Effectual Software Design using Software Matrices and COCOMO-II

Dheeraj Agrawal  
Department of computer science  
SSCET, Bhilai  
dheeraj_arg@rediffmail.com

Megha Mishra  
Department of computer science  
SSCET, Bhilai

Abstract—Software engineering means the application of a systematic, disciplined and quantifiable approach to the development, operation, and maintenance of software. As we are becoming more and more reliant on software systems, more effort is being put on development of efficient, reliable and cost-effective software. Software measurement helps us understand software, manage a software project and improve the software process itself. We believe a method of defining software data sets is necessary to ensure that software data are trustworthy. Software companies introducing a measurement program need to establish procedures to collect and store trustworthy measurement data. Without appropriate definitions it is difficult to ensure data values are repeatable and comparable. Software quality analysis of program modules with identifies defect or quality-based class labels. Hence it is vital to develop a tool for effectual software design. This paper presents our approach on design a web based tool for integrating the estimation (using COCOMO II model), planning & tracking, calibration process and software matrices used to evaluate software defects. Software design patterns based on C&K and Mood matrices.

Keywords—COCOMO-II, Matrices, Software design, Estimation, Software Defects.

1. INTRODUCTION

Software development time and cost estimation are the process of estimating the most realistic use of time and cost required for developing a software. Cost estimation methodology for web-based application is very important for software development as it would be able to assist the management team to estimate the cost. Furthermore, it will ensure that the development of cost is within the planned budget and provides a fundamental motivation towards the development of web-based application project [21] [22]. Software development effort typically includes human effort expended for high-level design, detailed design, coding, unit testing, integration testing, and customer acceptance testing. Effort is often regarded as a surrogate for software development cost since personnel cost is the dominant cost in software development [15]. When project is under development it is necessary to take feedback from the development process and analyse the status of project.

The COCOMO II capability for estimation of Application Generator, System Integration, or Infrastructure developments is based on Application Composition model (for early prototyping efforts) and two increasingly detailed estimation models for subsequent portions of the life cycle, Early Design and Post-Architecture.

In estimating a cost, the tool considers all variables in COCOMO II and requires expert judgment to key in the input of the variables such as project size, project type, cost adjustment factor and cost driven factor [22].

The measurement of software quality is traditionally based upon 1) complexity and 2) design quality Metrics. The first research contributions were aimed at providing operating definitions and metrics of software complexity, focusing on the analysis of the code’s information flow [14].

The object-oriented programming paradigm, coupling, cohesion, inheritance, and information hiding have been identified as the basic properties of software design quality [9], [14], [23], [24]. Based on these four basic properties, a number of metrics have been proposed to evaluate the design quality of object-oriented software. The most widely known metrics were first proposed by Chidamber and Kemerer [25] (WMC, NOC, DIT, RFC, LCOM, and CBO) and by F.B. Abreu [9] (COF, PF, AIF, MIF, AHF, and MHF).

Software defects play a key role in software reliability, and the number of remaining defects is one of the most important software reliability indexes. Observing the trend of the number of remaining defects during the testing process can provide very useful information on the software reliability [4].

Software tools can improve the quality and maintainability of software, but are expensive to acquire, deploy, and maintain, especially in large organizations individually [21].
II. LITERATURE REVIEW

A. COCOMO II

COCOMO II follows the openness principles used in the original COCOMO. Thus, all of its relationships and algorithms will be publicly available. Also, all of its interfaces are designed to be public, well-defined, and parameterized, so that Complementary pre-processors (analogies, case-based, or other size estimation models), post-processors (project planning and Control tools, project dynamics models, risk analysers), and higher level packages (project management packages, product Negotiation aids), can be combined straightforwardly with COCOMO II. To support the software marketplace sectors above, COCOMO II provides a family of increasingly detailed software cost estimation models, each tuned to the sectors’ needs and type of information available to support software cost estimation.

