

Ensemble Meta Classifier with Sampling and Feature Selection for Data with Multiclass Imbalance Problem

ABSTRACT

Ensemble learning by combining several single classifiers or another ensemble classifier is one of the procedures to solve the imbalance problem in multiclass data. However, this approach still faces the question of how the ensemble methods obtain their higher performance. In this paper, an investigation was carried out on the design of the meta classifier ensemble with sampling and feature selection for multiclass imbalanced data. The specific objectives were: 1) to improve the ensemble classifier through data-level approach (sampling and feature selection); 2) to perform experiments on sampling, feature selection, and ensemble classifier model; and 3) to evaluate the performance of the ensemble classifier. To fulfil the objectives, a preliminary data collection of Malaysian plants' leaf images was prepared and experimented, and the results were compared. The ensemble design was also tested with three other high imbalance ratio benchmark data. It was found that the design using sampling, feature selection, and ensemble classifier method via AdaboostM1 with random forest (also an ensemble classifier) provided improved performance throughout the investigation. The result of this study is important to the ongoing problem of multiclass imbalance where specific structure and its performance can be improved in terms of processing time and accuracy