

DYNAMIC TOPOLOGICAL DATA ANALYSIS FOR FUNCTIONAL BRAIN SIGNALS

Tananun Songdechakraiwut, Moo K. Chung

University of Wisconsin–Madison, USA

ABSTRACT

We propose a novel dynamic topological data analysis (TDA) framework that builds persistent homology over a time series of 3D functional brain images. The proposed method encodes the time series as a time-ordered sequence of Vietoris-Rips complexes and their corresponding barcodes in studying dynamically changing topological patterns. The method is applied to the resting-state functional magnetic resonance imaging (fMRI) of the human brain. We demonstrate that the *dynamic-TDA* can capture the topological patterns that are consistently observed across different time points in the resting-state fMRI.

Index Terms— Topological data analysis, persistent homology, barcodes, time series, resting-state fMRI

1. INTRODUCTION

Topological Data Analysis (TDA) [1, 2, 3] provides a general framework to analyze high-dimensional and noisy data using techniques from topology. Persistent homology, a branch of the TDA, is a method for measuring topological features inferred from a simplicial complex at different spatial resolutions. The general principle underlying the persistent homology is based on the persistence of k -dimensional holes. For instance, a 0-dimensional hole is a connected component and a 1-dimensional hole is a circular loop. The persistence of such topological features is quantified using a barcode [2].

The persistent homology has been widely used for various medical imaging applications. It is mainly applied to capture the static summary of dynamically changing images including functional magnetic resonance imaging (fMRI) data [4, 5] and electroencephalography data [6]. Nonetheless, there are a few notable studies from the growing literature that begin to apply the TDA to capture the dynamic patterns of time series including the application to financial data [7] and gene expression data [8]. Motivated by these studies, we leverage the application of the TDA to dynamically changing resting-state fMRI data and investigate whether this could lead to new topological characterization of the brain.

Correspondence should be sent to Tananun Songdechakraiwut (email: songdechakra@wisc.edu) or Moo Chung (email: mkchung@wisc.edu). This study was funded by NIH Grants R01 EB022856 and EB028753.

The main contributions of the paper are as follows. 1) We present a new *dynamic-TDA* framework that builds the persistent homology over a multivariate time series. 2) The method is applied to the resting-state fMRI of more than 400 human subjects in investigating the stationarity and sexual dimorphism of the resulting topological signals.

2. METHODS

2.1. Dynamic-TDA

Consider a d -dimensional multivariate time series $x[t] = (x_1[t], \dots, x_d[t]) \in \mathbb{R}^d$. For each sliding window of length w at time t , we have a point cloud consisting of w points $X[t] = \{x[t], \dots, x[t + w - 1]\}$ in \mathbb{R}^d . Given the point cloud $X[t]$, we then construct a Vietoris-Rips complex K_ϵ whose k -simplices correspond to unordered $(k + 1)$ -tuples of points, which are pairwise within distance ϵ [3]. For sufficiently small ϵ , the complex is a set consisting only nodes; for sufficiently large ϵ , the complex is a single connected $(w - 1)$ -dimensional simplex. The Vietoris-Rips complex K_ϵ only changes at a finite number of increasing filtration values ϵ allowing efficient computation.

The persistent homology looks for the topological features such as the k -dimensional holes that persist for the whole range of the parameter ϵ . The persistences of the k -dimensional holes can be represented using barcodes. The k -dimensional barcode (k D barcode) corresponding to the Vietoris-Rips complex K_ϵ is a collection I of intervals $[t_{start}, t_{end}]$ such that each interval tabulates the life-time of a k -dimensional hole that appears at the filtration value t_{start} and vanishes at the value t_{end} . Given the k D barcode for the sliding window at time t , we calculate the sum $\sum_{i \in I} (t_{end} - t_{start})_i$ of all the life-times to obtain the time series $B[t]$, which measures the total life-time of all the k D holes. Figure 1 displays the schematic of the proposed dynamic-TDA.

