A Framework for Analogy-based Software Cost Estimation using Multi-objective Genetic Algorithm

Sweta Kumari* and Shashank Pushkar

Abstract— **This paper presents a Multi-objective Genetic Algorithm (MOGA) based approach for project selection to improve the Analogy Based Estimation (ABE) system's performance. A multi-criteria project selection problem has been formulated that allows project interactions and incorporates the decision maker's preferences. The experimental results prove that the proposed approach improves the existing ABE process. The model has been experimented on two standard datasets (COCOMO 81 and COCOMONASA) and tested based on the criteria of Mean Magnitude of Relative Error (MMRE) and Prediction (PRED) for software cost estimation. The results show the suitability of the proposed method for improving the cost prediction using the ABE based estimation method. The paper also highlights that how interactive effects among projects change the cost prediction of the projects.**

Index Terms—**Multi-objective genetic algorithm, Multi-criteria decision making, Analogy based estimation, Genetic algorithm, Non-linear integer programming**

I. INTRODUCTION

The success of a software organization, mostly depends upon proper management activities such as planning, budgeting, scheduling, resource allocation and effort requirements for software projects. Software effort estimation is the process of making an approximate judgment of the costs of the software. Inaccurate cost estimation results problems such as project failure, budget overrun, delayed deliveries, etc. Software cost estimation methods can be divided into three main categories: Algorithmic method, Expert judgment and ABE method [28]. Algorithmic methods use a formula to calculate the software cost. Expert judgment [9] relies on expert experiences and understandings for estimating software cost. However, the accuracy of expert based prediction is found to be low. Estimation by analogy is a technique which is appropriate when previous projects in the same application domain are finished and their related past data are in place for use. Here, a critical role is played by the similarity measures among a pair of projects [8] [10] [13]. Here, the distance is calculated between the software project being estimated and each of the similar past projects. It then finds the most similar project to estimate the project cost.

Manuscript received Aug 28, 2015; revised May 29, 2016. This work was supported by UGC, India. (Grant no.-F./2015-16/NFO-2015-17-OBC-BIH-33244 (SA-III/Website)).

S. Kumari is with the Department of CSE, BIT, Mesra, Ranchi, INDIA (corresponding author e-mail : swetak44@gmail.com).

S. Pushkar is with the Department of CSE, BIT, Mesra, Ranchi, INDIA (corresponding author e-mail : shashank_pushkar@yahoo.com).

ABE process is basically a type of Case Based Reasoning [17]. However, as it is argued in [26] there are positive advantages in respect with rule based systems, for example the reality that users are eager to accept solutions from analogy based techniques, rather than solutions derived from uncomfortable chains of rules or neural nets. Naturally, there are some difficulties with this approach such as lack of appropriate analogues and issues with selecting and using them. Choosing an appropriate set of projects participating in the cost estimation process are very important for any organizations to achieve their goals. In this process, several reasons involved, such as the quantity of investment projects, the existence of various decision criteria for example value maximization and risk minimization and many management activities. Moreover, project selection is a difficult task if there are relations between projects based on various selection criteria and decision maker's preferences, mainly in the existence of a huge quantity of projects.

II. RELATED WORKS

Various methods and models for project selection have been introduced by many researchers. Lorie and Savage [15] uses a pioneering ranking method, whereas Nemhauser and Ullman [20] uses a dynamic programming for project selection. Rengarajan and Jagannathan [24] proposes a scoring method for the large R&D organization. This method is used to select and rank projects that is based on objectives and characteristics. Lockett *et. al.* [14] and Murahaldir *et. al.* [18] proposes a methodology for information system project selection using analytical hierarchy process. Tiryaki *et. al.* [27] and Ozkan *et. al.* [23] proposes a multi criteria decision making methodology using AHP in e-commerce, transport problem and portfolio Selection. Lee and Kim [12] uses an analytical network process and zero-one goal programming for information system project selection problem. Santhanam and Kyparisis [25] proposes a multiple criteria decision model for information system project selection using a non-linear 0-1 goal programming model. Aaker and Tyebjee [1] presents a R&D project selection model that provides interrelationships between three types of projects. Ghasemzadeh *et. al.* [6] proposes a zero-one integer linear programming model for the project portfolio selection that is based on the organization's objectives and constraints such as resource limitations and interdependency between projects. According to the literature review, it has been found that estimation by analogies requires a significant amount of computations as well as lack of appropriate analogous and issues with selecting and using them. And it has also been observed that interactive effects between projects are significant for project selection as it may give different results when interactive effects are considered [3]. In this work, we propose a MOGA based approach for project selection to improve the ABE system's performance. A multi-criteria project selection problem has been formulated for allowing project interactions and for incorporating the decision maker's preference information. Here, the performance of ABE system has been analyzed and compared in terms of MMRE and PRED (0.25). The rest of the paper is divided as follows: multi-criteria project selection problem is formulated in Sect. III. In Sect. IV, MOGA is used as the optimization technique for project selection problem for ABE. In Sect. V, Numerical examples are also given for illustration purpose. Experiments and comparison results are described in Sect. VI and in the end, the conclusion is discussed.

III. Formulation of the Problem

Often, decision-makers deal with the problem of selecting a subset of projects from a given large set. This selection is generally based on some criteria. This problem has been formulated using multi-criteria decision making (MCDM) process [11] [21]. In general, there are two types of MCDM problems: multi-attribute decision making (MADM) and multi-objective decision making (MODM). Based on this categorization, multi-criteria project selection problem can be seen as a distinctive MADM problem having the characteristics of project selection. In this paper N projects are considered for selection as well as evaluation, and decision variable x_n denotes whether the project is selected or not. Having no interactive effects between projects, the problem can be formulated as below:-

$$\text{Maximize } E = \sum_{n=1}^{N}\left(\sum_{j=1}^{J} p_j\, c_{nj}\right) x_n$$

s.t. $\sum_{n=1}^{N} x_n = R$

$$x_n = \{0,1\}, x_n = 1,2,\ldots,7 \qquad (1)$$

Whereas, the problem with interactive effects between projects can be formulated as below:-

$$\text{Maximize } E = \sum_{n=1}^{N}\left(\sum_{j=1}^{J} p_j c_{nj}\right) x_n + \sum_{j=1}^{J}\sum_{m=1}^{M}\left(p_j\left(b_j(c_m)\right)\left(\sum_{n=1}^{Q} c_{nj}\right)\right)\prod_{n=1}^{Q} x_n$$

s.t. $\sum_{n=1}^{N} x_n = R$

$$x_n = \{0,1\} \qquad (2)$$

Parameters

E = Total effects of selected projects.

