BrainStates: an fMRI data analysis toolbox

Iga Adamska[1,2], Karolina Finc[1]

[1]Centre for Modern Interdisciplinary Technologies, Nicolaus Copernicus University in Toruń, Poland

[2]Faculty of Philosophy and Social Sciences, Nicolaus Copernicus University in Toruń, Poland

The complexity of the brain architecture is accompanied by the dynamics of neuronal activity while performing a task. Brain's activity can be measured using functional magnetic resonance imaging (fMRI) - a method enabling to track the changes in blood oxygenation, known as a BOLD (blood oxygen-level dependent) signal. The most common analytical approaches, developed to localize brain regions active during certain task conditions, rely on the General Linear Model (GLM). Yet, the GLM-based methods do not capture the dynamics of time-varying brain activity; thus, new computational approaches are needed to describe these complex brain activity patterns.

Discrete states of brain activity can be identified using unsupervised machine learning methods, (Cornblath et al., 2020). Clustering techniques, such as K-Means clustering (MacQueen, 1967) or hierarchical clustering (Ward & Joe, 1963), can be used to group together fMRI time-series with similar activity patterns and to identify so-called *brain states*. These brain states can be characterised by taking into account a number of properties, such as: state duration, state's spatial architecture, persistence or transition probability.

So far, there are no available open-source and well documented tools for automatic detection and characterization of brain states derived from fMRI data that allow their analysis in an intuitive and easy way. Therefore, our main goal was to develop a new open-source toolbox dedicated to describe and visualize the dynamics of time-varying brain activity. Our motivation was to create a tool that would help young researchers and students start with the fMRI data analysis and get insight into the brain dynamics, rather than make sophisticated software that would be difficult to use. Moreover, professionals and advanced researchers can also benefit from its simplicity and robustness.

The tool is developed in the Python programming language that is well-known because of its readability, modularity, and comprehensive standard library (Muller et al., 2015, Oliphant, 2007). Ultimately, it will be openly available for researchers, together with the detailed documentation and tutorial.

The proposed tool implements an unsupervised machine learning approach to cluster both resting-state and task-based fMRI time-series into discrete brain states (Cornblath et al., 2020). Each brain state is characterized and visualized by taking into account different brain states' characteristics. The toolbox is divided into three modules, containing functions for: (1) clustering, (2) calculating brain states properties, and (3) visualization. The clustering module includes a set of functions that identifies brain states using K-Means clustering algorithm with Euclidean distance as a measure of similarity. Next, the module allowing brain states characteristics contains a set of functions calculating several measures which describe brain states, such as:

- fractional occupancy - the percentage of time points in each run classified as a particular state
- dwell times - the mean length of consecutive timepoints classified as a particular state
- persistence probability - probability in remaining in the same state
- transition probability - the probability of the transition from state to another

Finally, the visualization module, implementing Nilearn and Matplotlib functions, provides aesthetic visualizations of brain states.

The toolbox will be provided along with a tutorial and detailed documentation containing a full description of each module and its functions, making this tool easy to use and suited even for non-advanced programmers. It is the first tool to provide such fMRI data analysis along with comprehensive specification of every function. Additionally, the results of the project will be published on open source GitHub and Zenodo platforms, which makes it accessible to everyone.

The proposed tool provides an innovation for the neuroscience community, as no such tool was previously developed. The most important feature - developing a tool for detecting brain states and calculating their properties - will implement unsupervised machine learning, a state-of-the-art approach that enables to discover hidden structure in data (Vu et al., 2018). Usage of this toolbox will extend basic GLM-based approaches, commonly used to analyze fMRI data, by identifying distinct brain states with no explicit knowledge about the structure of the task. Thus, methods used in the toolbox will bring the scientific community closer to understanding complex brain dynamics. Representing brain activity in the form of dynamic brain states and their characteristics might be useful in clinical applications, such as brain stimulation or medical treatment of neuropsychiatric disorders (Deco & Kringelbach, 2014). Therefore, the usage of this toolbox should be perceived as one of the first steps towards the personalised treatment by understanding the patient's temporal activity. The tool finds use for resting state fMRI under unusual conditions such as altered states of consciousness induced by psychedelics (Singleton et al., 2021 preprint), abnormal brain activity in mental disorders (Du et al., 2021) or naturalistic stimuli (van der Meer et al., 2020). Extending the variety of the fMRI data analysis allows one to get a comprehensive overview of a subject's brain activity, distinguish between abnormal and healthy states of brain activity and consequently, provides detailed information about their physical and mental health, general well-being and cognitive performance. Hence, the brain states' analysis provided by this toolbox also shows potential use in clinical research.

References:

Cornblath, E.J., Ashourvan, A., Kim, J.Z. *et al.* (2020). Temporal sequences of brain activity at rest are constrained by white matter structure and modulated by cognitive demands. *Commun Biol*. 3, 261. doi: 10.1038/s42003-020-0961-x

Deco, G., Kringelbach, M. L. (2014). Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron*, 84(5), 892-905. doi: 10.1016/j.neuron.2014.08.034

Du, M., Zhang, L., Li, L. *et al*. (2021). Abnormal transitions of dynamic functional connectivity states in bipolar disorder: A whole-brain resting-state fMRI study. *Journal of affective disorders*, *289*, 7-15. doi: 10.1016/j.jad.2021.04.005

Joe H., Ward Jr. (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association,* 58:301, 236-244, doi: 10.1080/01621459.1963.10500845

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), 281-297

van der Meer, J.N.v.d., Breakspear, M., Chang, L.J. *et al.* (2020). Movie viewing elicits rich and reliable brain state dynamics. *Nat Commun.* 11, 5004 . doi: 10.1038/s41467-020-18717-w

Muller, E., Bednar, J. A., Diesmann, M. *et al*. (2015) Python in neuroscience. *Front. Neuroinform*. 9:11. doi: 10.3389/fninf.2015.00011.

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science & Engineering*, *9*(3), 10-20. doi: 10.1109/MCSE.2007.58.

Singleton, S. P., Luppi, A. I., Carhart-Harris, R. L. *et al*. (2021). LSD flattens the brain's energy landscape: evidence from receptor-informed network control theory. *bioRxiv*. doi: 10.1101/2021.05.14.444193

Vu, M. A. T., Adalı, T., Ba. *et al*. (2018). A shared vision for machine learning in neuroscience. *J Neurosci*. 38(7), 1601-1607. doi: 10.1523/JNEUROSCI.0508-17.2018