Time Series Classification using Topological Data Analysis

Sai Manoj Kumar, Degavath Devi Singh, Bhagya Lakshmi
Department of Computer Science and Automation, IISc Bangalore

1. Problem Description

Real-world time series data contains both chaotic and non-chaotic samples. Chaotic samples mean their values are unpredictable and random at certain time points; hence they are considered irrelevant for a classification model. So the problem is to devise a better data preprocessing step to extract essential and relevant features from the samples and give them as input to the classification model.

2. Project Goals

First, we construct quasi attractors from samples. Then extract persistent homology features from them, construct a Betti sequence, and implement a 1D CNN classification model that can work with these extracted feature vectors. Then we took chaotic and non-chaotic datasets, applied the proposed implementation, and explored further according to the problem description.

3. Our Contributions

- We have picked chaotic and non-chaotic datasets and explored them by visualizing their quasi attractors, persistence diagrams, barcodes, and betti sequences and then applying the proposed 1D CNN algorithm.
- We used an extra step of preprocessing the data (other than what was proposed in the paper), i.e., dimensionality reduction, to reduce the dimension of the high dimensional quasi attractor. This resulted in a reduction in some noise from the 'Human activity recognition dataset.
- Also, as part of our original work, we devised another modified version of the proposed classification model of the paper, which is a 2D CNN. We converted the extracted features vectors into Matrices and devised a 2D CNN which can work with these inputs. But the results are the same as the proposed 1D CNN, which adds up to the argument that the proposed 1D CNN algorithm is better in these scenarios.
- We also implemented other classification algorithms like SVM, KNeighboursClassifiers, and RandomForestClassifiers which suit the extracted topological features, and cross-verified their results with the proposed algorithm.

4. Key challenges faced

- We faced difficulties understanding the chaotic time series data as it is very random and unpredictable. We also faced challenges in figuring out the optimal values for the time delay and embedding dimensions parameters for the datasets.
- The higher dimensional quasi attractors of HAR dataset samples contained points spread more randomly. This is due to its chaotic nature. So we tried to use dimensionality techniques to reduce its dimensions by preserving the variance.

5. Responsibilities of individual team members

- **SaiManoj Kumar:** Computation of the Topological data analysis on the dataset, Implemented the proposed algorithms, and devised a new modified version of the proposed classification model.
- **Bhagya Lakshmi:** Identified the datasets and applied the algorithms to analyze them visually and analytically.
- **Devi Singh:** Implemented other classification models which suits the extracted features and compared and analyzed various datasets using them.

References

[1] https://www.sciencedirect.com/science/article/pii/S0957417421007557

[2] An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists, Frédéric Chazal and Bertrand Michel, 25 Feb 2021

[3] A topological data analysis based classification method for multiple measurements, Henri Riihimäki, Wojciech Chachólski, Jakob Theorell, Jan Hillert and Ryan Ramanujam