Australian Journal of Basic and Applied Sciences, 8(3) March 2014, Pages: 16-26

AENSI Journals
Australian Journal of Basic and Applied Sciences
ISSN:1991-8178
Journal home page: www.ajbasweb.com

A Combinational Approach for Software Development Effort using Graphical and Genetic Approach

Dr. K. Suresh Joseph, Dr. P. Dhavachelvan and Dr. Saleem Baasha, M.S.

Pondicherry University, Department of Computer Science School of Engineering and Technology, Pondicherry, India

ARTICLE INFO

Article history:
Received 25 January 2014
Received in revised form 8 March 2014
Accepted 10 March 2014
Available online April 2014

Keywords:
Genetic Algorithm, Software Project, Schedule, Effort Estimation, imputation.

ABSTRACT

One of the most exceptionally important tasks in software development life cycle is to produce precise effort estimations and schedule. Developers may use effort estimate as input to project planning, budgeting, pricing policies and investment analyses. Missing or Incomplete data are exasperation to a certain extent everywhere but especially in effort estimation it is a greater hindrance. Accuracy in effort estimation is decisive for developers and customers. It is very difficult to obtain such accurate information at the early development stage of a software project. These issues led to the introduction of variety of techniques to solve the relevant issues. There exists variety of techniques, imputation based models yield much attraction and focus of the researchers due to their enhanced performance. Among those models, Genetic Algorithm (GA) is stochastic especially in domains where a direct search method cannot reveal better outputs and of course in the large search spaces. They have an advantage to operate with incomplete data to extract significant rules. From the above perspectives, the objective of this proposed research is to formulate a comprehensive model using Graphical and Genetic Approaches. This hybrid effort estimation model produces accurate outputs even with incomplete and vague data. The performance of the proposed approach is validated in terms of standard performance factors and demonstrated the efficiency of the same.

To Cite This Article: Dr. K. Suresh Joseph, Dr. P. Dhavachelvan and Dr. Saleem Baasha, M.S.., A Combinational Approach for Software Development Effort using Graphical and Genetic Approach. ***Aust. J. Basic & Appl. Sci.,*** *8(3): 16-26, 2014*

INTRODUCTION

Software Engineering is a discipline which is majorly concerned about development of systematic large software applications that are used in digital computers. This involves various theories and methodologies in designing which includes not only technical issues like tools and technologies for professional software development but also management aspects like scheduling, staffing, organizing, planning, controlling and budgeting. Software Engineering is a complex process which solves the problem by modeling it as an abstract model or also referred as iconic model. When compared with science SE is closely related to good practices in business. This well-perceived scientific method of software development takes user requirement as input and a software product is delivered as output.

This systematic and organized approach in software development was performed with a sequence of stages referred as Software Development Life Cycle. Some of the popular and commonly used SDLC models are agile development, waterfall model, V model, feature driven development model, several hybrid models, object oriented models, and various prototyping models. The structured set of tasks that are involved to develop the software system is called as a process. Each stage in SDLC is a process involving resources, constraints and other tasks. They also impose consistency and help the programmers to improve the activities that are related with this. The stages include gathering requirement from client, a detailed functional spec about what the system does and whether all the user requirements are fulfilled. When the customer accepts the functional specification as per the design specification the software is being developed and tested. The completed software is also delivered and installed in the acquirer end. The major activities of a software process includes software requirement gathering, Software design, Software construction, Software testing, Software maintenance, Software configuration management, Software engineering management, Software engineering process, Software engineering tools and methods, Software Quality Assurance and Risk management.

Contribution of Cost Estimation Accuracy:

The appropriateness of estimation certainly contributes in project management arena to provide quality management, scope management, communication management, Risk management, Human resource management, procurement management and Time management.

Corresponding Author: Dr. K. Suresh Joseph, Pondicherry University, Department of Computer Science School of Engineering and Technology, Pondicherry, India

Australian Journal of Basic and Applied Sciences, 8(3) March 2014, Pages: 16-26

High level resource assessment is necessary to develop a project in a refined manner. Since this key role helps in predicting effort and time span for completion of project. Many researchers declare that without proper planning and unrealistic approaches the industry fails. Having these in the forefront accurate estimation persuades now a days. Effective and strong estimates help the estimators to plan and control successfully. It is principally necessary to estimate effort for software projects and set deadline to do it profitably without cost overruns, schedule slipping and with utmost customer satisfaction.

Overview of COCOMO Effort Estimation Scheme:

The COCOMO model is an empirical model that was derived empirically based on the data from large number of software projects. These data were analysed and a model was formulated empirically to the best fit of the observations. These formulae link the size of the system and product, project and team factors to the effort to develop the system. In COCOMO, effort is expressed in Person Months (PM) as

$$PM = a * (SIZE)\, b * \prod_{i=1}^{15} EMi$$

where,
"a" and "b" are empirical constants,
EM refers to the set of Effort Multipliers. It contains 15 effort multipliers.
SIZE refers to the size of the project in terms of Lines of Code (LOC).

