Feature Representation and Feature Matching for Heterogeneous Defect Prediction

Thae Hsu Hsu Mon
thaehsuhsumon@ucsm.edu.mm
University of Computer Studies, Mandalay, UCSM

Hnin Min Oo
hninminoo@ucsm.edu.mm
University of Computer Studies, Mandalay, UCSM

Software Defect Prediction (SDP) is one of the highly influential software engineering research topics. Early within-project defect prediction (WPDP) used intra-project data. However, it has limitations in prediction efficiency for newprojects and projects without adequate training data. Studies of prediction have been carried out on cross-project defect prediction models (CPDP), i.e. models that are trained using other projects historical data. Heterogeneous defect prediction (HDP) is the special case of CPDP with different metric sets of source and target project. Despite the effectiveness of existing HDP methods, they can be affected by the issue of class imbalance that may decrease prediction performance. The proposed framework aims to exploit cost-sensitive principal component analysis (PCA) and the feature matching to build highly effective prediction model.