Pattern Recognition for Neuroimaging Toolbox

Jessica Schrouff
University of Liège, Belgium
jschrouff@ulg.ac.be

Maria J. Rosa
University College London, UK
m.rosa@ucl.ac.uk

Jane Rondina
University College London, UK

Andre F. Marquand
Kings College, London, UK

Carlton Chu
NIMH, NIH, USA

John Ashburner
University College London, UK

Jonas Richiardi
Stanford University, USA

Christophe Phillips
University of Liège

Janaina Mourao-Miranda
University College London, UK

Abstract: In the past years, mass univariate statistical analyses of neuroimaging data have been complemented by the use of multivariate pattern analyses, especially based on machine learning models. While these allow an increased sensitivity for the detection of spatially distributed effects compared to univariate techniques, they lack an established and accessible software framework. Here we introduce the "Pattern Recognition for Neuroimaging Toolbox" (PRoNTo), an open-source, cross-platform and MATLAB-based software comprising many necessary functionalities for machine learning modelling of neuroimaging data.

Keywords: software, neuroimaging, machine learning

1 Introduction

Various imaging modalities, such as functional/structural Magnetic Resonance Imaging (fMRI/sMRI) and Positron Emission Tomography (PET), have been developed to record brain structure and activity. Until recently, such data were analysed using standard univariate statistics, for example by linking the time-series of the signal in each voxel with a regressor, such as in the General Linear Model (GLM) implemented in Statistical Parametric Mapping (SPM, [2]). Although univariate analyses have proven powerful for making regionally specific inferences on brain function and structure, there are limitations to the type of research questions that they can address. More recently, these mass univariate analyses have been complemented by the use of pattern recognition analyses, in particular using machine learning based predictive models [3]. These analyses focus on predicting a variable of interest (e.g. mental state 1 vs. mental state 2, or patients vs. controls) from the pattern of brain activation/anatomy over a set of voxels. Due to their multivariate properties, these methods can achieve relatively greater sensitivity and are able to detect subtle, spatially distributed patterns in the brain. Potentially, pattern recognition can also be used to perform computer-aided diagnostic of neurologic or psychiatric disorders. Currently, the existing implementations consist of small code snippets, or sets of packages, and lack a dedicated single, integrated, and flexible software framework. In addition, the use of existing packages often requires high-level programming skills.

2 Pattern Recognition for Neuroimaging Toolbox

The "Pattern Recognition for Neuroimaging Toolbox" (PRoNTo[1], [6]) is a user-friendly and open-source toolbox that makes machine learning modelling available to every neuroimager. In PRoNTo, brain scans are treated as spatial patterns and learning models are used to identify statistical properties of the data that can be used to discriminate between experimental conditions or groups of subjects (classification models) or to predict a continuous measure (regression models). In terms of neuroimaging modalities, PRoNTo accepts NIfTI files[2] and can therefore be used to analyse sMRI and fMRI, PET, SPM contrast images and potentially any other modality in NIfTI file format. Its framework allows fully flexible machine learning based analyses and, while its use requires no programming skills, advanced users can easily access technical details and expand the toolbox with their own developed methods. Each step of the analysis can also be reviewed via user-friendly displays. Figure 1 provides an overview of the toolbox framework.