Hence, in principle, the effort is calculated by multiplying the estimation variable with the constant 'a' in the first stage, and some effort can be added with or deducted from the calculated effort at the second stage. In addition to that, each user should calibrate the model and the attribute values in accordance to their own historical projects data, which will reflect local circumstances that greatly influences accuracy of the model. From these viewpoint, whenever using the algorithmic effort estimation models, it is preferred that the impacts of cost drives have to be quantified and assessed in a proper way. Since the significance of the vagueness and uncertainty features that are inhabited in the effort drives due to the cognitive judgments are less, this Imputation approach can be preferred and applied to change the estimation scheme of the COCOMO II, which can substitute values in the place of vague information.

It is observed that the accuracy of the COCOMO II is relied on three attributes; the value of estimation variable, the overall value of the effort multipliers and their impact in the estimation scheme. Here the estimation variable is the primary attribute and standard methods available for estimating the estimation variables (e.g. LOC, FP count, etc…)[10]. It is simple to estimate the overall value of the effort multipliers after assigning the appropriate values as per the requirements. But the complicated issue is to estimate impact of the effort multipliers, which plays a major role in the estimation scheme and causes for overestimation or underestimation of the software development effort.

Categorization of Effort Multiplier Using Graphical Approach:

In this view, this work is aimed at refining the cost drivers and scale factors handling mechanisms in COCOMO II estimation scheme. There are 17 effort multipliers as cost drivers and they are devised into two groups; Optimistic Group and Pessimistic Group.

Definition-1: Optimistic Group (OG): This group can be defined as a set of effort multipliers of whose range of values are inversely proportional to the overall effort to be predicted and it can be described as follows:
OG = {EM1, EM2, ….. EM9} = {ACAP, PCAP, PCON, AEXP, PEXP, LTEX, TOOL, SITE, SCED}

Definition-2: Pessimistic Group: This group can be defined as a set of effort multipliers of whose range of values are directly proportional to the overall effort to be predicted and it can be described as follows:
PG= {EM10, EM11, ….. EM17} = {RELY, DATA, CPLX, RUSE, DOCU, TIME, STOR, PVOL}

This classification makes a sense in the estimation scheme and plays a significant role in improving the accuracy of the COCOMO II estimation model. The three attributes model can be visualized as a three edged object in a graphical form. In this scheme the overall effort can be estimated in terms of the area(s) of three edged object(s). Let us consider the figure 2(a).

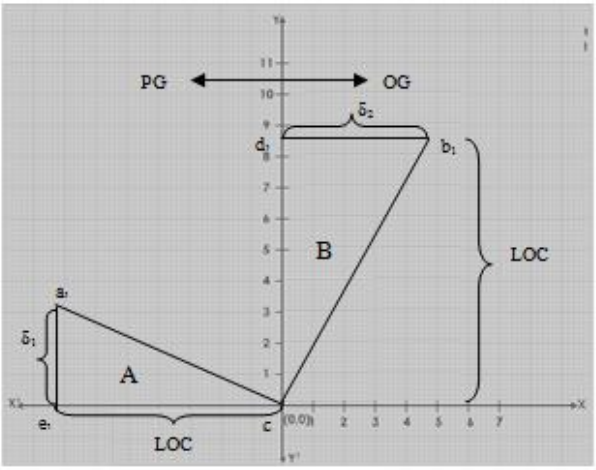

Fig 2(a)

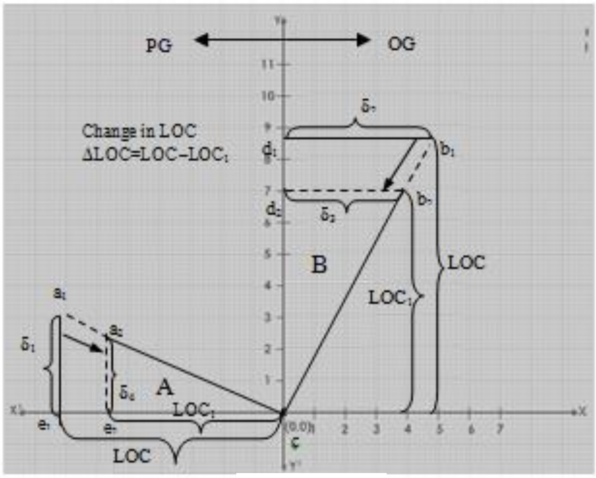

Fig 2(a)