2.2. Trend stationary model on barcodes

We model the time series $B[t]$ using the trend stationary model [9], which decomposes the time series as a nonstationary trend component and a zero-mean stationary noise. We

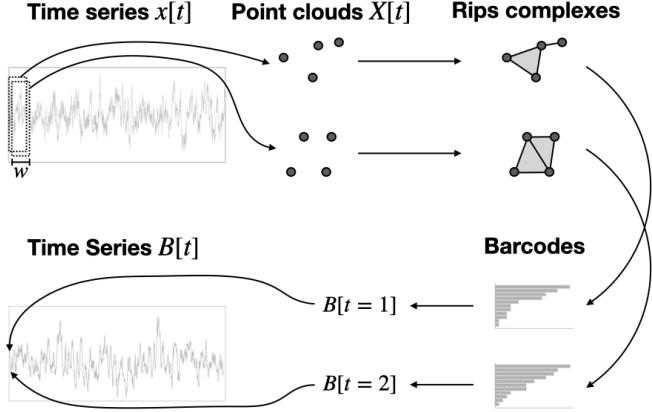


Fig. 1. Schematic of the dynamic-TDA.

write the model as

$$B[t] = \mu[t] + z[t]$$

where $\mu[t]$ denotes the unknown signal to be estimated and $z[t]$ is a stationary noise process. Since the trend is expected to smoothly vary, we assume the signal $\mu[t]$ to be a polynomial of the form $\sum_{i=0}^p c_i t^i$, which is estimated using the least squares method by solving $B = \mathbf{P}c$, where

$$\mathbf{B} = \begin{bmatrix} B[1] \\ B[2] \\ \vdots \\ B[m] \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 1^p & 1^{p-1} & \dots & 1^0 \\ 2^p & 2^{p-1} & \dots & 2^0 \\ \vdots & \vdots & \ddots & \vdots \\ m^p & m^{p-1} & \dots & m^0 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} c_p \\ c_{p-1} \\ \vdots \\ c_0 \end{bmatrix}.$$

We assume $p+1 \ll m$. The least squares estimation of \mathbf{c} is then given by $\hat{\mathbf{c}} = (\mathbf{P}^\top \mathbf{P})^{-1} \mathbf{P}^\top \mathbf{B}$.

We subtract the estimated signal $\hat{\mu}[t]$ from the time series $B[t]$ and obtain the residual process

$$\tilde{z}[t] = B[t] - \hat{\mu}[t].$$

Since the preceding data points effect the current data point through the overlapping sliding windows, the residual process $\tilde{z}[t]$ is further modeled as the autoregressive model AR(1):

$$\tilde{z}[t] = \rho \tilde{z}[t-1] + e[t],$$

where $e[t]$ is a Gaussian white noise process. If $|\rho| = 1$, the residual process $\tilde{z}[t]$ is non-stationary. If $|\rho| < 1$, we can continue to iterate backward and represent the AR(1) model as $\tilde{z}[t] = \sum_{i=0}^{t-1} \rho^i e[t-i]$, which converges to a stationary time series as $t \rightarrow \infty$. Since the estimated residual process is expected to be small, we do not consider the diverging case $|\rho| > 1$, which indicates that the signal $\mu[t]$ is estimated poorly from the start.

Subsequently, Dickey-Fuller (DF) test [10] is used to test for stationarity of the residual process $\tilde{z}[t]$. The DF test provides a procedure to test whether the residual process $\tilde{z}[t]$ is non-stationary ($H_0 : \rho = 1$) as opposed to stationary ($H_1 : |\rho| < 1$).

2.3. Validation

For validation, two simulation studies with known ground truth were performed. In each study, we performed 100 simulations. In each simulation, two blocks of signals were generated with different underlying topologies. The proposed dynamic-TDA was applied with the window length $w = 15$ and a time series $B[t]$ was quantified using its mean over time.