R = Number of the selected projects based on criteria.

N = Number of projects to be evaluated and selected.

p_j = Preference degree of decision makers on criterion $j, j = 1, 2. . . J$.

c_{nj} = The value of projects n in criteria j.

$b_j(c_m)$ =Value of interactive effects in a combination of m projects in $j, j = 1, 2. . . J$.

c_m = Combination of m projects, $m = 1, 2. . . M$.

Q = Number of variables with interactive effects.

Decision variable

$x_n = \begin{cases}1\\0\end{cases} = 1$ if project n is selected 0 otherwise.

It can be observed that, equation (2) takes the form of a 0-1 nonlinear programming problem which is difficult to be solved using traditional optimization techniques like a branch and bound algorithm and other existing method [7]. To solve this kind of problem, we have incorporated the MOGA based approach. However, as equation (1) is a standard binary integer programming it may be solved using binary integer programming software's like CPLEX.

IV. MOGA-Based Optimization Approach

MOGA deals with solving an optimization problem which involves more than one objective, such as cost minimization and value maximization. Unlike the single objective optimization, it gives a set of optimal solutions. Here, the main task is to find out the Pareto-front, which gives a set of non-dominated solution points, known as Pareto-optimal or non-inferior solutions. None of the solutions in the non-dominated set is extremely better than other solutions and are selected based on preference criteria [22]. Mathematically, we can formulate a multi-objective problem as:-

Maximize/Minimize $y = f(x) = \{f_1(x), f_2(x), \ldots, f_m(x)\}$

Subject to

$g(x) = \{g_1(x), g_2(x), \ldots, g_j(x)\} \le 0$

$h(x) = \{h_1(x), h_2(x), \ldots, h_k(x)\} = 0$

Where, $x = \{x_1, x_2, \ldots, x_n\} \in X$

$$y = \{y_1, y_2, \ldots, y_m\} \in Y \qquad (3)$$

Here, x represents the vector of decision variable, y represents the objective vector, X represents the decision space and Y represents the objective space. In this work, we use a weighted sum approach [4] [19] to combine more than one objective into a single objective as shown below:-

$$f(x) = w_1 f_1(x) + w_2 f_2(x) + \cdots + w_m f_m(x) \qquad (4)$$

Where $f_1(x), f_2(x), \cdots, f_m(x)$ are the objective functions and $w_1, w_2, \cdots, w_m$ are the weights of equivalent objectives are normalized that convince the following conditions:-

$w_i \ge 0 \quad \forall_i = 1, 2, \cdots, m$

$w_1 + w_2 + \cdots + w_m = 1$

In this case, $F = w_1 f_1 + w_2 f_2$, where, $w_1 + w_2 = 1$ (5)

Here, we altered weights to obtain Non-dominated or Pareto-optimal set. Generally, weights are random numbers within $(0, 1)$. The steps of MOGA algorithm are presented in fig. 1 and the detailed process of the project selection for ABE using MOGA is explained below:-

i. Encoding: - Binary encoding is used here in which chromosome is represented by 0 or 1. Here, 0 denotes that the interrelated project from the past dataset is not selected and 1 denotes that it is selected.

ii. Generate population randomly: - Population is a group of all possible solutions (chromosome). It is randomly generated. A chromosome represents a solution in the form of a set of genes ($x = x_1, x_2, \cdots x_R$) if R variables exist.

iii. Fitness function: - It is used to obtain the best solution in the optimization process. Here, our main objective is to minimize the MMRE and the other is to maximize the PRED. To maximize PRED, we

ISBN: 978-988-14047-1-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

take the reciprocal of PRED. Convert the multiple objectives into single-objective by using equation (5), for every chromosome in the population. This is the final fitness function.

iv. Selection: - Roulette wheel selection operator has been used here for the selection of higher fitness chromosomes.

v. Crossover: - Heuristic crossover is applied in this work to produce new chromosomes.

vi. Mutation: - Adaptive feasible mutation is applied here.

vii. Elite Strategy: - It is used here to survive individuals automatically for the next generation.

viii. Replacement: - The new population replaces the current one.

ix. Finding and Updating Non-dominated Solutions: - Find the non-dominated solutions in the current population and update the previous non-dominated solutions with the current ones.

x. Stopping criteria: - If the maximum number of generations reached, then stop, otherwise go to step iii.

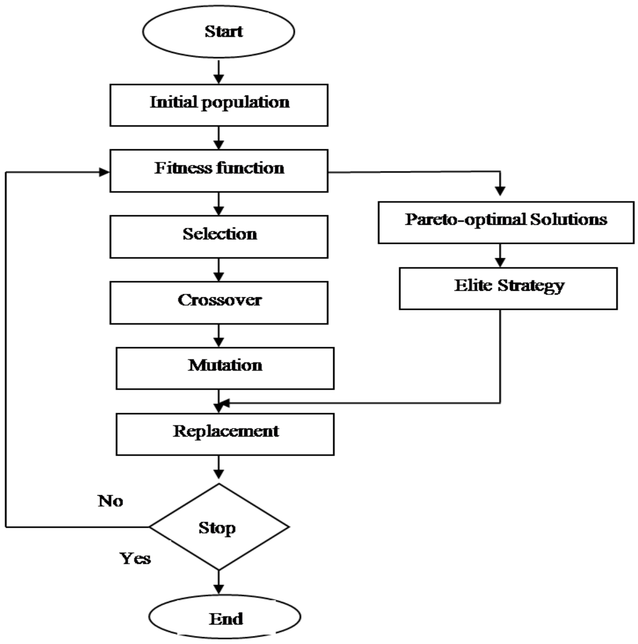

Fig. 1. Flow chart of MOGA Algorithm.