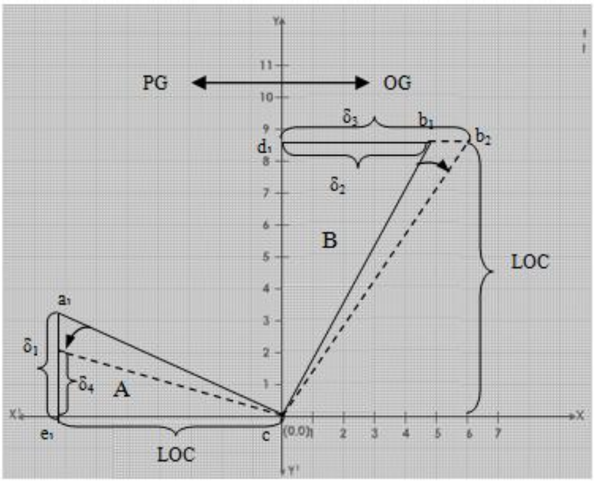

Fig 2(e)

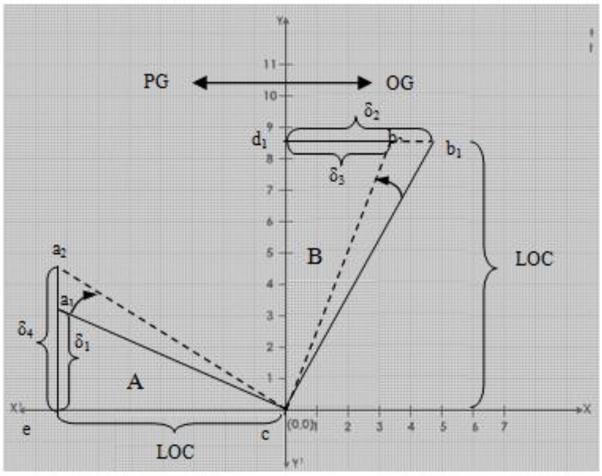

Fig 2(d)

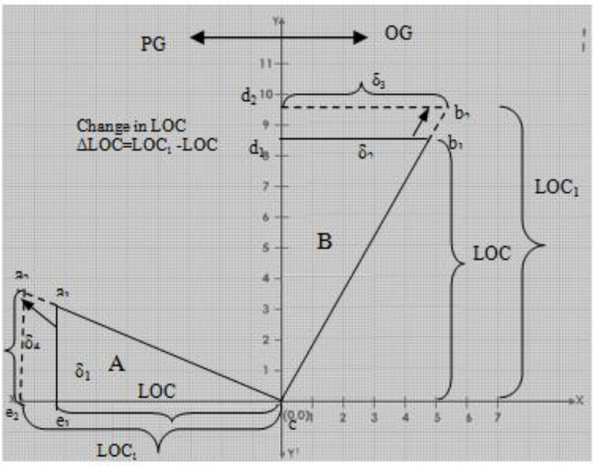

Fig 2(b)

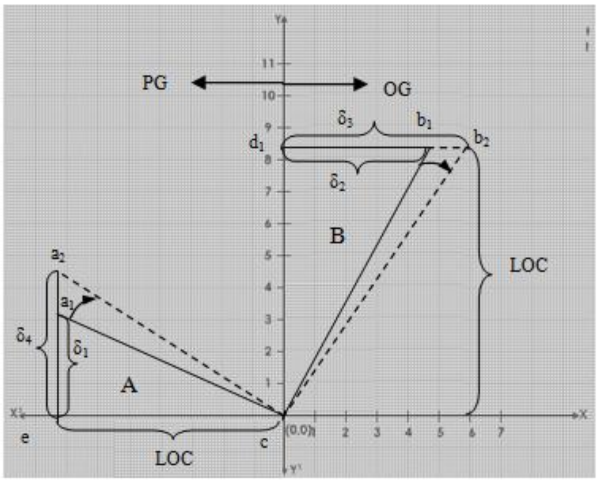

Fig 2(f)

Fig. 2: Model Formulation using Graphical Representations – A Proof for Size does matter.

In the triangle A, the slope 'ca1' represents the overall value of the effort multipliers of the Pessimistic Group (PG). The length of the line 'ce1' will be determined accordingly using the value of the estimation variable, here it is LOC measure. By having the lengths of 'ca1' and 'ce1', the length of 'δ1' can be determined automatically, which represents the impact of the effort multipliers over the overall effort to be estimated. The angle between the lines 'ca1' and 'ce1' will be determined by the slope of 'ca1', which will be determined by overall value of the effort multipliers of the Pessimistic Group (PG). For a constant LOC value, depending on

Australian Journal of Basic and Applied Sciences, 8(3) March 2014, Pages: 16-26

the values of the effort multipliers of the Pessimistic Group (PG), the area under the triangle 'ca1e1' will be varied accordingly.

