Chapter VII

Web Effort Estimation Using Classification and Regression Trees

Abstract

The use of realistic effort estimates is fundamental to both software and Web project management as they help project managers allocate resources, control costs and schedule, and improve current practices, leading to projects that are finished on time and within budget. Different effort techniques have been used to obtain effort estimates for Web projects. Two—stepwise regression and case-based reasoning—have already been presented in Chapters V and VI respectively. In this chapter we detail a third technique used to obtain effort estimates for Web projects, known as classification and regression trees (CART), that is considered a machine-learning technique. We detail its use by means of a case study where a real effort prediction model based on data from completed industrial Web projects is constructed step by step.
Introduction

The effort estimation technique that will be used in this chapter is called classification and regression trees (CART; Brieman, Friedman, Olshen, & Stone, 1984). This is, in addition to case-based reasoning (CBR), one of the techniques proposed by the machine-learning community, and has also been one of the techniques used in a previous study to estimate effort for Web applications.

CART uses independent variables (predictors) to build binary trees where each leaf node either represents a category to which an estimate belongs, or a value for an estimate. The former situation occurs with classification trees and the latter occurs with regression trees; that is, whenever predictors are categorical (e.g., Yes or No), the tree is called a classification tree, and whenever predictors are numerical, the tree is called a regression tree.

In order to obtain an estimate, one has to traverse the tree nodes from root to leaf by selecting the nodes that represent the category or value for the independent variables associated with the case to be estimated.

For example, assume we wish to obtain an effort estimate for a new Web project using as its basis the simple regression tree structure presented in Figure 4 (deemed a regression tree because effort is a numerical variable measured on a ratio scale). In this example, we assume the tree was generated from data obtained from past finished Web applications, taking into account their existing values of effort and independent variables (e.g., new Web pages [WP], new images [IM], and new features/functions [FN]). The data used to build a CART model are called a learning sample.

Figure 1. Example of a regression tree for Web effort estimation
Marketing Strategies in the Age of Web 3.0
www.igi-global.com/chapter/marketing-strategies-in-the-age-of-web-30/127298?camid=4v1a