

Abstract

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

For solving those problems, classifiers that are used for classifying multi-label data are Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN). Both of these methods are based on statistics with good accuracy in some research [4,7,10]. SVM is a machine learning method which always tries to find the best of hyperplane to separate two classes in input space. k-NN is instance-based learning that is sometimes referred to as “lazy” learning methods, because this method classifies data based on “k” nearest neighbors and counts maximum a posteriori to determine the label set of the test set. The effectiveness of both methods is evaluated based on evaluation matrices including accuracy, precision, recall, hamming loss, one-error and ranking loss.

In this final project, the analysis is done by comparing the results of computing evaluation matrices of both classifiers. It will find which classifier is the most reliable to handle the classification of multi-label data. Results of enumeration evaluation matrices of these methods show the comparison of good classification methods to solve multi-label problems. Besides that, the analysis also compares between specific multi-label classifiers and general classifiers.

Keywords: *multi-label classification, SVM, k-NN, hyperplane, maximum a posteriori, evaluation matrices*