Unsupervised Identification of Behaviorally-relevant States in Biological and Artificial Neural Systems

Arman Behrad (arman.behrad@tu-dresden.de)
Computational Neuroscience, Dept. of Child and Adolescent Psychiatry,
Faculty of Medicine, TU Dresden, Germany

Mohammad Taha Fakharian (fakharian.taha@ut.ac.ir)
School of Electrical and Computer Engineering,
University of Tehran, Iran

Christian Beste (christian.beste@ukdd.de)
Cognitive Neurophysiology, Dept. of Child and Adolescent Psychiatry,
Faculty of Medicine, TU Dresden, Germany

Shervin Safavi (research@shervinsafavi.org)
Computational Neuroscience, Dept. of Child and Adolescent Psychiatry,
Faculty of Medicine, TU Dresden, Germany
Dept. of Computational Neuroscience,Max Planck Institute for Biological Cybernetics,
Tübingen,Germany

Abstract

Identifying distinct neural dynamics corresponding to cognitive states and their transitions is crucial for understanding the neural machinery of cognitive functions in both biological and artificial intelligent systems. However, conventional methods for state identification constrain the analysis by relying on predefined state labels. Here, we introduce a novel unsupervised approach to robustly detect behaviorally-relevant state transitions without prior assumptions or knowledge about behavioral labels. We assume that each state has a characteristic dynamics within each state (with minimal variation), but triggered by behavioral demands, transitions to other states (with different characteristic dynamics). Therefore, comparing neural dynamics across time, should provide us key information about state transitions. Based on this idea, we developed *Moving Window Dynamical Similarity Analysis (MoDSA)* for an unbiased detection of state transitions in neural systems. We validated our method on biological neural data recorded from macaque area V4 during selective attention tasks, and data from diverse recurrent neural networks trained on context-dependent decision-making tasks. We demonstrate that our method can identify behaviorally meaningful states purely based on neural dynamics, in both domains of artificial and biological neural systems.

Keywords: brain state, neural dynamics, dynamical similarity analysis, recurrent neural network, brain-behavior embedding

State in biological and artificial neural systems

A state in cognitive neural systems, whether biological or artificial, should be defined based on neural dynamics that informs about the behavior. The state can be a characteristic and sufficiently stable neural activity that dynamically changes over time depending on the behavioral demands. Such state transitions are shaped by internal factors (e.g., attention, Flavell et al., 2022) and/or environmental input (e.g., task cues, Gonzalez-Castillo & Bandettini, 2018). Despite the significance of identifying these behaviorally relevant and distinct dynamical states in an *unsupervised* manner, this remains a challenge for the analysis of data from both artificial and biological neural networks. Thus, we develop an **unsupervised** approach for identifying behaviorally relevant states, purely based on network dynamics without behavioral labels.

Identifying states through temporal evolutions

We assume each state has a stable characteristic neural dynamics for distinct behavioral phases (Figure 1a); thus, neural dynamics within each state vary minimally, but across different states vary drastically. Therefore, comparing neural dynamics across time should provide us with key information about state transitions. Based on this idea, we developed *Moving Window Dynamical Similarity Analysis (MoDSA)* for an unbiased detection of state transitions in neural systems. MoDSA extends Dynamical Similarity Analysis (DSA, Ostrow et al., 2023), a method that compares the dynamics of two networks by embedding them into high-dimensional linear systems, capturing essential features of original nonlinear dynamics. Unlike fixed-point analyses, DSA evaluates global dynamic structures, efficiently identifying similarities despite differing geometries or inapplicable perturbations. Importantly, DSA utilizes Procrustes analysis to quantify vector field similarity. To improve efficiency, we introduce *fastDSA*, an optimized variant of DSA.