Similarly, in the triangle B, the slope 'cb1' represents the overall value of the effort multiplier of the Optimistic group (OG). The length of the line 'cd1' will be determined accordingly by the value of the estimation variable, here it is LOC measure. By having the length of 'cb1' and 'cd1', the length of 'δ1' can be determined automatically, which represents the impact of the effort multipliers over the overall effort to be estimated. The angle between the lines 'cb1' and 'cd1' will be determined by the slope of 'cb1', which will be determined by overall value of the effort multipliers of the Optimistic Group (OG). For a constant LOC value, depending on the values of the effort multipliers of the Optimistic Group (OG), the area under the triangle 'cb1d1' will be varied accordingly. The Graphical representations are shown below in Fig.2. Let LOC be the height of the triangle, δ1 be the impact of the effort multipliers in the PG set and δ2 be the impact of the effort multipliers in OG set. The effort can be calculated as the areas of the triangles for the PG and OG sets.

The corresponding areas can be calculated respectively as follows:

$$^{PG}E_{2a} = 0.5 * \delta_1 * height \tag{1}$$

$$^{OG}E_{2a} = 0.5 * \delta_2 * height \tag{2}$$

where,
PGE2a is the effort due to the Pessimistic Group in the figure 2(a),
OGE2a is the effort due to the Optimistic Group in the figure 2(a).

Now based on the insights of the deductive and additive theories of impact of effort multipliers over the overall effort, the values of δ1 and δ2 can be assumed accordingly. By following the deductive principle, the value of δ1 can be taken as LOC/EM and by following the additive principle, the value of δ2 can be taken as LOC*EM. This cognitive assumption improves the accuracy of the COCOMO estimation scheme to a larger extent.

By substituting the values of δ1 and δ2, the equations (1) and (2) can be rewritten as follows:

$$\left.\begin{aligned} ^{PG}E_{2a} &= 0.5 * LOC * EM * LOC \\ ^{OG}E_{2a} &= 0.5 * LOC / EM * LOC \end{aligned}\right\} \tag{3}$$

Then the total effort is defined as the product of efforts due to OG and PG and it can be calculated as follows:

$$E_{2a} = {}^{PG}E_{2a} * {}^{OG}E_{2a} \tag{4}$$

Hence the overall effort includes the corresponding impacts of both the types of effort multipliers. An important and interesting point to be noted in this scheme is, though by keeping the values of effort multipliers as constants, the corresponding impact may be varied in accordance to the variations in the value of LOC measure. This interesting property of this scheme can be explained in detail under various conditions as follows:

CASE1 – If the LOC value is increased:

As already discussed above, the value of LOC has the direct impact on the overall effort. i.e. overall effort increases along with the increments in the LOC value and the overall effort decreases along with the decrements in the LOC value by keeping the other parameters as constant. In the figure 2(b), the LOC value is increased from LOC to LOC1 and these increments are clearly represented in both the triangles. In the figure 2(b), the variations are done only based on the LOC values, but these variations are automatically reflected in the impacts of the effort multipliers. i.e. the value of δ1 is increased to δ4 and the value of δ2 is increased to δ3. Since the overall effort is the function of areas of the triangles, the increments in the values of δ1 and δ2 (as δ4 and δ3) will be correspondingly reflected in the overall effort.

From the figure 2(b), the corresponding overall effort can be estimated as follows:

$$E_{2b} = {}^{PG}E_{2b} * {}^{OG}E_{2b} \tag{5}$$

where,
E2b is the overall effort from the figure 2(b),
PGE2b is the effort due to the Pessimistic Group in the figure 2(b)
OGE2b is the effort due to the Optimistic Group in the figure 2(b).
The effort of Pessimistic Group can be written as follows:

$$^{PG}E_{2b} = 0.5 * (\delta_1 + \Delta\delta_{PG}) * (LOC + \Delta LOC) \quad (6)$$

where,
ΔδPG is the change in the impact of the effort multiplier of
PG and it can be calculated as δ4-δ1
ΔLOC is the change in the LOC measure and it can be calculated as LOC1-LOC.
Similarly, effort of Optimistic Group can be written as

$$^{OG}E_{2b} = 0.5*(\delta_2 + \Delta\delta_{OG})*(LOC + \Delta LOC) \quad (7)$$

where,
ΔδOG is the change in the impact of the effort multiplier of OG and it can be calculated as δ3-δ2
ΔLOC is the change in the LOC measure and it can be calculated as LOC1-LOC.

Hence, in this case, the increment is done in the LOC measure only. But, though there is no change in the values of effort multipliers, the corresponding impact has been taken into account in accordance to the variation in the value of LOC and accordingly the overall has been estimated. This significant property can be realized through the equations (6) and (7) as in the above case.

