3643

PRoNTo v3.1: Elastic-Net Multiple Kernel Learning for Multisource Prediction

Professor Janaina Mourao-Miranda PhD[1,2], Konstantinos Tsirlis MSc[3], Professor Christophe Phillips PhD[4]
[1]Centre for Artificial Intelligence, Department of Computer Science, University College London, London, United Kingdom. [2]UCL Hawkes Institute, University College London, London, United Kingdom. [3]Department of Computer Science, London, United Kingdom. [4]GIGA-CRC Human Imaging, University of Liège, Liège, Belgium

Introduction

The Pattern Recognition for Neuroimaging Toolbox (PRoNTo)[1] is an open-source MATLAB toolbox that brings machine learning methods to neuroimaging, supporting classification and regression. A key feature of PRoNTo is its use of regularised linear kernel methods, which enable computationally efficient, generalisable, and interpretable models for high-dimensional data, even when the number of features exceeds the number of samples, a common situation in neuroimaging. Kernel methods represent data implicitly via a kernel function that encodes pairwise similarity; they are computationally efficient in high-feature low-sample-size settings because they operate on the kernel matrix (samples × samples) rather than the raw data matrix (samples × features).

PRoNTo also integrates multiple kernel learning (MKL) to combine information from several kernels. Each kernel can represent a different imaging modality or a different feature grouping (for example, regions of interest from a brain atlas). MKL learns an optimal weighted linear combination of these base kernels under a chosen regularisation to maximize predictive performance, estimating kernel weights and model feature weights simultaneously to enable principled integration and improved interpretability.

Different penalties have been used to regularize kernel weights estimation in MKL, including L1, L2, and general Lp norms. The elastic-net penalty, which combines L1 and L2, encourages both sparsity and selection of correlated kernels, making elastic-net MKL (ENMKL) especially useful when interpretability matters and kernels capture correlated information, as in neuroimaging. We recently introduced a computationally efficient ENMKL formulation that yields a simple analytical update for the kernel weights[2]. Here, we evaluated three approaches available in PRoNTo v3.1 for combining multiple data sources for prediction: ENMKL[2], SimpleMKL (L1-MKL)[3] , and baseline models (SVM and KRR) trained on the unweighted sum of kernels.

Methods

Data description

IXI dataset: The IXI subset (https://brain-development.org/ixi-dataset/) included structural Magnetic Resonance Imaging (MRI) volumes from Guy's Hospital (102 subjects, ages 60–90). Preprocessing in SPM8 followed standard steps: segmentation into grey and white matter, normalization to MNI space, and spatial smoothing with a Gaussian kernel (FWHM = [10 10 10] mm). For the IXI dataset the task was to predict the subject's age from their pre-processed structural MR images.

Haxby dataset: Haxby is a block-design functional MRI visual paradigm in which participants passively viewed greyscale images from eight categories (faces, cats, houses, chairs, scissors, shoes, bottles, and control nonsense images)[4]. We used data from a single subject (participant 1): 12 runs, each with eight 24 s blocks (each image shown for 500 ms with a 1500 ms inter-stimulus interval (ISI)); whole-brain volumes were acquired with repetition time (TR) = 2.5 s. Preprocessing in SPM8 included motion correction, segmentation, and normalization to MNI space; no spatial smoothing was applied. For the Haxby data the task was classifying pre-processed fMRI volumes during the visualisation of faces versus visualisation of houses.

Experimental Protocol

The experiments were implemented using the PRoNTo v3.1 batch framework. For each dataset we used the Automated Anatomical Labeling (AAL)[5] template to partition images into 116 anatomically defined regions. The full AAL atlas served as a mask to select voxels or features from each dataset. For each region we computed a linear kernel from the regional pattern of tissue or activation by using all voxels within that region. Each regional kernel was then mean-centred and normalized prior to combination. Mean-centring the kernel is equivalent to subtracting the feature mean across samples (computing the mean based on the training data). Normalising the kernel corresponds to dividing each sample's feature vector by its norm, which compensates for differences in region or modality size and ensures kernels computed from regions with different numbers of voxels are comparable when used in MKL.

