Topological Data Analysis

Software Packages

Ulderico Fugacci

CNR - IMATI







Topological Data Analysis allows for assigning to (almost) *any dataset* a collection of features representing a *topological summary* of the input data



Goal:

Today, we address one main question:

What software packages are available for computing persistent homology?

Several software packages for computing persistent homology have been developed:

javaPlex

- + jHoles
 - + Dionysus
 - + Perseus
 - + PHAT
 - + DIPHA
 - + Gudhi
 - SimpPers
 - + Ripser
 - + giotto-tda
 - + TDAstats
 - + ...

javaPlex:

- Language:
 - Java
- + Algorithms:
 - * Standard, Dual, Zigzag
- Coefficient Fields:
 - * Q, \mathbb{Z}_p
- + Homology:
 - * Simplicial, Cellular

- + Accepted Inputs:
 - Simplicial complexes, Zigzag, CW complexes
- Computed Filtrations:
 - Vietoris-Rips complexes, (parametrized) Witness complexes
- Visualizations:
 - Persistence Barcodes
- Additional Features:
 - * Homology Generators

jHoles:

- Language:
 - Java
- + Algorithms:
 - * Standard (uses javaPlex)
- Coefficient Fields:
 - * Z2
- + Homology:
 - * Simplicial

- Accepted Inputs:
 - •
- Computed Filtrations:
 - * Weight Rank Clique filtration
- Visualizations:
 - ***** -
- Additional Features:
 - *

Dionysus:

- + Language:
 - C++ (with Python bindings)
- + Algorithms:
 - * Standard, Dual, Zigzag
- Coefficient Fields:
 - * \mathbb{Z}_2 (standard, zigzag), \mathbb{Z}_p (dual)
- + Homology:
 - * Simplicial
- + Accepted Inputs:
 - * Simplicial complexes, Zigzag

Computed Filtrations:

- Vietoris-Rips and Čech complexes, Alpha-Shapes, Lower Star of Cubical complexes
- Visualizations:
 - •*•
- Additional Features:
 - Bottleneck and Wasserstein distances, Vineyards, Circle-Valued functions, Homology Generators

Perseus:

- + Language:
 - C++ (with Python bindings)
- + Algorithms:
 - * Standard, Morse reductions
- Coefficient Fields:
 - * Z2
- + Homology:
 - * Simplicial, Cubical

- Accepted Inputs:
 - * Simplicial and Cubical complexes
- Computed Filtrations:
 - Vietoris-Rips complexes,
 Lower Star of Cubical complexes
- + Visualizations:
 - * Persistence Diagrams
- Additional Features:
 - * Weighted Points for VR



- + Language:
 - * C++ (with Python bindings)
- + Algorithms:
 - * Standard, Dual, Twist, Chunk, Spectral Sequences
- Coefficient Fields:
 - * Z2
- + Homology:
 - * Simplicial, Cubical

- + Accepted Inputs:
 - * Boundary Matrices
- Computed Filtrations:
 - * -
- + Visualizations:
 - * -
- Additional Features:
 - * •

DIPHA:

- + Language:
 - C++ (with Python bindings)
- + Algorithms:
 - * Dual, Twist, Distributed
- Coefficient Fields:
 - * Z2
- + Homology:
 - * Simplicial, Cubical

- Accepted Inputs:
 - * Boundary Matrices
- Computed Filtrations:
 - Vietoris-Rips complexes,
 Lower Star of Cubical complexes
- + Visualizations:
 - * Persistence Diagrams
- Additional Features:

*

Gudhi:

- + Language:
 - C++ (with Python bindings)
- + Algorithms:
 - * Dual, Annotation, Multifield
- + Coefficient Fields:
 - ✤ ℤ_p
- + Homology:
 - * Simplicial, Cubical
- + Accepted Inputs:
 - * Simplicial and Cubical complexes

Computed Filtrations:

- Vietoris-Rips and Čech complexes, Alpha-Shapes, Witness complexes, Lower Star of Cubical complexes
- Visualizations:
 - Persistence Diagrams, Barcodes, Landscapes, Persistence Images

Additional Features:

 Bottleneck and Wasserstein distances, Sliced Wasserstein kernel, Heat kernel, Weighted Gaussian kernel

SimpPers:

- + Language:
 - * **C++**
- + Algorithms:
 - * Simplicial Maps
- Coefficient Fields:
 - ✤ ℤ₂
- + Homology:
 - * Simplicial

- Accepted Inputs:
 Maps of simplicial complexes
- Computed Filtrations:
 - * -
- Visualizations:
 - * -
- Additional Features:
 - •

Ripser:

