Topological Analysis Research

Spring Semester, 2022

Nathan Edwards
A01849553
CS 6950 Directed Readings
Dr. Steve Petruzza

Downloaded and using ParaView 5.10.0
Code can be found here: https://github.com/nathansedwards/Topological_Analysis
(Make sure to have the Underscore in the URL)

1. **Source**:

   Computational Topology for Data Analysis
   Tamal Krishna Dey, Yusu Wang
   https://www.cs.purdue.edu/homes/tamaldey/book/CTDAbook/CTDAbook.pdf

   Summary:

   (a) Preface

   Topological Data Analysis was popular in the 80's and 90's but has yet to become very popular and mainstream. This book will cover all aspects of TDA. This book is also said to teach Graduate students the math they need to know that is involved in TDA (which is perfect for me). This book is for Computer Science students and teachers.

   (b) Prelude

   Topological Data Analysis can be seen in Euler's problem: Seven Bridges of Kinogsberg. He had to find a route through the city in which you could only cross each of the 7 bridges once. What was important about the problem was the connectivity between the nodes (parts of the cities). TDA is more of a study of the way space is connected and how it can be manipulated without be added to or removed from. A few examples of fields in which TDA has been useful: visualization, material science, neuroscience, computer graphics, computational biology...

   (c) Chapter 1: Basics

   The main concepts in this chapter were the following: manifolds, homeomorphism, isotopy, and other maps. They are preceded in the chapter by first discussing topological space and metric space.

Topological space is basically a set of points that connect in a specified way. This is called a neighborhood or an open-set. Topology refers to the connectivity of multiple neighborhoods. A metric space is a topological space that has a specified distance between points that is specified. The limit point and accumulation point are also defined in this section.

Maps are used to map topologies into different shapes. They are functions that do so while preserving the topology. This is used to find the equivalence of two topologies. Two topologies are said to be homeomorphic if they have continuity and inverse. This is a truer test of equivalence.

Another concept is the concept of isotopy: "When two topological spaces are subspaces of the same larger space, a notion of similarity called isotopy exists which is stronger than homeomorphism. If two subspaces are isotopic, one can be continuously deformed to the other while keeping the deforming subspace homeomorphic to its original form all the time. For example, a solid cube can be continuously deformed into a ball in this manner."

The next topic covered is a Manifold. A manifold is a topological space that is locally connected in a way. They are defined as m-manifolds. m refers to the dimension of the manifold. Manifolds can be separated into boundary and internal points. A smooth manifold has no geometry.

(d) Chapter 2:

This chapter is about Simplicial Complexes and Homology Groups.

(e) Chapter 3:

2. **Source**:

   Topological Data Analysis
   Maddie Weinstein
   https://www.youtube.com/watch?v=nG$_V$*eme7bqw*

   Summary:

   Really just an intro to what Topological Data Analysis is. Shapes are discussed but rather than using shapes, you use dots to represent a shape that can be manipulated by twisting and stretching. These shapes can be in any dimensions. Topological Data Analysis can be useful for studying tumors in the medical field.

3. **Source**:

   Topological Data Analysis Intro

Calculus Humor
https://www.youtube.com/watch?v=VFOhup8YuqM

Summary:

We can represent data as nodes with links that then become a shape. So data is not unstructured. We can then analyze the shape of our data. This is what Topological Data Analysis is. Topology is very much based in math as started by Euler.

3 Main Concepts of Topology:

1) Coordinate Invariance... The shape doesn't change when you rotate the shape or change its coordinate system.

2) Deformation Invariance... The shape doesn't change even if you stretch or squash the shape. For example: the same letter in different fonts are the same letter.

3) Compressed Representation... Representing a sphere as an octahedron. You now have a sphere represented as a list of nodes, edges, and faces.

Very large datasets can be analyzed using topological data analysis. This will very much be based in Machine Learning.

4. **Source**:

Gunnar Carlsson: "Topological Modeling of Complex Data"
Gunnar Carlsson, Joint Mathematics Meetings
https://www.youtube.com/watch?v=8nUBqawu41k

Summary:

Size is not just the problem with big data, it's also the complexity. Modeling complex data can buy you Compression, understanding... Clustering is also a way to model. These are algebraic modeling though that have problems. There a quite a few different problems that can't be solved by algebraic modeling.

Another modeling principle is that "Data has shape, and the shape matters." The idea is to take datasets and represent them as networks. Building a network for a dataset requires: applying a projection to a datset, put the data into overlapping bins, cluster each bin using clustering and create a node for each cluster, create an edge between any two nodes that overlap.

