Abstract

One of problem in document categorization and bioinformatic is the characteristic data which have more than one label (multi-label). It can be solved by classifying data using classification method. In multi-label classification, each instance in training set is associated with a set of labels[4]. The training set will be training using one of certain classification method (classifier) and the task will result a model that used for predicting the set label for unlabeled data (test set).

For solving those problem, classifier that is used for classifying multi-label data are Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN). Both of methods are based on statistic with good accuracy in some research [4,7,10]. SVM is machine learning method which always trying to find the best of hyperplane to separate two classes at input space. k-NN is Instance-based learning that sometimes referred to as "lazy" learning methods, because this method classify data base on "k" nearest neighbor and count maximum a posterior to determine the label set of test set.The effectivness both of methods are evaluated base on evaluation matrics include accuracy, precission, recall, hammingLoss, one-error and rankingloss.

In this final project, the analysis is done by comparing the result of computing evaluation matrics of both classifier. It will find which classifier that the most reliable to handle the classification multi-label data. Result of enumeration evaluation matrics of methods show the comparison of good classification method to solve multi-label problem. Beside that, the analysis also compare between specific multi-label classifier and general classifier.

Keywords: *multi-label classification*, SVM, k-NN, *hyperplane, maximum a posterior,* evaluation matrics