Software Cost Estimation Techniques Using Soft Computing

Shiva Suryanarayana Ch. and Dr. Satya Prakash Singh

Computer Science & Engineering

1BIT, Mesra, Ranchi, Noida Campus India, India, 2Ras Al Khaimah, Dubai Campus

Abstract: - The software cost estimation is the process of predicting the most realistic amount of effort required to develop or maintain software based on incomplete, uncertain and noisy input. Effort estimates may be used as input to project plans, iteration plans, budgets, and investment analyses, pricing processes and bidding rounds. For example, estimation of value estimation and danger examination is a real issue in programming undertaking administration. As programming advancement has turned into a fundamental speculation for some associations, precise programming expense estimation models are expected to adequately foresee, screen, control and evaluate programming improvement. Programming expense estimation is a testing and cumbersome errand. Then again, the approach utilized for the estimation of programming exertion by relationship is not ready to handle the unmitigated information in an unequivocal and exact way. The nature of an expense estimation model is less ascribed to the introductory assessment, but instead the pace at which the appraisals merges to the genuine expense of the task. COCOMO is a well known algorithmic model for expense estimation whose expense variables can be customized to the individual improvement environment, which is imperative for the precision of the expense gauges. More than one strategy for expense estimation ought to be carried out so that there is some correlation accessible for the evaluations. This is particularly imperative for extraordinary undertakings. Cost estimation must be done more diligently throughout the project life cycle so that in the future there are fewer surprises and unforeseen delays in the release of a product. In this paper, we present a soft computing framework to tackle this challenging problem. Estimating the work-effort and the schedule required to develop and/or maintain a software system is one of the most critical activities in managing software projects.

Keywords: Software Engineering, Software Cost Estimation, COCOMO model, Soft Computing Techniques.

I. Introduction

Software cost estimation is process of predicting the effort required to develop a software engineering project. While the software cost estimation may be simple in concept, it is troublesome and perplexing truly. The genuine bit of cost of programming change is a direct result of human-effort and most cost estimation frameworks focus on this viewpoint and give assesses in regards to individual month. It chooses the measure of effort imperative to complete an item wander in terms of its booking, securing of resources, and meeting of arrangement essentials. The effective and viable change of the programming requires exact gages. Programming researchers are giving various cost estimation frameworks for a couple of decades yet the essential issue hang on in programming building field. Early programming estimation models can't avoid being constructed in light of the backslide examination or exploratory derivations. Among those systems, COCOMO II is the most normally used
model. Today's models are concentrated around reenactment, neural framework, genetic computation, fragile handling, the soft method of reasoning showing etc. In this paper, COCOMO II model utilized the most much of the time and generally utilized hereditary calculation approach for upgrading the current coefficients that gauge the streamlined prescient exertion obliged for the advancement of programming undertaking.

1.1 Cost Estimation Process: To comprehend the finished result or the yields of the product cost estimation process we should first comprehend what is programming expense estimation process. By definition, programming expense estimation methodology is a situated of strategies and methods that is utilized to infer the product expense gauge. There is normally a situated of inputs to the methodology and afterward the procedure utilizes these inputs to produce or ascertain a set of yields.

1.2 Classical View: Most of the product cost estimation models sees the estimation transform just like a capacity that is registered from a situated of expense drivers. Also in most cost estimation methods the essential expense driver or the most critical expense driver is accepted to be the product necessities. As showed in figure 1, in an established perspective of programming estimation prepare, the product necessities are the essential information to the procedure furthermore structure the premise for the expense estimation. The expense assessment will then be balanced likewise to various other expense drivers to land at the last gauge. So what is expense driver? Fetched driver is anything that may or will influence the expense of the product. Taken a toll driver are things, for example, outline procedure, ability levels, hazard evaluation, staff experience, programming dialect or framework unpredictability.

In a traditional perspective of the estimation process, it will produce three yields - endeavors, term and stacking. The accompanying is a concise depiction of the yields:

- Manpower stacking - number of work forces (which likewise incorporates administration faculty) which are distributed to the venture as a capacity of time.

- Project term - time that is expected to finish the task.

- Effort - measure of exertion needed to finish the undertaking and is generally measured in units as man-months (MM) or individual months (PM).

