TDA Reading Group MT 2018

October 12

Oliver Vipond

Introduction to Persistent Homology

This talk will serve as an introduction to those less familiar with TDA. I'll introduce the key theory for one dimensional persistence including the decomposition theorem, stability theorem and isometry theorem. I will then introduce a range of methods for producing filtered complexes built on data sets and discuss the relative advantages of each construction.

References: Ghrist, Oudot, Edelsbrunner and Harer

October 19

Oliver Vipond

Reeb Graphs, Extended Persistence and Mapper

This week we will look at Extended Persistence, Reeb Graphs and Mapper. The Reeb Graph and Mapper algorithm provide a visualisation of high dimensional data sets as graphs. We will cover the Reeb graph construction and the Mapper algorithm, and show how extended persistence can capture the features of the resulting graphs. If time permits we will look at applications of Mapper to real world data sets.

Notes: Mapper Examples

References: Structure and Stability of the 1-Dimensional Mapper, Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition

October 26

Jacob Leygonie

Learning with Persistence Diagrams

During this session, we will see how the topological features we are now able to compute thanks to the technics of session 1 may allow designing learning models. That is, we want to use classical models for classification or regression on top of the persistent diagrams we obtained. I attached the slides of the presentation.

The first two sessions ended up in applications with unsupervised models: the Reeb graph computed by Mapper (Session 2) provides a way of visualizing and clustering our data. In a similar spirit, Mode-seeking can make use of Persistent Homology we explored in Session 1 to obtain unsupervised representation of the data (with advantageous guarantees and flexibility compared to competitors such as Mean-Shift, k-means, etc.): Persistent Homology Clustering.

In this session, we will assume we know what we want to learn and hence the setting is supervised learning. Given a dataset of features and target values, we want to build a model, a "predictor", that will efficiently correlate features and target. It happens that popular models in supervised learning usually deal with data represented as finite dimensional vectors. Hence there is some work needed if we want to make use of our topological descriptors, the barcodes which are multi-sets, in order to design a predictor. I will motivate the need of "vectorizing" barcodes and then explore the main methods developed for doing so. We will explore five methods: Persistent Images, Finite Metric Space, Polynomial Evaluation, Landscapes and Convolutional Weighted Kernel. The presentation follows the logic of the course: Learning On Persistent Diagrams.

Notes: Learning on Persistence Diagrams

November 2

Ambrose Yim

Clustering with Persistent Homology

I will discuss the ToMATo (Topological Mode Analysis Tool) clustering algorithm [1] which is inspired by recent advances in persistent homology. ToMATo is of interest to our reading group for two reasons. Firstly, its phenomenal success in clustering challenging datasets where other methods fail merits our attention. Secondly, ToMATo is a rare case where algebraic methods - in this case, the interleaving of persistent modules [2] - have been used to prove results about clustering algorithms. I will sketch how interleaving is used to prove that ToMATo recovers clusters with high probability. This is also an opportune moment to introduce the concept of interleaving and the isometry theorem to the reading group.

References:

[1] ToMATo: Persistence-Based Clustering in Riemannian Manifolds

[2] The Structure and Stability of Persistence Modules

November 9

Oliver Vipond

Generalised Persistence Modules, Interleavings and Gromov Hausdorff Distance

In our previous meetings we have looked at single parameter persistent homology, and the associated persistence module. The interleaving distance encapsulates the notion of approximate isomorphism, and is used to compare persistence modules. I will show how the notion of interleaving admits a generalisation to modules over arbitrary posets and can be formulated as the Gromov-Hausdorff distance on weighted categories. This will enable us to describe and compare a richer class of algebraic objects derived from data. If time permits, I will describe a universality result which establishes the interleaving distance as the most discriminative stable distance on multiparameter persistence modules.

Notes: Interleaving Distance as Gromov Hausdorff Distance

References: Metrics for generalised persistence modules, Theory of Interleavings on Categories with a Flow, Interleaving and Gromov Hausdorff distance

November 16

Naya Yerolemou

The Algebraic Stability Theorem

This week took a detailed look at the stability of persistence barcodes, through an explicit map from category of p.f.d. persistence modules to category of barcodes. We looked at the proof in the paper by Bauer and Lesnick which was distilled by Naya into its essential components.

Reference: Algebraic Stability of Persistence Barcodes

November 23

William Matlock

Bounding Recombination through Topological ARGs

This week Will showed us how TDA tools have been used to provide a new bound for the number or recombination events in an Ancestral Recombination Graph. This method enjoys lower computational complexity than other methods and bars in the H_1 persistence barcodes can be interpreted as recombination events.

Notes: tARGs

Reference: Inference of Ancestral Recombination Graphs through Topological Data Analysis

November 30

Jacob Leygonie & Oliver Vipond

Persistent Homology and Euler Characteristic Transforms

This week our one hour talk will be divided in two parts held by Oliver and I around a recent paper: https://arxiv.org/abs/1805.09782 by J. Curry, S. Mukherjee and K. Turner.

Broadly speaking, one has the intuition that 1-parameter persistent homology, e.g when we filter a space with a Rips filtration and compute the corresponding barcode, encapsulates more information than purely topological ones. As an example, two homotopy equivalent spaces might have different barcodes. Hence one can try to characterise which type of information about the space is kept by this transformation. If one wants to retain as many information as possible, it could be natural to take many filters and compute all the corresponding barcodes, instead of just one. It happens that for a simple family of filters, we can discriminate between all spaces inside a broad class using these barcodes. I will explain the interest of this result and how to obtain it. Another version of this theorem is available in https://arxiv.org/abs/1804.04740 by R. Ghrist, R. Levanger and H. Mai.

Then we will address a more challenging question which is the second part of the paper: can we make use of only finitely many filters to discriminate between all spaces? Oliver will explain how we can positively answer that question. This is of interest in application: if we have a dataset of shapes, we know there exists a finite number of filters, such that computing PH according to those filters will produce distinct barcodes. Hence we can have discriminative summary -although how discriminative is hard to evaluate- of the dataset.

In case of metric graphs, a similar analysis in spirit is available: https://arxiv.org/abs/1712.03630

Also, the crucial ingredients of the proofs are the tameness of the sets and the triangulation theorem. The large concept of o-minimal structure is the right situation in one want to be in order to use those results: https://www.cambridge.org/core/books/tame-topology-and-o-minimal-structures/7AC940248AF2B05DA4D33E4FB05C97A2

Notes: Finitely Many Directions Suffice - Slides