9

Estimation of the Effort Component of the Software Projects Using Heuristic Algorithms

Mitat Uysal
Dogus University, Turkey

Abstract

In this study, a multivariate interpolation model was developed to estimate the effort component of the software projects. A COCOMO based equation was used to represent the effort function. The data set that was used consists of two independent variables, first is Lines of Code (LOC) and second is Methodology (ME) and one dependent variable Effort (CE). Data set is taken from NASA projects and the results that are obtained in this work are compared with the results of A.F.Sheta who is produced a similar model for estimating the effort component of software projects.
In this paper, it has been shown that Simulated Annealing algorithm can be used to estimate the optimal parameters of the effort components of software projects. The upper and lower bounds of the search space should be considerately given by designer or be cited from other reference papers, if possible.
***Keywords*—** Curve fitting, heuristic optimization, software cost estimation, software engineering.

1. Introduction

The goal of an optimization method is to find the optimum solution of a given problem. If the goal function is a cost function , then optimum solution can be obtained by finding the minimum value of the function.
An optimization process proceeds with two important actions:

a) Search
b) Determining the improvement.

The goal function represents the objective in terms of the parameters and variables involved in the optimization problem.
Several search methods are used in the related optimization studies.
The purpose of this paper is to compare three of these search methods with respect to their applicability to the prediction of the effort component of the software projects.
There are three main classes of search techniques for problems of combinatorial optimization:

1. Enumerative techniques
2. Implicit enumerative techniques
3. Calculus-based techniques

Implicit enumerative techniques can be separated into two groups:
1. Guided search techniques
2. Neural networks

Guided search techniques consist of three algorithms:
1. Simulated annealing
2. Evolutionary algorithms (evolutionary strategies , genetic algorithms)
3. Tabu search (Affenzeller et al.)

Particle swarm optimization(PSO) is a new method in evolutionary computation that is similar to genetic algorithm in that the system is initialized with a population of random solutions.

In this study,simulated annealing algorithm is used as heuristic optimization algorithm.

The importance of software cost estimation is well documented. Good estimation techniques serve as a basis for communication between software personnel and non-software personnel such as managers sales people or even customers (Knafe , 1995).

Resource models consist of one or more empirically derived equations that predict effort (in person-months) , project duration (in chronological months) or the other pertinent project data.

Basili (1980) described four classes of resources models:

- Static single – variable models
- Static multi – variable models
- Dynamic multi – variable models
- Theoretical models

The static single – variable model takes the form:

Resource = P_1 * (Estimated Characteristics) ^{P2}

Where the resources could be effort , project duration , staff size or requisite lines of software documentation. The parameters P1 and P2 are derived from data collected from past projects. The basic version of the Constructive Cost Model or COCOMO is an example of a static single variable model.

Static multi – variable model has the following form :

Resources = P11 e1 + P21 e2 + ------ + Pn1 en

Where e i is the i'th software characteristics and P11 , P21 are empirically derived optimal parameters for the i'th characteristics (Pressmann , 1992).

A dynamic multi – variable model projects resource requirements as a function of time.

A theoretically approach to dynamic multi – variable modeling hypothesizes a continuous resource expenditure curve and from it , derives equations that model the behaviour of the resource. The Putnam estimation model is a theoretical dynamic multi – variable model.

Some new models are proposed for software cost estimation. One of them is Peters and Ramanna Model based on an application of the Choquet integral (Peters and Rommano , 1996).

Fuzzy logic and neural networks are the other tools to develop software cost estimation models.

In the recent studies , evolutionary algorithms and genetic algorithms are widely used to estimate the optimal parameters of the software cost models. Typical examples are (Sheta , 2006) and (Huang and Chiu , 2006).

Dillibabu and Krish-naiah have developed an effort estimation model using COCOMO II. 2000 reference.
Burgess and Lefley have concluded that Genetic Programming can offer significant improvements in accuracy of effort parameters but this depends on the measure and interpretation of accuracy used.
Shin and Goel described a detailed Radial Basis Function modeling study for software cost estimation using well-known NASA dataset.
Mantere and Alander have required the work applying computational evolutionary methods in software engineering.
Kaczmarek and Kucharski have presented a methodology for estimation of software size and effort at early stages of software development.
(Uysal , 2006) has developed a multivariate interpolation model to estimate the effort component using Lagrange interpolation model.