- + Language:
 - C++ (with Python bindings)
- + Algorithms:
 - * Dual, Twist
- Coefficient Fields:
 - ✤ ℤ_p
- + Homology:
 - * Simplicial
- + Accepted Inputs:
 - * Point Clouds, Distance Matrices

Computed Filtrations:

- Vietoris-Rips and Čech complexes, Alpha-Shapes, Lower Star of Cubical complexes
- Visualizations:
 - Persistence Diagrams (through Persim: Persistence Images)

Additional Features:

 Representative Cocycles (through Persim: Bottleneck distance, modified Gromov–Hausdorff distance, Sliced Wasserstein kernel, Heat kernel)

giotto-tda:

- + Language:
 - * **Python** (built on scikit-learn)
- + Algorithms & Coefficient Fields:
 - ✤ (does it run Ripser?)
- + Homology:
 - Simplicial, Cubical, Directed
 Simplicial
- + Accepted Inputs:
 - Point Clouds, Distance Matrices, Images, Graphs

- Computed Filtrations:
 - Vietoris-Rips and Čech complexes, Alpha-Shapes
- Visualizations:
 - Persistence Diagrams, Landscapes, Heat Representations, Persistence Images
- Additional Features:
 - Bottleneck and Wasserstein distances, Heat kernel, L_p-distance between representations

TDAstats:

- + Language: * R
- + Algorithms:
 - * **Dual, Twist** (uses Ripser)
- Coefficient Fields:
 - ✤ ℤ_p
- + Homology:
 - * Simplicial

- + Accepted Inputs:
 - * Point Clouds
- Computed Filtrations:
 - * Vietoris-Rips complexes
- + Visualizations:
 - Persistence Diagrams, Barcodes
- Additional Features:
 - *

Computation Times:

| Data set | (a) Computations on cluster: wall-time seconds | | | | | | | |
|-----------------|--|---------------------|-----------------------|-----------------------|-----------------------|-----------------------|--|--|
| | eleg | Klein | HIV | drag 2 | random | genome | | |
| Size of complex | 4.4×10^{6} | 1.1×10^{7} | 2.1 × 10 ⁸ | 1.3 × 10 ⁹ | 3.1 × 10 ⁹ | 4.5 × 10 ⁸ | | |
| Max. dim. | 2 | 2 | 2 | 2 | 8 | 2 | | |
| JAVAPLEX (st) | 84 | 747 | - | - | - | - | | |
| Dionysus (st) | 474 | 1,830 | - | - | - | - | | |
| DIPHA (st) | 6 | 90 | 1,631 | 142,559 | - | 9,110 | | |
| Perseus | 543 | 1,978 | - | - | - | - | | |
| Dionysus (d) | 513 | 145 | - | - | - | - | | |
| DIPHA (d) | 4 | 6 | 81 | 2,358 | 5,096 | 232 | | |
| Gudhi | 36 | 89 | 1,798 | 14,368 | - | 4,753 | | |
| Ripser | 1 | 1 | 2 | 6 | 349 | 3 | | |

Computation Times:

| Data set | (b) Computations on cluster: CPU seconds | | | | | | | |
|-----------------|--|---------------------|-----------------------|-----------------------|-----------------------|-----------------------|--|--|
| | eleg | Klein | HIV | drag 2 | random | genome | | |
| Size of complex | 4.4 × 10 ⁶ | 1.1×10^{7} | 2.1 × 10 ⁸ | 1.3 × 10 ⁹ | 3.1 × 10 ⁹ | 4.5 × 10 ⁸ | | |
| Max. dim. | 2 | 2 | 2 | 2 | 8 | 2 | | |
| JAVAPLEX (st) | 284 | 1,031 | - | - | - | - | | |
| Dionysus (st) | 473 | 1,824 | - | - | - | - | | |
| DIPHA (st) | 68 | 1,360 | 25,950 | 1,489,615 | - | 130,972 | | |
| Perseus | 542 | 1,974 | - | - | - | - | | |
| Dionysus (d) | 513 | 145 | - | - | - | - | | |
| DIPHA (d) | 39 | 73 | 1,276 | 37,572 | 79,691 | 3,622 | | |
| Gudhi | 36 | 88 | 1,794 | 14,351 | - | 4,764 | | |
| Ripser | 1 | 1 | 2 | 5 | 348 | 2 | | |