CASE2 – If the LOC value is decreased:

In the figure 2(c), the LOC value is decreased from LOC to LOC1 and these decrements are clearly represented in both the triangles. In the figure 2(c), the variations are done only based on the LOC values, but these variations are automatically reflected in the impacts of the effort multipliers. i.e. the value of δ1 is decreased to δ4 and the value of δ2 is decreased to δ3. From figure 2(c), the corresponding overall effort can be estimated as follows:

$$E_{2c} = {}^{PG}E_{2c} * {}^{OG}E_{2c} \quad (8)$$

where,
E2c is the overall effort from the figure 2(c),
PGE2c is the effort due to the Pessimistic Group in the figure2(c)
OGE2c is the effort due to the Optimistic Group in the figure 2(c).
The values of PGE2c and OGE2c can be calculated as follows:

$$^{PG}E_{2c} = 0.5 * LOC_1 * EM * LOC_1 \quad (9)$$

$$^{PG}E_{2c} = 0.5*(\delta_1 - \Delta\delta_{PG})*(LOC - \Delta LOC) \quad (10)$$

where,
ΔδPG is the change in the impact of the effort multiplier of PG and it can be calculated as δ1-δ4.
ΔLOC is the change in the LOC measure and it can be calculated as LOC - LOC1.
Similarly, the equation (10) can also be written as

$$^{OG}E_{2c} = 0.5*(\delta_2 - \Delta\delta_{OG})*(LOC - \Delta LOC) \quad (11)$$

where,
ΔδOG is the change in the impact of the effort multiplier of OG and it can be calculated as δ2-δ3
ΔLOC is the change in the LOC measure and it can be calculated as LOC- LOC1.

CASE3 – If the EM range is increased:

It is also discussed that the range of EM also has the impact on the overall effort. i.e. overall effort increases along with the increments in the range of PG set by keeping the other parameters as constant. In the figure 2(d), the EM range of both the groups are increased. The corresponding overall effort can be estimated as follows:

$$E_{2d} = {}^{PG}E_{2d} * {}^{OG}E_{2d} \tag{12}$$

where,
E2d is the overall effort from the figure 2(d),
PGE2d is the effort due to the Pessimistic Group
2(d) and OGE2d is the effort due to the Optimistic Group in the figure 2(d).
The values of PGE2d and OGE2d can be calculated as follows:

$${}^{PG}E_{2d} = 0.5*(\delta_1 + \Delta\delta_{PG})*LOC \tag{13}$$

$${}^{OG}E_{2d} = 0.5*(\delta_2 - \Delta\delta_{OG})*LOC \tag{14}$$

where,
ΔδOG is the change in the impact of the effort multiplier of OG and it can be calculated as δ2-δ3.
ΔδPG is the change in the impact of the effort multiplier of PG and it can be calculated as δ4-δ1.

CASE 4 – If the EM range is decreased:

In the figure 2(e), the EM range of both the groups is decreased. When the range of OG group decreased from EM to EM1 then the slope of the line cb1 decreases and falls to the point b2, i.e δ2 to δ3, thus the effort increases in OG group, similarly for PG group if the range decreases from EM to EM1, then the slope of the line ca1 decreases to point a2 as shown in the figure 2(e). The corresponding effort can be estimated as follows:

$$E_{2e} = {}^{PG}E_{2e} * {}^{OG}E_{2e} \tag{15}$$

where,
E2e is the overall effort from the figure 2(e),
PGE2e is the effort due to the Pessimistic Group
OGE2e is the effort due to the Optimistic Group in the figure 2(e).
The values of PGE2e and OGE2e can be calculated as follows:

$${}^{PG}E_{2e} = 0.5*(\delta_1 - \Delta\delta_{PG})*LOC \tag{16}$$

$${}^{OG}E_{2e} = 0.5*(\delta_2 + \Delta\delta_{OG})*LOC \tag{17}$$

where,

- ΔδOG is the change in the impact of the effort multiplier of OG and it can be calculated as δ3-δ2.
- ΔδPG is the change in the impact of the effort multiplier of PG and it can be calculated as δ1-δ4.

CASE 5 – if the EM Range of OG is decreased and EM range of PG is increased:

In the figure 2(f), the EM range of OG group is decreased and the EM range of PG group is increases. When the range of OG group decreased from EM to EM1 then the slope of the line cb1 decreases and falls to the point b2, i.e δ2 to δ3, thus the effort increases in OG group, similarly for PG group if the range increases from EM to EM1, then the slope of the line ca1 increases to point a2 as shown in the figure 2(f). From the Figure 2(f), the corresponding overall effort can be estimated as follows:

$$E_{2f} = {}^{PG}E_{2f} * {}^{OG}E_{2f} \tag{18}$$

where,
E2f is the overall effort from the figure 2(f),
PGE2f is the effort due to the Pessimistic Group in the figure 2(f)
OGE2f is the effort due to the Optimistic Group in the figure 2(f).
The values of PGE2f and OGE2f can be calculated as follows:

$$^{PG}E_{2f} = 0.5 * LOC * EM_1 * LOC \tag{19}$$

$$^{OG}E_{2f} = 0.5 * LOC / EM_1 * LOC \tag{20}$$

The equation (19) can also be written as follows:

$$^{PG}E_{2f} = 0.5 * (\delta_1 + \Delta\delta_{PG}) * LOC \tag{21}$$

The equation (24) can also be written as

$$^{OG}E_{2f} = 0.5 * (\delta_2 + \Delta\delta_{OG}) * LOC \tag{22}$$

where,
ΔδPG is the change in the impact of the effort multiplier of PG and it can be calculated as δ4-δ1.
ΔδOG is the change in the impact of the effort multiplier of OG and it can be calculated as δ3-δ2.

