4th International Conference on Applied Mathematics, Modeling and Computer Simulation (AMMCS 2024)

Review on Topological Data Analysis (TDA) for Complex Systems Modeling

GOSWAMI Shankha Shubhra,*

[a]Abacus Institute of Engineering and Management, Hooghly, West Bengal, India, 712148

*Corresponding author's email: ssg.mech.official@gmail.com

Supported by Abacus Institute of Engineering and Management, India

Introduction

In the era of big data, analyzing and modeling complex systems present a significant challenge due to the sheer volume and intricacy of high-dimensional data. Classical statistical methods often fall short in capturing the geometric and topological structures inherent in these datasets. TDA provides a novel approach, employing tools from algebraic topology to extract topological features such as connected components, loops, and voids. These features are critical for understanding the behavior of complex systems across a wide range of fields, from biological systems to financial markets.

What is TDA?

Topological Data Analysis (TDA) is a method in data science that focuses on the study of the shape (topology) of data. TDA provides tools to extract meaningful information from complex, high-dimensional data by analyzing its underlying structures, patterns, and relationships.

- Topology: The study of spaces and their properties that are preserved under continuous deformations (e.g., stretching or bending but not tearing or gluing).
- Persistent Homology: A key tool in TDA that tracks the evolution of "features" such as connected components, loops, and voids across multiple scales.
- Simplicial Complexes: Structures that represent data points and their relationships as vertices, edges, triangles, etc., helping to capture the shape of data.

Research objectives

- To explore the theoretical foundations and mathematical frameworks of Topological Data Analysis (TDA), specifically focusing on its key tools such as persistent homology and simplified complexes, and how they provide novel insights into the shape and structure of high-dimensional data.
- To review and evaluate the application of TDA in complex systems modeling across various fields, including biology, neuroscience, and finance, identifying both the strengths and limitations of TDA in capturing intricate relationships within complex datasets.

Theoretical Foundations

Algebraic Topology

Simplified complexes: These are generalizations of graph structures, formed by connecting points (vertices) with edges, triangles, and higher-dimensional analogs.

Homology: Homology captures topological features like connected components, loops (1-dimensional holes), and voids (2-dimensional holes).

Persistent Homology

Filtration: Given a dataset, we construct a family of simplified complexes indexed by a scale parameter ϵ.

Persistence: Topological features (e.g., loops or voids) appear and disappear as the scale changes.

Mathematical Formulation of Persistent Homology

Let X be a finite metric space, such as a set of points in $\mathbb{R}^n$. We build a family of simplified complexes $\{K(\epsilon)\}_{\epsilon \geq 0}$ via a filtration process, where $K(\epsilon)$ is the simplified complex at scale ϵ.

$K(\epsilon_1) \subseteq K(\epsilon_2) \subseteq \cdots \subseteq K(\epsilon_m)$

$H_k(K(\epsilon_1)) \to H_k(K(\epsilon_2)) \to \cdots \to H_k(K(\epsilon_m))$

Computation of Persistent Homology

Construct the boundary matrix: For a simplified complex $K(\epsilon)$ of dimension d, define the boundary operator $\partial: C_d \to C_{d-1}$ that maps d-chains (linear combinations of d-simplexes) to their boundaries, which are $(d-1)$ chains.

Compute homology groups: The k-th homology group is calculated using the relation.

$H_k(K(\epsilon)) = \ker(\partial_k)/\mathrm{im}(\partial_{k+1})$

Persistent homology calculation: Construct the persistence matrix, which records the birth and death of features as the filtration progresses.

Example: Persistent Homology of 2D Point Cloud

Filtration: Create a sequence of simplified complexes for increasing values of ϵ (e.g., from 0 to the maximum distance between points).

Boundary calculation: At each step, calculate the boundary matrices for the simplexes formed.

Homology groups: Identify H_0 (connected components) and H_1 (loops) for each complex.

Persistence diagram: The resulting persistence diagram will show two points in H_0 (the two connected components) and a long-lived feature in H_1 represents the circular loop.

Computational Aspects

Algorithms for Persistent Homology

- Vietoris-rips complex: For a given dataset, a vietoris-rips complex is constructed by connecting points that are within a specified distance ϵ from each other.
- Čech complex: The čech complex provides an exact representation of the underlying topology of a dataset.

Complexity and Challenges

- The primary computational challenge in TDA arises from the fact that the size of simplified complexes grows exponentially with the number of points in the dataset.

Applications in Complex Systems Modeling

Biological Systems

TDA has been widely used in biology to study the structure of protein folding, cell morphology, and neuronal activity.

Example: Protein Structure Analysis

Neuroscience

In neuroscience, TDA has been applied to study the connectivity of brain networks.

Example: Brain Connectivity

Financial Systems

TDA is used to model the shape of market data, identify persistent structures, and detect anomalies, such as the onset of financial crises.

Example: Stock Market Analysis

Material Science

The structure of porous materials or the organization of crystalline structures in solid materials can be studied through persistent homology.

Example: Porous Materials

Conclusion

- Topological Data Analysis is a promising method for studying the complex shapes and structures within high-dimensional data.
- Its robustness to noise, multi-scale analysis, and ability to capture non-linear structures make it a powerful tool for modeling complex systems.
- As computational techniques continue to improve, TDA is likely to become increasingly relevant across various fields, from biology and neuroscience to finance and beyond.

Acknowledgement

We would like to acknowledge the support of Abacus Institute of Engineering and Management for their contributions to this research.

References

[1] Smith, A. D., Dłotko, P., & Zavala, V. M. (2021). Topological data analysis: concepts, computation, and applications in chemical engineering. Computers & Chemical Engineering, 146, 107202. https://doi.org/10.1016/j.compchemeng.2020.107202

[2] Uray, M., Giunti, B., Kerber, M., & Huber, S. (2024). Topological Data Analysis in smart manufacturing: State of the art and future directions. Journal of Manufacturing Systems, 76, 75-91. https://doi.org/10.1016/j.jmsy.2024.07.006

[3] Carlsson, G. (2020). Topological methods for data modelling. Nature Reviews Physics, 2(12), 697-708. https://doi.org/10.1038/s42254-020-00249-3