The yields (stacking, term and exertion) are generally registered as settled number with or without resilience in the traditional perspective. However actually, the expense estimation procedure is more unpredictable than what is indicated in figure 1. Large portions of the information that are inputs to the methodology are altered or refined amid the product cost estimation process.
1.3 Actual View: In the actual cost estimation process there are other inputs and constraints that needed to be considered besides the cost drivers. One of the primary constraints of the software cost estimate is the financial constraint, which are the amount of the money that can be budgeted or allocated to the project. There are other constraints such as manpower constraints, and date constraints. Other input such as architecture, which defines the components that made up the system and the interrelationships between these components. Some company will have certain software process or an existing architecture in place; hence for these companies the software cost estimation must base their estimates on these criteria.

There are only very few cases where the software requirements stay fixed. Hence, how do we deal with software requirement changes, ambiguities or inconsistencies? During the estimation process, an experienced estimator will detect the ambiguities and inconsistency in the requirements. As part of the estimation process, the estimator will try to solve all these ambiguities by modifying the requirements. If the ambiguities or inconsistent requirements stay unsolved, this will correspondingly affect the estimation accuracy.

1.4 Cost Estimation Accuracy: The cost estimation accuracy helps to determine how well or how accurate our estimation is when using a particular model or technique. We can assess the performance of the software estimation technique by:

- Absolute Error (Epred - Eact)
- Percentage or Relative Error (Epred - Eact) / Eact
- Mean Magnitude of Relative Error

Each of the error calculation techniques has advantages and disadvantages. For example, absolute error fails to measure the size of the project, and mean magnitude of relative error will mask any systematic bias.

Method of Cost Estimation: There is a lot of software cost estimation methods or techniques in the software industry. Here are a few techniques that will be discussed in this Paper:
- Algorithmic model
- Expert Judgment (Expertise Based)
II. Algorithmic Model

These product cost estimation procedures utilize the numerical mathematical statements to perform the product estimation. The numerical correlations are concentrated around chronicled data or theory. SLOC (source line of code), limit centers, and other cost drivers are the inputs. For most algorithmic model, arrangement to the specific programming environment can be performed to improve the estimation. Delineations of the parametric models are COCOMO (Constructive Cost Model), COCOMO II, Putnam's item life-cycle model (SLIM).

As per paper, COCOMO II post structural engineering system ascertains the product advancement exertion (in individual months) by utilizing the accompanying comparison:

\[ \text{Effort} = A \times (\text{SIZE})^E \times \prod_i EM_i \]  
\[ \text{Duration: } TDEV = C \times (\text{Effort})^F \]  

Where, \( A \) - multiplicative constant with value 2.94 that scales the effort according to specific project conditions. \( \text{Size} \) - Estimated size of a project in Kilo Source Lines of Code or Unadjusted Function Points.
\( E \) - An exponential factor that accounts for the relative economies or diseconomies of scale encountered as a software project increases its size
\( EM_i \) - Effort Multipliers.

The coefficient \( E \) is determined by weighing the predefined scale factors (SF_i) and summing them using following equation:

\[ E = 0.91 + 0.01 \sum_i SF_i \]  

The development time (TDEV) is derived from the effort according to the following equation:

\[ TDEV = C \times (\text{Effort})^F \]  

Latest calibration of the method shows that the multiplier \( C \) is equal to 3.67 and the coefficient \( F \) is determined is a similar way as the scale exponent by using following equation:

\[ F = 0.28 + 0.002 \sum_i SF_i \]  

According to paper [6], when all the factors and multipliers are taken with their nominal values, then the equations for effort and schedule are given as follows:

\[ \text{Effort} = 2.94 \times (\text{Size})^{1.1} \]  
\[ \text{Duration: } TDEV = 3.67 \times (\text{Effort})^{3.18} \]  

COCOMO II is clear and effective calibration process by combining Delphi technique with algorithmic cost estimation techniques. It is tool supportive and objective. This model is repeatable, versatile. But its limitation is that most of extensions are still experimental and not fully calibrated till now.

Advantages:
- Generate repeatable estimations
- Easy to adjust information
- Easy to refine and modify equations
- Objectively aligned to experience

Disadvantages:
- Unable to manage remarkable conditions
- Some experience and components cannot be measured
- Sometimes calculations may be restrictive
2.1 Expert Judgment

This method catches the experience and the information of the estimator who gives the appraisal focused around their experience from a comparative task to which they have partaken. Illustrations are the Delphi, Wideband Delphi and Work Breakdown Structure (WBS).

Preferences:
- Useful without measured, experimental information.
- Can figure in contrasts between past venture encounters and necessities of the proposed undertaking
- Can calculate in effects brought on by new innovations, applications and dialects.