V. Numerical Examples

A popularly used dataset known as COCOMO data set [2] has been chosen for the experiments and comparisons. This data set consists of two variables. They are size, effort adjustment factor and effort. Size is described in a thousand lines of codes and effort is in man-months. Here, the idea is to reduce the complete set of projects into an appropriate subset participating in the cost estimation process. This saves the computation time and produces more accurate results. This is because; it eliminates unrelated projects and considers only related projects. Here, we also compare the efficiency of MOGA-based optimization approach to the traditional approaches. ILOG CPLEX barrier and GA are typical nonlinear optimizers which are used for computation and comparison purpose. This dataset is divided into two categories of projects. The first category of projects contains only seven variables and our aim is to select two best projects from the given project and the second category of projects contains twenty-one variables in the project selection. Here, we want to select eleven best projects based on criteria which are decided by decision makers. Let the two criteria are j = {1: Magnitude of Relative Error (MRE), 2: Absolute Relative Error (ARE)}. The MRE and ARE can be calculated by the following equation:

$$MRE = \frac{|Actual\ Effort - Estimated\ Effort|}{Actual\ Effort} \qquad (6)$$

$$ARE = |Actual\ Effort - Estimated\ Effort| \qquad (7)$$

Let P_j represents the preference degree which is determined by decision makers in terms of criteria j. In general, preference can be calculated by the following equation:

$$p_j' = \frac{p_j}{\sum_{j=1}^{J} p_j} \qquad (8)$$

Table I shows the MRE and ARE value of seven different projects and Table II contains data which was obtained by using the equation (8). Here, the interactive effects between projects are assumed and tabulated in Table III. This assumed value is only for the purpose of simplicity and does not make any difference in the methodology proposed here. However, it can be calculated using analysis of variance [5]. Without considering the interactive effects between projects, the optimization problem as per equation (1) takes the following form

Maximize $E = 20.793x_1 + 3.897x_2 + 8.378x_3 + 1.214x_4 + 344.213x_5 + 99.798x_6 + 107.477x_7$

s.t. $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 = 2$

$x_n = \{0, 1\}$, n = 1, 2, ...7 (9)

Whereas, after considering the interactive effects between projects, the problem formulation based on equation (2) takes the following form:

Maximize $E = 20.793x_1 + 3.897x_2 + 8.378x_3 + 1.214x_4 + 344.213x_5 + 99.798x_6 + 107.477x_7 + 0.02x_1x_2 + 17.420x_1x_3 + 6.544x_1x_4 + 72.591x_1x_5 + 60.174x_1x_6 + 25.612x_1x_7 + 5.392x_2x_3 + 1.974x_2x_4 + 255.195x_2x_5 + 46.558x_2x_6 + 38.88x_2x_7 + 5.184x_3x_4 + 140.358x_3x_5 + 64.605x_3x_6 + 63.483x_3x_7 + 137.251x_4x_5 + 65.525x_4x_6 + 48.796x_4x_7 + 243.261x_5x_6 + 269.688x_5x_7 + 62.034x_6x_7$

s.t. $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 = 2$

$x_n = \{0, 1\}$, n = 1, 2, ...7 (10)

The above equation (9) and (10) has been solved by the ILOG CPLEX barrier optimizer, by simple GA and by MOGA-based method. The results of the above methods are shown in Table IV. Here, it has been also noticed from the equation (9) and (10) that when interactive effects are not taken into consideration, the optimal solution comes as x = (0, 0, 0, 0, 0, 1, 1) and x = (0, 0, 1, 0, 1, 0, 0) if considered. This indicates that interactive effects are important for the project selection.

ISBN: 978-988-14047-1-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I: SEVEN DIFFERENT PROJECTS OF COCOMO DATASET:-

Criteria (j)	Preference (P_j)	Projects (n)						
		e_1	e_2	e_3	e_4	e_5	e_6	e_7
ARE	2	31.44	5.82	12.55	1.73	519.11	151.03	162.59
MRE	1	0.13	0.17	0.29	0.22	4.85	0.36	0.51

TABLE II: DATA WHICH WAS OBTAINED BY USING THE EQUATION (8) FOR SEVEN DIFFERENT PROJECTS OF COCOMO DATASET:-

Criteria (j)	Preference (P_j)	Projects (n)						
		e_1	e_2	e_3	e_4	e_5	e_6	e_7
ARE	0.66	31.44	5.82	12.55	1.73	519.11	151.03	162.59
MRE	0.33	0.13	0.17	0.29	0.22	4.85	0.36	0.51

TABLE III (A): THE INTERACTIVE EFFECTS BETWEEN PROJECTS BASED ON CRITERIA:-

Criteria (j)	Projects Pairs(C_m)									
	e_1e_2	e_1e_3	e_1e_4	e_1e_5	e_1e_6	e_1e_7	e_2e_3	e_2e_4	e_2e_5	e_2e_6
ARE	0.55	0.60	0.30	0.20	0.50	0.20	0.45	0.40	0.65	0.45
MRE	-0.15	0	-0.20	-0.05	-0.25	0	-0.42	-0.15	0	-0.15

TABLE III(B): THE INTERACTIVE EFFECTS BETWEEN PROJECTS BASED ON CRITERIA:-

Criteria (j)	Projects Pairs(C_m)										
	e_2e_7	e_3e_4	e_3e_5	e_3e_6	e_3e_7	e_4e_5	e_4e_6	e_4e_7	e_5e_6	e_5e_7	e_6e_7
ARE	0.35	0.55	0.40	0.60	0.55	0.40	0.65	0.45	0.55	0.60	0.30
MRE	-0.10	0	0	-0.80	-0.35	-0.15	-0.05	-0.03	0	-0.15	-0.22

TABLE IV: COMPARISON OF OBJECTIVE VALUE AND COMPUTATIONAL TIME FOR GA AND MOGA-BASED METHOD WITH CPLEX OPTIMIZER FOR SEVEN VARIABLES:-

Equation No.	ILOG CPLEX barrier optimizer based objective value	GA based objective value	MOGA based objective value	Computational Time(seconds)		
				CPLEX	GA	MOGA
9	451.691	728.6767	1026.0016	9	11	13
10	721.378	857.2201	1612.4166	13	15	18

TABLE V: COMPARISON OF OBJECTIVE VALUE AND COMPUTATIONAL TIME FOR GA AND MOGA-BASED APPROACH WITH CPLEX OPTIMIZER FOR TWENTY-ONE VARIABLES:-

Equation No.	ILOG CPLEX barrier optimizer based objective value	GA based objective value	MOGA based objective value	Computational Time		
				Minutes	Seconds	
				CPLEX	GA	MOGA
11	1139.955	1579.107	2035.952	1.07	15	11
12	6699.812	8812.295	9146.858	1.16	22	18

From the table IV, we can see that the CPLEX barrier optimizer takes less computational time than the GA and MOGA-based approach. This is because GA and MOGA-based optimization is a heuristic optimization method that can change the final result, once a small size problem is identified.