COCOMO II provides the following three-stage series of models for estimation of Application Generator, System Integration, and Infrastructure software projects:
1. The earliest phases or spiral cycles will generally involve prototyping, using the Application Composition model capabilities. The COCOMO II Application Composition model supports these phases, and any other prototyping activities occurring later in the life cycle.
2. The next phases or spiral cycles will generally involve exploration of architectural alternatives or incremental development strategies. To support these activities, COCOMO II provides an early estimation model called the Early Design model. This level of detail in this model is consistent with the general level of information available and the general level of estimation accuracy needed at this stage.
3. Once the project is ready to develop and sustain a fielded system, it should have a life-cycle architecture, which provides more accurate information on cost driver inputs, and enables more accurate cost estimates. To support this stage, COCOMO II provides the Post-Architecture model.

COCOMO II is a fully documented and widely accepted model, updated from the original COCOMO. COCOMO II uses TOOL, one of the Effort Multipliers in its Post-Architecture model to capture the impacts on software productivity and provides a TOOL rating scale based just on the completeness of tool coverage as a guideline for the evaluation of software tools [6].

B. Metrics and Measures

A metric is a quantitative measure of the degree to which a system, component, or process. Software measurement applies to a software engineering process there by measuring numerous entities encountered along the way. According to Dumke [28], software measurement is directed to three main components in the object-oriented software development.

The process measurement for understanding, evaluation and improvement of the development method, the product measurement for the quantification of the product (quality) characteristics and validation of these measures, the resource measurement for the evaluation of the supports (CASE tools, measurement tools etc.) and the chosen implementation system [7].

C. Chidamber & Kemerer Metrics Suite

This metrics suite was proposed in [25] by S. R. Chidamber and C. F. Kemerer. The structural design metrics proposed by them are explained here.

- Weighted Method per Class (WMC)
It is sum of complexities of all methods in a class. Consider a class C1 with methods M1, . . . , Mn that are defined in the class. Let c1, . . . , cn be complexities of each of these methods. For this work, complexity of each method is assumed to be unity and so WMC is simply sum of all defined methods.

- Depth of Inheritance Tree (DIT)
Depth of inheritance of the class is the DIT metric for the class.

- Number of Children (NOC)
Number of immediate sub-classes subordinated to a class in class hierarchy. C&K-Java Binding: It is the number of immediate subclasses of a class. For an interface it is the number of classes implementing it plus number of other interfaces extending this interface.

- Coupling between Objects (CBO)
CBO for a class is count of the number of other classes to which it is coupled. Two classes are coupled together if methods of one use methods or instance variables of other. Excessive coupling between object classes is detrimental to modular design and prevents reuse. The more independent a class is, the easier it is to reuse it in another application.

- Response for a Class (RFC)
RFC = | RS | where RS is response set for the class. This is a set of methods that can potentially be executed in response to a message received by an object of that class. Since it specifically includes methods called from outside the class, it is also a measure of the potential communication between the class and other classes.

- Lack of Cohesion in Methods (LOCM)
The LOCM is a count of the number of method pairs whose similarity is 0 minus the count of method pairs whose similarity is not zero. The degree of similarity for two methods M1 and M2 in class C1 is given by: s() = |{I1} \{I2} where |{I1} and |{I2} are the sets of instance variables used by M1 and M2. The larger the number of similar methods, the more cohesive the class, which is consistent with traditional
notions of cohesion that measure the inter-relatedness between portions of a program. A high cohesion is favoured in class designs.

C&K-Java Binding: Instance variables are the ones with any access modifier.

D. MOOD Metrics Set

F. B. Abreu proposed these system-level metrics in [27]. This set of six metrics measures four main structural mechanisms of object-oriented design that is encapsulation (Method Hiding Factor and Attribute Hiding Factor), inheritance (Method Inheritance Factor and Attribute Inheritance Factor), polymorphism (Polymorphism Factor) and message-passing (Coupling Factor). An explanation of the metrics with Java bindings follows except for coupling factor which was not measured. Common Java Binding Note: This set of metrics applies to system level. While measuring these metrics for standard libraries like J2SE etc.