Disadvantages:
- Estimate is just as great master's supposition
- Hard to archive the components utilized by the masters

2.2 Top-Down

This strategy is additionally called Macro Model, which use the worldwide perspective of the item and afterward parcelled into different low level segments. Illustration of this procedure is the Putnam model.

Points of interest:
- requires negligible undertaking subtle element
- usually quicker and simpler to execute
- focus on framework level exercises

Drawbacks:
- tend to ignore low level part

III. Literature Review

Numerous programming expense estimation models have been produced in the course of the most recent decades. A late study by Jorgensen gives a point by point survey of diverse studies on the product improvement exertion. Numerous analysts have connected the neural systems approach to gauge programming advancement exertion. A wide range of models of neural systems have been proposed. They may be assembled in two noteworthy classifications. Initial one is nourishing forward systems where no circles in the system way happen. Another is criticism arranges that have recursive circles. Understanding the affliction in applying neural systems, Nasser Tadayon has proposed a dynamic neural system that will at first use COCOMO II Model. COCOMO, be that as it may, has some limits. It can’t adequately bargain with uncertain and indeterminate data, and alignment of COCOMO is a standout amongst the most essential errands that need to be carried out so as to get precise estimations. Thus, there is dependably scope for creating exertion estimation models with better prescient precision.
**Some example of estimation methods inside each class:-**

<table>
<thead>
<tr>
<th>Estimation approach</th>
<th>Category</th>
<th>Examples of support of implementation of estimation approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analogy-based</td>
<td>Formal estimation</td>
<td>ANGEL, Weighted Micro Function Points</td>
</tr>
<tr>
<td>estimation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WBS-based (bottom</td>
<td>Expert estimation</td>
<td>Project management software, company specific activity templates</td>
</tr>
<tr>
<td>up) estimation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parametric models</td>
<td>Formal estimation</td>
<td>COCOMO, SLIM, SEER-SEM, True Planning for Software</td>
</tr>
<tr>
<td>Size-based estimation models</td>
<td>Formal estimation</td>
<td>Function Point Analysis, Use Case Analysis, SSU, Story points-based estimation in Agile software development</td>
</tr>
<tr>
<td>Group estimation</td>
<td>Expert estimation</td>
<td>Planning poker, Wideband Delphi</td>
</tr>
<tr>
<td>Mechanical combination</td>
<td>Combination-based estimation</td>
<td>Average of an analogy-based and a Work breakdown structure-based effort estimate</td>
</tr>
<tr>
<td>Judgmental combination</td>
<td>Combination-based estimation</td>
<td>Expert judgment based on estimates from a parametric model and group estimation</td>
</tr>
</tbody>
</table>

**IV. COCOMO**

COCOMO is no doubt the most popular method for doing software cost estimation. The estimations are relatively easy to do by hand. There also are tools available which allow you to calculate more complex estimation. Calibration of COCOMO is one of the most important things that needs to be done in order to get accurate estimations. Even though COCOMO may be the most popular estimation method it is recommended that you always use another method of estimation to verify your results. The other method should differ significantly from COCOMO. This way your project is examined from more than one angle and something that you may have overlooked when using COCOMO is not overlooked again.

COCOMO stands for Constructive Cost Model, it is a software cost estimation model that was first published in 1981 by Barry Boehm (2001). It is an algorithmic approach to estimating the cost of a software project. By using COCOMO you can calculate the amount of effort and the time schedule for projects. From these calculations you can then find out how much staffing is required to complete a project on time. COCOMO's main metric used for calculating these values is lines of code (denoted KLOC for COCOMO II, or KDSI for COCOMO 81 and measured in thousands), function points (FP), or object points (OP).

COCOMO also lets you check out ‘what if’ scenarios where by adjusting certain factors in COCOMO you can see how a projects time and effort estimates change as well (Bohem 2001). There have been a few different versions of COCOMO; the two that are discussed in this report are COCOMO 81 and COCOMO II.
V. COCOMO 81

COCOMO 81 was the first version of COCOMO. It was modeled around software practices of the 1980’s. It has been found that on average it is able to produce estimates that are within 20% of the actual values 68% of the time. COCOMO 81 has three different models that can be used throughout a projects life cycle (Bohem,2001)

Basic Model – this model would be applied early in a projects development. It will provide a rough estimate early on that should be refined later on with one of the other models.
- Intermediate Model – this model would be used after you have more detailed requirements for a project.
- Advanced Model – when your design for a project is complete you can apply this model to further refine your estimate.