In this section, we compare the effectiveness of our MOGA-based optimization method with other existing nonlinear optimization methods. Assume there are twenty-one projects and more than two evaluations are presented and we want to select eleven best projects from the given selection problem. Without considering the interactive effects between projects, the optimization problem as per equation (1) takes the following form

Maximize $E = 20.793x_1 + 3.897x_2 + 8.378x_3 + 1.214x_4 + 344.213x_5 + 99.798x_6 + 107.477x_7 + 27.166x_8 + 27.888x_9 + 1.647x_{10} + 9.016x_{11} + 8.326x_{12} + 2.996x_{13} + 1.512x_{14} + 133.983x_{15} + 85.022x_{16} + 157.615x_{17} + 112.854x_{18} + 12.851x_{19} + 23.146x_{20} + 0.128x_{21}$

s.t. $\sum_{i=1}^{21} x_n = 11$

$x_n = \{ 0 , 1 \} ,\ n = 1 , 2 , ... 21$ (11)

Whereas, after considering the interactive effects between projects, the problem formulation based on equation (2) takes the following form and the results are shown in Table V.

Maximize $E = 4.903x_1x_2 + 13.065x_1x_3 + 8.729x_1x_4 + 236.186x_1x_5 + 54.194x_1x_6 + 44.792x_1x_7 + 26.316x_1x_8 + 19.427x_1x_9 + 14.775x_1x_{10} + 16.338x_1x_{11} + 8.674x_1x_{12} + 5.902x_1x_{13} + 11.092x_1x_{14} + 108.212x_1x_{15} + 79.24x_1x_{16} + 35.618x_1x_{17} + 86.748x_1x_{18} + 18.428x_1x_{19} + 15.287x_1x_{20} + 6.244x_1x_{21} + 4.83x_2x_3 + 3.226x_2x_4 + 69.208x_2x_5 + 46.56x_2x_6 + 66.624x_2x_7 + 17.021x_2x_8 + 7.915x_2x_9 + 1.901x_2x_{10} + 1.903x_2x_{11} + 7.258x_2x_{12} + 3.388x_2x_{13} + 1.048x_2x_{14} + 41.311x_2x_{15} + 39.946x_2x_{16} + 104.826x_2x_{17} + 69.881x_2x_{18} + 13.257x_2x_{19} + 14.777x_2x_{20} + 1.378x_2x_{21} + 3.769x_3x_4 + 122.593x_3x_5 + 86.349x_3x_6 + 75.135x_3x_7 + 15.925x_3x_8 + 10.831x_3x_9 + 2.474x_3x_{10} + 3.433x_3x_{11} + 10.751x_3x_{12} + 6.744x_3x_{13} + 4.372x_3x_{14} + 35.503x_3x_{15} + 18.634x_3x_{16} + 107.735x_3x_{17} + 42.267x_3x_{18} + 6.281x_3x_{19} + 7.827x_3x_{20} + 3.781x_3x_{21} + 85.721x_4x_5 + 10.076_{x4x6} + 32.535_{x4x7} + 22.492_{x4x8} + 20.273_{x4x9} + 0.548_{x4x10} + 3.495_{x4x11} + 7.988_{x4x12} + 1.844_{x4x13} + 0.887x_4x_{14} + 26.992x_4x_{15} + 43.002x_4x_{16} + 87.233x_4x_{17} + 45.475x_4x_{18} + 8.301x_4x_{19} + 13.255x_4x_{20} + 0.483x_4x_{21} + 154.803x_5x_6 + 359.76x_5x_7 + 258.72x_5x_8 + 314.114x_5x_9 + 68.849x_5x_{10} + 105.471x_5x_{11} + 227.731x_5x_{12} +$

ISBN: 978-988-14047-1-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$189.702x_5x_{13} + 68.809x_5x_{14} + 95.295x_5x_{15} + 192.041x_5x_{16} + 188.341x_5x_{21} + 82.724x_6x_7 + 76.018x_6x_8 + 70.124x_6x_9 + 45.593x_6x_{10} + 32.582x_6x_{11} + 37.715x_6x_{12} + 46.183x_6x_{13} + 60.667x_6x_{14} + 93.416x_6x_{15} + 101.514x_6x_{16} + 128.518x_6x_{17} + 180.45x_6x_{18} + 84.343x_6x_{19} + 24.542x_6x_{20} + 34.91x_6x_{21} + 53.721x_7x_8 + 81.018x_7x_9 + 21.762x_7x_{10} + 52.319x_7x_{11} + 46.167x_7x_{12} + 71.673x_7x_{13} + 48.934x_7x_{14} + 84.368x_7x_{15} + 105.718x_7x_{16} + 105.908x_7x_{17} + 132.011x_7x_{18} + 66.048x_7x_{19} + 39.029x_7x_{20} + 26.83x_7x_{21} + 27.463x_8x_9 + 20.114x_8x_{10} + 27.031x_8x_{11} + 7.056x_8x_{12} + 19.541x_8x_{13} + 15.681x_8x_{14} + 56.299x_8x_{15} + 33.561x_8x_{16} + 83.056x_8x_{17} + 48.924x_8x_{18} + 29.913x_8x_{19} + 30.081x_8x_{20} + 10.881x_8x_{21} + 19.138x_9x_{10} + 7.349x_9x_{11} + 16.215x_9x_{12} + 18.449x_9x_{13} + 16.088x_9x_{14} + 40.42x_9x_{15} + 39.421x_9x_{16} + 55.554x_9x_{17} + 84.316x_9x_{18} + 20.282x_9x_{19} + 10.16x_9x_{20} + 8.382x_9x_{21} + 4.767x_{10}x_{11} + 6.423x_{10}x_{12} + 2.747x_{10}x_{13} + 2.396x_{10}x_{14} + 74.524x_{10}x_{15} + 30.285x_{10}x_{16} + 63.639x_{10}x_{17} + 40x_{10}x_{18} + 11.524x_{10}x_{19} + 16.032x_{10}x_{20} + 0.614x_{10}x_{21} + 3.424x_{11}x_{12} + 5.361x_{11}x_{13} + 4.12x_{11}x_{14} + 92.832x_{11}x_{15} + 42.248x_{11}x_{16} + 58.217x_{11}x_{17} + 66.904x_{11}x_{18} + 8.694x_{11}x_{19} + 19.191x_{11}x_{20} + 1.811x_{11}x_{21} + 5.044x_{12}x_{13} + 5.754x_{12}x_{14} + 56.819x_{12}x_{15} + 51.236x_{12}x_{16} + 82.815x_{12}x_{17} + 102.756x_{12}x_{18} + 15.775x_{12}x_{19} + 6.257x_{12}x_{20} + 5.006x_{12}x_{21} + 2.396x_{13}x_{14} + 82.091x_{13}x_{15} + 26.34x_{13}x_{16} + 32.084x_{13}x_{17} + 57.824x_{13}x_{18} + 3.147x_{13}x_{19} + 6.491x_{13}x_{20} + 1.538x_{13}x_{21} + 47.353x_{14}x_{15} + 38.82x_{14}x_{16} + 47.668x_{14}x_{17} + 51.363x_{14}x_{18} + 7.777x_{14}x_{19} + 4.873x_{14}x_{20} + 1.008x_{14}x_{21} + 164.093x_{15}x_{16} + 72.823x_{15}x_{17} + 110.862x_{15}x_{18} + 58.655x_{15}x_{19} + 70.592x_{15}x_{20} + 87.087x_{15}x_{21} + 48.454x_{16}x_{17} + 118.582x_{16}x_{18} + 34.189x_{16}x_{19} + 32.387x_{16}x_{20} + 42.488x_{16}x_{21} + 67.528x_{17}x_{18} + 127.681x_{17}x_{19} + 31.519x_{17}x_{20} + 94.544x_{17}x_{21} + 62.721x_{18}x_{19} + 88.175x_{18}x_{20} + 84.621x_{18}x_{21} + 12.494x_{19}x_{20} + 10.962x_{19}x_{21} + 8.101x_{20}x_{21}$