Hence, in this case, the changes are made in the EM measure only. But, though there is no change in the values of LOC, the corresponding impact has been taken into account in accordance to the variation in the value of EM and accordingly the overall has been estimated. This significant property can be realized through the equations (21) and (22). By using the equations (3) and (4) the overall effort can be estimated. Since there is no appropriate method available for estimating the impact of the effort multipliers, this scheme may be the first to estimate the impact of the effort multipliers as a function of LOC. This scheme yields a considerable degree of underestimation in overall effort estimation. Hence the output is given as input for further optimization.

$$f(x) = (\prod_{i=1}^{\alpha} 0.5 * LOC^2 * {}^{PG}EM_i * \prod_{j=1}^{\beta} 0.5 * LOC^2 / {}^{OG}EM_j) / LOC^x \tag{23}$$

Where $0 < x < 1$ and $LOC \geq 1$

Further, the above equation is given as input to the Genetic Part for further optimization. This Genetic approach is used for optimization problems and it curtails the residual values with the Hybrid Imputational approach.

Effort Estimation Optimization using Genetic approach:

Genetic algorithm is used in this research to enhance the performance of effort estimation. This approach is useful to select the most relevant features of effort estimates and cost drivers which have certainly significant influence on the COCOMO effort estimation model even if the input level is incomplete. If uncertainty exists in the input parameters, errors in the estimates are more or accuracy may not be good. A brief sketch of improving Effort estimation using GA is shown in figure 3. The figure depicts the flow is a cyclic process of Reproduction, Crossover and Mutation for N populations. Estimation process starts with an initial population of N project (chromosomes) with 17 EM's (genes) as input which follows genetic operations of Reproduction, Crossover and Mutation. In reproduction initial population chromosomes are grouped based on project mode and each chromosomes in each group is ranked based on fitness function. For the opted pair of parental chromosomes crossover non-nominal effort multipliers weightage through adjustment. Weightage of operation is performed for repairing OG each non-nominal multiplier are adjusted by adding or deducting weightage based on, tracking made on all the chromosomes for specific effort multiplier.

PSEUDO CODE FOR EFFORT ESTIMATION

Pseudo code 1 // Effort Estimation using Genetic Algorithm

```
N ← Population size
Pmax← Sub population size
Pn← Number of iteration for generating sub population from N populations
Pmax1 ← 0
Pstart← 1
Pmax← N / 2
Pn←  N /  Pmax
For j ← 1 to (Pn)
For i ←Pstart  toPmax
```

P ← Initialize sub population ( )
P ← Perform Selection ( )
(Offi, Offi+1) ← Crossover operation (Pi, Pi+1)
Offspring Evaluation (Offi, Offi+1)
(Offi', Offi+1') ← Mutation operation on nominal valued genes (Offi, Offi+1)
Offspring Evaluation (Offi', Offi+1')
Minimizing the $f(x)$ for the offsprings by finding the value of 'x'

$$f(x) = \left(\prod_{i=1}^{\alpha} 0.5 * LOC^2 * {}^{PG}EM_i * \prod_{j=1}^{\beta} 0.5 * LOC^2 / {}^{OG}EM_j\right) / LOC^x$$

where, $0 < x < 1$ and $LOC \geq 1$
If j ←PmaxAnd i ≠ Pn then
  Pstart←Pmax +1
  Pmax←Pstart + Pmax
 Else
  Pstart←Pmax +1
  Pmax1 ← N - Pmax
  Pmax←Pstart + Pmax1
 End if
 End For
End For

The general structure of the effort estimation is shown in Pseudo code 1. This codes a solution from underestimation or overestimation of effort multipliers for accurate effort. To code a solution, projects with 17 effort multipliers are chromosomes, where genes are effort multipliers used as initial population. The Schema theorem proposes that "fit individuals having chromosomes with particular patterns of gene that perform well, termed schemata" (Goldberg D.E. 1989). The fit individuals find more chances in reproduction phase these are proliferated into the population and this increase the opportunity of better solutions. GA implicitly conducts a parallel search over a large number of schemata. This implicit parallelism is one of the reasons for the good performance of GA (Beasley D., Bull 1993). GA's find facility to have good building block (Goldberg D.E. 1989) which has short length schemata which tends to enhance the fitness of an individual when integrated into it. Since fit individuals have more chances of reproduction, these schemata are propagated in the population, and the chance of finding better solution increases. Also, since every chromosome has a large number of gene patterns, GA absolutely conducts a parallel search over a huge amount of schemata.

