World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:9, No:06, 2015

Determining Optimal Number of Trees in Random Forests

Authors : Songul Cinaroglu

Abstract : Background: Random Forest is an efficient, multi-class machine learning method using for classification, regression and other tasks. This method is operating by constructing each tree using different bootstrap sample of the data. Determining the number of trees in random forests is an open question in the literature for studies about improving classification performance of random forests. Aim: The aim of this study is to analyze whether there is an optimal number of trees in Random Forests and how performance of Random Forests differ according to increase in number of trees using sample health data sets in R programme. Method: In this study we analyzed the performance of Random Forests as the number of trees grows and doubling the number of trees at every iteration using “random forest” package in R programme. For determining minimum and optimal number of trees we performed Mc Nemar test and Area Under ROC Curve respectively. Results: At the end of the analysis it was found that as the number of trees grows, it does not always means that the performance of the forest is better than forests which have fever trees. In other words larger number of trees only increases computational costs but not increases performance results. Conclusion: Despite general practice in using random forests is to generate large number of trees for having high performance results, this study shows that increasing number of trees doesn’t always improves performance. Future studies can compare different kinds of data sets and different performance measures to test whether Random Forest performance results change as number of trees increase or not.

Keywords : classification methods, decision trees, number of trees, random forest
Conference Title : ICADMA 2015 : International Conference on Advanced Data Mining and Applications
Conference Location : Istanbul, Türkiye
Conference Dates : June 18-19, 2015