Enhanced Software Defect Prediction: NSGA-II-based K-member Clustering and Ensemble Learning

Shima Javadimoghdam¹, Seyed Mojtaba Sabagh-Jafari², and Amid Khatibi Bardsiri³

¹Islamic Azad University Kerman Branch
²Rafsanjan University of Vali Asr
³Islamic Azad University

May 20, 2024

Abstract

Predicting software defects can improve software assurance reliability and reduce development costs. Traditional predictions usually lack precision. The early detection of fault-prone modules helps software project managers allocate developers’ limited time for testing the defect-prone modules more intensively. Traditional methods for predicting software defects often lack precision, highlighting the need for enhanced techniques. The proposed solution introduces an ensemble learning algorithm based on clustering for software defect prediction, aiming to handle imbalanced datasets and manage redundant features. The method incorporates clustering and oversampling for data preparation, followed by cost-sensitive feature selection and ensemble learning for improved classification. By employing clustering to group similar data and addressing imbalanced datasets through K-member fuzzy clustering algorithms, the strategy enhances the accuracy of defect prediction. The optimization stage involves selecting the best classifier, features for each cluster, and their hyper-parameters using the NSGA-II algorithm. Experiments conducted on real-world datasets demonstrate better performance of the suggested approach in contrast to other established approaches in the software defect detection literature.

Hosted file