2. Effort Estimation Model That Is Used In This Study

Effort estimation can be used for several purposes. One of them is to use for project management purposes. Effort estimation methods can be classified in many groups.
The following software effort model is used in the present study:

$$E = f\ (LOC, ME)$$

Where E is effort, LOC is the number of lines of the developed Code and ME is methodology used in the software project.
f is a nonlinear function in terms of LOC and ME.
We present two different functions for f :

I. First function for f is expressed as below:

$$E = a.LOC^{b} + c.ME^{d} + e \tag{1}$$

The presented model contains five parameters a, b, c, d and e. This model is slightly different than the model that is proposed in (Sheta , 2006).

II. Second function for f is expressed as below :

$$E = a.LOC^{b} + c.ME^{d} + e.\ln(ME) + f.\ln(LOC) + g \tag{2}$$

Above presented model is original and firstly proposed in this study.
Model contains 7 parameters; they are a, b, c, d, e, f and g.

3. Solution Method

In multidimensional global optimization ,the main goal is to find a global optimum of an objective function defined in a given space. Deterministic and heuristic methods are used to solve multidimensional optimization problem.In some cases,hybrid techniques can be used to.

The following solution method is used to find optimal values of the model parameters:

$$\text{Minimize} \quad \sum_{i=1}^{n} (E_{meas} - E_{comp})^2$$

Where E_{meas}, is measured value of effort, E_{comp} is computed value of effort according to the model used.
In order to minimize the total squared error given above, simulated annealing algorithm is used changing the parameter values of the model.

4. Evolutionary Computation and Simulated Annealing Algorithms

Evolutionary algorithms, simulated annealing and tabu search are widely used heuristic algorithms for combinatorial optimization.
The term evolutionary algorithm is used to refer to any probabilistic algorithm whose design is inspired by evolutionary mechanisms found in biological species (Youssef et al., 2000).
One of the most widely known of heuristic algorithms is simulated annealing (SA) algorithm. SA exploits an analogy between the way in which a metal cools and freezes into a minimum energy crystalline structure (the annealing process) and the search for a minimum in a more general system(Xianghua et al.,2007).
Simulated annealing is a generalization of a Monte Carlo method for examining the equations of state and frozen states of n-body systems that is found by Metropolis et al. in 1953.
In the optimization process, the solution randomly walks in its neighborhood with a probability determined by Metropolis principle while the system temperature decreases slowly; when the annealing temperature is closing zero, the solution stays at the global best solution in a high probability.(Xianghua et al.,2007)
An original MATLAB code is developed for simulated annealing algorithm in this work. This code is used to estimate the optimal values of the parameters of model 1 and model 2.

5. Results

Optimization algorithm have been applied on NASA software project data like (Shin and Goel , 2000) and (Sheta , 2006).
The data set consist of two independent variables , Lines Of Code (LOC) and the Methodology (ME) and one dependent variable , effort. LOC is described in Kilo Line of Code and effort is in man-months. Data set is given in Table 1.

Project No	KDLOC	ME	Measured Effort
1	90.2000	30.0000	115.8000
2	46.2000	20.0000	96.0000
3	46.5000	19.0000	79.0000
4	54.5000	20.0000	90.8000
5	31.1000	35.0000	39.6000
6	57.5000	29.0000	98.4000
7	12.8000	26.0000	18.9000
8	10.5000	34.0000	10.3000
9	21.5000	31.0000	28.5000
10	3.1000	26.0000	7.0000
11	4.2000	19.0000	9.0000
12	7.8000	31.0000	7.3000
13	2.1000	28.0000	5.0000
14	5.0000	29.0000	8.4000
15	78.6000	35.0000	98.7000
16	9.7000	27.0000	15.6000
17	12.5000	27.0000	23.9000
18	100.8000	34.0000	138.3000

Table 1. NASA software project data Measured Effort

Figure 1 shows same data in 3D space.

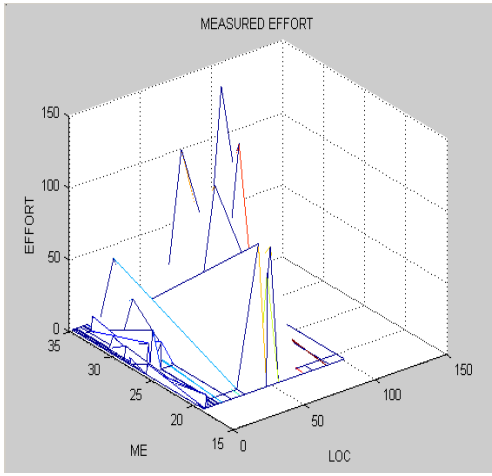

Fig. 1. NASA Software Projects Data in 3D.