In the first study, we simulated pairs of 1D time series

$$y_1[t] = e_1[t] \quad 1 \leq t \leq 200,$$

$$y_2[t] = \begin{cases} e_2[t] & 1 \leq t \leq 40, 161 \leq t \leq 200 \\ e_2[t] - 200 & 41 \leq t \leq 96, 106 \leq t \leq 160 \\ e_2[t] + 200 & 97 \leq t \leq 105 \end{cases}$$

where $e_1[t], e_2[t]$ are Gaussian white noises with standard deviation 40 (Figure 2a). The jump continuities were added to $y_2[t]$ to introduce different numbers of connected components (0D holes). The pairs of time series $B[t]$ were constructed and the paired t -test was performed (Figure 2b). In all simulations, p -values were less than 0.0001 indicating that the dynamic-TDA captures the different underlying topologies.

In the second study, we used the Lorenz system [11] to generate 1D holes (circular loops):

$$\frac{dx_1}{dt} = 10(x_2 - x_1), \frac{dx_2}{dt} = x_1(28 - x_3) - x_2, \frac{dx_3}{dt} = x_1x_2 - \frac{8}{3}x_3,$$

where (x_1, x_2, x_3) are 3D coordinates with initial value $(1, 1, 1)$. The solution $\Phi[t] = (x_1[t], x_2[t], x_3[t])$ yields the well-known strange attractor [12] with two circular loops (1D holes) (Figure 3a). We simulated pairs of multivariate time series in \mathbb{R}^3

$$y_1[t] = e_1[t] \quad 1 \leq t \leq 100,$$

$$y_2[t] = \Phi[t] + e_2[t] \quad 1 \leq t \leq 100$$

where $e_1[t], e_2[t]$ are 3D noise vectors whose components are independent and identically distributed Gaussian white noises with standard deviation of 3 (Figures 3b and 3c). The pairs of time series $B[t]$ were constructed and the paired t -test was performed (Figure 3d). In all simulations, p -values were less than 0.0001 indicating that the dynamic-TDA differentiates underlying topology differences.

3. APPLICATION

3.1. Resting-state fMRI

We used the resting-state fMRI of $n = 412$ subjects collected as part of the Human Connectome Project [13]. fMRI data had undergone spatial and temporal preprocessing including motion and physiological noise removal. We employed the Automated Anatomical Labeling (AAL) to parcellate the brain volume into 116 regions [14]. The fMRI were then averaged across voxels in each region, resulting in $d = 116$ average fMRI time series with 1200 time points for each subject.

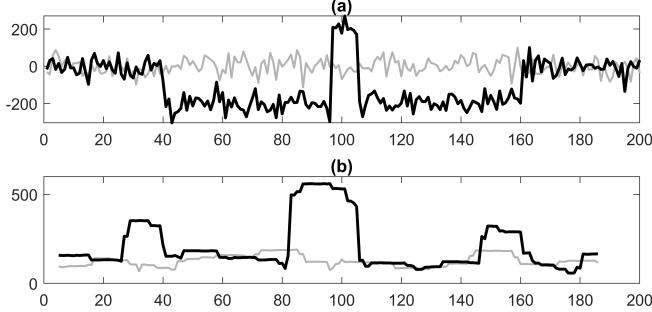


Fig. 2. First simulation. (a) displays an example of a pair of time series $y_1[t]$ (gray) and $y_2[t]$ (black). (b) displays the time series of the life-time of 0D holes corresponding to y_1 and y_2 .

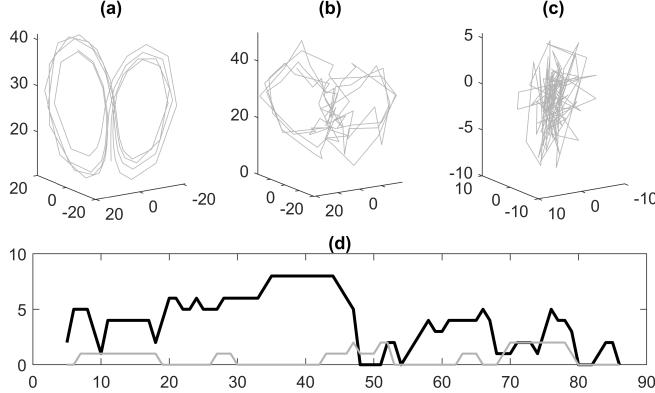


Fig. 3. Second simulation. (a) displays the Lorenz attractor. (b) and (c) display an example of a pair of time series $y_2[t]$ and $y_1[t]$ respectively. (d) displays the time series of the life-time of 1D holes corresponding to y_2 (black) and y_1 (gray).

