An Introduction to Persistent Cohomology

Yuqing Xing
Huazhong University of Science and Technology

September 2, 2021

1 Motivation

Persistent Cohomology is a developing method in Topological Data Analysis. Persistent homology, a commen method for studying topological features over changing scales, can reflect the geometry or topological information of data clouds, but other important non-geometric information of datasets may be neglected. For example, considering the point cloud data of chemical atoms, different points represent different atomic types. Two point cloud data may have the same geometric structure but different atomic types, and such differences cannot be identified by persistent homology.

Can we use topological methods to analyze non-geometric information of datasets? We utilize the persistent cohomology, which adds multiple non-geometric information into functions fully or partially defined on simplicial complexes locally associated with the cohomology generators. Thus other important information in a dataset is embedded in topological invariants generated from the geometric information of the dataset.

2 Persistent Cohomology Theory

2.1 Persistent Homology

2.1.1 Homology

let X be a simplicial complex of dimension n. X^k is the set of all k-simplices in X.

A k-chain on X is a linear combination $\sum_{\sigma_i \in X^k} a_i \sigma_i$ of k-simplices in X, where $a_i \in \mathbb{Z}/2\mathbb{Z}$ and σ is a k-simplex represented by $[v_0, v_1, ..., v_k]$.

The k-th chain group $C_k(X)$ is the set of all k-chains in with the addition given by the addition of coefficients in $\mathbb{Z}/2\mathbb{Z}$.

The boundary operator $\partial_k : C_k(X) \to C_{k-1}(X)$ is a linear mapping that maps a k-simplex to the alternating sum of its k-1 simplexs generated by removing one vertex of the k-simplex,

$$\partial_k [v_0, v_1, ..., v_k] = \sum_{i=0}^{k} (-1)^i [v_0, ..., \hat{v_i}, ..., v_k]$$

A k-chain c is a **k-boundary** if c is in the image of the boundary map ∂_{k+1}.

A k-chain c is a **k-cycle** if $\partial_k(c) = 0$.

$\partial_p \circ \partial_{p+1} = 0$ follows from the definition of boundary operator. Thus $Im(\partial_{k+1})$ is a subgroup of $ker(\partial_k)$.

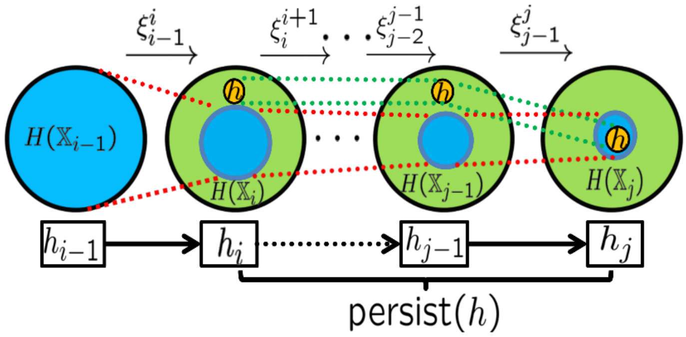
The kth homology group is the quotient group $H_k(X) = ker(\partial_k)/Im(\partial_{k+1})$ containing equivalence classes of k-cycles.

Two k-cycles are in the same equivalence class in $H_k(X)$, i,e, called **homologous** if they differ by the boundary of a (k+1)-chain.

2.1.2 Persistence

Let S be a finite set of points, r be a positive real number, define the **Rips complex** $R(r) = \{\sigma \subset S \mid diam\ \sigma \leq r\}$, where $diam\ \sigma$ is the diameter of σ, i.e. the maxiamal distance between any two points of σ. **A filtration** is a finite increasing sequence of sets $X = \{X_i \mid X_i \subset X_j \quad for\ i < j,\ i, j \in R\}$. The inclusion $f^{i,j} : X_i \to X_j$ induces a homomorphism $f_p^{i,j} : H_p(X_i) \to H_p(X_j)$. **The p-th persistent homology group** $H_p^{i,j}$ is defined as the image of $f_p^{i,j}$.

The p-th persistent Betti number is defined as the rank of $H_p^{i,j}$. Letγ be a class in $H_p(X_i)$, γ is born at X_i if $\gamma \notin H_p^{i-1,i}$; furthermore, γ dies entering at $X_j (j > i)$ if $f_p^{i,j-1}(\gamma) \notin H_p^{i-1,j-1}$ and $f_p^{i,j}(\gamma) \in H_p^{i-1,j}$. If homology class γ is born at X_i and dies entering X_j , define **the persistence of** γ as $j - i$.



