

Tree Ensemble Methods for Computer Aided Diagnosis (CAD) Systems



M. Wehenkel^{1,2}, P. Geurts², C. Phillips^{1,2}

¹ Cyclotron Research Centre, University of Liège, Belgium

² Dept. of Electrical Engineering and Computer Science & GIGA-R, University of Liège, Belgium



Introduction

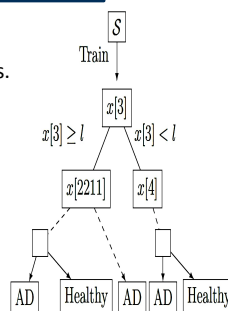
Over the last few years, machine-learning (ML) techniques have been increasingly popular in the field of neuroimaging through the development of « mind reading » approaches and of CAD systems for neurodegenerative diseases [1,2]. The most common ML methods are the « support vector machine » (SVM) and other similar kernel methods.

However, non-kernel approaches such as the « tree ensemble » methods, e.g. « Random Forests » or « Extra Trees », have only been investigated briefly to our knowledge. They might be promising pattern recognition techniques and should be studied more deeply.

Decision tree for neuroimaging data

Characteristics :

- Class labels are predicted, e.g. healthy vs. Alzheimer's disease;
- There are as many attributes as voxels;
- Voxel space is partitioned by the voxel intensities;
- Each node tests an attribute (voxel), a branch corresponds to a voxel intensity value and each leaf is a terminal node with a class label.



Growing the tree :

Consider a learning sample $S = \{ (x_i, y_i) \in (X \times Y) \}_{i=1, \dots, n}$, where $X = \mathbb{R}^d$ is the input space (each image contains d voxels) and $Y = \{0, 1\}$ is the output space (classification problem; e.g. 0 = healthy and 1 = diseased).

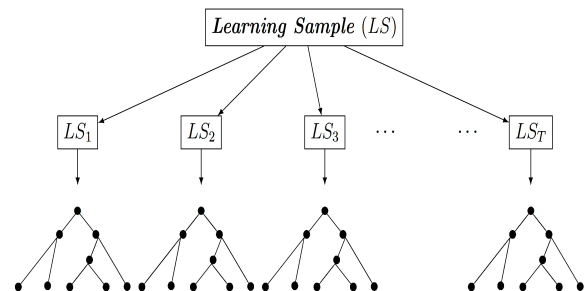
Given one impurity measure, at each step, while the samples in the learning sample have not the same label,

1. split a node by choosing the best attribute (the j^{th} voxel $x[j]$) and the best cut-point leading to the maximum expected reduction of impurity,
2. for each value a of this attribute, build a new smaller learning sample $S_s = \{ x_i \in S \mid x_i[j] = a \}$ from the original one,
3. grow a subtree from S_s .

Tree-based Ensemble Method

Characteristics of Tree Ensemble Methods :

- Combine the prediction of several trees;
- Trees are grown either independently (Randomized tree ensembles) or sequentially (Boosting ensembles);
- Improve the bias-variance trade-off of single trees (mostly by reducing their variance).
- E.g.: Randomized tree ensembles:



Examples of randomized tree ensemble methods:

- « Tree Bagging » grows each tree from a bootstrap sample¹ of the initial learning set [3].
- « Random Forests », an improved version of « Tree Bagging », selects the best split of a node between K randomly selected attributes (voxels) instead of choosing in the whole set [4].
- « Extra Trees » differ from other methods by the absence of bootstrap sampling and the choice of both K features and cut-point at random to split a node [5].

Some advantages & drawbacks :

- ✓ interpretability (through variable importance scores);
- ✓ very good predictive performance;
- ✓ robustness to irrelevant variables;
- ✓ ease of use (no need of data pre-processing and very light parameter tuning);
- ✓ can handle multiclass and multi-label classification;
- ✗ known bias in the presence of heterogeneous, continuous and categorical, features (e.g., clinical features and voxels);
- ✗ blind with respect to feature correlation and structure (e.g., 3D organization of voxels).

¹A bootstrap sample is obtained by random sampling from the original training set uniformly and with replacement.

Conclusion

To date tree-based approaches have not been sufficiently explored in the field of neuroimaging and neurodegenerative diagnosis methods. Yet, they exhibit many advantages in comparison with the common ML techniques and might prove interesting alternatives, notably if we can handle the bias problem with heterogeneous data and deal with the usual encountered dimensionality issue of neuroimaging data (less samples than features).

REFERENCES

- [1] Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K. R. (2011). Introduction to machine learning for brain imaging. *Neuroimage*, 56(2), 387-399.
- [2] Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, 45(1), S199-S209.
- [3] Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- [4] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [5] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1), 3-42.



ACKNOWLEDGEMENTS & SPONSORS: Belgian National Fund for Scientific Research (F.R.S-FNRS) – University of Liège (ULg) – Cyclotron Research Centre (CRC)