The optimal values of parameters in the first function (1) were estimated using Simulated Annealing algorithm as below:

a=3.3275 b=0.8202 c= - 0.0874 d= 1.6840 e=18.0550

So, function can be written as follows :

$E=3.3275 \cdot LOC^{0.8202} - 0.0874 \cdot ME^{1.6840} + 18.0550$

The required iteration number is 1200.

Figure 2 shows the measured data (NASA projects data) and the predicted values obtained according to the function given above.

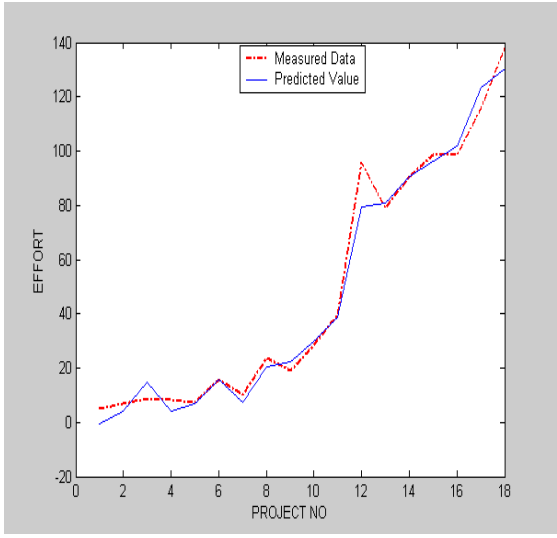

Fig. 2. Measured Data and Predicted Values according to the first model

Applying the two variables (LOC and ME) mentioned above , the effort model surface as a function of LOC and ME was obtained as shown in Figure 3.

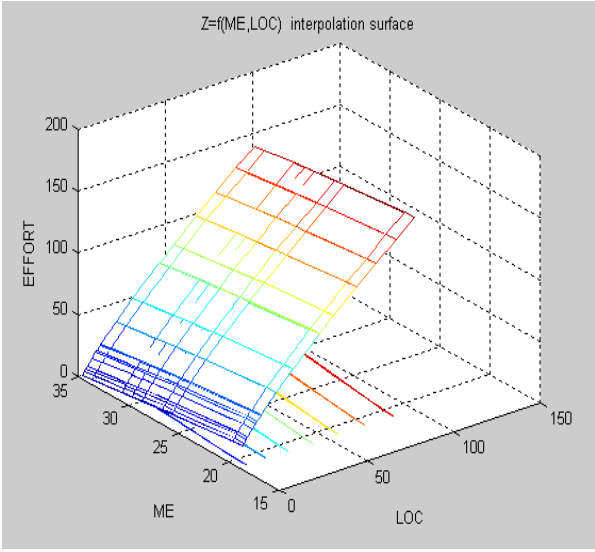

Fig. 3. Surface of Effort E=f(LOC,ME)

The optimal values of parameters in the second function (2) were estimated using Simulated Annealing Algorithm as below:

a=3.8930 b=0.7923 c= - 0.2984 d= 1.3863 e=2.8935
f= -1.2346 g=15.5338

Second function can be written using these optimal values as follows:

$$E=3.8930 \ .LOC^{0.7923} \ - 0.2984 \ .ME^{1.3863} + 2.8935.\ln(ME) \ -1.2346 \ .\ln(LOC) + 15.5338$$

The required iteration number for the second model is 1910.
For the optimal solutions , the total squared error in the second function is less than the total squared error in the first function.
Figure 4 shows the measured data and the predicted values obtained according to function given above.

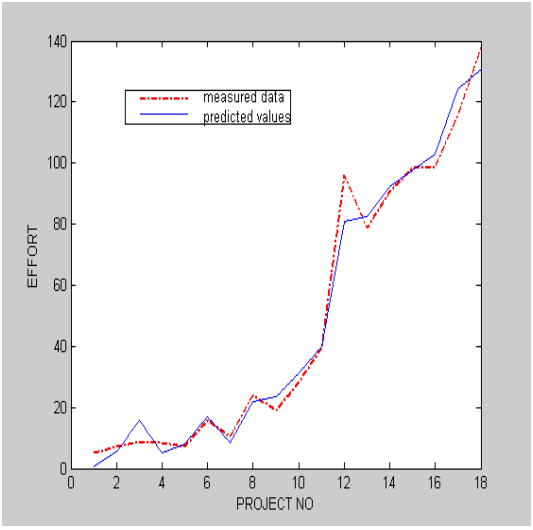

Fig. 4. Measured Data and Predicted values according to the second model.

