Lecture 10: Feb 9, 2017

Lecturer: Prof. Bei Wang <beiwang@sci.utah.edu> *Scribe: Adam Conkey*

10.1 Introduction

The goal of persistent homology is to discover topological invariants in the data. The motivation is that the data may have structural information that is only present when viewed in the appropriate dimension. For example, if the point cloud of your data is circular in two dimensions, but you are looking at a projection onto a one-dimensional line, the data will appear sporadic on a line and you lose the coherent structure it had in two dimensions. Tunnels, voids, and higher dimensional topological features that are invariant in the data may be useful in making conclusions from it, and the goal of persistent homology is to uncover these invariants.

Persistent homology will differentiate between noise and essential features of the data. The basic idea is to sample points from continuous space (where computation is infeasible) and operate on the resulting point cloud using simplicial complexes (where computation is feasible). Balls of a certain radius determine when a simplex is born, but it is generally not known what radius should be used to accurately distinguish between a topological invariant of the data and noise. Thus, persistent homology considers all possible radii and tracks when topological features are born and die. This is visualized through persistence diagrams and persistence barcodes, which will be illustrated in examples that follow.

10.2 Application: Persistent Homology and Brain Data

Paper discussed: [W2016]

Brain networks are generated from fMRI data and establish functional connections between regions of the brain. In this paper 264 regions of interest (ROIs) were considered. A timeseries is associated with each ROI corresponding to resting state brain activity of 30 control subjects and 57 subjects falling somewhere on the Autism Spectrum. There is no existing diagnostic test for Autism Spectrum Disorder (ASD) and is typically evaluated behaviorally by a specialist rating an individual using the Autism Diagnostic Observation Schedule (ADOS). The goal of the paper was to leverage topological features present in the data to predict the ADOS classification for individuals with ASD.

TDA using persistent homology proceeds as follows:

1. Create a point cloud of the fMRI timeseries for each subject by allowing each point to correspond to a ROI (264 points in total).
2. Compute the pairwise distance between points as a function of the correlation between the corresponding timeseries associated with the ROIs:
$$d(x, y) = \sqrt{1 - corr(x, y)}$$
where $corr$ is Pearson Correlation and x and y are timeseries.

3. Run filtration using radius r from 0 (minimum distance) to $\sqrt{2}$ (maximum distance) such that two points are considered connected if the balls of radius r around the two points overlap.

4. Generate persistence diagrams and persistence barcodes for the filtration.

5. Do regression with persistence data as input to predict ASOD score.

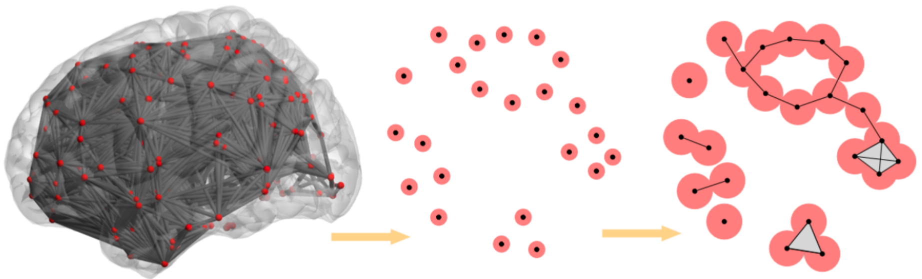


Figure 10.1: Brain network to metric space

Regression in this case was done with Kernel Partial Least Squares Regression which uses the kernel trick with the following kernel:

$$K_\sigma^{TDA}(A,B) = \frac{1}{8\pi\sigma} \sum_{p\in A, q\in B} e^{-\frac{||p-q||^2}{8\sigma}} - e^{-\frac{||p-\bar{q}||^2}{8\sigma}}$$

where A and B are barcodes and for $q = (a,b) \in B$, $\bar{q} := (b,a)$.

The conclusion of this research was that the features from TDA alone were insufficient for predicting ADOS score accurately, but when TDA information was used in conjunction with standard correlation information, classification accuracy improved over using standard correlation information alone.

10.3 Persistent Cohomology

Persistent cohomology is a dual of persistent homology, in the sense that notions like chain and cycle in homology have corresponding notions like cochain and cocycle in cohomology. Cohomology is often easier to compute than homology which is one reason why it is often favored in some applications.

