Abstract

Software defect prediction using classification algorithms was advocated by many researchers. Moreover the classifier ensemble can effectively improve classification performance compared to a single classifier. The research on defect prediction using classifier ensemble methods are motivated since they have not been fully exploited. Software defects leads to failure of many defense systems. A comparative study of various classification methods was performed to classify software defects. The methods include Random Tree, Random Forest, Bayesian Network, Naive Bayes, K-Nearest Neighbour and Instance Based Classifier. Random Forest algorithm was found to give more accurate prediction than other classifiers. To enhance the classification accuracy the new algorithm &quot;Improved Random Forest&quot; is proposed. It works by incorporating best feature selection algorithm with the Random Forest to gives better accuracy. Correlation based Feature Subset Selection algorithm selects the optimal subset of features. The optimal features are fed as a part of Random Forest classification to give better accuracy in software defect prediction. The six optimal subset of features were selected for PC1 dataset. The features are selected by the CFS and utilized by Random Forest to improve the accuracy of existing Random Forest. The experiments were carried on public-NASA datasets of PROMISE repository.

Index Terms

Computer Science

Software Engineering
Keywords
Software Defect Prediction  Feature Selection  Classification  Classifier Evaluation.