s.t. $\sum_{i=1}^{21} x_n = 11$

$x_n = \{0, 1\},\ n = 1, 2, \ldots 21$ (12)

Similarly, from equation (11) and (12), it has been found that results are different. Without considering the interactive effects between projects, the best solution as x = (0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1) and x = (1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1) if considered. Here, it has been found that the computational times are increased for both GA and MOGA-based optimization method with CPLEX barrier optimizer when the number of different projects and quadratic terms are increased, but the results shows that MOGA is better than GA and CPLEX for solving this type of problem.

VI. Experimental Results and Comparisons

Two popularly used datasets known as COCOMO 81 [2] and COCOMONASA [2] has been chosen for the experiments and comparisons. Popularly used criteria for measuring the accuracy of software cost are MMRE and PRED [16]. They are defined as below:-

MMRE is an average percentage of the absolute value of the relative errors over a whole data set. It can be calculated by the following equation:-

$$MMRE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Predicted\ Value - Actual\ Value}{Actual\ Value}\right| \quad 13)$$

PRED (0.25) is defined as the percentage of predictions falling within 25% of the actual known value. It can be calculated by the following equation:-

$199.68x_5x_{17} + 295.958x_5x_{18} + 159.925x_5x_{19} + 127.888x_5x_{20} +$

$$PRED = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Predicted\ Value - Actual\ Value}{Actual\ Value} \leq 0.25\right) \quad (14)$$

Here, n is the number of projects. These two performance measures are considered as the objective function for MOGA to search optimal parameter of COCOMO 81 and COCOMONASA datasets. Project portfolio selection for an ABE system becomes a bi-objective problem where, MMRE has to be minimized while PRED is being maximized. The graphs and tables below summarize the experimental results. Fig. 2 to 6 and fig. 7 to 11 represents the best three Pareto-optimal solutions obtained from COCOMO 81 and COCOMONASA datasets respectively by varying weights to the two objectives. It can be observed that the best compromised MMRE is 0.52 and PRED is 0.72 for COCOMO 81 dataset and the best compromised MMRE is 0.48 and PRED is 0.72 for COCOMONASA dataset. This is the case where the preference is to find maximum prediction with lower MMRE value. It has been found that MMRE and PRED values using MOGA based approach are better than Basic COCOMO, Intermediate COCOMO and GA method. The results are summarized in Table VI. Table VII shows the comparison of results among various project categories (having interacting and non-interacting effects). It shows more accurate MMRE and PRED values for effort estimation when a subset of projects is selected rather than using the complete set of historical projects. It also indicates that the MMRE and PRED values are changing when interaction between the projects are considered. However, the project selection based approach gives better MMRE and PRED in both the above cases.

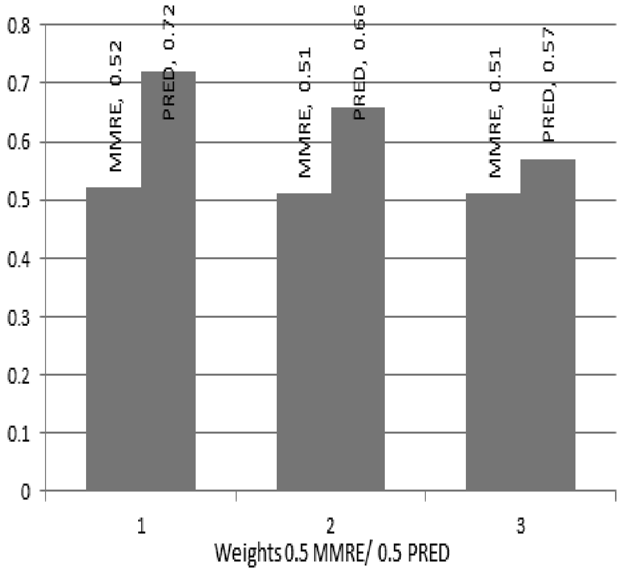

Fig. 2. Best three Pareto-optimal solutions obtained by weights 0.5 MMRE/ 0.5 PRED.

