Learning to Rank Algorithm for Constructing Defect Prediction Models

تعلم تصنيف الخوارزمية لبناء نماذج التنبؤ بالعيوب

By:

Sara Ahmad Alomari

Supervisor:

Dr. Yousef Elsheikh

Co-Supervisor:

Prof. Mohammad Al-Azzeh

Master Thesis

This thesis was submitted in Partial Fulfilment of the Requirements for the master's degree In Computer Science

Deanship of Scientific Research and Graduate Studies

Applied Science Private University

Amman-Jordan

Feb 2022

Abstract

Software defect prediction is one of the most active research fields in software development. The outcome of defect prediction models provides a list of the most likely defect-prone modules that need a huge effort from quality assurance teams. It can also help project managers to effectively allocate limited resources to verifying software products and invest more effort in defect-prone modules. As the size of software projects grows, fault prediction models can play an important role in assisting developers and shortening the time it takes to create more reliable software products by ranking software modules based on their defects. Therefore, we need a learning to rank technique that can prioritize and rank defective modules to reduce testing effort, cost, and time. In this thesis, in this thesis, we construct a new learning to rank technique that helps the quality assurance team with the ranking of the most likely defect-prone modules by using regression technique. The proposed technique will be evaluated on a set of benchmark data sets using well-known evaluation measures such as the Fault-Percentile Average (FPA), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and the Cumulative Lift Chart (CLC).

Keywords: Software Defect Prediction (SDP), Learning to Rank (LTR), Software Metrics.