructive Cost Model)Seminar onSoftware Cost Estimation" WS 2002 / 2003.
[10] Pushpendra K Rajput, Geeta Sikka, and Aarti "CGANN-Clustered Genetic Algorithm with Neural Network for Software Cost Estimation" International Conference on Advances in Engineering and Technology (ICAET'2014) March 29-30, 2014 Singapore.
[11] Anna Galinina1, Olga Burceva2, Sergei Parshutin3, 1-3Riga Technical University "The Optimization of COCOMO Model CoefficientsUsing Genetic Algorithms"2015.
[12]] Brajesh Kumar Singh , A. K. Misra Software Effort Estimation by Genetic Algorithm Tuned Parameters of Modified Constructive Cost Model for NASA Software Projects".
[13] Astha Dhiman1, Chander Diwaker2 "Optimization of COCOMO II Effort Estimation using Genetic Algorithm" American International Journal of Research in Science, Technology, Engineering & Mathematics, 3(2), June-August, 2013.
[14] Vishali * Anshu Sharma Suchika Malik " COCOMO model Coefficients Optimization Using GA and ACO" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 1, January 2014.
[15] Alaa F. Sheta " Estimation of the COCOMO Model Parameters Using Genetic Algorithms for NASA Software Projects" Journal of Computer Science 2 (2): 118-123, 2006.
[16] Efi Papatheocharous1, Harris Papadopoulos2 and Andreas S. Andreou3, "Software Effort Estimation with Ridge Regression and Evolutionary Attribute Selection".
[17] Dr. Devesh Kumar Srivastava "Measure Performance of VRS Model using Simulation Approach by Comparing COCOMO Intermediate Model in Software Engineering" International Journal of Computer Science & Engineering Technology (IJCSET)2013.

BIOGRAPHY

Sonam Bhatia received B.Tech degree in Information Technology from Punjab Technical University, Jalandhar, India, in 2013, pursuing M.tech in Computer Science Engineering from Guru Nanak Dev University, Amritsar, India. Her research area includes software engineering.