ISBN: 978-988-14047-1-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

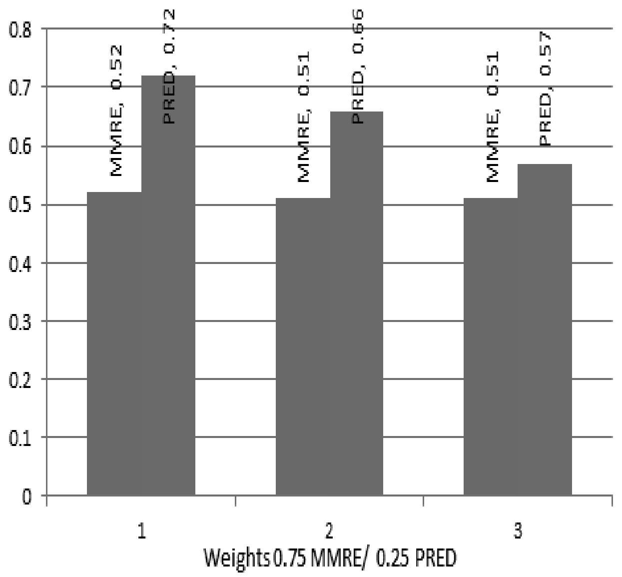

Fig. 3. Best three Pareto-optimal solutions obtained by weights 0.75 MMRE/ 0.25 PRED.

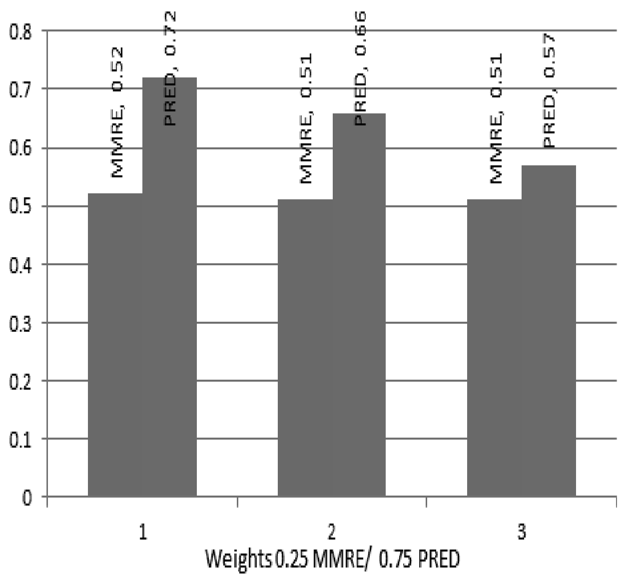

Fig. 4. Best three Pareto-optimal solutions obtained by weights 0.25 MMRE/ 0.75 PRED.

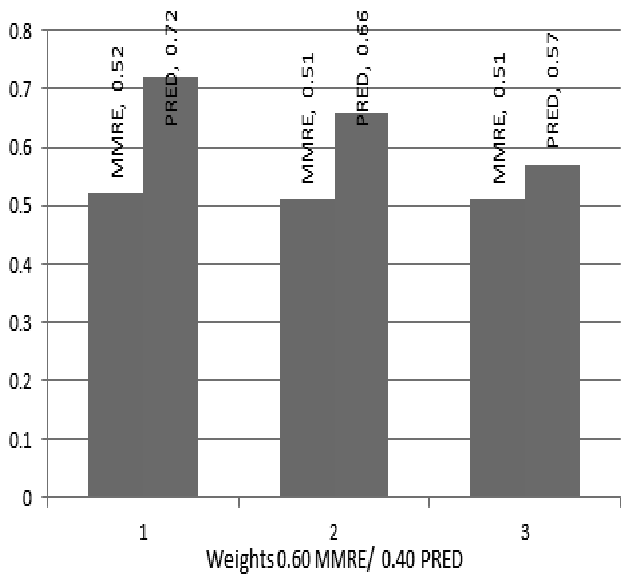

Fig. 5. Best three Pareto-optimal solutions obtained by weights 0.60 MMRE/ 0.40 PRED.

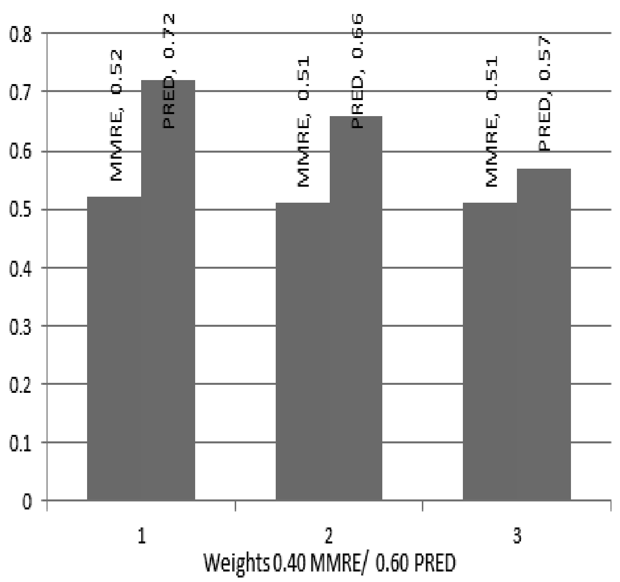

Fig. 6. Best three Pareto-optimal solutions obtained by weights 0.40 MMRE/ 0.60 PRED.

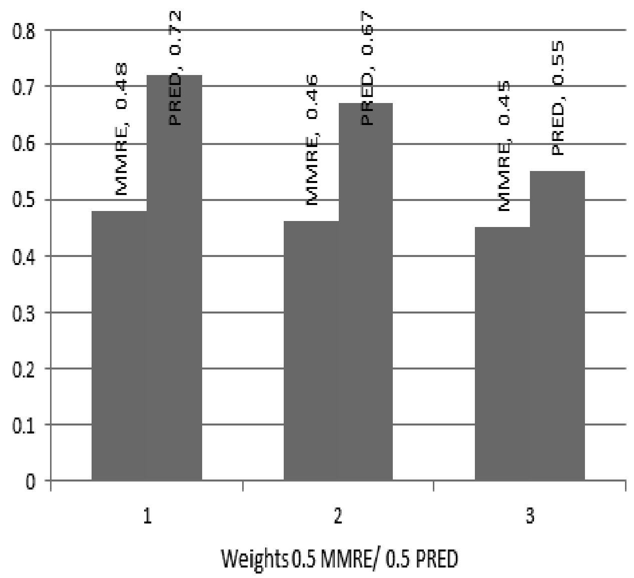

Fig. 7. Best three Pareto-optimal solutions obtained by weights 0.5 MMRE/ 0.5 PRED.

