1 List of works

1.1 Introduction and Summary

1. [1] Roadmap for computation of PH, 2017
   Thorough review of computation and some theory of PH.

1.2 Original work

1.2.1 TDA

1.2.2 Persistent Homology

1.3 Applications

2 Questions

2.1 Theory

2.1.1 Homology

1. What other "features" can be revealed by homology?
   It seems that although Homology groups are informational
   Betti numbers, which are the basis for barcode and other diagrams, do not tell us anything more that holes and voids.

2.2 Computation

1. What is the bottleneck of the general computation process?
   (a) Is it the same for all the libraries?
   (b) Possible optimization? Meaningful?
2. What is the bottleneck for a specific data type?
3. For a specific purpose?

2.3 Application

1. Search for a problem?
   It seems that the methods is a not problem-solving-oriented.
   But rather developed from math intuition.
   Therefore seems bound to be less powerful than DL, which is clearly developed to solve problems.

3 TODO List

3.1 Classic Dimensionality Reduction

1. Learn about the following:

(a) PCA
(b) MDS
(c) isomap

2. Maybe need to implement some of them
3. Know the difference from a practical aspect

3.2 Clustering

1. Need to learn about the different approaches
   (a) Variational
   (b) Spectral
   (c) Hierarchical
   (d) Density thresholding
   (e) Mode seeking
   (f) Valley Seeking
2. Need to know the specific applications
3. Need to delve into one of them to see if possible for improvement
4. Mathematical details about Mode Seeking
   (a) Hypotheses on the function of f and estimator $\hat{f}$
   (b) Exact conclusions about the prominence gap.
   (c) How to specify prominence parameter τ
   (d) Algorithm details

3.2.1 Persistent Homology

1. How to compute a filtration from point cloud?
   (a) Various ways to define simplices, Cech, VR, etc
   (b) What if data not necessarily point cloud?
      i. Image?
      ii. 3D objects?
2. Need to delve into the details of the PH algorithms
   (a) Like reading off the intervals.
   (b) Different implementations.

4 Clustering

Point cloud (with coordinates)

Distance / dissimilarity matrix

Note: this seems to be a different idea of the distance function used in TDA, which is called lens in that context.

Barcode →merge tree →dendrogram

4.1 Mode Seeking Paradigm

Problems:

1. Noisy estimator
2. Neighborhood graph

Solutions:

1. Be proactive: smooth
2. Be reactive: merge clusters after clustering
   This leads to "topological persistence"

Persistence for Model Seeking:
Probability density function f

- Nested family (filtration) of of inverse images, or superlevel-sets $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$
- Track evolution of "topology"

Similar stability theorem.
Seems to have relation with Morse theory. "If f is Morse, then..."

5 Topological Persistence

Persistence diagram shows the "persistence" of the topological features. Slight perturbation causes slight difference in persistence diagrams.

6 Homology

Definition: $h : X \mapsto Y$ is a homeomorphism if there exists a map $h^{-1} : Y \mapsto X$, s.t.

- both are continuous
-

7 Computation of Persistence Homology

This part is largely cited from [1].
How to keep track of how one feature "merges" to another?
Boundary matrix: the matrix representation of boundary operator.
We also need a total ordering compatible with the "filtration" in the following sense:

- a face of a simplex precedes the simplex.
- a simplex in K_i precedes simplices in K_j for $j > i$, and not in K_i.
  - this essentially means that we place the simplices by the order of "appearing".

Standard Algorithm:

- Form the boundary matrix from the ordering.
- Reduction, which is essentially Gaussian elimination.
- Reading off intervals.
  1. some details to do.
  2. degree: $\mathrm{dg}(\sigma) =$ smallest number l s.t.$\sigma \in K_l$
  3. pair (σ_i, σ_j) gives $[\mathrm{dg}(\sigma_i), (\sigma_j))$
  4. unpaired extends to infinity.
-

8 Mapper

9 Why TDA?

9.1 Theoretically well understood

9.2 Qualitative data features

9.3 Computable via linear algebra

9.4 Robust under perturbation

10 Possible Topics and Data Sets to Try

1. Sensor network coverage.
2. Proteins.
3. 3-dimensional structure of DNA
4. Robotics
5. Signals in images

6. Periodicity in time series
7. Cancer
8. Phylogenetics
9. Natural images
10. Self-similarity in geometry
11. Materials science
12. Financial networks
13. Neuroscience
14. Other networks
15. Time-series output of dynamical systems
16. Natural language analysis

References

[1] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, "A roadmap for the computation of persistent homology," *EPJ Data Science*, vol. 6, p. 17, Aug 2017.