Signal Separation using Topological Data Analysis for Machine Learning Applications

Jing Yang, Khushboo Mittal and Shalabh Gupta

Department of Electrical and Computer Engineering, University of Connecticut

Signal separation deals with the recovery of original signals from an observed mixture of these signals. Signal separation is required in various domains, the most common being audio communication, where sound signals are corrupted by background noise, recordings are made using multiple microphones, etc. Some other applications of signal separation include medical signal processing, telecommunications, etc. Lack of prior knowledge about the signal sources makes this problem challenging.

In this work, a novel methodology to extract the original signals from a mixed signal, using topological data analysis is proposed. Persistent homology [1] based topological features [2], such as Betti numbers, which provide information about the cycles present in data are extracted and used to determine the individual components of the mixed signal. Along with Betti numbers, several other features such as diameter of a cycle, centroid of a cycle, etc. are also utilized in this work. The proposed approach exploits the fact that even after mixing the signals, their topological information is preserved, which can be used to extract meaningful information from the mixed signal. The overall approach is illustrated in Fig.1. Phase portrait should first be constructed with selected delay parameter (k) equal to one. Then, persistent intervals, such as Betti number can be computed from the phase portrait using persistent homology. Last, with extracted information from Betti number, several features are used to reconstruct sine and square waveform.

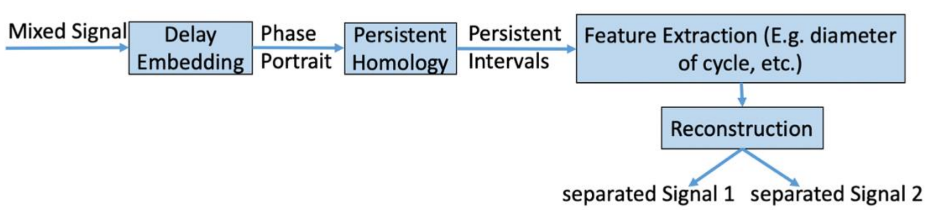

Fig 1. Methodology framwork

Signal separation problem has been studied by several researchers in the past. Comon [3] formulated the mathematical framework of the Independent Component Analysis (ICA) to separate multivariate signal into statistically independent non-Gaussian signals. Hyvarinen and Oja [4] proposed a fixed-point algorithm for ICA called FastICA, which uses fixed-point iteration for orthogonal rotation to achieve faster convergence. ICA has the limitation that it requires at least two mixed signals with different mixing coefficients to identify the signal sources. Moreover, it fails to correctly recover the magnitude of original signal. Another statistical procedure to separate signals using eigenvalue decomposition is Principle component analysis [5]. It is used to first preprocess the signal before performing ICA (PCA-ICA).

This paper shows one example of the mixture of one sinusoid waveform and one square waveform with different frequency and magnitude. Let the delay parameter (k) be 1 and construct the phase portrait shown in figure 2. The persistent homology used simplicial complexes with a certain space to extract topological information such as birth and death time for a cycle. This information is encapsulated in a structure called the persistent barcode. Figure 3 is the barcode for the sine with square example which shows the birth and death time for each cycle.

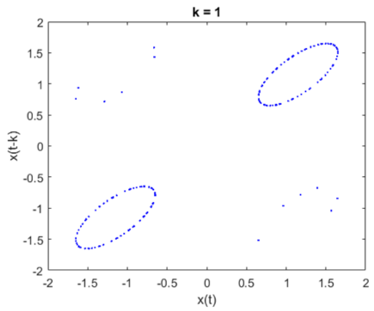

Fig 2. Phase portrait of sine on square

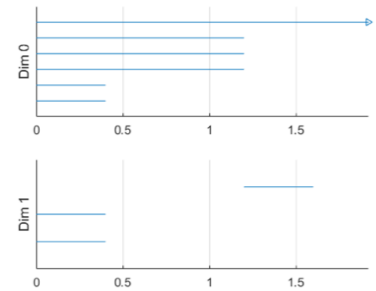

Fig 3. Barcode with dimension 0/ dimension 1

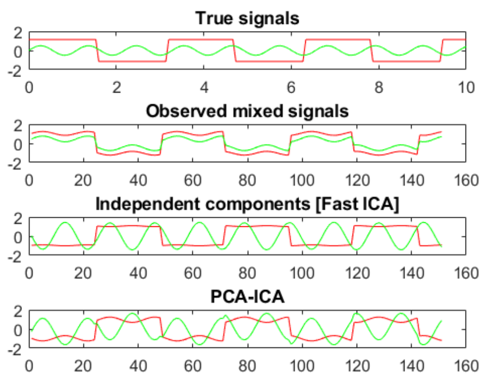

Fig 4. Sine on square separation with ICA

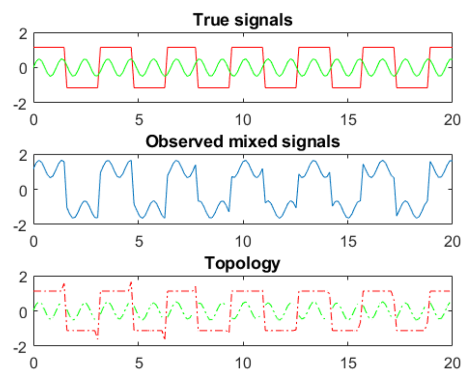

Fig 5. Sine on square separation with Topology

Using the information in barcode dimension one, points in both cycles can be determined and stored as sine data points with shifted zero-mean. Some features as diameter of a cycle, centroid of a cycle can be extracted to reconstruct the square data point by shifting every point in the cycles to the center point of the cycles [2]. Figure 4 illustrates the original signals, mixed signals and separated signals with FastICA and PCA-ICA. FastICA can recover frequency and with the shape slightly distorted while PCA-ICA fail to recover the shape of signals. Figure 5 shows the original signals, linear mixture signal and the plot of sampling time with reconstructed signals. As shown, topological data analysis can recover the signals with exact frequency and magnitude.

References

[1] Edelsbrunner, H., Letscher, D. and Zomorodian, A., 2000. Topological persistence and simplification. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on* (pp. 454-463). IEEE.

[2] Mittal, K and Gupta, S. "Topology characterization and early detection of bifunctions and chaos in complex systems using persistent homology". Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(5):051102,2017

[3] Comon, P. "Independent component analysis, A new concept?". Signal Processing, 36 (1994) 287-314.

[4] Hyvärinen, A and Oja, E. "A Fast Fixed-Point Algorithm for Independent Component Analysis". Neural Computation, 9:1483-1492, 1997.

[5] Jolliffe, I. T. "Principal Component Analysis". Springer-Verlag. New York. 1986