A Short Introduction on Topological Machine Learning Methods

Yifan Chen

Southern University of Science and Technology

April 4th,2022

Introduction

Utility of Topology

- Encode the overall shape of data.
- Capture multi-scale, global, and intrinsic properties of data sets.

Methods of TDA

Extrinsic Topological Features

- Extract topological features of given data.
- Feed downstream machine learning models with these features.
- Achieved by vectorisation of topological features.
- Or achieved by layers of neural networks to handle them.

Intrinsic Topological Features

- Topological analysis of aspects of the machine learning model.

Methods of TDA

Observational Methods

- Analysis the topology of the data or model.
- Do not directly influence the model training or architecture.

Interventional Methods

- Inform the architectural design and/or model training.
- By topological properties of the data and post-hoc analysis of topological features of machine learning models.

Graphic Conclusion of Methods of TDA

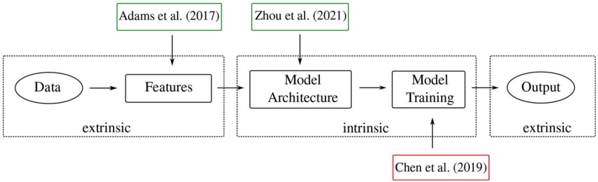

FIGURE 4 | This overview figure shows examples of methods discussed in the survey and their range of influence. Green (red) boxes signify *observational* (*interventional*) methods. **Table 1** provides a more in-depth classification of all methods.

Figure 1: Methods used in TDA and their influences

Extrinsic Topological Features in Machine Learning

Strategies on Vectorising Persistence Diagrams (PDs)

- Different representations that ideally give rise to feature vectors.
- Kernel-based methods that permit the integration into certain classifiers.

General Schemes

- Representation: map persistence diagrams into an auxiliary vector space by discretisation, or mapping into a (Banach- or Hilbert-) function space.
- Kernel-based methods: measure similarity between persistence diagrams.

Representation Methods of Persistence Diagrams

Summary Statistics of Topological Descriptors of a Persistence Diagram

- Total persistence.
- p-norm.
- Persistent entropy.

They give rise to hypothesis testing based on topological information.

Representation techniques

- Betti curves.
- Directly generates a high-dimensional feature vector.
- Persistence landscapes.
- Persistence images (PIs).

Betti Curves

Definition

Given a persistence diagram $\mathcal{D}$, and a weight function $w\colon \mathbb{R}^2 \to \mathbb{R}$, its Betti curve is the function $\beta\colon \mathbb{R} \to \mathbb{R}$ defined by

$$\beta := \sum_{(b,d)\in\mathcal{D}} w(b,d) \cdot \mathbb{1}_{[\mathrm{b},\mathrm{d}]}(t)$$

Strengths and Weaknesses

- Permit the calculation of a mean curve, distance and kernel.
- The mapping from a diagram to a curve is not injective.
- The curve only contains counts of topological features and does not permit tracking single features.

Betti Curve

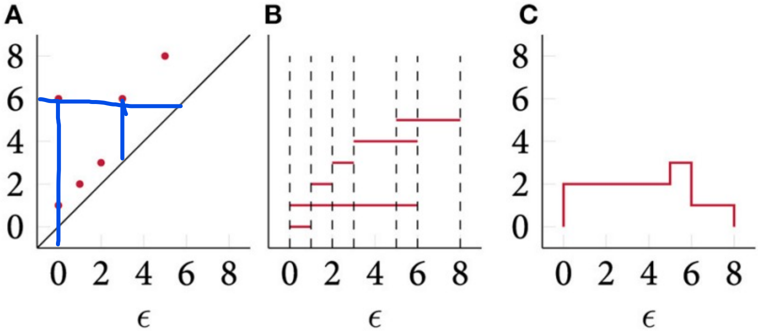

FIGURE 5 | A persistence diagram **(A)**, its persistence barcode **(B)**, and its corresponding Betti curve **(C)**. Notice that the *interpretation* of the axes of different plots is different, hence we exclude labels for the barcode representation.

Figure 2: Betti curve calculation example

Directly Generate a High-dimensional Feature Vector

- For each pair (p, q) of points in persistence diagram $\mathcal{D}$, compute $m(p, q) := min\{d_\infty(p, q), d_\infty(p, \Delta), d_\infty(q, \Delta)\}$, here $\Delta := \{(x, x)|x \in \mathbb{R}\} \subset \mathbb{R}^2$ refers to points on the diagonal.
- Associate to $\mathcal{D}$ the vector of these values, sorted in descending order.
- Obtain a vector representation of $\mathcal{D}$ based on the distribution of pairwise distances of its elements.
- Provide a good baseline to furnish any machine learning classifier with simple topology based feature vectors.

