Re-estimating software effort using prior phase efforts and data mining techniques

Pichai Jodpimai1 · Peraphon Sophatsathit1 · Chidchanok Lursinsap1
Received: 22 November 2016 / Accepted: 24 April 2018

Abstract
Software effort estimation has played an important role in software project management. An accurate estimation helps reduce cost overrun and the eventual project failure. Unfortunately, many existing estimation techniques rely on the total project effort which is often determined from the project life cycle. As the project moves on, the course of action deviates from what originally has planned, despite close monitoring and control. This leads to re-estimating software effort so as to improve project operating costs and budgeting. Recent research endeavors attempt to explore phase level estimation that uses known information from prior development phases to predict effort of the next phase by using different learning techniques. This study aims to investigate the influence of preprocessing in prior phases on learning techniques to re-estimate the effort of next phase. The proposed re-estimation approach preprocesses prior phase effort by means of statistical techniques to select a set of input features for learning which in turn are exploited to generate the estimation models. These models are then used to re-estimate next phase effort by using four processing steps, namely data transformation, outlier detection, feature selection, and learning. An empirical study is conducted on 440 estimation models being generated from combinations of techniques on 5 data transformation, 5 outlier detection, 5 feature selection, and 5 learning techniques. The experimental results show that suitable preprocessing is significantly useful for building proper learning techniques to boosting re-estimation accuracy. However, there is no one learning technique that can outperform other techniques over all phases. The proposed re-estimation approach yields more accurate estimation than proportion-based estimation approach. It is envisioned that the proposed reestimation approach can facilitate researchers and project managers on re-estimating software effort so as to finish the project on time and within the allotted budget.

Keywords Re-estimating software effort · Prior phase effort · Data transformation · Outlier detection · Feature selection · Learning