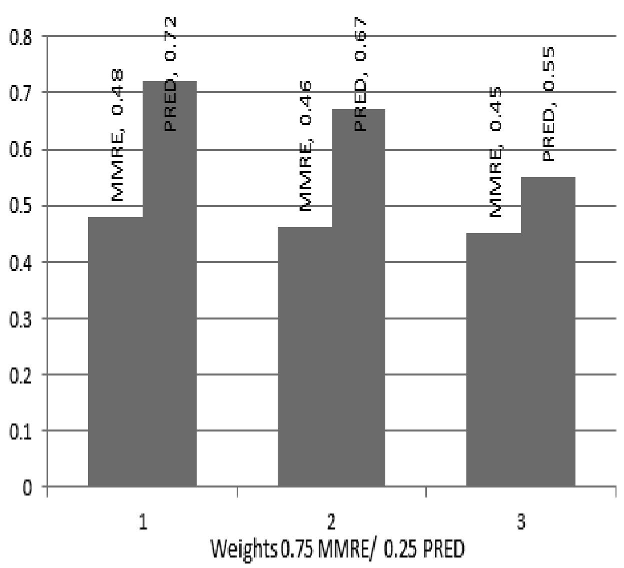

Fig 8. Best three Pareto-optimal solutions obtained by weights 0.75 MMRE/ 0.25 PRED.

ISBN: 978-988-14047-1-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

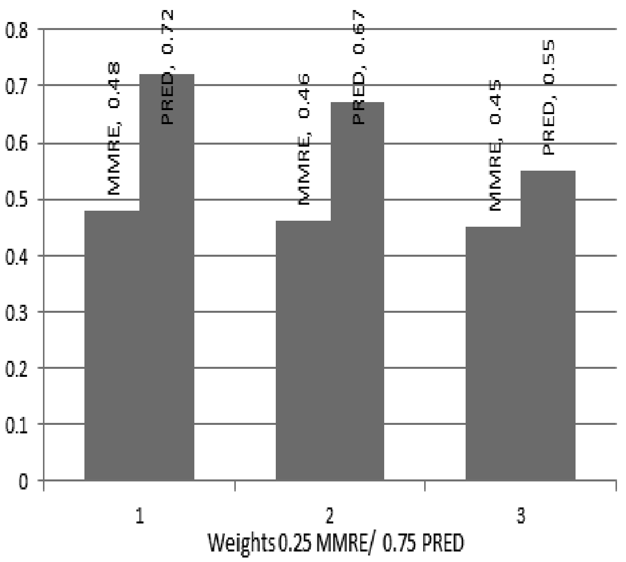

Fig. 9. Best three Pareto-optimal solutions obtained by weights 0.25 MMRE/ 0.75 PRED.

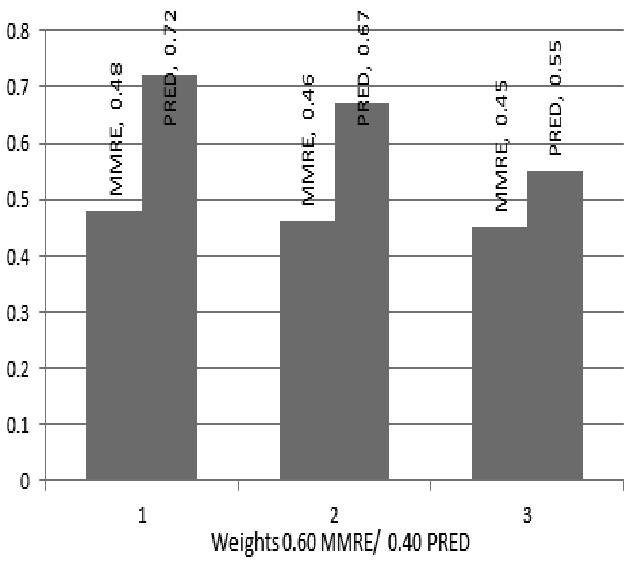

Fig. 10. Best three Pareto-optimal solutions obtained by weights 0.60 MMRE/ 0.40 PRED.

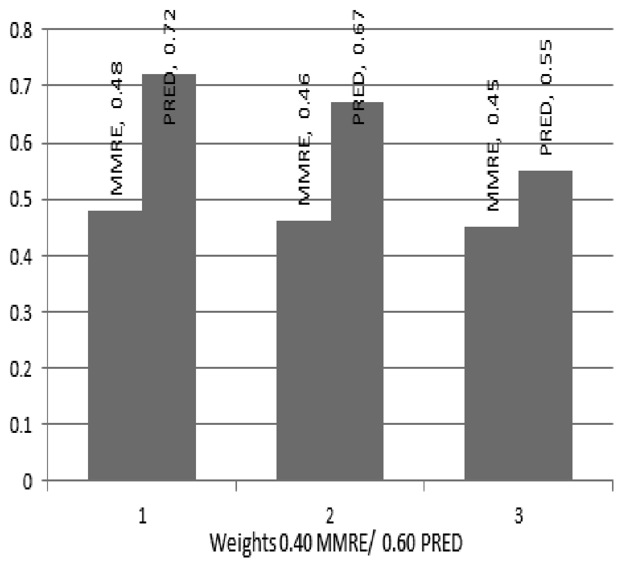

Fig. 11. Best three Pareto-optimal solutions obtained by weights 0.40 MMRE/ 0.60 PRED.

VII. Conclusion

This paper proposes a MOGA based approach for project selection to improve the ABE system's performance. A multi-criteria project selection problem has been formulated for allowing project interactions and for incorporating the decision maker's preference information. It has been shown here through experimental results that the proposed approach enhances the existing ABE process. The model has been experimented on two standard datasets (COCOMO 81 and COCOMONASA) and tested based on the criteria of MMRE and PRED for software cost estimation. The results show that the MOGA based project selection approach has lowest MMRE with maximum PRED value, therefore provides good estimation capabilities for ABE system. The proposed approach has also been compared with other existing methods. The results show the suitability of the proposed method for improving the cost prediction using the ABE based estimation method. It has also been shown here that how interactive effects among projects changes the project cost prediction. However, as it is a meta-heuristic based approach, it is possible that this method would not able to find the exact global optimum in some cases. However, this limitation is for any such meta-heuristic technique used for optimization. Therefore, the future direction can be experimenting with some more methods for project selection that may help to overcome the above limitation and can further improve the process of software cost estimation.