- Method Hiding Factor (MHF)

\[ MHF = \frac{\sum_{i=1}^{TC} M_i(C_i) - (1 - V(M_m))}{\sum_{i=1}^{TC} M_i(C_i)} \]

where:

\[ V(M_m) = \frac{\sum_{i=1}^{TC} M_i(C_i)}{TC} \]

and:

\[ is\_visible(M_m, C_j) = \begin{cases} 1 & \text{iff } \# \neq i \text{ and } C_j \text{ may call } M_m \smallskip \text{0 otherwise} \end{cases} \]

- Attribute Hiding Factor (AHF)

\[ AHF = \frac{\sum_{i=1}^{TC} A_i(C_i) - (1 - V(A_a))}{\sum_{i=1}^{TC} A_i(C_i)} \]

where:

\[ V(A_a) = \frac{\sum_{i=1}^{TC} A_i(C_i)}{TC - 1} \]

and:

\[ is\_visible(A_a, C_j) = \begin{cases} 1 & \text{iff } \# \neq i \text{ and } C_j \text{ may reference } A_a \smallskip \text{0 otherwise} \end{cases} \]

MOOD-Java Binding: Ad (Ci) – number of all attributes with any access modifier including abstract and inherited methods.

- Method Inheritance Factor (MIF)

\[ MIF = \frac{\sum_{i=1}^{TC} M_i(C_i)}{\sum_{i=1}^{TC} M_a(C_i)} \]

Where Ma (Ci) = Md (Ci) + Mi (Ci)

The numerator is the sum of inherited methods in all classes of the system. The denominator is the total number of available methods in all classes.

MOOD-Java Binding: Mi (Ci) – number of inherited methods but not overridden, Md (Ci) – number of defined non-abstract methods with any access modifier, Ma (Ci) – number of methods that class Ci can call.

- Attribute Inheritance Factor (AIF)

\[ AIF = \frac{\sum_{i=1}^{TC} A_i(C_i)}{\sum_{i=1}^{TC} A_a(C_i)} \]

Where, Aa (Ci) = Ad (Ci) + Ai (Ci)

It is defined analogous to MIF.

MOOD-Java Binding: Ai (Ci) – number of inherited attributed, Ad (Ci) – number of defined attributes with any access modifier, (Ci) – number of attributes that class Ci can reference.

E. Software Defect Estimation

Software defects play a key role in software reliability, and the number of remaining defects is one of most important software reliability indexes. Observing the trend of the number of remaining defects during the testing process can provide very useful information on the software reliability. However, the number of remaining defects is not known and has to be estimated [9]. Therefore, it is important to study the trend of the remaining software defect estimation (ISCA algorithm describe next section).

There has been some research on the trend of the software defects. Early studies of defect occurrences suggest that it follows a Rayleigh curve [2], [30], [31] roughly proportional to project staffing. McConnell [32] discusses the relationship between defect rate and development time, indicating that the projects achieving the lowest defect rates also achieve the shortest schedules.

F. Design Patterns

Design patterns describe good solutions to common and recurring problems in software design. They have been widely applied in many software systems in industry. However, pattern related information is typically not available in large system implementations. Recovering these design pattern instances in software systems can help not only to understand the original design decisions and tradeoffs but also to change the systems with quality assurance. Design patterns using an XML file, which include their structural, behavioural and semantic characteristics. These pattern characteristics are used in different phases. During structural analysis phase, our tool extracts the structural information of the pattern and encodes it into a matrix and weights in a similar way as we encode the system. Thus, the structural analysis can be reduced to the matching of the design pattern matrix with the system matrix as well as the weights of the design pattern classes with the weights of the system classes [1], [11].

III. APPROACH OVERVIEW

The system will automate the process of the estimation using the COCOMO II model [21] for effort estimation. The
system will also help in tracking the status of project by tracking daily input from each developer in the organization and will show the status in the form of a Gantt chart. The system will generate the reports for the projects. Identify Software defects play a key role in software reliability, investigates an approach to incorporate the time dependencies between the fault detection, and the number of remaining defects is one of most important software reliability indexes. Administrator control overall system module.

