Can we interpret linear kernel machine learning models using anatomically labelled regions?

Submission Number:

3112

Authors:

Jessica Schrouff[1], Joao Monteiro[2], Maria Joao Rosa[3], Liana Portugal[4], Christophe Phillips[5], Janaina Mourao-Miranda[6]

Institutions:

[1]Laboratory of Behavioral and Cognitive Neurology, Stanford University, Palo Alto, USA, [2]University College London, London, United Kingdom, [3]King's College London, London, United Kingdom, [4]Computer Science Department, University College London, London, United Kingdom, [5]Cyclotron Research Centre, University of Liege, Sart Tilman, Liege, Belgium, [6]Computer Science Department - University College London, London, United Kingdom

First Author:

Jessica Schrouff - Lecture Information | Contact Me
Laboratory of Behavioral and Cognitive Neurology, Stanford University
Palo Alto, USA

Introduction:

Recently, pattern recognition models have been applied to neuroimaging data [1], enabling predictions about a variable of interest based on patterns of activation or anatomy over a set of voxels. These machine learning based methods present undeniable assets over classical (univariate) techniques, by providing predictions for unseen data, as well as accounting for correlations in the data due to their multivariate nature.
However, the obtained weight map (i.e. the model's parameter) does not allow regionally specific inference, leading to difficulties in terms of interpretability. In cognitive and clinical neuroscience applications it is important to identify the contribution of different brain regions to the predictive models. In the present work, we used previous knowledge about brain anatomy and compared two different approaches to describe the machine learning models in terms of anatomically labelled regions.

Methods:

More specifically, anatomically labelled regions (as defined by the AAL atlas [2]) were used to:

a) Summarize whole brain model weights
In the present case, a whole brain model is built and the weights per voxel computed. The weights are then averaged within anatomically defined regions by taking the sum of their absolute values and dividing by the size of the region. This measure further referred to as Normalized Weights (NW, [4,5]) can then be used to rank the considered regions. No thresholding can be performed.
b) Combine the information from different brain regions hierarchically through Multiple Kernel Learning (MKL, [6])
MKL aims at simultaneously learning the kernel weights and the decision function in supervised learning settings. Here, each anatomically labelled region corresponds to a different kernel and decision function. These decision functions are then weighted [7] to obtain the final model. This approach therefore corresponds to a hierarchical model, in which the models from each individual brain region are assembled to form the whole brain model. Regions can then be ranked according to their weighting parameter (i.e. the kernel weights, dm). In this work, we used the MKL version of [7], which enforces sparsity via a L1-regularization on dm.

These two approaches were tested on the Haxby dataset [8], a single subject fMRI dataset in which a subject viewed pictures of 8 categories of objects, during 12 runs. To illustrate the approaches, we focused on the comparison between viewing houses and buildings. All models were based on SVM classifiers and accuracy was evaluated using a leave-one-run out cross-validation.

Results:

Model performance was 96.76% for the whole brain SVM model (a) and 97.69% for the region-based MKL model (b). The MKL model identified 13 regions that contributed to the model, within which 8 were selected in at least 50% of the folds (and 4 in all folds). The two approaches led to overlapping lists of regions, with the fusiform areas ranked in the top 10 (Table 1, Figure 1). It therefore seems that both approaches were able to identify the signal of interest within visual areas in this particularly clean dataset, although the fusiform regions were not ranked first or the only one selected by the MKL model. MKL therefore provided a thresholded list of regions consistent with the literature, from a hierarchical whole-brain multivariate approach.

Rank	SVM whole brain (NW, %)	MKL (dm, %)
1	Parietal_Sup_L (2.23)	Lingual_L (28.12)
2	Occipital_Mid_R (2.10)	Occipital_Mid_L (26.17)
3	Fusiform_L (1.99)	Fusiform_L (15.80)
4	Occipital_Mid_L (1.83)	Lingual_R (8.33)
5	Parietal_Sup_R (1.74)	Fusiform_R (6.80)
6	Occipital_Sup_L (1.72)	Cingulum_Post_L (6.60)
7	Lingual_L (1.63)	Frontal_Inf_Oper_R (4.52)
8	Precuneus_L (1.60)	Parietal_Sup_L (1.07)
9	Lingual_R (1.59)	Caudate_L (0.68)
10	Fusiform_R (1.55)	Occipital_Mid_R (0.66)
11	Vermis_4_5 (1.47)	Calcarine_L (0.59)
12	Cerebelum_7b_R (1.39)	Occipital_Sup_L (0.59)
13	Calcarine_L (1.38)	Paracentral_Lobule_R (0.07)

Table 1: Ranking of the anatomically defined regions for (a) the whole brain SVM model summarized and (b) the sparse MKL model. The 13 regions displaying non-null weight parameter dm in the MKL are displayed on the right while regions are ranked according to NW for the SVM model. Regions common to both rankings are displayed in red. Please note that the ranking for the SVM model comprises all considered regions.