TABLE VI: RESULTS AND COMPARISONS ON COCOMO 81 AND COCOMONASA DATASET:-

Dataset	Model	MMRE	PRED
COCOMO 81	Basic	0.89	0.21
	Intermediate	0.38	0.43
	GA	0.49	0.54
	MOGA	0.52	0.72
COCOMONASA	Basic	0.34	0.53
	Intermediate	0.25	0.66
	GA	0.47	0.52
	MOGA	0.48	0.72

TABLE VII: RESULTS AND COMPARISONS ON COCOMO DATASET WITH AND WITHOUT CONSIDERATION OF PROJECT INTERACTIONS -

Project Category	MMRE & PRED before project selection		MMRE & PRED after project selection without consideration of project interactions		MMRE & PRED after project selection with consideration of project interactions.	
	MMRE	PRED	MMRE	PRED	MMRE	PRED
Seven Variable	1.06	0.28	0.34	0.50	0.16	0.50
Twenty-one Variable	0.49	0.53	0.22	0.64	0.17	0.73

REFERENCES

[1] Aaker, D.A. and Tyebjee, T.T. (1978) 'Model for the selection of interdependent R&D projects' ,*IEEE Transactions on Engineering Management*, Vol. 25, No. 2, pp. 30–36.

[2] Boehm, B.W. (1981) *Software Engineering Economics*, Prentice- Hall, Englewood Cliffs, NJ. USA.

[3] Carlsson, C. and Fuller, R. (1995) 'Multiple criteria decision making: the case for interdependence' , *Computers & Operations Research*, Vol. 22, pp. 251–260.

[4] Chang, P. C., Hsieh, J. C. and Wang, Y. W. (2004) 'Development of Multi-objective Genetic Algorithms for Scheduling' , Proceedings of the Fifth Asia Pacific Industrial Engineering and Management Systems Conference.

[5] Cox, D. R. (1984) *Interaction International Statistical Review*, 52, 1–25.

[6] Ghasemzadeh, F., Archer, N. and Iyogun, P (1999) 'Zero-one model for project portfolio selection and scheduling' , *Journal of the Operational Research Society*, Vol. 50, No. 7, pp. 745-755.

[7] Hammer, P. and Rudeanu, S. (1968) *Boolean methods in operations research and related areas*, Berlin, Springer.

[8] Huang, S. and Chiu, N. (2006) 'Optimization of analogy weights by genetic algorithm for software effort estimation' , Inf. Softw. Technol., 48(11), pp. 1034-1045.

[9] Jorgenson, M. (2004) 'A review of studies on expert estimation of software development effort' , *The Journal of Systems and Software*, 70, pp. 37-60.

[10] Keung, J. and Kitchenham, B. (2008) 'Analogy-X: Providing statistical inference to analogy-based software cost estimation' , IEEE Trans. Softw. Engg., 34(4), pp. 471-484.

[11] Korhonen, P., Moskowitz, H. and Wallenius, J. (1992) 'Multiple criteria decision support—a review' , *European Journal of Operational Research*, 63, pp. 361–375.

[12] Lee, J. W. and Kim, S. H. (2000) 'Using Analytic network and goal programming for interdependent information systems project selection' , *Computers and Operations research*, Vol. 19, pp. 367-382.

[13] Li, J. and Ruhe, G. (2008) ' Analysis of attribute weighting heuristics for analogy-based software effort estimation method AQUA+ ', Empirical Softw. Engg., 13(1), pp. 63-96.

[14] Lockett, G., Hetherington, B. and Yallup, P. (1986) 'Modelling a research portfolio using AHP: A group decision process' , *R&D Management*, Vol. 16, No. 2.

[15] Lorie, J. H. and Savage, L. J. (1955) 'Three problems in rationing capital' , *The Journal of Business*, Vol. 28, No. 4, pp. 229-239.

[16] Moløkken-Østvold, K. et al. (2004) 'Project Estimation in the Norwegian Software Industry - A Summary' , Simulation Research Laboratory.

[17] Mukhopadhyay, T., Vicinanza, S. S. and Prietula, M. J. (1992) 'Examining the feasibility of a case-based reasoning model for software effort estimation' , *MIS Quarterly*, 16, pp. 155–171.

[18] Murahaldir, K., Santhanam, R. and Wilson, R. L. (1990) 'Using the analytical hierarchy process for information system project selection' , *Information & Management*, Vol. 17, No. 1, pp. 87-95.

[19] Narzisi, G. (2008) 'Classic Methods for Multi-Objective Optimization' ,3rd ed. Courant Institute of Mathematical Sciences, New York University.

[20] Nemhauser, G. L. and Ullman, Z. (1969) 'Discrete dynamic programming and capital allocation' , *Management Science*, Vol 15, No. 9, pp. 494-505.

[21] Ostermark, R. (1997) 'Temporal interdependence in fuzzy MCDM problems' , *Fuzzy Sets and Systems*, Vol. 88, pp. 69–79.

[22] Osyczka, A. (1985) 'Multicriteria optimization for engineering design' , in Design Optimization, J. S. Gero, Ed. New York: Academia, pp. 193-227.

[23] Ozkan, B., Bashgil, H. and Sahin, N. (2011) 'Supplier Selection using Analytic Hierarchy Process: An Application from Turkey' , Proceedings of the World Congress on Engineering, Vol. II.

[24] Rengarajan, S. and Jagannathan, P. (1997) 'Projects selection by scoring for a large R&D organization in a developing country' , *R&D Management*, Vol. 27.

[25] Santhanam, R. and Kyparisis, G. J, (1996) 'A decision model for interdependent information system project selection' , *European Journal of Operational Research*, Vol. 89, pp. 380-399.

[26] Shepperd, M. J. and Schofield, C. (1997) 'Estimating software project effort using analogies' , *IEEE Transaction on Software Engineering*, 23, pp. 736–743.

[27] Tiryaki, F. and Ahlatcioglu B. (2009) ' Fuzzy Portfolio Selection using Fuzzy Analytic Hierarchy Process' , Information Sciences, 179, pp. 53-69.

[28] Walkerden, F. and Jeffery, R. (1999) 'An empirical Study of Analogy-based Software Effort Estimation' , *Kluwer Academic Publishers*, Boston.