

Evaluation Method in Random Forest as Applied to Microarray Data

Ng Ee Ling and Yahya Abu Hasan
*School of Mathematical Sciences,
Universiti Sains Malaysia,
11800 Minden, Pulau Pinang
E-mail: eeneling@gmail.com*

ABSTRACT

Unlike other decision tree classifiers, Random Forest grows multiple trees which create a forest-like classification. Thus, Random Forest produces better performance as compared to that of a single tree classifier. We consider several evaluation methods which include the 10-fold cross validation, leave-one-out cross validation and bootstrap estimation. These evaluation methods are to assess the performance of the Random Forest classifier. The usage of different evaluation methods certainly shows the durability of Random Forest. To help illustrate the problem better, the four microarray datasets of binary-class and multi-class are used as experimental datasets. The evaluation method is a subjective issue and it is bound to the researcher and his study scope when selecting an evaluation method. However, we have shown that Random Forest is best evaluated using 10-fold cross validation and bootstrap estimation.

Keywords: Microarray; Random Forest; Classification; Evaluation Methods