Web Page Classification using Random Subspace Ensemble

Htwe Htwe Su
University of Computer Studies, Yangon, Myanmar
opposite2010@gmail.com

Abstract

More or more data and content are published on the web on the daily basis. The categorization of classification of web content according to their types or category is essential for organization such as web portal, news portal and search engine.Many machines learning approach can be used for the classification of web pages.Decision trees are one of the most popular machines learning technique for web page classification. Ensemble method provides high accuracy in classification system. Random subspace is used different feature to construct individual decision tree in ensemble, so they are suitable for the high dimensional data.This proposed system used random subspace approach for classification of web pages.C4.5 algorithm will be used for building the tree in the random subspace ensemble and majority voting will be used for combining the output of the classifier.