---

[1]PRoNTo, and all its documentation, are available to download freely from: http://www.mlnl.cs.ucl.ac.uk/pronto/
[2]http://nifti.nimh.nih.gov/nifti-1/
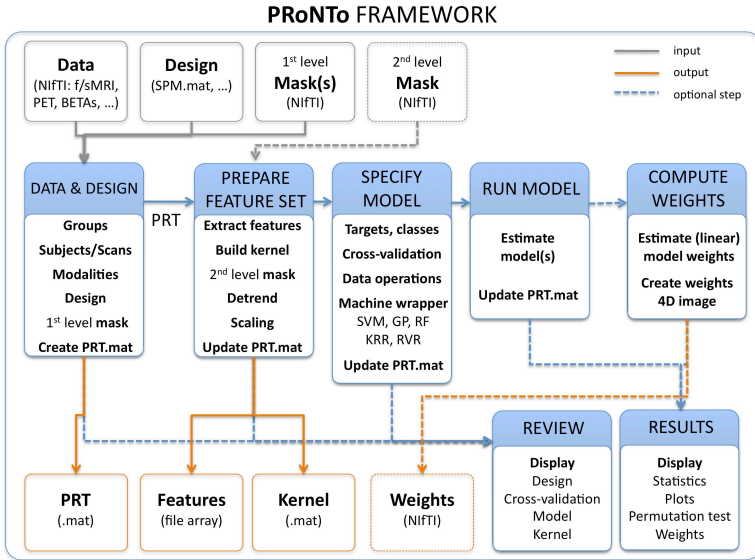
PRoNTo FRAMEWORK

Fig. 1: PRoNTo framework. PRoNTo consists in five main analysis modules (blue boxes in the centre): dataset specification, feature set selection, model specification, model estimation and weights computation. In addition, it provides two main reviewing and displaying facilities (model, kernel and cross-validation displays, as well as, results display). PRoNTo receives as input any NIfTI image (comprising the data and a first-level mask, while an optional second-level mask can also be entered). The outputs of PRoNTo include: a data structure called PRT.mat, a data matrix (with all features), one or more kernels, and (optionally) images with the classifier weights.

PRoNTo can be used in three ways: through a graphical user-interface requiring no programming skills, using the MATLAB-batch system , or by scripting function calls. In the current version of PRoNTo, two linear kernel classification algorithms are embedded in the framework: Support Vector Machines ([1], LIBSVM implementation) and (binary and multiclass) Gaussian Process classification ([4], GPML toolbox). Regression can be performed using Kernel-Ridge Regression (KRR, [5]), Relevance Vector Regression (RVR, [7]) or Gaussian Processes Regression [4]. All algorithms are wrapped into what is called a "machine", which is independent from the design definition and cross-validation procedure. This allows an easy integration of new machine learning algorithms, enhancing the exchange of newly developed methods within the community, and the possibility to develop more advanced validation frameworks (e.g. nested cross-validation).

Several data sets were analysed with PRoNTo, showing the breadth of questions it can address [6]. As examples: Do patterns of brain activation recorded in fMRI encode information about a mental state? Can groups of subjects be distinguished based on features derived from their sMRI? Could their age be predicted with these neuro-anatomical features?

3 Discussion and Conclusions

In this work, we presented PRoNTo, a freely available software which addresses neuroscientific questions using machine learning based modelling. Although in its first version, PRoNTo provides both graphical interfaces for an easy use and a flexible programming framework. The authors therefore hope to facilitate the interaction between the neuroscientific and machine learning communities. On one hand, the machine learning community should be able to contribute to the toolbox with novel machine learning models. On the other hand, the toolbox should provide a variety of tools for the (clinical) neurosciencists, enabling them to ask new questions that cannot be easily investigated using existing statistical analysis tools.

References

[1] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[2] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny. *Statistical Parametric Mapping: the analysis of functional brain images*. Elsevier Academic Press, London, 2007.

[3] T. Hastie, R. Tibshirani, and J. Friedman. *Elements of Statistical Learning*. Springer, 2003.

[4] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.

[5] C. Saunders, A. Gammerman, and V. Vovk. Ridge regression learning algorithm in dual variables. In *Proceedings of the 15th International Conference on Machine Learning*, pages 515–521, 1998.

[6] J. Schrouff, M. Rosa, J. Rondina, A. Marquand, C. Chu, J. Ashburner, C. Phillips, J. Richiardi, and J. M. ao Miranda. PRoNTo: Pattern Recognition for Neuroimaging Toolbox. *Neuroinformatics*, pages 1–19, 2013.

[7] M. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.