

Abstract

A current data can have a lot of features. The number of features that can be owned by a single object instance is not necessarily the relevant information required by the data mining system. Feature selection is a process of selecting a subset of features / attributes that are relevant to using certain criteria. By doing feature selection is able to reduce the number of irrelevant features, eliminating data redundancy, and improve the accuracy of learning.

Classification is one of the stages in data mining, whose function is to predict membership or classes of data. In some studies indicated that an ensemble (set) of some of the classifier is generally more accurate than a single classifier. One way to generate an ensemble is to choose several different subset of features from the original dataset and for each feature subset is then performed classification. This approach is known as ensemble feature selection. Here, the author will try to implement a genetic algorithm for optimization of feature selection in the formation of ensembles, namely Genetic Algorithm-Sequential Ensemble Feature Selection (GA-SEFS). Conventional feature selection algorithms aim to find the best feature subset, while the ensemble feature selection has the objective to find the best subset of the set of features that can improve the accuracy in classification.

In GA-SEFS contained six important parameters. Parameters of population size, number of generations, and the offspring do not directly affect the resulting accuracy of the classification ensemble. Ensemble size parameter can help to increase the accuracy of vote due to a variety of feature subset that can help improve accuracy. Alpha parameter can help to provide improved accuracy obtained by the combination of the above 4 parameters (ensemble size, population, number of generations, and the number of offspring). Beta parameter in this Final trial for three different datasets were further provide high accuracy values on the value of a negative beta.

Keywords: *subset feauret selection, ensemble, genetic search*