Weighted Consensus Clustering for Unbiased Feature Importance in Random Forests

Ndèye Niang, Mory Ouattara

Weighted Consensus Clustering for Unbiased Feature Importance in Random Forests

Ndèye Niang and Mory Ouattara

Abstract Ranking the importance of features in Random Forests (RF) has been shown to be biased in the presence of highly correlated features, especially for high-dimensional data when the number of features is much larger than the sample size. Several methods have been proposed for unbiased ranking. Among them, the Fuzzy Forest (FF) method [1] combines feature clustering and recursive feature elimination random forests (RFE-RF) and provides relatively unbiased rankings. RFE-RF is performed on each block of features leading to the selection of a percentage of features that will be kept in each block. Finally, a RF is applied on the selected variables. In this work, through simulation studies, we show that applying different clustering algorithms yields different feature groups of unequal quality and thus different results concerning important variables. This may lead to an issue for the choice of the feature clustering algorithm. To overcome this issue, we propose to use new weighted consensus clustering method to get an unique partition [2] on which RFE-RF is performed. The experimental results on simulated data as well as real ones show better performances and stability for the recovery of important variables.

Keywords: Random forest, Feature importance, Weighted consensus

References

1. Conn, D., Ngun, T., Li, G., Ramirez, C. M. Fuzzy Forests: Extending Random Forest Feature Selection for Correlated, High-Dimensional Data. Journal of Statistical Software, (2019) 91(9), 1–25. https://doi.org/10.18637/jss.v091.i09
2. Niang Ndèye and Ouattara Mory : Weighted consensus clustering for multiblock data. In : SFC 2019. https://cnam.hal.science/hal-02471611

Ndèye Niang
CEDRIC- CNAM, 2 rue Conté 75003 Paris, France, e-mail: ndeye.niang_keita@cnam.fr

Mory Ouattara
Université de San Pédro, Côte d'Ivoire e-mail: ouattara.mory@usp.edu.ci