Each of the average fMRI time series was then normalized by subtracting its sample mean over time [4].

For our study, we used the window length $w = 12$. We discarded the first five time points to avoid any artifacts in the fMRI data including large variance at the beginning of each scan [15]. We only considered 0D and 1D holes since higher-dimensional holes are rarely observed resulting in mostly empty barcodes across the time series. The time serie $B[t]$ was constructed for each subject. We then fitted the trend stationary model with $p = 1$ (linear trend) and performed the DF test for each subject (Figure 4). For both 0D and 1D holes, all the subjects show p -values less than 0.001 indicating the topological patterns are trend stationary.

3.2. Sexual dimorphism on barcodes

Time series $B[t]$ of 0D and 1D holes exhibits the trend stationarity with an estimated linear signal $\mu[t] = c_0 + c_1 t$ for each subject. We investigated if the estimated linear topological pattern can be a biomarker for discriminating 240 females and 172 males in the data. Figure 5 displays the group level average time series showing clear group separation. We per-

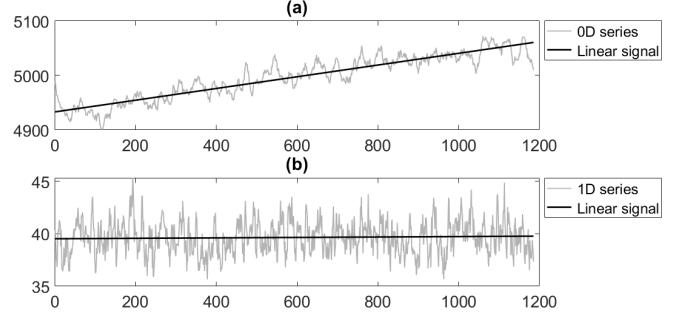


Fig. 4. (a) and (b) display the average time series of the life-time of 0D and 1D holes (gray) across $n = 412$ subjects with their estimated linear signals (black). The estimated linear signals in (a) and (b) are $\mu_{0D}[t] = 0.11t + 4932.71$ and $\mu_{1D}[t] = 0.0002t + 39.5289$, respectively.

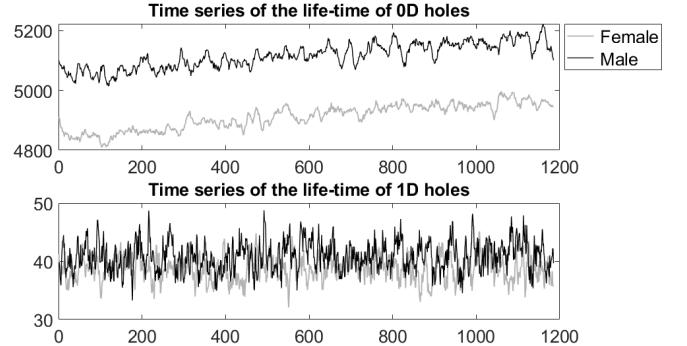


Fig. 5. The average time series of the life-time of 0D and 1D holes for 240 females (gray) and 172 males (black). The time series display separation clear group separation between female and male groups.

formed the two-sample t -test on c_0 and c_1 (Table 1). The results for both 0D and 1D holes indicate that 1) the clear separation of the topological pattern is statistically significant and 2) the separation between female and male groups appears to be consistent over time.

We further checked the proposed pipeline is not detecting false positives by generating 10000 null datasets by randomly mixing the half of males and the half of females to create two new groups with no signal. The average p -values for 0D and 1D holes are all 0.5 ± 0.28 and 0.5 ± 0.29 respectively indicating that the pipeline is not detecting signals as expected.

