Abstract

Problems that are often encountered in the document clustering is to determine the appropriate algorithm or method with a number of specific documents. Sometimes the method that has quite accurate results takes a long time for processing. There are several approaches used to solve this clustering problem, namely partitioning with clustering approach and hierarchical clustering approach. Each has advantages and disadvantages. One of the many clustering methods that can be used is the sequential information bottleneck (SIB) method [5]. The algorithm of sequential information bottleneck in document guarantees the discovery of solution which is a local maximum of the target function.

This paper applied sequential information bottleneck algorithm as a clustering method. The clustering results are then measured its accuracy using micro-averaged precision and micro-averaged recall by considering changes in the input parameters used, such as maximum loop (maxL), the number of random initialization of the cluster, and the error value. The test result obtained that sequential information bottleneck algorithm is best used as a clustering method, it can be seen from the results of experiments in which the value of accuracy is achieved on average over 70%.

Accuracy of the cluster increases with increasing parameter values until the maximum loop to some extent where the looping condition has stopped because the other parameters (error value) has fulfilled its value. as a result of documents that have been processed using sequential information bottleneck method has a very high precision value, then the clustering results can be used as a training set for supervised classification method [5]. supervised classification method to be used is the naive Bayes method clasification. Naive Bayes clasification method that uses the results of the sIB as a training set then compared its accuracy with Naive Bayes clasification method that is not using clustering with the sIB as a result of training.

Keywords: *Sequential Information Bottleneck Method, Naive bayes classification method, clustering, classification*