Persistence Landscapes

Definition

Given a persistence diagram $\mathcal{D} = \{(b_\mathrm{i}, d_\mathrm{i})\}_{\mathrm{i}\in\mathcal{I}}$, for $b < d$, and an auxiliary function $f_{(\mathrm{b},\mathrm{d})}(t) := max\{0, min\{t - b, d - t\}\}$, its persistence landscape is the function $\lambda\colon \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ defined by

$$\lambda(k, t) := kmax\{f_{(\mathrm{b_i},\mathrm{d_i})}(t)\}_{\mathrm{i}\in\mathcal{I}}$$

Here, kmax denotes the k-th largest element of the set.

- Map persistence diagrams into a (Banach or Hilbert) function space invertibly.
- Do not require any choice of auxiliary parameters.
- Afford various summary statistics.
- Applications in time series analysis:
  PLLay (Persistence Landscape Based Topological Layer)

Persistence Landscapes

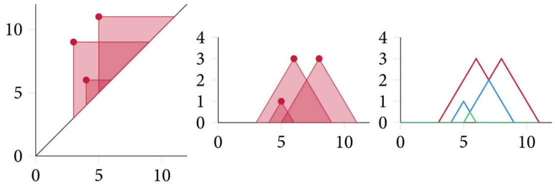

FIGURE 6 | Computing a *persistence landscape* involves calculating the “area of influence” of each topological feature in a persistence diagram. Each connected shaded region with at least *k* intersections forms the basis of the *k*-th persistence landscape, which can be obtained by “peeling off” layers in an iterative fashion.

Figure 3: Persistence Landscapes Computation

Persistence Images (PIs)

Procedure

- Transform PD $\mathcal{D}$ from "birth–death"-coordinates into "birth–persistence"-coordinates: $T\colon \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto (x, y - x)$.
- For each $u \in \mathbb{R}^2$, choose a differentiable probability density ϕ_{u} on $\mathbb{R}^2$, and a weighting function $f\colon \mathbb{R}^2 \to \mathbb{R}^2_{\geq 0}$ satisfying $f|_{\{0\}\times\mathbb{R}} \equiv 0$.
- Choose a discretisation of a relevant subdomain of $\mathbb{R}^2$ by a standard grid, each region $\mathcal{R}$ of this grid then corresponds to a pixel in the persistence image with value given by $\int_{\mathcal{R}} \sum_{u\in\mathcal{T}(\mathcal{D})} f(u)\phi_{\mathrm{u}}(z)dz$.

Persistence Images (PIs)

Non-canonical Choices

- Choice of the weighting function.
- Distribution ϕ_{u}, with standard choice being a normalised symmetric Gaussian with $\mathbb{E}[\phi_{\mathrm{u}}] = u$.
- The resolution of the discretisation grid.

Strength

- Being stable with respect to the 1-Wasserstein distance between persistence diagrams.
- Highly flexible, often employed to make a classifier "topology-aware".

Drawbacks

- Quadratic storage and computation complexity.
- Choice of appropriate parameters.
- No guidelines for picking such hyperparameters

Persistence Images (PIs)

FIGURE 7 | A persistence image arises as a discretisation of the density function (with appropriate weights) supported on a persistence diagram. It permits the calculation of an increasingly better-resolved sequence of images, which may be directly used as feature vectors.

Figure 4: Persistence Images Computation

Kernel-based methods

Motivation

- The space of persistence diagrams can be endowed with metrics, but there is no natural Hilbert space structure on it.
- Implicitly introduce such a Hilbert space structure to which persistence diagrams can be mapped via the feature map of the kernel.

Definition

Given a set $\mathcal{X}$, a function $k\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a (positive definite) kernel if there exists a Hilbert space $\mathcal{H}_\mathrm{k}$ together with a feature map $\phi\colon \mathcal{X} \to \mathcal{H}_\mathrm{k}$ such that $k(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle_{\mathcal{H}_\mathrm{k}}$ for all $x_1, x_2 \in \mathcal{X}$.

Kernel-based methods

Categories

- 1-Wasserstein distance kernel.
- Sliced Wasserstein kernel.
- Persistence weighted Gaussian kernels.

Strength and Weaknesses

- Not limited with respect to the input data.
- Good performance for shape classification or segmentation tasks, as well as in orbit classification.
- Suffer from computational complexity, which scales quadratically in the number of points.

Kernels based on topological information

Weisfeiler–Lehman (WL) Procedure

- An iterative scheme in which vertex label information is aggregated over the neighbours of each vertex, resulting in a label multiset.
- Can be repeated until a pre-defined limit has been reached or until the labels do not change any more.
- Useful for assessing the dissimilarity between two graphs in polynomial time, enjoying popularity for graph learning tasks.

Persistent Weisfeiler–Lehman (P-WL) Kernel for Graphs

- Compute topological features during a Weisfeiler–Lehman (WL) procedure.
- Extract topological information of the graph with respect to the current node labelling for each WL iteration.
- Imbues data-based labels into the calculation of persistent homology.