UML Developments:
Cost Estimation from Requirements

Philippe Larvet[1], Frédérique Vallée[2]

[1] Alcatel CIT,MDL, 10 rue Latécoère, BP 57,
78141 Vélizy-Villacoublay Cedex, France
Philippe.Larvet@alcatel.fr
[2] Mathix, 19 rue du Banquier,
75013 Paris, France
frederique.vallee@mathix.fr

Abstract. This paper presents the result of a model developed by Alcatel CIT to be able to predict, early in the life cycle, the development cost of UML projects. It demonstrates that textual metrics extracted from the requirements document are able to provide a rather precise information of the total development cost of a UML project.

1   Introduction

Being able to estimate accurately the cost of software development early in the life cycle is both an important and a difficult issue. Management can use cost estimation at different level, for example to set up a project proposal or to control the development process more efficiently. Consequently, a lot of effort has been dedicated since years to cost estimation ([1], [2], [3]).

In the usual process of cost modeling, one important thing is to have historical project data on which the model is constructed and validated. The useful data consist generally in a measure of size, whether evaluated in terms of LOC or of functional size measure, and a number of productivity characteristics that are dealing with the project and the process.

For Alcatel, in the precise case of UML[4] object-oriented software development projects, the need was to be able to master the cost aspect the earliest as possible, i.e. as soon as the text of requirements is available. Relevant work already exists in the area ([5], [6], [7]) but most of it is based upon metrics based on the UML models that are available too late to fulfill the need of Alcatel. Therefore Alcatel has undertaken a study with the Consultant Company Mathix, in order to develop an adequate cost estimation model.

To reach this objective, Alcatel and Mathix have decided to build a model that shall be able to give directly a good initial estimation of the total development cost. The model construction should be based on a sample of thirty-three UML projects, real industrial projects, already done.

The process followed to build this model is constituted of 4-main steps:

− refinement of the cost variable to be evaluated ;
− definition of the well-known, potentially useful, project characteristics ;
− gathering of the data ;
− data analysis and cost model building.

Remark 1. The model built in this study is mainly pragmatic and come from lessons learned. Consequently, it is specific to the scope of this work. Especially the incredibly low variance in the data is likely due to the fact that the projects on which the modeling is based were all of the same nature and were developed with similar techniques and skills.

This paper presents the modeling process and the results obtained:

− Section 2 defines the cost variable to be modeled ;
− Section 3 describes which explicative project characteristics have been selected to enter into the model construction ;
− Section 4 shows the available data that have been used as reference ;
− Section 5 gives the formulas obtained with their performances of prediction on the reference sample.

2   Refinement of the Cost Variable

The cost variable targeted is the Total Development Cost. In the Alcatel historical database, this cost was asked to be decomposed into 5 partial costs as follows :

$$\text{Total Development Cost } = \text{Specification Cost} + \text{Design Cost} + \text{Coding Cost} + \\ \text{Documentation Cost} + \text{Management Cost.} \tag{1}$$

In formula (1), the verification, validation and testing costs are distributed between specification, design and coding.

Our feeling was that these partial costs were not independent and that the Coding Cost was probably correlated to the Total Development Cost. This feeling has been verified through a graphical analysis of the distribution of the ratio: Coding Cost / Total Development Cost on our modeling sample.

Figure 1 gives the distribution of this ratio for the 22 projects of the historical database on which the cost decomposition was given.

This analysis allowed us to assert that it was possible to calculate the Total Development Cost from the Coding Cost, by using the following formula :

$$\text{Total Development Cost} = \text{Coding Cost} / 17\% \tag{2}$$