JACN 2016 Vol.4(1): 24-27 ISSN: 1793-8244
DOI: 10.18178/JACN.2016.4.1.198

Clustering and Feature Selection Technique for Improving Internet Traffic Classification Using K-NN

Trianggoro Wiradinata and Adi Suryaputra Paramita

Abstract—This research will use the algorithm K-Nearest Neighbour (K-NN) to classify internet data traffic, K-NN is suitable for large amounts of data and can produce a more accurate classification, K-NN algorithm has a weakness takes computing high because K-NN algorithm calculating the distance of all existing data. One solution to overcome these weaknesses is to do the clustering process before the classification process, because the clustering process does not require high computing time, clustering algorithm that can be used is Fuzzy C-Mean algorithm, the Fuzzy C-Mean algorithm does not need to be determined in first number of clusters to be formed, clusters that form on this algorithm will be formed naturally based datasets be entered, but the algorithm Fuzzy C-Mean has the disadvantage of clustering results obtained are often not the same even though the same input data, this is because the initial dataset that of the Fuzzy C-Mean is not optimal, to optimize initial datasets in this research using feature selection algorithm, after main feature of dataset selected the output from fuzzy C-Mean become consistent. Selection of the features is a method that is expected to provide an initial dataset that is optimum for the algorithm Fuzzy C-Means. Algorithms for feature selection in this study used are Principal Component Analysis (PCA). PCA reduced non significant attribute to created optimal dataset and can improve performance clustering and classification algorithm. Results in this study is an combining method of classification, clustering and feature extraction of data, these three methods successfully modeled to generate a data classification method of internet bandwidth usage that has high accuracy and have a fast performance.

Index Terms—Clustering, classification, feature, bandwidth.