Computational topology - group project

Sliding Window Persistence for time series

Introduction: Time series is a sequence of values, often representing a signal, a measurement, etc. One of the more challenging and important tasks is to determine the extent to which a given time series is periodic. Discrete Fourier Transform, wavelets, and related approaches work well in signal processing, but may fail to detect periodicity in more irregular settings. In this project you will explore how to use topological data analysis to detect and evaluate periodicity. Your starting point will be the the Sliding Window pipeline for TDA, described in `https://arxiv.org/pdf/1307.6188` and `https://arxiv.org/pdf/2103.04540`.

Goal: Experiment with the Sliding Window Persistence Pipeline to discover how it captures the periodicity of time series.

Methodology: The Sliding Window Persistence Pipeline:

- Given a time series $f\colon \{1, 2, \ldots, n\} \to \mathbb{R}$, choose a dimension $M \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ to define a map
$$SW_{M,\tau} f(t) = (f(t), f(t+\tau), \ldots, f(t+M\tau)).$$
- Compute the 1-dimensional persistent homology of the image of the said map. The geometric intuition dictates that for an approximately periodic time series and appropriate choices of M and τ, persistent homology should exhibit a long 1-dimensional bar.
- Once the computation of persistence diagrams becomes slow, modify your approach:
  - Compose $SW_{M,\tau} f$ with a dimension reduction (for example, PCA) to $\mathbb{R}^2$ or $\mathbb{R}^3$ before computing persistence.
  - Fix a value α and compute the 1-dimensional persistence of $SW_{M,\tau} f(\{m, m+1, m+2, \ldots, m+\alpha\})$ for all feasible values of m. See how persistence diagrams change with increasing m. If a long 1-dimensional bar is appearing throughout all choices of m, this might indicate periodicity.

Tasks:

1. Implement the Sliding Window Persistence Pipeline described in the methodology.

2. Test your approach on various time series. First, generate synthetic time series which are periodic (say $\sin(k \cdot t)$, etc.) sort of periodic (say $\cos(k \cdot t) + t$ or $\cos(k \cdot t) + \log(t)$, etc.) or not periodic at all, try to add a small amount of noice. Then test your approach on datasets from `https://www.timeseriesclassification.com/dataset.php` or `https://machinelearningmastery.com/time-series-datasets-for-machine-learning/`. Explore how the output depends on M and τ.

3. Determine whether long 1-dimensional bars in persistence diagrams correspond to periodic time series.

4. Try to detect periodicity in functions such as $\cos(k \cdot t) + t$ or $\cos(k \cdot t) + \log(t)$.

5. Try to derive a reasonable score evaluating degree to which time series is periodic.

Results: The report should include a description of methodology and tasks undertaken, a pseudocode, methods of computation, results of experiments, and division of work.

Students are encouraged to take the initiative and possibly implement their own ideas on the theme of the project.