Exploiting a hierarchy of brain regions for Alzheimer's disease classification

Helena Barros and Margarida Silveira for the Alzheimers Disease Neuroimaging Initiative
Institute for Systems and Robotics, Instituto Superior Técnico
Universidade de Lisboa, Lisbon, Portugal

I. Introduction

Sparse methods have gained popularity as an effective way to alleviate the curse of dimensionality in neuroimaging applications such as Alzheimer's disease (AD) and Mild Cognitive Impairment (MCI) classification. By imposing sparsity inducing regularization terms these methods are able to perform feature selection jointly with classification.

The simplest of these methods is the Lasso which uses L_1 norm regularization to induce sparsity. It has been used for AD classification in [1] and [2]. It is effective but the selected features may be sparsely distributed throughout the whole brain and unstable. To overcome this, one possibility is to consider groups of features. This is the approach of Group Lasso that uses $L_{2,1}$ norm to promote group sparsity. It was used for AD in [3]. It can also be generalized to allow for overlapping groups, including tree structured groups where a hierarchy of relationship between features can be defined. This approach has been proposed in [4] for AD and MCI classification using a pyramid tree hierarchy.

In this paper we propose an alternative tree structure, more consistent with disease related atrophy, where neighboring features are grouped according to anatomically defined regions of the brain and in a hierarchy that joins regions in the left and right hemispheres of the brain to take into account bilateral symmetry which typically occurs in AD.

We apply these methods to MRI images from ADNI [5] and evaluate their classification performance and the stability of the obtained feature weights when several runs are performed.

II. Methods

We use the tree structured group Lasso which is based on the group-Lasso penalty, and defines each node in the tree as a group and leaf nodes as individual features (voxels). Our goal is to calculate a parameter vector of weights associated with each feature which reflect the importance of a given cerebral voxel for AD and MCI classification.

A set of m image samples are used to train the model, $\{a_i, b_i\}_{i=1}^m$ where $a_i \in \Re^n$ represents the n-dimensional features and $b_i \in \{-1, 1\}$ the class of each sample.

The parameter vector $x \in \Re^n$ is found by solving the optimization problem in Eq. (1). The first term is the Logistic Regression loss function, the second term is the regularization term (penalty) and $\lambda > 0$ is the regularization parameter that regulates the trade-off between the two terms.

$$\min_x \sum_{i=1}^{m} w_i \log(1 + e^{-b_i(x^T a_i + c)}) + \lambda \sum_{i=0}^{d} \sum_{j=1}^{n_i} w_j^i \|x_{G_j^i}\|_2 \qquad (1)$$

In the penalty term, d is the number of tree levels, n_i is the number of nodes for a given level, G_j^i is the group of voxels in node j of the i^{th} level of the tree and w_j^i is the weight assigned to that group.

Two types of tree structured Group Lasso were tested.

- Pyramid Tree - This tree adopts the pyramid structure of [4] which is based on the assumption that neighbour voxels are spatially correlated but does not explore brain anatomy. It has 3 levels; the first level divides the brain into 4^3 cubic regions, the second level divides each region in the previous level into another 4^3 cubes and the third level is formed by single voxel leaves.
- Atlas Tree - This tree groups voxels according to the brain regions in the Harvard-Oxford (HO) atlas and exploits the fact that AD is typically bilateral. It has 3 levels; the first one groups atlas regions with their counterpart in the opposing brain hemisphere, when applicable; the second level contains the 21 cortical regions and 48 subcortical regions in the HO atlas and the third level contains all the leaf nodes corresponding to individual brain voxels. This tree is represented in Fig. 1.

For classification we use the Logistic Regression Classifier, which computes the posterior probability of class $y \in \{-1, 1\}$ and assigns each test sample to the more likely class.

III. Experimental Results

We used 1.5T MRI data from ADNI [5] (details in Table I), which we warped into the MNI152 space, the same space as the HO atlas. We tested both Tree methods and Lasso in two problems, AD vs. Control Normals (CN) and MCI vs. CN. The value of the regularization parameter was estimated with 5-fold nested cross-validation. The values for λ were specified as a ratio $\alpha \times \lambda_{max}$ with $\alpha \in \{10^{-3}, 10^{-2}, ...10^{0}\}$, where λ_{max} denotes the maximum value of λ above which the solution to Eq. (1) is zero. All methods were implemented with the SLEP toolbox [6].

Group	CN (75)	MCI (135)	AD (58)
Age(mean±sd)	75.9±4.9	75.1±6.6	76.0±6.6
Sex (M/F)	49/26	88/47	34/24
MMSE (mean±sd)	29.1±1.0	27.2±1.6	23.5±1.9
CDR (mean±sd)	0±0	0.5±0.0	0.8 ±0.2

TABLE I
Demographic and clinical characteristics of each group. The number of images is shown in parentheses.

Fig. 2 displays the ROC curves obtained with the three methods for both classification tasks. These curves show the possible trade-off between sensitivity and specificity. The same figure presents the Area Under the ROC Curve (AUC) for each curve. As can be seen, in both cases, the performance of the three methods is close, and worse for MCI vs CN than AD vs CN, as expected. For MCI vs CN there is some improvement of the both Tree methods over Lasso and of the Atlas Tree over the Pyramid Tree. For AD vs CN only Atlas Tree was able to outperform Lasso, this is probably due to the rigidity of the Pyramid tree groups.