Pseudo code 2 // initial population POP of size $[Pmax \times CMax]$

Start
Step 1: Set$Pmax$, $CMax$
Step 2: Initialize $Pcur = 0$
Step 3: Obtain the range of values for each EM
Step 4: Repeat through Step 6 Until $Pcur <= Pmax$
Step 5: Generate $Indiv$ of size $[1 \times CMax]$
$\forall i\ [1 \leq i \leq CMax], \quad Indiv[i] = UniformRandom(range(EM_i))$
Step 6: Add the generated individual into the Population
  $Pcur = Pcur + 1$; $POP[Pcur] = Indiv$;
Step 7: return POP
Stop
Initial Population is represented as shown below:

Two point crossover operations are performed in generating offspring's. Mutation operation is carried on each nominal gene of offspring's generated which undergoes a series of process for offspring's effort estimation. Effort estimation is made succeeding nominal weightage adjustment of both pessimistic and optimistic group in each offspring. Mutation of 0.01% is carried over nominal genes in the offspring by means of weightage adjustment by adding or deducting from the original weightage. Termination check is made if N population has been completed inorder to terminate the process else the cyclic process from reproduction, crossover and mutation operation will be carried on till completions of effort estimation of N projects. The algorithmic approach and its elements, such as chromosomal representation, initial population, GA operators (such as crossover and mutation) and control parameters used in gene repairing. This codes a solution from underestimation or overestimation of effort multipliers for accurate effort person month estimation. To code the solution project with 17 EM's are chromosome, where genes are effort multipliers used as initial population. To evaluate real value encoding for representing chromosomes has been used a straight forward way to work on real-time data's. Direct proportionality is used in fitness is evaluation considering four genes (LEXP, PCAP, CPLX and ACAP) as a base for best pair of chromosomes as parents. In selection integration of grouping and ranking technique has been followed apart for best parents pairing apart from fitness evaluation. Two point crossover operation ( ) have been used for offspring's generation which used the method of gene Tran-infection and repairing for gene weightage adjustment. Mutation operation ( ) is carried further on offspring's in order to improve the accuracy of effort estimation with 0.01 percentage. Effort estimation through offspring evaluation () will be carried out for whole population.

RELY	DATA	CPLX	RUSE	DOCU	TIME	STOR	PVOL	ACAP	PCAP	PCON	APEX	PLEX	LTEX	TOOL	SITE	SCED
1.09	1.12	1.33	0.97	0.96	1.59	1.11	0.97	0.98	1.32	1.09	1.15	0.89	0.96	1.15	1.16	1.25
0.98	1.14	1.23	1.06	1.08	1.45	1.36	1.14	1.16	0.79	0.95	1.02	0.97	1.10	0.89	1.13	1.36
1.19	1.11	1.29	0.96	0.94	1.46	1.09	0.98	1.00	1.32	1.19	1.25	0.99	1.06	1.05	1.19	1.26
0.99	1.13	1.24	1.05	1.07	1.50	1.33	1.13	0.96	0.79	0.96	1.20	0.87	0.98	1.19	1.03	1.34
1.10	1.12	1.30	0.98	0.96	1.52	1.19	0.99	0.99	1.32	1.10	1.05	0.90	0.99	1.04	1.06	1.28
0.97	1.14	1.25	1.02	1.07	1.47	1.26	1.12	1.06	0.79	1.05	1.08	0.95	1.01	1.11	1.07	1.35
1.08	1.11	1.32	1.00	0.99	1.57	1.11	0.99	0.97	1.32	1.19	1.17	0.93	0.99	1.13	1.08	1.27
0.99	1.12	1.25	1.06	1.08	1.40	1.12	1.11	1.19	0.79	1.15	1.09	1.05	1.11	0.99	1.08	1.36

Experimentation and Result Analysis:

The ability of the COCOMO II and proposed estimation models are examined using standard performance validation measures like Magnitude of Relative Error (MRE), Magnitude of Absolute Error (MAE), Correlation, Root Mean Square (RMS), Mean Magnitude of Relative Error (MMRE), Relative Root Mean Square (RMSE) and PRED. The outcomes of the analyses are recorded in the Table I. It is understood that, the accuracy of COCOMO II is uncertain because 45 projects' scale factor is not in the agreement with the model manual. Only one projects' scale factor lies away from the range in the proposed hybrid model. Due to the above said scale factor, the rest of the performance measures are uncertain in the COCOMO II. Except for the under estimated result all others are satisfactory with the proposed model. In the proposed model 60 projects were underestimated were as in COCOMO II is 49. But comparing with the results of other measures and the amplitude of under estimation it is accepted.

Table I: Overall Comparison Between the COCOMO II and Proposed Model.