Topological modeling will produce a graph or network as its output. It can be viewed on a screen. This can be used for hotspot analysis, phenomena explanation, and feature selection. Using a network that showed clusters/groups, they found that there are 3 different kinds of type 2 diabetes.

Time can not represent something as well as a network/model can sometimes. The states of getting sick and getting better is shown well on a looped network that varies in time based off of individuals and different illnesses. The time doesn't matter when considering the loop. The same concept can be used to model breast cancer. Different colors can show the survival and shape shows different paths.

TDA can be used to better model data that has many features. We only care about the shape so we can take the data as a matrix and then transpose it to build the topological model. This can give us insight into high-dimensional data that traditional algorithms cannot.

The study of Crohn's disease modeled using TDA can show the 2 different kinds of Crohn's compared to healthy subjects. The coloring of the data nodes shows the differences. Pancreatic Cancer can also be modeled using TDA.

Free text can be modeled using TDA as did Lockheed Martin. Unstructured data like free text can be modeled with TDA. We can make predictive models that use coloring to represent a range of values.

You have to watch out for the quality of your data. You need to watch for patterns in your Topological model to actually be a useful result rather than a problem with your data. The errors could come from your data gathering that don't actually show you any results in your data.

Exploratory data analysis was started by John Tukey. 3d visualizations are coming for topological analysis (This lecture was from 2018 so this might have already happened).

5. **Source**:

   Reeb graphs for shape analysis and applications
   S. Biasotti, D. Giorgi, M. Spagnuolo, B. Falcidieno
   https://www.sciencedirect.com/science/article/pii/S0304397507007396

   Summary:

   SUMMARY

6. **Source**:

   Julien Tierny (5/19/20): An Introduction to the Topology ToolKit
   Julien Tierny
   Applied Algebraic Topology Network
   https://www.youtube.com/watch?v=5m5sZ3DFtmAt=3s

   Summary:

Overview of TTK and its general usage. TTK was released open source in 2017. It's written in C++. It's specifically for low dimensional continuous data. It is not good for vector/tensor data. It's also not good for high dimensional data. There is a plugin for ParaView. There is also a Python API for fast scripting.

A tour of ParaView is provided. Paraview is very much a Visual programming with a pipeline. You can also get a python script from the state of paraview state. You can run tutorial data examples from the ttk website. They are state files into ParaView.

TTK tour: The input data is represented by a piecewise linear scalar field. The simplicial complex can be any mesh or a regular grid. Features for Scalar Data: Extraction of Critical Points (ScalarFieldCriticalPoints). It outputs points with index and vertex IDs. Persistence Diagrams (PersistenceDiagram, PersistenceCurve). Extremum and saddle pairs. The persistence curve is also available.

Topological Data Simplification (TopologicalSimplification). Extremum removal and default multiscale mechanism. It outputs the "flattened" data. Topology aware Compression (TopologicalCompression). Preserves the persistence diagram. Only for dimension 0 or d-1. Outputs the compressed image data. Merge and Contour Trees (FTMTree). Skeletion extraction and levelset based segmentation. They output the nodes of the trees, the arcs, and the data segmentation.

He then applies merge and contour trees to biomedical imaging. Reeb Graphs (FTMGraph). Skeleton Extraction, levelset based segmentation, for non simply connected domains. Morse smale complexes (MorseSmaleComplex). Extremal curves, separating surfaces. Outputs critical points, 1D and 2D separatrices and data segmentation.

He applies application to quantum chemsitry. Distance between persistence diagrams (Bottlnec, PersistenceDiagramClustering). Bottleneck and Wasserstein. Outputs Diagrams and Matching. Barycenters / clusters of persistence diagrams (PersistenceDiagramClustering). Wasserstein barycenters + clustering. Outputs diagrams and matching, barycenter (cluster centers).

Time Tracking (TrackingFromFields, TrackingFromPersistenceDiagrams). Minimmizes Wasserstein distance. Dimension Reduction (DimensionReduction). Wrapper to scikit-learn features. Persistence based clustering (PersistenceDiagram, TopologicalSimplification, MorseSmaleComplex). Outputs clustered point cloud. Mapper (FTRGRaph). Density based Reeb graph. Outputs graph and clustering.

There are many features on the tutorial page of the website.

TTK is officially integrated with Paraview 5.10.

7. **Source**:

   The Topology ToolKit
   Julien Tierny, Guillaume Favelier, Joshua A. Levine, Member, IEEE, Charles Gueunet, and Michael Michaux

https://arxiv.org/pdf/1805.09110.pdf

Summary:

$_{S}UMMARY$

8. **Source**:

$_{T}ITLE$
$_{A}UTHOR$