Within each of these models there are also three different modes. The mode you choose will depend on your work environment, and the size and constraints of the project itself. The modes are:
- Organic – this mode is used for “relatively small software teams developing software in a highly familiar, in-house environment”.
- Embedded – operating within tight constraints where the product is strongly tied to a “complex of hardware, software, regulations and operational procedures”.
- Semi-detached – an intermediate stage somewhere in between organic and embedded. Projects are usually of moderate size of up to 300,000 lines of code.

Equations Used

There are two main equations that are used to calculated effort and schedule time (measured in months). They are:

Equation 1 \[ PM = a(KDSI)b \times EAF \]
Equation 2 \[ TDEV = c(PM)d \]

Where:
- \( PM \) is effort in person-months
- \( EAF \) is the effort adjustment factor
- \( TDEV \) is the schedule time
- \( KDSI \) is the number of lines of code (in thousands)
- \( a, b, c, \) and \( d \) are all constants based on the mode you are using (refer to Table 1)

<table>
<thead>
<tr>
<th>Model</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organic</td>
<td>2.4</td>
<td>1.05</td>
<td>2.5</td>
<td>0.38</td>
</tr>
<tr>
<td>Semi-detached</td>
<td>3.0</td>
<td>1.12</td>
<td>2.5</td>
<td>0.35</td>
</tr>
<tr>
<td>Embedded</td>
<td>3.6</td>
<td>1.20</td>
<td>2.5</td>
<td>0.32</td>
</tr>
</tbody>
</table>

5.1 Cost Drivers

The EAF is used to tailor your estimate based on conditions of the development environment. For the basic model it is not used and just set to 1. For the intermediate model there are 15 different cost drivers that can be used to calculate your EAF. They are grouped into 4 different categories; product attributes, computer attributes, personal attributes, and project attributes (see Table 2). Each cost driver is rated on a scale Very Low to Extra High depending on how that cost driver will affect your development. These ratings are based on a statistical analysis of historical data collected from 83 past projects.
To calculate the EAF from the cost drivers you simply choose values for each cost driver and multiply them all together. The resulting number is your EAF.

<table>
<thead>
<tr>
<th>Category</th>
<th>Cost Driver</th>
<th>Very Low</th>
<th>Low</th>
<th>Nominal</th>
<th>High</th>
<th>Very High</th>
<th>Extra High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Attributes</td>
<td>RELY Required Software Reliability</td>
<td>0.75</td>
<td>0.88</td>
<td>1.00</td>
<td>1.15</td>
<td>1.40</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DATA Database Size</td>
<td>-</td>
<td>0.94</td>
<td>1.00</td>
<td>1.08</td>
<td>1.16</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CPLX Product Complexity</td>
<td>0.70</td>
<td>0.85</td>
<td>1.00</td>
<td>1.15</td>
<td>1.30</td>
<td>1.65</td>
</tr>
<tr>
<td>Computer Attributes</td>
<td>TIME Execution Time Constraint</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>1.11</td>
<td>1.30</td>
<td>1.66</td>
</tr>
<tr>
<td></td>
<td>STOR Main Storage Constraint</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>1.06</td>
<td>1.21</td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td>VIRT Virtual Machine Volatility</td>
<td>-</td>
<td>0.87</td>
<td>1.00</td>
<td>1.15</td>
<td>1.30</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>TURN Computer Turnaround Time</td>
<td>-</td>
<td>0.87</td>
<td>1.00</td>
<td>1.07</td>
<td>1.15</td>
<td>-</td>
</tr>
<tr>
<td>Personnel Attributes</td>
<td>ACAP Analyst Capability</td>
<td>1.46</td>
<td>1.19</td>
<td>1.00</td>
<td>0.96</td>
<td>0.71</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>AEXP Applications Experience</td>
<td>1.29</td>
<td>1.13</td>
<td>1.00</td>
<td>0.91</td>
<td>0.82</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>PCAP Programmer Capability</td>
<td>1.42</td>
<td>1.17</td>
<td>1.00</td>
<td>0.86</td>
<td>0.70</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>VEXP Virtual Machine Experience</td>
<td>1.21</td>
<td>1.10</td>
<td>1.00</td>
<td>0.90</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LEXP Language Experience</td>
<td>1.14</td>
<td>1.07</td>
<td>1.00</td>
<td>0.95</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Project Attributes</td>
<td>MODP Modern Programming Practices</td>
<td>1.24</td>
<td>1.10</td>
<td>1.00</td>
<td>0.91</td>
<td>0.82</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>TOOL Use of Software Tools</td>
<td>1.24</td>
<td>1.10</td>
<td>1.00</td>
<td>0.91</td>
<td>0.83</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SCED Required Development Schedule</td>
<td>1.23</td>
<td>1.08</td>
<td>1.00</td>
<td>1.04</td>
<td>1.10</td>
<td>-</td>
</tr>
</tbody>
</table>