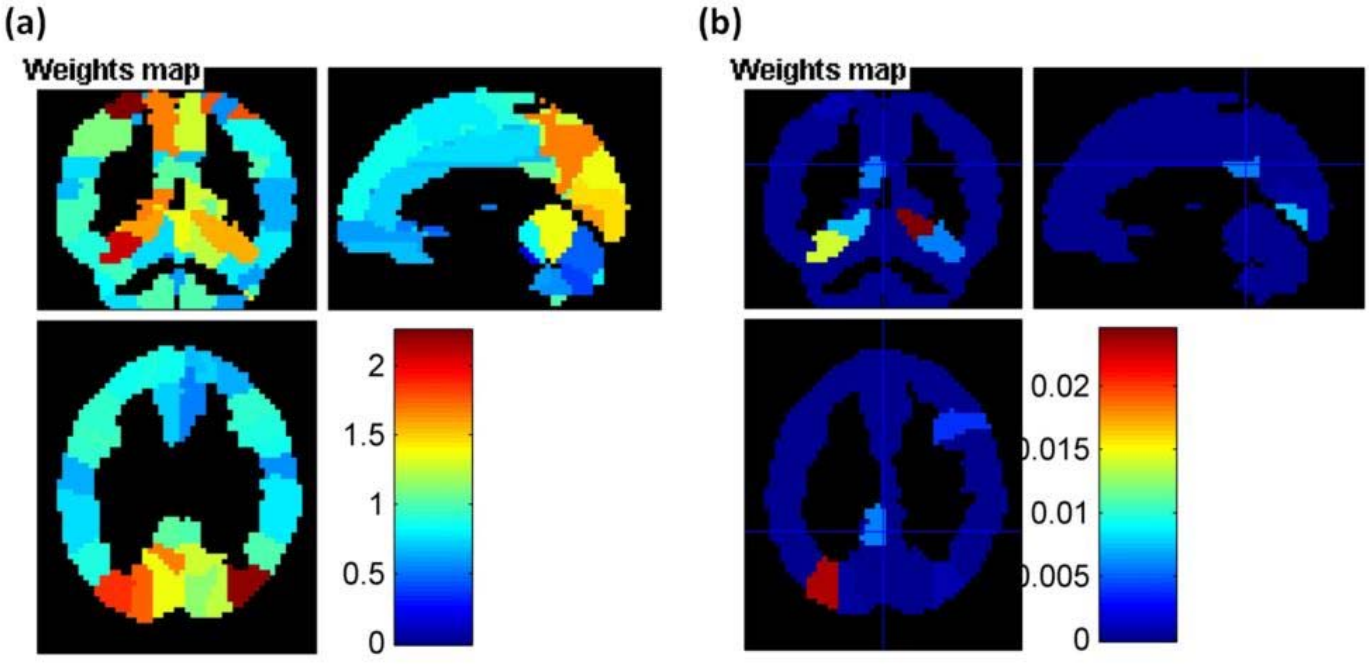

Figure 1: Maps of the weights per region (a) as summarized by NW after whole brain SVM modelling, (b) as learned from the MKL model (dm). The maps show that (b) is sparse, while (a) is not. Figures generated by PRoNTo.

Conclusions:

While machine learning models allow the prediction of a variable of interest, localizing the information leading to the prediction is complex due to their multivariate nature. In this work, we propose to use a priori anatomical information to build sparse hierarchical multivariate models and thereby facilitate model interpretation. Although the proposed approach depends on the precision and resolution of the anatomical template, the framework is general and can be applied to different templates. The methods were implemented in PRoNTo [9], which is a Matlab-based, SPM compatible toolbox.

Modeling and Analysis Methods:

Classification and Predictive Modeling

Reference

[1] Pereira, F., et al. (2009) 'Machine Learning Classifiers and fMRI: a tutorial overview', NeuroImage, vol 45, pp. S199-S209.
[2] Tzourio-Mazoyer, N., et al. (2002), 'Automated Anatomical Labeling of activations in SPM using a Macroscopic Anatomical Parcellation of the MNI MRI single-subject brain' NeuroImage, vol. 15, pp. 273-289.
[3] Kriegeskorte, N. et al. (2006) 'Information-based functional brain mapping', PNAS, vol. 103, pp. 3863-3868.
[4] Schrouff, J., et al. (2013) 'Discriminant BOLD Activation Patterns during Mental Imagery in Parkinson's Disease', Proceedings of Machine Learning Interpretration for NeuroImaging (MLINI 2012).
[5] Schrouff, J. et al. (2013) 'Localizing and comparing weight maps generated from linear kernel machine learning models.' Third International Workshop on Pattern Recognition in NeuroImaging (PRNI 2013): proceedings.
[6] Bach, F., et al. (2004) 'Multiple kernel learning, conic duality, and the SMO algorithm', Proceedings of the 21st International Conference on Machine Learning, pp. 41-48.
[7] Rakotomamonjy, A. et al. (2008) 'SimpleMKL', Journal of Machine Learning, vol. 9, pp. 2491-2521.
[8] Haxby, J., et al. (2001). 'Distributed and overlapping representations of faces and objects in ventral temporal cortex', Science, vol. 293, pp. 2425-2430.
[9] Schrouff, J., et al. (2013) 'PRoNTo: Pattern Recognition for Neuroimaging Toolbox',Neuroinformatics, vol. 2013, pp. 1-19.
www.mlnl.cs.ucl.ac.uk/pronto.