A. System Overview

![Fig.1 A Proposal of System Design Tool](image)

B. Development Phase

The proposed System has three development phases.

1. Phase I

Phase I was dedicated to the database design, designing the system and for developing the part which estimates the size, effort and schedule for the project along with the programs for inserting the data into the backend and for its manipulation. An interactive and user friendly interface with an accurate estimation model was the goal of this phase [12].

1. Estimation of the size of the intended project. This results in either source lines of code (SLOC) or function point counts (FPC) or new object points (NOP) for the project but other measures for the size are also available.
2. Estimation of the effort for the project in man-months or man-hours.
3. Estimation of the schedule in calendar-months.

The information source for estimation can be the project proposal, system specification or software requirement specification. If the size estimation is being done in the later stages such as design or during coding, then design specifications and other work products can be used as information source for estimation [12], [7].

1. By Analogy: If similar projects have been experienced by the organization then with the help of past experience the size for the new project can be estimated. This is performed by dividing the new project into small modules and comparing those modules with the past project data. This method can give almost the accurate estimate for the project size if the past projects were similar to the new one [19].

2. By Parametric Measurement: The size could be estimated by counting features of the project and using them as parameter for any parametric measurement approach like object point analysis or function point analysis. Even if the organization has no experience of the intended project, the features of the project can be used for parametric measurement.

2. Phase II

Phase II develop a Metrics Attributes Calculation Module (MACM). These modules are used to Calculate Attributes of Mood and Ck matrices.

![Fig. 2 Design a tool for calculating the Matrices attribute](image)

3. Phase III

Phase III Design an Algorithm Integrated Software Calibration Algorithms (ISCA). Using ISCA algorithm to identify software Defects. Major work was done in this phase.

```plaintext
Procedure ISCA ()
STEP-1 Gen_Matric ()
STEP-2 Gen_Multidimensional_Matric ()
STEP-3 Cal_Defect ()
End procedure
```

**Algorithm- Gen_Matric ()**

```plaintext
/* WMC= Total number of Methods/total number of class
MDF, AHF, MIF, AIF, CIF, CF
Programmer 1, 2, 3, 4, 5..............................N.
Mat [1........N][1........N]
*/

Step1 i ← 1
Step2 loop i <= N
Step3 j ← 1
Step4 Mat[i][j] ← LOC i,j
j ← j+1
Mat[i][j] ← WMC i,j
j ← j+1
Mat[i][j] ← MDF i,j
j ← j+1
Mat[i][j] ← AHF i,j
j ← j+1
Mat[i][j] ← AIF i,j
j ← j+1
Mat[i][j] ← CIF i,j
If j ≤ N Then
break
Step6 end Step4 loop
Step7 i ← i+1
Step8 end Step3 loop
End procedure
```
IV. RESULT OF ISCA

Anecdotal and empirical evidence reported in the literature suggest, such as reduced time to market, reduced development costs, improved quality of the software, reduced costs of planning, and enhanced trust, motivation, and information and knowledge transfer among developers and project leader.

Fig 3 provides the comparison of earlier and our approach. It shows that the ISCA significantly reduce software defects and reduce software cost.

![Fig. 3 Comparison of earlier Approach and ISCA.](image)

V. CONCLUSION

Software estimation helps project management to plan the project. Tools available for the project estimation are great helps in the process. But estimating the project and then planning it without caring about the status of project at any instant of time is a problem worth to be considered. The process known as tracking is an important process that needs to be integrated with the estimation and planning process. The core of software crisis starts with the wrong estimation. We are introducing software matrices approach to calibrate software. ISCA algorithm tracks the developer and with various software matrices attribute are used to find out software defects cohesion, coupling etc.

REFERENCES

[1] Jing Dong, Senior Member, IEEE, Vajing Zhao, and Yongtao Sun, "A matrix based approach to recovering design pattern," IEEE