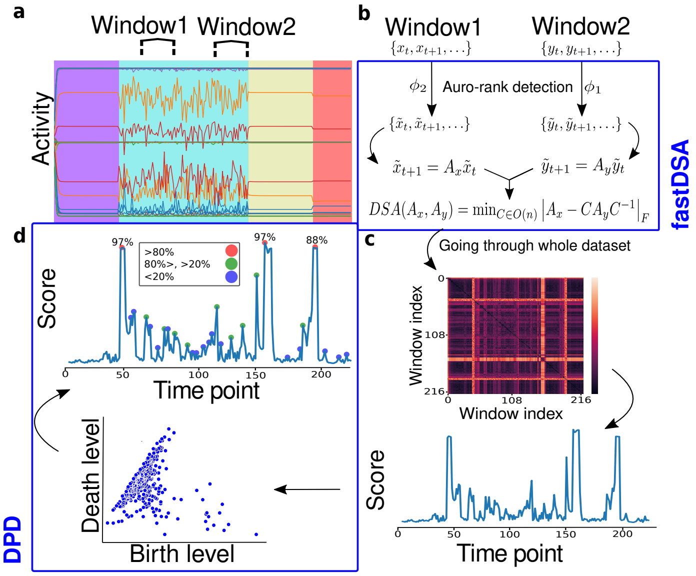
a
Window1
Window2
Activity
b
Window1
$\{x_t, x_{t+1}, \dots\}$
Window2
$\{y_t, y_{t+1}, \dots\}$
ϕ_2
Auro-rank detection
ϕ_1
$\{\tilde{x}_t, \tilde{x}_{t+1}, \dots\}$
$\{\tilde{y}_t, \tilde{y}_{t+1}, \dots\}$
$\tilde{x}_{t+1} = A_x \tilde{x}_t$
$\tilde{y}_{t+1} = A_y \tilde{y}_t$
$DSA(A_x, A_y) = \min_{C \in O(n)} \left| A_x - C A_y C^{-1} \right|_F$
fastDSA
c
Going through whole dataset
Window index
0
108
216
0
108
216
Window index
Score
0
50
100
150
200
Time point
d
97%
97%
88%
>80%
80%>, >20%
<20%
Score
0
50
100
150
200
Time point
Death level
Birth level
DPD

Figure 1: (a) Activity of individual neurons of an RNN performing a contextual decision-making task. Color shades indicate different behavioral phases (fixation, processing stimulus, delay, and decision). (b) Two example windows (noted in (a)) undergo processing by *fastDSA*, where dynamics in each window are first represented by linear models. ϕ indicates SVD on the delay embedding matrix. These linear systems are then quantitatively compared using Procrustes analysis to evaluate their dynamical similarity (further detailed in the text). (c) MoDSA systematically compares all pairs of windows (sliding through the entire length of neural activity), generating a comprehensive distance matrix. Averaging this matrix columnwise yields a scoreline that reflects dynamical similarity across the time course. (d) Democratic Peak Detection (*DPD*) then processes the scoreline, calculating (based on TDA) the state transition probability in an unsupervised manner.

FastDSA enhances DSA efficiency through two innovations: First Automatic optimal rank determination via Optimal Singular Value Hard Thresholding (SVHT, Gavish & Donoho, 2014), minimizing computational overhead without sacrificing essential information on the dynamics. On the next step, a hybrid similarity transformation approach, optimizing orthogonality constraints through regularized gradient descent followed by a singular value decomposition-based orthogonality enforcement step. Specifically, we minimize the regularized loss function (λ stands for penalty term, also see, Figure 1b):

$$L(C) = |A - CBC^T|_F^2 + \lambda|C^TC - I|_F^2 \, , \tag{1}$$

and iteratively update the transformation matrix C using gradient descent:

$$C^{(t+1)} = C^{(t)} - \eta\left[-4(A - CBC^T)BC + 4\lambda(CC^TC - C)\right] . \tag{2}$$

MoDSA employs a moving-window approach, repeatedly calculating fastDSA across pairs of time windows (Figure 1c). The size of the window is automatically detected as the smallest window that can capture meaningful dynamics. For windows within each state, neural dynamics will remain similar (thus small distances based on DSA) whereas transitions significantly increase this metric (Figure 1d, top, aligned with Figure 1a). These *scoreline* reflect a form of state transition probability.

To identify when the network elicits a state transition (peaks in the scorelines, see Figure 1d), we developed *Democratic Peak Detection (DPD)* to robustly capture peaks in an unsupervised manner. To this end, we apply Topological Data Analysis (TDA) to extract robust, noise-insensitive topological features from the scoreline data (Figure 1d bottom). Persistent homology, a fundamental TDA method, captures how topological features emerge and vanish across varying scales, through a process called *filtration*. Intuitively, persistent homology is like watching a landscape slowly flood. Small puddles (features) appear at first (birth), and as water rises, some merge or vanish (death). The longer a feature lasts before disappearing, its persistence, the more meaningful it is. Thus, we transform the extracted topological features into a geometric framework within the Birth-Death coordinate system. It has been shown that in this topological coordinate system, data points away from the diagonal are the peaks of time-series (Bois et al., 2024). To identify off-diagonal data points robustly and objectively, we employ an ensemble learning strategy comprising 13 distinct unsupervised outlier detection algorithms (Zhao et al., 2019, and references therein). The results of these algorithms are aggregated through a voting procedure, assigning each point a probability reflecting its likelihood as a genuine peak (Figure 1d top, numbers reflect the votes). This democratic approach ensures robust, reliable, and domain-agnostic identification of dynamical state transitions.

