Abstract

Class imbalance problem is a problem where the data distribution is very unbalance. The data of one class has significantly a lot more in quantity (majority class) if compare to another class (minority class). In general classification, this minority class can not be classified correctly; because, if the number of instances of this minority class is too few, they are often will be directly predicted to be the data of the majority class. There are several ways to handle this problem; and boosting is one of those solutions. There is a feature in boosting which is called iteration. By using this iteration in the learning process, a model will be built from certain training set and then the data will be resampled to be used in the next iteration. Generally, those iterations will improve the accuracy of the classification. The boosting algorithms which are used in this undergraduate thesis are AdaBoost, MultiBoost, LogitBoost, RareBoost-1, AdaC1, AdaC2, and AdaC3. Each of these algorithms will be tested into two kind of datasets. One of them is a dataset without noise, and another is a dataset with noise. The results show that these boosting algorithms can improve the accuracy of the *base classifier*, regardless there are noises or not.

Keywords: *imbalance*, *boosting*, AdaBoost, MultiBoost, LogitBoost, RareBoost-1, AdaC1, AdaC2, AdaC3.