Software Engineering Seminar

Cross-Project Defect Prediction

Description

Defect prediction has been one of the most popular topics in software engineering, which usually consists of code history mining, instance generation, data pre-processing, training, and prediction [1]. Different from traditional within-project defect prediction approaches which is based on history data of different versions from the same project, cross-project defect prediction collects training data from other projects to enable prediction for new projects.

Cross-project defect prediction is usually considered as a transfer learning problem [2]. Recent studies have leveraged different models to improve the performance: Zhang et al. [5] proposed an unsupervised spectral clustering based classifier; In Xia et al.'s approach [4], a genetic algorithm and ensemble learning are applied to create a hybrid model; Ryu et al. [3] proposed a value-cognitive boosting with support vector machine approach to alleviate data imbalance problems.

The goal of this seminar topic is to examine and discuss the current state of the art cross-project defect prediction approaches based on the given literatures, which includes technical summarization, comparison of different approaches, and critical discussion.

References


Contacts

Minxing Tang (tanminxi@informatik.hu-berlin.de)
Software Engineering Group
Institut für Informatik
Humboldt-Universität zu Berlin