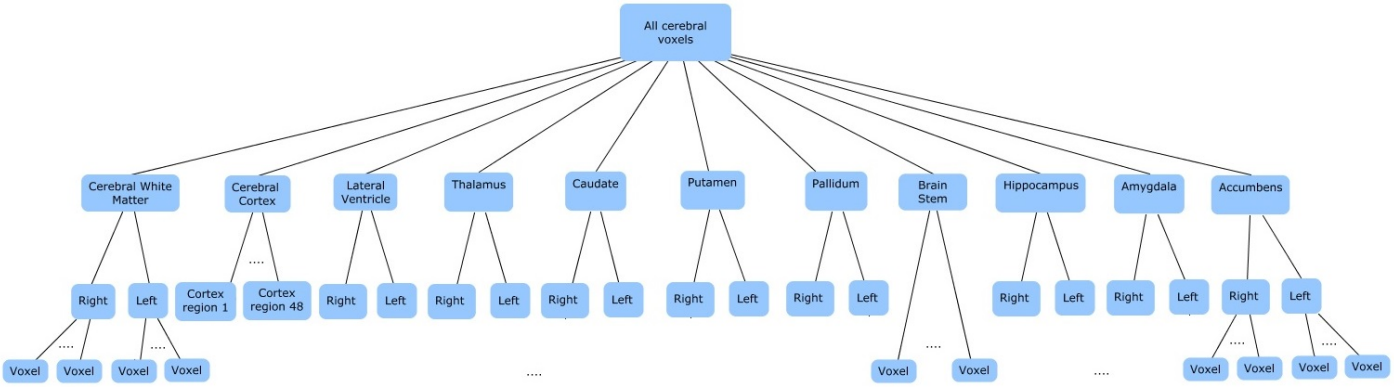

Fig. 1. Tree based on the cortical and subcortical regions of the Harvard-Oxford atlas.

We also analyzed the stability of feature weights across the different folds. Our stability metric is based on Pearsons correlation coefficient to measure the similarity between two weight images. We evaluate the pairwise similarities over all the possible pairs and then compute the average. Fig. 3 shows the stability obtained for AD vs CN and MCI vs CN classification. In both cases the Tree methods obtained dramatic improvements in stability when compared with Lasso. The Tree Atlas method consistently obtained the best results while Lasso obtained the worse.

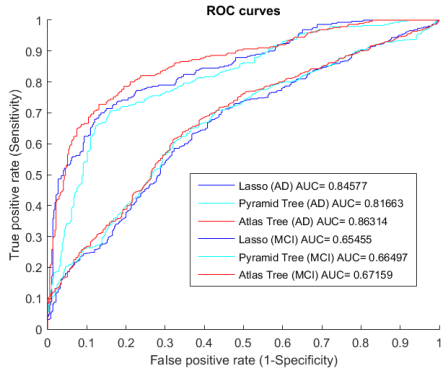

Fig. 2. Receiver operating characteristic curve (ROC).

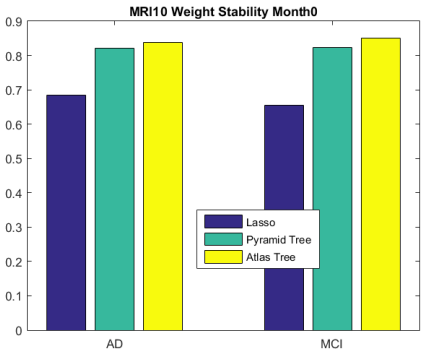

Fig. 3. Weight stability.

Regarding the regional distribution of the features selected by Tree Atlas, i.e. those with weights different from zero, we verified that most of the features are selected in the amygdala, the hippocampus and the parahippocampal gyrus. This is in agreement with previous studies [4]. We also confirmed that in these regions the weights are, in general, symmetrically distributed.

IV. Conclusion

We applied sparse logistic regression to AD and MCI classification. Two types of tree-structured sparse penalties were investigated and were shown to outperform L1-regularization. The proposed method using tree structured group sparsity based on the Harvard Oxford atlas brain regions and on bilateral symmetry not only attained the best classification performance but also generated more stable and therefore more interpretable feature patterns.

Acknowledgment

This work was supported by FCT project UID/EEA/50009/2013. Data used in the preparation of this article were obtained from the ADNI database (http://www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

References

[1] A. Rao, Y. Lee, A. Gass, and A. Monsch, "Classification of Alzheimer's disease from structural MRI using sparse logistic regression with optional spatial regularization," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, Aug 2011, pp. 4499–4502.

[2] J. Ye, M. Farnum, E. Yang, R. Verbeeck, V. Lobanov, N. Raghavan, G. Novak, A. DiBernardo, and V. Narayan, "Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data," *BMC Neurology*, vol. 12, no. 1, 2012. [Online]. Available: http://dx.doi.org/10.1186/1471-2377-12-46

[3] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895 – 907, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S105381191101144X

[4] M. Liu, D. Zhang, P.-T. Yap, and D. Shen, "Tree-guided sparse coding for brain disease classification," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. Springer, 2012, pp. 239–247.

[5] "Alzheimers disease neuroimaging initiative." [Online]. Available: http://www.loni.ucla.edu/ADNI

[6] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections*, Arizona State University, 2009. [Online]. Available: http://www.public.asu.edu/ jye02/Software/SLEP