The advanced model of COCOMO 81 goes one step further than the intermediate model in that it uses costs drivers that are rated differently depending on the current phase that a project is in. One of the problems with using a model like COCOMO 81 today is that it does not match the development environment of the late 1990’s and 2000’s. It was created in a time when batch were the norm, programs were run on mainframes and compile times were measured in hours not seconds. It is outdated for use in today’s development environment (rapid application development, 4th generation languages etc) so in 1997 COCOMO II was published and was suppose to solve most of these problems.

5.2 COCOMO II

COCOMO II was published in 1997 and is an updated model that addresses the problems with COCOMO 81. The main objectives of COCOMO II were set out when it was first published. They are:

- To develop a software cost and schedule estimation model tuned to the life cycle practices of the 1990’s and 2000’s.
- To develop software cost database and tool support capabilities for continuous model improvement.
- To provide a quantitative analytic framework, and set of tools and techniques for evaluating the effects of software technology improvements on software life cycle costs and schedules.

For the most part estimates are obtained in pretty much the same way as COCOMO 81. The main changes have been in the number and type of cost drivers and the calculation of equation variables rather then the use of constants (for a detailed look at the specific differences between COCOMO 81 and COCOMO II see). The equations still use lines of code as their main metric, you can however also using

function points and object points to do estimates. The line of code metric used is now the LOC. There are standards set out by SEI for proper counting of lines, things like if/then/else statements would be counted as one line (there are automated tools that will do the counting for you when you want to collect data from your own code).

COCOMO II again has three models, but they are different from the ones for COCOMO 81. They are:

- **Application Composition Model** – this would be used for projects built using rapid application development tools. Normally you would use object points for size estimates. It “involves prototyping efforts to resolve potential high-risk issues such as user interfaces, software/system interaction, performance, or technology maturity.”.

- **Early Design Model** – This model can provide you with estimates early in a projects design before the entire architecture has been decided on. Normally you would use function points as a size estimate. It “involves exploration of alternative software/system architectures and concepts of operation. At this stage, not enough is generally known to support fine-grain cost estimation.”.

- **Post-Architecture Model** – The most detailed on the three, used after the overall architecture for the project has been designed. You could use function points or LOC’s for size estimates. It “involves the actual development and maintenance of a software product”.

### 5.2.1 Cost Drivers

In COCOMO II there are 17 cost drivers that are used in the Post-Architecture model. They are used in the same way as in COCOMO 81 to calculate the EAF. The cost drivers are not the same ones as in COCOMO 81; they are better suited for the software development environment on the 1990’s and 2000’s. They are grouped together as shown in table 3. We will not go into specific details on all of the cost drivers here as that information can be found in the paper “Cost Models for Future Software Life Cycle Processes: COCOMO 2.0”. The cost drivers for COCOMO II are again rated on a scale from Very Low to Extra High in the same way as in COCOMO 81.

**Example:-** Assume that the size of an organic type software product has been estimated to be 32,000 lines of source code. Assume that the average salary of software engineers be Rs. 15,000/- per month. Determine the effort required to develop the software product and the nominal development time. From the basic COCOMO estimation formula for organic software:

Effort = $2.4 \times (32)^{1.05} = 91$ PM

Nominal development time = $2.5 \times 91^{0.38} = 14$ months

Cost required to develop the product = $14 \times 15,000 = \text{Rs. 210,000/-}$

### VI. Conclusion

The paper suggests soft computing approach for estimating of software project development cost and time. Most important issue in software project management is accurate and reliable estimation of software development effort and cost. This is more essential particularly in the early period of programming improvement so that the chief may submit his assets for on time conveyance of the product. Programming advancement is famous for going after some time and plan. This issue is because of the way that product improvement is a complex methodology due to the quantity of components included, including the human element, and the many-sided quality of the item that is created. Besides, the business is exceedingly aggressive. COCOMÓ is a prominent exact estimation display that has been joined into a few devices. Any model ought to be adjusted to the improvement environment on the grounds that all advancement situations are distinctive. In the end, an accurate estimate cannot be guaranteed and so using more than one method of estimation is recommended for verification of an estimate.
VII. References