Behaviorally-relevant state identification

We validated our method for identifying state transitions using biological data recorded from macaque monkeys and data from various recurrent neural network (RNN) architectures. For the biological data set, we analyze the spiking data from area V4 during a selective attention task (Engel et al., 2016; Zeraati et al., 2023). Monkeys were trained to detect changes in visual stimuli and report antisaccadic responses. The spiking data was collected using 16-channel linear array electrodes converted into peri-stimulus time histograms. The V4 neural activity was recorded during different phases of the task, allowing us to examine the state dynamics associated with the attention and perceptual decision-making processes.

To investigate states in artificial neural networks, we employed several RNN architectures, including long- and short-term memory (LSTM) networks, gated recurrent units (GRUs) and vanilla RNNs. These networks were trained in context-dependent decision-making (CDM, Mante et al., 2013) tasks consisting of four distinct phases: fixation, stimulus, delay, and decision phases. During the stimulus phase, the network simultaneously received input from two sensory modalities (e.g., color and motion). After a delay, the network was trained to make perceptual decisions based solely on one modality (e.g., dominant motion direction or dominant color), ignoring the irrelevant modality.

Our method computes the probability of behaviorally-relevant state transitions at each time point (the time of the peaks corresponds to state transition times). **As the method is built based on neural dynamics principles, it is applicable to monkey neural recordings and RNN data.** Overall, our unsupervised approach demonstrates a robust method for state identification in both biological and artificial cognitive systems, highlighting its versatility in exploring the neural dynamics underlying complex cognitive processes.

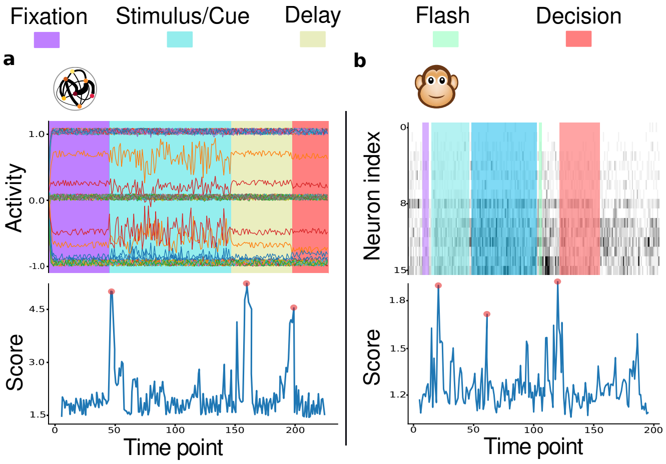

Figure 2: Application of MoDSA on (a) The hidden layers of an RNN (with LSTM units) and (b) peri-stimulus time histograms (PSTH) of V4 neurons. The red dots indicate time points with high (>%80) probability of state transition.

References

Bois, A., Tervil, B., & Oudre, L. (2024). A persistent homology-based algorithm for unsupervised anomaly detection in time series. *Transactions on Machine Learning Research*.

Engel, T. A., Steinmetz, N. A., Gieselmann, M. A., Thiele, A., Moore, T., & Boahen, K. (2016). Selective modulation of cortical state during spatial attention. *Science*, *354*(6316), 1140–1144.

Flavell, S. W., Gogolla, N., Lovett-Barron, M., & Zelikowsky, M. (2022). The emergence and influence of internal states. *Neuron*, *110*(16), 2545–2570.

Gavish, M., & Donoho, D. L. (2014). The optimal hard threshold for singular values is $4/\backslash$sqrt $\{3\}$. *IEEE Transactions on Information Theory*, *60*(8), 5040–5053.

Gonzalez-Castillo, J., & Bandettini, P. (2018). Task-based dynamic functional connectivity: Recent findings and open questions. neuroimage, 180 (pt b), 526–533.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, *503*(7474), 78–84.

Ostrow, M., Eisen, A., Kozachkov, L., & Fiete, I. (2023). Beyond geometry: Comparing the temporal structure of computation in neural circuits with dynamical similarity analysis. *Advances in Neural Information Processing Systems*, *36*, 33824–33837.

Zeraati, R., Shi, Y.-L., Steinmetz, N. A., Gieselmann, M. A., Thiele, A., Moore, T., Levina, A., & Engel, T. A. (2023). Intrinsic timescales in the visual cortex change with selective attention and reflect spatial connectivity. *Nature communications*, *14*(1), 1858.

Zhao, Y., Nasrullah, Z., & Li, Z. (2019). Pyod: A python toolbox for scalable outlier detection. *Journal of Machine Learning Research*, *20*(96), 1–7.