Persistent homology can be represented in two equivalent forms: either as a **persistence barcode** or as a **persistence diagram**. **A persistence barcode** is a collection of bars whose left end points and right end points are birth and death respectively. Each interval in the persistence barcode is represented in the persistence diagram by a point in the plane, with its **birth coordinate** on the horizontal axis and with its **death coordinate** on the vertical axis. Points of a persistence diagram all lie above the diagonal line y = x. A persistence diagram is given by of birth coordinate and death coordinate .

Bottleneck distance can measure the similarity of two persistent diagrams. The bottleneck distance between digram A and digram B is defined as follows:

$$d(A, B) = \underset{\alpha}{inf}\ \underset{x}{sup} \mid x - \alpha(x) \mid$$

where $x \in A, u = (x, y), |u| = max\{|x|, |y|\}, \alpha$ traverses all the bijections from A to B.

The bottleneck distance is a robust metric of similarity between two leaf structures. The more similar the diagrams are, the smaller it is.

When persistent homology theory is applied to topological data analysis, we just need to consider the case k=0,1,2.

2.2 Persistent Cohomology

Steps of persistence cohomology:
input datasets $\Rightarrow$ **get persistent intervals**$[x_i, x_j)$ **and persistence barcode** $\Rightarrow$ **add representative cocycles** $\Rightarrow$ **get smoothed cocycles** f $\Rightarrow$ **get persistent cohomology enriched barcodes** $[x_i, x_j, f^*)$

2.2.1 Cohomology

Let X be a simplicial complex of dimension n. X^k is the set of all k-simplices in X. Let R be a commutative ring(such as $\mathbb{Z}$, F_p, here p is a prime).

A k-cochain on X is a function $\alpha : X^k \to R$.

The k-th cochain group $C_k(X)$ is the set of all k-cochains with the addition given by the addition of coefficients in R.

The coboundary operator $\delta^k : C^k(X) \to C^{k+1}(X)$ is a linear map given by

$$\delta^k(\alpha)([v_0, v_1, ..., v_k, v_{k+1}]) = \sum_{i=0}^{k} (-1)^i \, [v_0, ..., \hat{v_i}, ..., v_k, , v_{k+1}]$$

where α is a k-cochain.

A k-cochain α is a **k-coboundary** if α is in the image of the coboundary map δ^{k-1}.

A k-cochain α is a **k-cocycle** if $\delta^k(\alpha) = 0$.

$\partial_k \circ \delta^{k-1} = 0$ follows from the definition of boundary operator. Thus $Im(\delta^{k-1})$ is a subgroup of $ker(\delta^k)$.

The kth cohomology group is the quotient group $H^k(X) = ker(\delta^k)/Im(\delta^{k-1})$.

Two k-cocycles are in the same equivalence class in $H^k(X)$, i,e, called **cohomologous** if they differ by the coboundary of a (k-1)-chain.

When persistent cohomology theory is applied to topological data analysis, we just need to consider the case k=0,1,2. The detailed process is as follows.

0-cochain group $C^0 = C^0(X;R) = \{f : X^0 \to R\}$

1-cochain group $C^1 = C^1(X;R) = \{g : X^1 \to R\}$

2-cochain group $C^2 = C^2(X;R) = \{h : X^2 \to R\}$

$\delta^0 : C^0 \to C^1$ is given by $(\delta^0 f)(ab) = f(a) - f(b)$

$\delta^1 : C^1 \to C^2$ is given by $(\delta^1 g)(abc) = g(bc) - g(ac) + g(ab)$

$\delta^1 \circ \delta^0 = 0, \delta^2 \circ \delta^1 = 0$ arise from the definition of coboundary map.

2.2.2 Representative Cocycle

Given a functinon $f : X^k \to R$ $(0 \leq k \leq n)$, we seek a method to embed the information of f on the persistence barcodes obtained with a given filtration of X. In other words, we seek a representation of f on cohomology generators.

proposition 1. Let X be a topological space, $K(G,n)$ is a Eilenberg-MacLane space(n>0), then

$$[X, K(G,n)] \cong H^n(X;\mathbb{Z})$$

.

Corallary. $[X, S^1] \cong H^1(X;\mathbb{Z})$

For a simplicial complex X with a nontrivial topology, if there exists a nonzero cohomology class $[g] \in H^1(X;\mathbb{Z})$, thus we can construct a map $F : X \to S^1$ to represent [g].

2.2.3 Smoothed Cocycle

However, some representative cocycles in persistent cohomology may not reflect the overall location and structure related to their cohomology generators.To better embed the additional information of datasets in cohomology generators, we look for a smoothed representative cocycle in each cohomology class.