We compared two elastic-net algorithms (ENMKL-SVM and ENMKL-KRR) with the SimpleMKL (classification and regression), which corresponds to a L1-norm regularized MKL algorithm. In addition we also compared the MKL classification models with an SVM trained on the unweighted sum of the kernels and the MKL regression models with a KRR trained on the unweighted sum of the kernels.

Results

IXI results: The lowest mean squared error (MSE) and highest correlation between actual and predicted age were obtained with ENMKL-KRR (MSE = 25.25; r = 0.54), followed by KRR on the unweighted kernel sum (MSE = 29.18; r = 0.48) and SimpleMKL regression (MSE = 29.63; r = 0.44). ENMKL-KRR selected all kernels, whereas SimpleMKL retained 59 of the 116 kernels. These results suggest a distributed predictive pattern, with most brain regions contributing information about age.

Figure 1 (a-c) shows the weight maps for the IXI dataset: ENMKL-KRR (a), SimpleMKL regression (b), and KRR (c). The top five regions according to the ENMKL-KRR kernel weights (values in brackets) were: left Heschl (2.63), right Heschl (2.26), left hippocampus (1.50), left superior temporal (1.39), and left fusiform (1.32). The top five regions according to the SimpleMKL regression kernel weights were: left Heschl (11.87), right middle orbital frontal (9.13), left putamen (6.65), left hippocampus (6.07), and left fusiform (4.18).

Haxby results: Both MKL models, ENMKL-SVM and SimpleMKL classification, achieved identical performance: balanced accuracy = 97.22% and AUC = 100%, outperforming the standard SVM (balanced accuracy = 81.02%, AUC = 98%). ENMKL-SVM selected 56 kernels, compared with 18 kernels selected by SimpleMKL; both models selected fewer kernels than in the previous (IXI) task. These findings suggest that the discriminative patterns in the Haxby dataset are sparser than in the IXI dataset, which is consistent with the tasks: age influences widespread brain structure, whereas visual processing relies on a more localised network of regions.

Figure 1 (d-f) shows the weight maps for the Haxby dataset: ENMKL-SVM (d), SimpleMKL classification (e), and SVM (f). The top five regions according to the ENMKL-SVM kernel weights (values in brackets) were: left fusiform (16.87), right lingual (14.81), left middle occipital (12.12), right fusiform (11.02), and left lingual (10.16). The top five regions according to the SimpleMKL classification kernel weights were: right lingual (24.00), left fusiform (22.29), left middle occipital (17.42), left lingual (10.25), and right fusiform (8.22).

Results Figure (Optional)