4. DISCUSSION

Using the proposed dynamic-TDA, we demonstrated that the resting-state fMRI has stationary topological patterns. Our study also reveals that sexually dimorphic topological patterns exhibit clear and consistent difference in the life-time of 0D and 1D holes across the scan durations while maintaining stable topological patterns over time. We believe that the pro-

		c_0		c_1	
		0D	1D	0D	1D
Mean	F	4845.37	38.94	0.11	-0.0002
	M	5054.57	40.35	0.10	0.0008
SD	F	397.04	5.81	0.13	0.0064
	M	392.92	5.57	0.14	0.0075
p-value		1.92×10^{-7}	0.014	0.54	0.13

Table 1. t -test results comparing males and female.

posed method will yield the new direction of the TDA applied to other dynamically changing medical images. By applying the proposed pipeline to each brain region separately, it may be further possible to localize brain regions. This is left as a future study.

There is a notable end-to-end TDA processing pipeline based on the conventional time-delay embedding [16]. While our method decomposes a d -dimensional multivariate time series $x[t]$ into a dynamic sequence of point clouds $X[t]$, [16] converts d individual univariate time series into d separate point clouds, discarding the dynamic aspect. Further, the dimension d in which the point clouds $X[t]$ reside is provided by the dataset in our method. On the other hand, the time-delay embedding procedure requires an extra fine-tuning of dimension estimation.

5. ACKNOWLEDGEMENTS

We thank Hernando Ombao of KAUST, Yuan Wang of University of South Carolina, Yasu Wang of Ohio University and Taniguchi Masanobu of Waseda University and other participants of the KAUST workshop on TDA in January 2020 for valuable discussions on the Rips filtration and dynamic-TDA.

6. REFERENCES

- [1] H. Edelsbrunner, D. Letscher, and A. Zomorodian, “Topological persistence and simplification,” in *Proceedings 41st annual symposium on foundations of computer science*. IEEE, 2000, pp. 454–463.
- [2] G. Carlsson, A. Zomorodian, A. Collins, and L. J. Guibas, “Persistence barcodes for shapes,” *International Journal of Shape Modeling*, vol. 11, no. 02, pp. 149–187, 2005.
- [3] R. Ghrist, “Barcodes: the persistent topology of data,” *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.
- [4] M. K. Chung, S. G. Huang, A. Gritsenko, L. Shen, and H. Lee, “Statistical inference on the number of cycles in brain networks,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 113–116.
- [5] D. S. Bassett and O. Sporns, “Network neuroscience,” *Nature neuroscience*, vol. 20, no. 3, pp. 353, 2017.
- [6] Y. Wang, H. Ombao, and M. K. Chung, “Topological data analysis of single-trial electroencephalographic signals,” *The annals of applied statistics*, vol. 12, no. 3, pp. 1506, 2018.
- [7] M. Gidea and Y. Katz, “Topological data analysis of financial time series: Landscapes of crashes,” *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 820–834, 2018.
- [8] J. A. Perea, A. Deckard, S. B. Haase, and J. Harer, “SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data,” *BMC bioinformatics*, vol. 16, no. 1, pp. 257, 2015.
- [9] C. R. Nelson and C. R. Plosser, “Trends and random walks in macroeconomic time series: some evidence and implications,” *Journal of monetary economics*, vol. 10, no. 2, pp. 139–162, 1982.
- [10] D. A. Dickey and W. A. Fuller, “Distribution of the estimators for autoregressive time series with a unit root,” *Journal of the American statistical association*, vol. 74, no. 366a, pp. 427–431, 1979.
- [11] E. N. Lorenz, “Deterministic nonperiodic flow,” *Journal of the atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.
- [12] D. Ruelle and F. Takens, “On the nature of turbulence,” *Les rencontres physiciens-mathématiciens de Strasbourg-RCP25*, vol. 12, pp. 1–44, 1971.
- [13] D. C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. W. Curtiss, et al., “The Human Connectome Project: a data acquisition perspective,” *Neuroimage*, vol. 62, no. 4, pp. 2222–2231, 2012.
- [14] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain,” *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
- [15] J. Diedrichsen and R. Shadmehr, “Detecting and adjusting for artifacts in fMRI time series data,” *Neuroimage*, vol. 27, no. 3, pp. 624–634, 2005.
- [16] L. M. Seversky, S. Davis, and M. Berger, “On time-series topological data analysis: New data and opportunities,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 59–67.