

Abstract

Generally, data have big amounts of features, feature can also be named as attribute or dimension. Using big amounts of features can affect the learning process becomes less effective, even produce worst cluster quality. Therefore, feature selection is considered to decrease the data features. In this Final Task, the feature selection is done in clustering. The feature selection process can maintain or even increase the cluster quality. There are two Feature Selection methods: wrapper method and filter method. Wrapper method selects feature with the assistance of learning algorithm, whereas, filter method selects feature first and then does the learning algorithm. Feature selection is done with wrapper method. There are three processes in Wrapper method: searching process, clustering process, and feature evaluation process. This Final Task uses sequential forward search strategy which is greedy. The clustering process uses EM algorithm and the cluster produced is evaluated with log likelihood cluster evaluation method.

Keyword: *feature selection, clustering, log likelihood, sequential forward search, wrapper.*