6. Conclusion and Future Work

In this paper, it has been shown that Simulated Annealing algorithm can be used to estimate the optimal parameters of the effort components of software projects. The upper and lower bounds of the search space should be considerately given by designer or be cited from other reference papers, if possible. Generally speaking, if a larger search space is built, it would be more time of computations and convergence of search may become very slow. Conversely, if the search space is set too small, the optimal parameters probably could not been found.
A new model has been proposed (model 2) to estimate the software effort.
It can be seen that this new model provides better results than the previous studies.
The effectiveness of SA's tend to depend on implementation details, how the problem is encoded etc. (Mantere et.al, 2005).

All three heuristic search algorithms(GeneticAlgorithms,TabuSearch,SimulatedAnnealing) work well and produce an acceptable sub-optimal solution within a reasonable amount of time.
The SA algorithm outperforms the TS and GA algorithms in all the simulation runs.
In the future work, a GUI will be developed for the two variables estimation model.

7. References

Basili, V, "Models and Metrics for Software Management and Engineering", IEEE Computer Society Press, 1980

Burgess, C.J., Lefley, M., "Can Genetic Programming Improve Software Effort Eestimation? A Comperative Evaluation", Information and Software Technology 43, 863-873, 2001

Cunha,M.C.,Ribeiro,L.,"Tabu search algorithms for water network optimization",Europ.J. Operational Res. 157,2004,746-758

Dillibabu, R., Krashnaiah, K., "Cost Estimation of a Software Product Using COCOMO II.2000 Model – a case study", Int. Journal of Project Management 23, 297-307, 2005

Houch, C., Joines, J., Kay, M.G., "A Genetic Algorithm for Function Optimization: A Matlab Implementation", ACM Trans. On Math. Software, 1996

Huang, S.J., Chiu, N.H., "Optimization of Analogy Weights by Genetic Algorithm for Software Effort Estimation", Information and Software Tech. 48, 1034-1045, 2006

Kaczmarek, J., Kucarski, M.,"Size and Effort Estimation for Applications Written in Java", Information and Software Tech. 46, 589-601, 2004

Knafl, G.J., Morgan, J.A., Follenweider, R.L., Korcich, R.M., "Software failure data analysis using the least squares approach and the time per failure concept", Int. J. Reliability, Quality, Safety Eng., 2, 161-175

Mantere, T., Alender, J.T., "Evolutionary Software Engineering, A review", Appl. Soft. Computing 5, 315-331, 2005

Peters, J.F., Ramanna, S., "Application of the Choquet Integral in Software Cost Estimation", IEEE, 2, 862-866, 1996

Pressman, R.S., Software Engineering: A Practitioner's Approach, The McGraw Hill, 1992

Sheta, A.F., "Estimation of the COCOMO Model Parameters Using Genetic Algorithms for NASA Software Project", Journal of Computer Science 2 (2), 118-123, 2006

Shin, M., Goel, A.L., "Empirical Data Modeling in Software Engineering Using Radial Basis Functions", IEEE Trans. and Software Eng. Vol.26 No.6, June 2000

Uysal, M., "Multivariate Interpolation Model to Estimate the Effect Component of Software Project", Inf. Tech. Journal 5 (6), 1143-1145, 2006

Xianghua,X.,Xingang,L.,Jianxun,R.,"Optimization of heat conduction using combinatorial optimization algorithms",Int.J. of Heat and Mass Transfer,50(2007) 1675-1682

Youssef, H., Sait, S.M., Adiche, H., "Evolutionary Algorithms, Simulated Annealing and Tabu Search: A Comparative Study", Eng. Appl. of Artificial Int. 14, 167-181, 2001

New Trends in Technologies

Edited by Blandna ramov

ISBN 978-953-7619-62-6
Hard cover, 242 pages
Publisher InTech
Published online 01, January, 2010
Published in print edition January, 2010

This book provides an overview of subjects in various fields of life. Authors solve current topics that present high methodical level. This book consists of 13 chapters and collects original and innovative research studies.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Mitat Uysal (2010). Estimation of the Effort Component of the Software Projects Using Heuristic Algorithms, New Trends in Technologies, Blandna ramov (Ed.), ISBN: 978-953-7619-62-6, InTech, Available from: http://www.intechopen.com/books/new-trends-in-technologies/estimation-of-the-effort-component-of-the-software-projects-using-heuristic-algorithms

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821