Software Packages

Computation Times:

| Data set | (c) Computations on shared-memory system: wall-time seconds | | | | | | |
|-----------------|---|---------------------|-----------------------|-----------------------|-----------------------|-----------------------|--|
| | eleg | Klein | HIV | drag 2 | genome | fract r | |
| Size of complex | 3.2 × 10 ⁸ | 1.1×10^{7} | 2.1 × 10 ⁸ | 1.3 × 10 ⁹ | 4.5 × 10 ⁸ | 2.8 × 10 ⁹ | |
| Max. dim. | 3 | 2 | 2 | 2 | 2 | 3 | |
| JAVAPLEX (st) | 13,607 | 1,358 | 43,861 | - | 28,064 | - | |
| Perseus | - | 1,271 | - | - | - | - | |
| Dionysus (d) | - | 100 | 142,055 | 35,366 | - | 572,764 | |
| DIPHA (d) | 926 | 13 | 773 | 4,482 | 1,775 | 3,923 | |
| Gudhi | 381 | 6 | 177 | 1,518 | 442 | 4,590 | |
| Ripser | 2 | 1 | 2 | 5 | 3 | 1,517 | |

Software Packages

Memory Usage:

| Data set | (a) Computations on cluster | | | | | | | |
|-----------------|-----------------------------|---------------------|-----------------------|-----------------------|-----------------------|-----------------------|--|--|
| | eleg | Klein | HIV | drag 2 | random | genome | | |
| Size of complex | 4.4 × 10 ⁶ | 1.1×10^{7} | 2.1 × 10 ⁸ | 1.3 × 10 ⁹ | 3.1 × 10 ⁹ | 4.5 × 10 ⁸ | | |
| Max. dim. | 2 | 2 | 2 | 2 | 8 | 2 | | |
| JAVAPLEX (st) | <5 | <15 | >64 | >64 | >64 | >64 | | |
| Dionysus (st) | 1.3 | 11.6 | - | - | - | - | | |
| DIPHA (st) | 0.1 | 0.2 | 2.7 | 4.9 | - | 4.8 | | |
| Perseus | 5.1 | 12.7 | - | - | - | - | | |
| Dionysus (d) | 0.5 | 1.1 | - | - | - | - | | |
| DIPHA (d) | 0.1 | 0.2 | 1.8 | 13.8 | 9.6 | 6.3 | | |
| Gudhi | 0.2 | 0.5 | 8.5 | 62.8 | - | 21.5 | | |
| Ripser | 0.007 | 0.02 | 0.06 | 0.2 | 24.7 | 0.07 | | |

Software Packages

Memory Usage:

| Data set | (b) Computations on shared-memory system | | | | | | |
|-----------------|--|---------------------|---------------------|-----------------------|-----------------------|-----------------------|--|
| | eleg | Klein | HIV | drag 2 | genome | fract r | |
| Size of complex | 3.2 × 10 ⁸ | 1.1×10^{7} | 2.1×10^{8} | 1.3 × 10 ⁹ | 4.5 × 10 ⁸ | 2.8 × 10 ⁹ | |
| Max. dim. | 3 | 2 | 2 | 2 | 2 | 3 | |
| JAVAPLEX (st) | <600 | <15 | <700 | >700 | <700 | >700 | |
| Perseus | - | 11.7 | - | - | - | - | |
| Dionysus (d) | - | 1.1 | 16.8 | 134.2 | - | 268.5 | |
| DIPHA (d) | 31.2 | 0.9 | 17.7 | 109.5 | 37.3 | 276.1 | |
| Gudhi | 15.4 | 0.5 | 10.2 | 62.8 | 21.4 | 134.8 | |
| Ripser | 0.2 | 0.03 | 0.07 | 0.2 | 0.07 | 155 | |

Software Packages

Supported Maximal Size:

| JAVAPLEX
st | DIONYSUS | | DIPHA | | Perseus | GUDHI | Ripser |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|---------------------|-----------------------|-----------------------|
| | st | d | st | d | st | d | d |
| 4.5 · 10 ⁸ | 1.1 • 10 ⁷ | 2.8 × 10 ⁹ | 1.3 · 10 ⁹ | 3.4 · 10 ⁹ | 1 • 10 ⁷ | 3.4 · 10 ⁹ | 3.4 · 10 ⁹ |

Bibliography

General References:

- Books on TDA:
 - A. J. Zomorodian. *Topology for computing*. Cambridge University Press, 2005.
 - * H. Edelsbrunner, J. Harer. *Computational topology: an introduction*. American Mathematical Society, 2010.
 - R. W. Ghrist. *Elementary applied topology*. Seattle: Createspace, 2014.

Papers on TDA:

• G. Carlsson. *Topology and data*. Bulletin of the American Mathematical Society 46.2, pages 255-308, 2009.

Today's References:

- Software Packages:
 - N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, H.A. Harrington. A roadmap for the computation of persistent homology. EPJ Data Science, 6.1, 2017.
 - G. Tauzin, U. Lupo, L. Tunstall, J. B. Pérez, M. Caorsi, A. M. Medina-Mardones, A. Dassatti, K. Hess. *giotto-tda: a topological data analysis toolkit for machine learning and data exploration.* Journal of Machine Learning Research, 22.39, pages 1-6, 2021.