A new classification model for a class imbalanced data set using genetic programming and support vector machines: case study for wilt disease classification

Abstract

Class imbalanced data set is a state where each class of the given data set is not evenly distributed. When such case happens, most standard classifiers fail to recognize examples that belong to a minority class. Hence, several methods have been proposed to solve this problem such as resampling, modification on classifier optimization problem or introducing a new optimization task on top of the classifier. This work proposes a new optimization task based on genetic programming, built on top of support vector machine, in order to improve the classification rate for minority class without significant reduction on accuracy metric. The experimentation carried out on wilt disease data set shows the new classifier, support vector based on genetic programming machine, gives a more balanced accuracy between classes compared to various classification techniques in solving the imbalanced classification problem.