The elements in persistent cohomology are functions over simplicial complexes and serve as indicator functions for the presence of an element. For example, for every vertex v_i of a simplicial complex there exists a function $v_i^* : \mathbb{V} \to \mathbb{Z}_2$ such that

$$\begin{aligned} &v_i^*(v_i) = 1 \\ &v_i^*(v_j) = 0 \quad \forall i \neq j \end{aligned}$$

The functions v_i^* are the elementary 0-cochains of cohomology and are the dual notion corresponding to vertices which are elementary 0-chains in homology. Similarly, for every edge e_i of a simplicial complex there exists an indicator function $e_i^* : \mathbb{E} \to \mathbb{Z}_2$ such that

$$e_i^*(e_i) = 1$$
$$e_i^*(e_j) = 0 \quad \forall i \neq j$$

The functions e_i^* are the elementary 1-cochains of cohomology and are the dual notion corresponding to edges which are elementary 1-chains in homology. Formally, we have for a simplicial complex X with X^0, X^1, and X^2 denoting the sets of vertices, edges, and triangles of X, respectively:

$$C^0 = \{\beta : X^0 \to \mathbb{Z}_2 | \beta = \sum_i g_i v_i^*\}$$
$$C^1 = \{\alpha : X^1 \to \mathbb{Z}_2 | \alpha = \sum_i g_i e_i^*\}$$
$$C^2 = \{\gamma : X^2 \to \mathbb{Z}_2 | \gamma = \sum_i g_i \Delta_i^*\}$$

where $g_i \in \{0, 1\}$ and Δ represents a triangle of the simplicial complex.

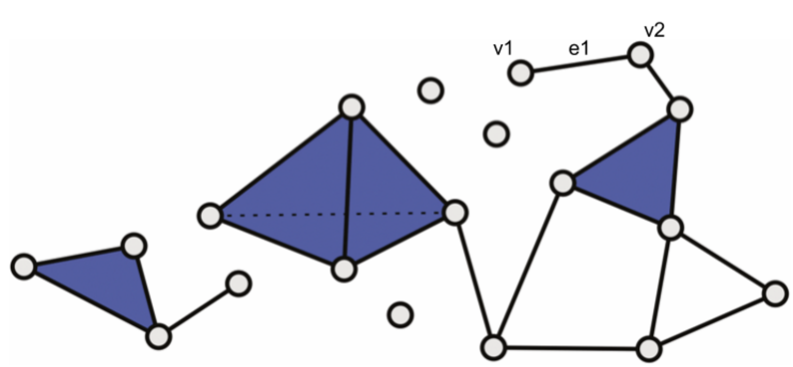
Definition 10.1. *Coboundary maps $\delta_0 : C^0 \to C^1$ and $\delta_1 : C^1 \to C^2$ are defined as*

$$(\delta_0\beta)(ab) = \beta(b) - \beta(a)$$
$$(\delta_1\alpha)(abc) = \alpha(bc) - \alpha(ac) + \alpha(ab)$$

Definition 10.2. *A p-cocycle is an $\alpha \in C^p$ such that $\delta_p\alpha = 0$.*

Definition 10.3. *A p-coboundary is an $\alpha \in C^p$ such that there exists a $\beta \in C^{p-1}$ such that $\delta_{p-1}\beta = \alpha$.*

Example: Consider the following simplicial complex:



We have the indicator functions $e_1^* \in C^1$ and $v_1^*, v_2^* \in C^0$ and

$$\delta_0\beta = e_1^* = \beta(\partial e_1) = \beta(v_1 + v_2) = \beta(v_1) + \beta(v_2) = v_1^* + v_2^*$$

10.4 Application: Persistent Cohomology and Circular Coordinates

Paper discussed: [D2011]

The paper regards the problem of non-linear dimensionality reduction (NLDR). The goal of NLDR is to reduce the dimensionality of data while preserving its intrinsic structure. For point cloud data $X \in \mathbb{R}^d$ the data is to be mapped into a lower dimension by a map $\phi : X \to \mathbb{R}^m$ such that $m < d$.

This paper dealt specifically with computing circular value coordinates for a statistical data set with $\phi : X \to \mathbb{S}^1$, and this map is computed through the use of persistent cohomology. Their conclusion is that they are able to perform NLDR analysis for a broader class of data sets than traditional techniques such as Isomap and Laplacian Eigenmaps.

References

[W2016] E. Wong, S. Palande, B. Wang, B. Zielinski, J. Anderson, P.T. Fletcher, "Kernel partial least squares regression for relating functional brain network topology to clinical measures of behavior", *International Symposium on Biomedical Imaging (ISBI)*, 2016

[D2011] v. De Silva, D. Morozov, M. Vejdemo-Johansson, "Persistent cohomology and circular coordinates", *Discrete & Computational Geometry*, 2011, 45(4), 737-759.