Sl. No.	Parameters	COCOMO II Effort Estimation	Proposed Effort Estimation	Impact Status of Proposed Model
1	No. of Cost drives	17	17	+ve
2	No. of Scale Factors	5	5	+ve
3	Scale Factors range	1.01 to 31.62	0.087 to 2.899	+ve
4	No. of Project lie within the SF range	45	95	+ve
5	No. of project does not lie within the SF range	51	1	+ve
6	Maximum MRE (%)	1,073.51	31.46	+ve
7	MMRE (%)	33.65	5.79	+ve
8	RMS	2254.99	51.1603	+ve
9	RRMS	4.43237	0.100559	+ve
10	Correlation Coefficient	0.6952	0.9985	+ve
11	No. of Project over estimated	47	39	+ve
12	Standard deviation for over estimated projects	168.8	10.17	+ve

Wilcoxon Signed-Rank Test:

The Wilcoxon signed-rank test is a simple, nonparametric test that determines the level of bias. A nonparametric test may be thought of as a distribution-free test; i.e. no assumptions about the distribution are made. The better results that can be achieved by the model estimates is to show no difference between the number of estimates that over estimated versus those that under estimated. Another performance measure of a model predicting numeric values is the correlation between predicted and actual values. Correlation ranges from +1 to -1 and a correlation of +1 means that there is a perfect positive linear relationship between variables. It can be calculated as follows:

$$\left.\begin{aligned} &P= \frac{\Sigma_i^T \text{Predicted}_i}{T}, \quad a= \frac{\Sigma_i^T \text{Actual}_i}{T}, \\ &S_{p=} \frac{\Sigma_i^T (\text{Predicted}_i - p)^2}{T-1}, \quad S_{a=} \frac{\Sigma_i^T (\text{Actual}_i - a)^2}{T-1}, \\ &S_{pa=} \frac{\Sigma_1^T (\text{Predicted}_i - p)(\text{Actual}_i - a)}{T-1}, \\ &\text{Corr}= S_{pa} / \sqrt{S_p * S_a} \end{aligned}\right\} \quad (24)$$

From the above assessments, it is clearly understood that the correlation coefficient for COCOMO II is 0.6952 and the correlation coefficient for proposed model is 0.9985, which shows better implication in the proposed model rather than the existing.

Conclusion and Future Work:

The purpose of this research was to provide insight into how sophisticated imputation techniques are and this facilitates the understanding. It also integrates between statisticians and software engineers to make succession effort estimation.

This research has described a variant approach for COCOMO II effort estimation model, by redefining the effort multipliers and the scale factors and thereby the overall accuracy of the COCOMO II has been improved. A unique form of graphical representation schemes and Genetic approach has been used for enhancing the estimation schemes of COCOMO II. The enhancement has been proved and objectively compared with the traditional COCOMO II in terms of the performance validation factors such as MRE, MAE, Pred, MMRE, RMS and RRMS. The observations and analyzes over the obtained results may encourage the researchers to enhance other effort estimation schemes to further levels. All decision making process requires quality data which is of greater importance for any validation process. Furthermore this work can be extended to software specific approach.

REFERENCES

Capers Jones, 2002. “Software cost estimation in 2002”, CROSSTALK - The Journal of Defense Software Engineering, pp: 4-8.

Iris Fabiana de Barcelos Tronto, Jose Demı sioSimoes da Silva, 2008. An investigation of artificial neural networks based prediction systems in software project management, , The Journal of Systems and Software, 81: 356-367 Elsevier.

Jovan Živadinović, Ph.D, Zorica Medić, Dragan Maksimović, Aleksandar Damnjanović, M.Sc, Slañana Vujčić, R.J.A. Little, D.B. Rubin, 1987. “Statistical analysis with missing data”, Wiley series in probability and statistics, 1st edn. Wiley, New York.

“Methods of effort estimation in software engineering", International Symposium Engineering Management And Competitiveness 2011 (EMC2011)June 24-25, 2011, Zrenjanin, Serbia.

Parag C. Pendharkar, Girish H. Subramanian and James A. Rodger, 2005. “A Probabilistic Model for Predicting Software Development Effort.”IEEE Transactions on Software Engineering.

Rajib Mall, 2009. “Fundamentals of Software Engineering”, PHI, 3rd Edition.

Roger S. Pressman, “Software Engineering – A Practitioner’s Approach”, McGraw-Hill International, 6th Edition.

Satyananda, 2009. “An improved fuzzy approach for COCOMO’s effort estimation using Gaussian Membership Function” Journal of Software.

Chulani, S., B. Boehm and B. Steece, 1999. Bayesian Analysis of Empirical Software Engineering Cost Models, IEEE Trans. Software Eng., 25(4).

Tim Menzies, Member, IEEE, zhihacchen, Jairus Hihn and Karen Lum, 2006. Selecting Best Practices for Effort Estimation, IEEE Transaction on Software Engineering, 32(11).

Witten, I.H. and E. Frank, 2005. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco, CA, 2nd edition.