Consider a filtration of X: $X_1 \subseteq X_2 \subseteq \cdots \subseteq X_n = X$. The filtration can be represented as a dragram $X_1 \to X_2 \to \cdots \to X_n$, where X_i is the complex formed by filtration at x_i. For a coefficient field F, we can get a homomorphisim between cohomology group $H^k(X_1;F) \leftarrow H^k(X_2;F) \leftarrow \cdots \leftarrow H^k(X_2;F)$

Accoding to the persistence decomposition theorem , the diagram decomposes as a direct sum of 1-dim terms indexed by intervals with the form $[x_i, x_j)$. The collection of intervals can be represented by a persistent diagram or a persistent barcode. Long intervals reveals valuable topological information. However, the persistent decomposition works for diagrams of vector spaces over a field F_p, but fails for diagrams of vector spaces over the rings of integer $\mathbb{Z}$. Fortunately, F_p cocycle can be lifted to integer coefficients cocycle.

In fact, the short exact sequence of coefficient rings $0 \longrightarrow Z \xrightarrow{\cdot p} Z \longrightarrow F_p \longrightarrow 0$ induce a long exact sequence, called the Bockstein sequence. The following is the relevant of the sequence:

$$\cdots \longrightarrow H^1(X,Z) \longrightarrow H^1(X,F_p) \longrightarrow H^2(X,Z) \xrightarrow{\cdot p} H^2(X,Z) \longrightarrow \cdots$$

By exactness, the cokernel of $H^1(X,F_p) \longrightarrow H^2(X,Z)$ is isomorphic to the kernel of $H^2(X,Z) \xrightarrow{\cdot p} H^2(X,Z)$. Since the kernel is the set of p-torsion elements of $H^2(X,Z)$. If there is no p-torsion, then the cokernel of $H^1(X,Z) \longrightarrow H^1(X,F_p)$ is zero, then it is surjective, thus any cocycle in $C^1(X,F_p)$ can be lifted to a cocycle in $C^1(X,Z)$. If there exist a p-torsion, then pick another prime and recompute the kernel. It is enough to pick a prime that does not divide the order of torsion subgroup of $H^(X,Z)$.

Let w be a representative cocycle for a persistence interval $[x_i, x_j)$ of dimension n. firstly, a lifting of w with interger coefficients $w\prime$ satisfies that $w(\sigma) \equiv w\prime(\sigma)(modp)$ and $w\prime \in \{i \in Z : -(p-1)/2 \leq i \leq (p-1)/2\}$ for all $\sigma \in X^k$. If $w\prime$ is not an interger cocycle($dw\prime \neq 0$), write $dw\prime = p\eta$ where $\eta \in C^(k+1)(X,Z), \eta = d\lambda$ where $\lambda \in C^(k)(X,Z)$, , then $w\prime - p\lambda$ is an integer cocycle.

Given an interger cocycle $\hat{w}$, find the minimum solution to this problem,

$$\bar{\alpha} = {}_{\alpha \in C^{k-1}(X,R)}^{\arg min} ||L(w + d\alpha)||_2^2,$$

A smoothed cocycle is obtained by letting $\bar{w} = w + d\bar{\alpha}$.

2.2.4 Persistent Cohomology Enriched Barcodes

Considering the interger cocycle $\hat{w}$ at filtration x, $\bar{w}$ is a smoothed cocycle corresponding to $\hat{w}$ obtained by solving the minimum solution defined above. let $\mu_x(\sigma) = |\bar{w}(\sigma)|$ be a measure on X^k at flitratin x. A function $f^* : [0,1) \to R$ for persistence interval $[x_i, x_j)$ is defined as

$$f^*(t) = \int_{X^k(t)} f d\mu_x / \int_{X^k(t)} d\mu_x, \quad wherex = (1-t)x_i + tx_j,\ t \in [0,1)$$

A persistent cohomology enriched barcode includes three elements: birth x_i, death x_j, function F^* for persistene interval $[x_i, x_j)$.

