

Jordan University of Science and Technology

Predicting Software Projects Cost Estimation Based on Mining Historical Data

Authors: Hassan Najadat, Izzat Alsmadi, and Yazan Shboul

Abstract: In this research, a hybrid cost estimation model is proposed to produce a realistic prediction model that takes into consideration software project, product, process, and environmental elements. A cost estimation dataset is built from a large number of open source projects. Those projects are divided into three domains: communication, finance, and game projects. Several data mining techniques are used to classify software projects in terms of their development complexity. Data mining techniques are also used to study association between different software attributes and their relation to cost estimation. Results showed that finance metrics are usually the most complex in terms of code size and some other complexity metrics. Results showed also that games applications have higher values of the SLOCmath, coupling, cyclomatic complexity, and MCDC metrics. Information gain is used in order to evaluate the ability of object-oriented metrics to predict software complexity. MCDC metric is shown to be the first metric in deciding a software project complexity. A software project effort equation is created based on clustering and based on all software projects' attributes. According to the software metrics weights values developed in this project, we can notice that MCDC, LOC, and cyclomatic complexity of the traditional metrics are still the dominant metrics that affect our classification process, while number of children and depth of inheritance are the dominant from the object-oriented metrics as a second level.