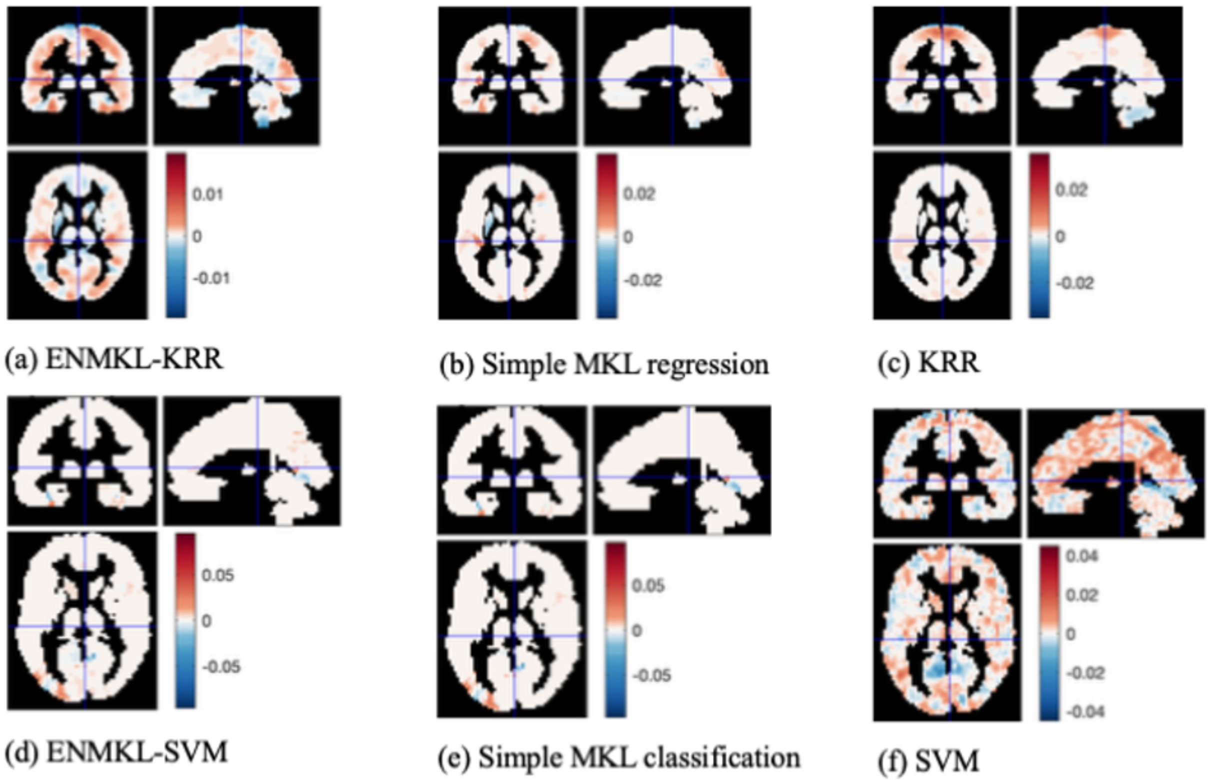

Figure 1. Voxel-wise weight maps — top row (IXI dataset): ENMKL-KRR (a), SimpleMKL regression (b), KRR (c); bottom row (Haxby dataset): ENMKL-SVM (d), SimpleMKL classification (e), SVM (f).

Conclusion

We compared three approaches available in PRoNTo v3.1 for combining multiple data sources for prediction: Elastic-Net Multiple Kernel Learning (ENMKL), SimpleMKL, and baseline models (SVM and KRR) trained on the unweighted sum of kernels. Evaluations used two neuroimaging datasets and two tasks: a regression task (predicting age from structural MRI) and a classification task (classifying patterns of brain activation during visualisation of faces versus houses). Across both tasks, ENMKL performed as well as or better than SimpleMKL and the baseline models. Unlike KRR or SVM trained on the unweighted sum of kernels, SimpleMKL enforces sparsity in the kernel weights via an L1 penalty, while ENMKL enforces an elastic-net penalty in the kernel weights that combines L1 (sparsity) and L2 (grouping) regularisation. This enables ENMKL to identify sets of correlated regions (for example, bilateral regions) and to reveal the effective level of sparsity in the predictive model, providing improved performance and greater interpretability.

While we demonstrate the benefits of ENMKL for combining brain regions for prediction, it can also be applied to combine features from different data sources, such as multiple imaging and non-imaging modalities for prediction.

References/Citations

1. Schrouff J, Rosa MJ, Rondina JM, Marquand AF, Chu C, Ashburner J, Phillips C, Mourão-Miranda J. PRoNTo: pattern recognition for neuroimaging toolbox. Neuroinformatics. 2013;11(3):319-337. doi:10.1007/s12021-013-9178-1
2. Mourão-Miranda J, Hussain Z, Tsirlis K, Phillips C, Shawe-Taylor J. Elastic-Net Multiple Kernel Learning: Combining Multiple Data Sources for Prediction. Preprint. Posted online December 12, 2025. doi:10.48550/arXiv.2512.11547
3. Rakotomamonjy A, Bach F, Canu S, Grandvalet Y. SimpleMKL. J Mach Learn Res. 2008;9:2491-2521.
4. Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 2001;293(5539):2425-2430.
5. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated Anatomical Labeling of activations in SPM using a Macroscopic Anatomical Parcellation of the MNI MRI single-subject brain. NeuroImage. 2002;15(1):273-289.

Primary Parent Category & Sub-Category

Modeling and Analysis Methods: Classification and Predictive Modeling

Secondary Parent Category & Sub-Category

Neuroinformatics and Data Sharing: Informatics Other

Keywords

Multivariate, Open-Source Code, Open-Source Software