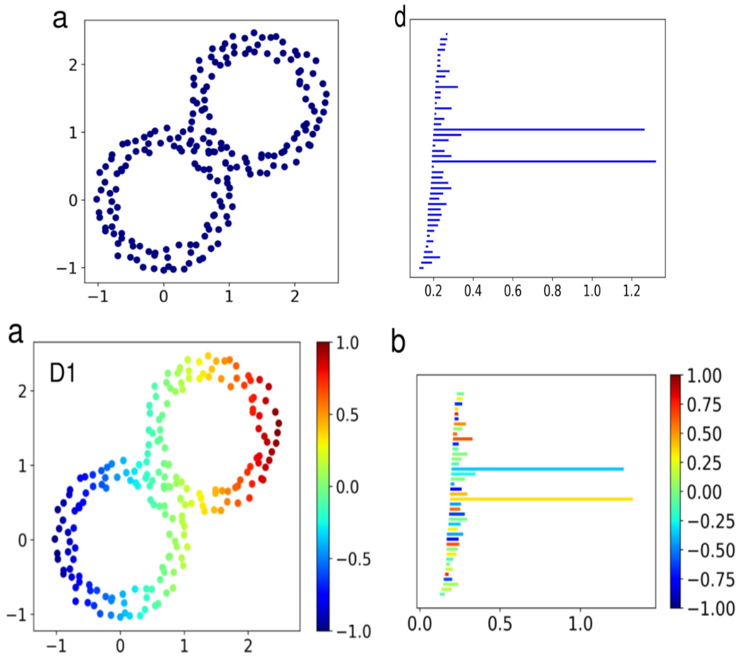
3 Application Examples of Persistent Cohomology

Take a annulus and a cyclic structure cucurbit uril molecule as examples. The traditional persistent homology barcodes will only show the structure of the molecule without the element type information. If we take subsets of atoms of selected element types, the resulting barcode does not faithfully represent the original structure. By using the enriched barcodes, we can quantify the element type composition of each bar while retaining the original structure. Specifically, if one type of automs(such as oxygen atom) is of interest, we construct an input function f defined on the points representing the atoms and outputs 1 on oxygen atoms and 0 elsewhere.

$$f(x) = \begin{cases} 1 & \text{x represents the oxygen autom} \\ 0 & \text{else} \end{cases} \tag{1}$$

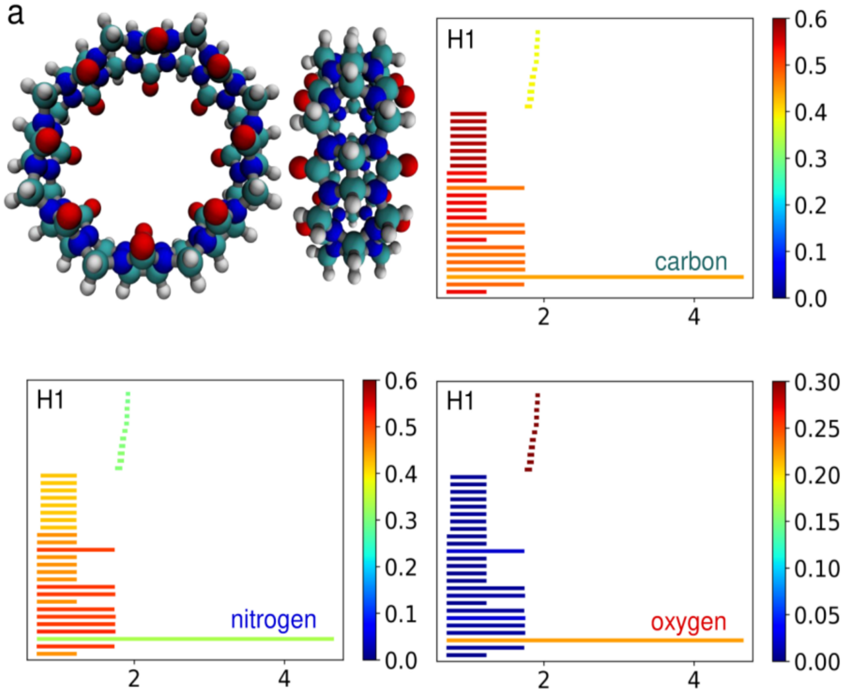
Then, we can get the value of F^* corresponding to each interval $[x_i, x_j)$, thus we get a persistence enriched barcode given by $[x_i, x_j, F^*)$.

Example 1 annulus



The above graphs show the persistent enriched barcode of two annulus with same values on the nodes and different values on the nodes respectively. The datasets is of similar geometry but different nongeometric information. We can use enriched barcodes to distinguish between them.

Example 2 cucurbit uril molecule



A cucurbit uril molecule contains eight 6-membered rings and sixteen 5-membered rings consisting of carbon and nitrogen atoms. The rings form a big cyclic structure with a relatively tighter opening surrounded by oxygen atoms. The enriched barcode shows that there are eight medium-sized H_1 bars that mainly consist of carbon and nitrogen atoms, which can be confirmed by observing the molecular structure .