Proactive Forest for Supervised Classification

Nayma Cepero Pérez
Luis Alberto Denis Miranda
Rafael Hernández Palacio
Mailyn Moreno Espino
Milton García Borroto

Universidad Tecnológica de la Habana José Antonio Echeverría, CUJAE, Cuba.
{ncepero,ldenis,rhernandezpe,my,mgarciab}@ceis.cujae.edu.cu

Abstract

Random Forest is one of the most used and accurate ensemble methods based on decision trees. Since diversity is a necessary condition to build a good ensemble, Random Forest selects a random feature subset for building decision nodes. This generation procedure could cause important features to be selected in multiple trees in the ensemble, decreasing the diversity of the entire collection. In this paper, we introduce Proactive Forest, an improvement of Random Forest that uses the information of the already generated trees to induce the remaining trees. Proactive Forest calculates the importance of each feature for the constructed ensemble in order to modify the probabilities of selecting those features in the remaining trees. In the conducted experiments, Proactive Forest increases the diversity of the obtained ensembles with a significant impact in the classifier accuracy.

Keywords: *Decision forests, Random Forest, Diversity*