Chapter V

Web Effort Estimation Using Regression Analysis

Abstract

Software effort models and estimates help project managers allocate resources, control costs, and schedule and improve current practices, leading to projects that are finished on time and within budget. In the context of Web development and maintenance, these issues are also crucial, and very challenging, given that Web projects have short schedules and a highly fluidic scope. Therefore, this chapter presents a case study where a real effort prediction model based on data from completed industrial Web projects is constructed step by step using a statistical technique called regression analysis.

Case Study

The case study we present herein describes the construction and further validation of a Web effort estimation model using multivariate regression techniques and data
from industrial Web projects, developed by Web companies worldwide from the Tukutuku database (Mendes, Mosley, & Counsell, 2003, 2005). It should be noted that the raw data cannot be presented due to a confidentiality agreement with those companies that volunteered data on their projects. This database is part of the ongoing Tukutuku project (http://www.cs.auckland.ac.nz/tukutuku), which collects data on Web projects for the development of effort estimation models and to benchmark productivity across and within Web companies.

The data set used in this chapter contains data on 87 Web projects: 34 and 13 are from two single Web companies respectively and the remaining 40 projects come from another 23 companies. The Tukutuku database uses 6 variables to store specifics about each company that volunteered projects, 10 variables to store particulars about each project, and 13 variables to store data about each Web application (see Table 1). Company data is obtained once, and both project and application data are gathered for each project a Web company volunteers.

All results presented were obtained using the statistical software package SPSS 12.0 for Windows produced and sold by SPSS Inc. Further details on the statistical methods used throughout this case study are given in Chapter 10. Finally, all the statistical tests set the significance level at 95% ($\alpha = 0.05$). Note that the different types of measurement scales are also detailed in Chapter 10.

The following sections describe our data analysis procedure, adapted from Maxwell (2002), which consists of the following:

1. Data validation
2. Variables and model selection
3. Model inspection
4. Extraction of effort equation
5. Model validation

We also explain step by step how to use SPSS to carry out the analyses described in this chapter. Although you may not have access to SPSS, other commercial statistical tools use a similar methodology of data input and the same statistical algorithms. As such, the detailed instructions provided by this chapter can also be carried out with other statistical software as the options they provide are overall very similar to those offered in SPSS.

**Data Validation**

Data validation (DV) performs the first screening of the data that have been collected. In general, this involves understanding what the variables are (e.g., their
Formal Specification of Adaptable Semantic Web Services Composition
International Journal of Information Technology and Web Engineering (pp. 14-34).
www.igi-global.com/article/formal-specification-of-adaptable-semantic-web-
services-composition/209719?camid=4v1a