Abstract

Random forest can achieve high classification performance through a classification ensemble with a set of decision trees that grow using randomly selected subspaces of data. The performance of an ensemble learner is highly dependent on the accuracy of each component learner and the diversity among these components. In random forest, randomization would cause occurrence of bad trees and may include correlated trees. This leads to inappropriate and poor ensemble classification decision. In this paper an attempt has been made to improve the performance of the model by including only uncorrelated high performing trees in a random forest. Experimental results have shown that, the random forest can be further enhanced in terms of the classification accuracy.

Index Terms

Computer Science Algorithms

Keywords

Strength Correlation Tree Performance Decision trees.