Topology combined machine learning for consonant recognition

Pingyao Feng

Southern University of Science and Technology

annfeng233@gmail.com

January 17, 2024

Overview

Background

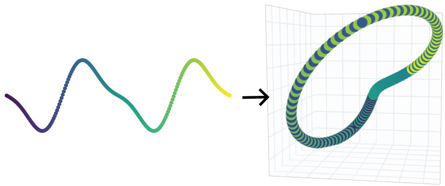


Figure 1: Time-delay embedding (dimension=3, delay=10, skip=1) of a periodic function. Resulting point clouds lay on a closed curve in 3-dimensional Euclidean space. The colour indicates their original locations in the time series.

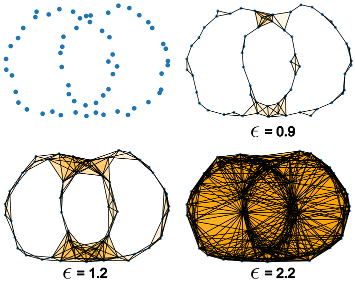


Figure 2: Computing Persistent Homology (persistent homology). The four plots consecutively show how a diagram or a barcode is computed.

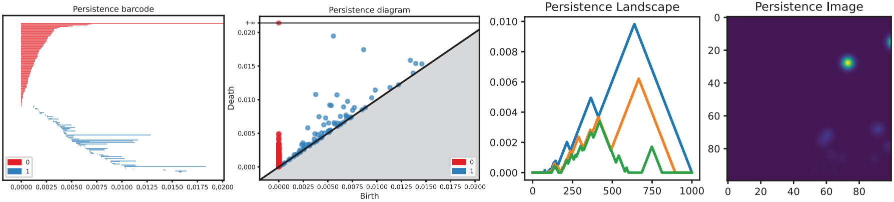


Figure 3: Commonly used representations for persistent homology.

Detection of vibration patterns

What kind of information can be spotted by topological methods?

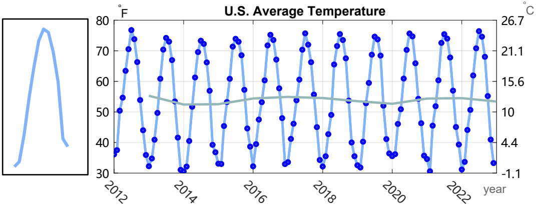


Figure 4: Average temperature in the U.S. with monthly values (dark blue dots) and yearly values (green curve). The left panel shows a single-year section of average temperature.

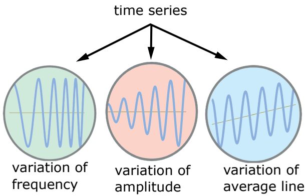


Figure 5: Characterising the vibration of a time series in terms of its variability of frequency, amplitude, and average line.

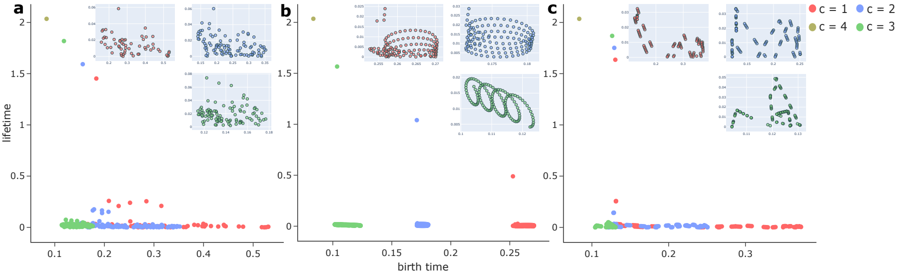
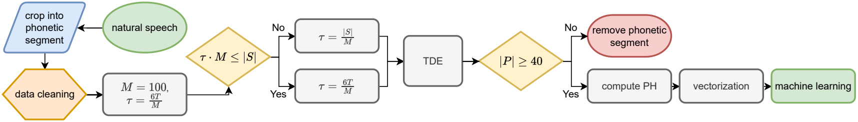


Figure 6: 1-dimensional persistent homology reveals three fundamental variations. **a**, Detecting variation of frequency. Upper-right panels zoom in to show the barcode distribution in the lower dense region, where the position and colour of each value of c in the main legend corresponds to those of its panel. Note that when $c = 4$, there is a single point, and so the panel for this value is omitted. **b**, Detecting variation of amplitude. **c**, Detecting variation of average line.

Abstract

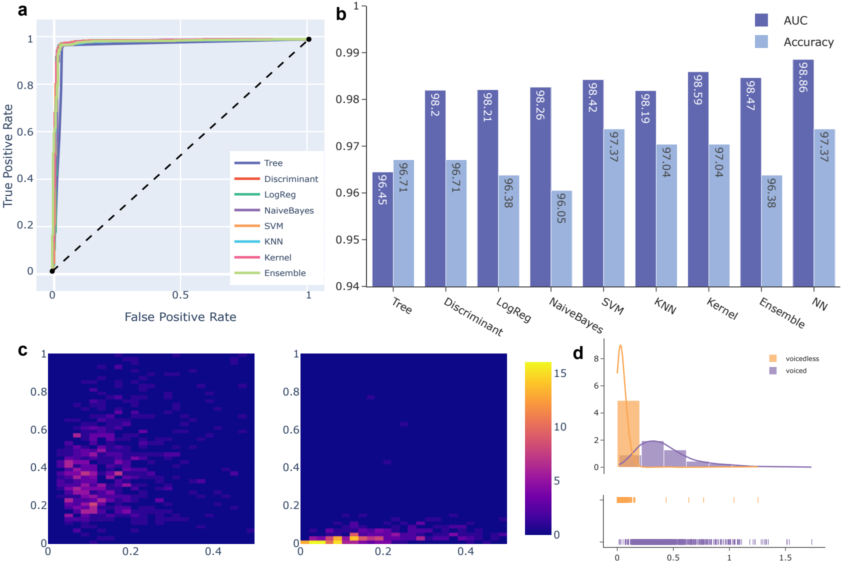
Here, we provide a transparent and broadly applicable methodology, TopCap, to capture the most salient topological features inherent in time series for machine learning. This information is then vectorised and fed to multiple machine learning algorithms such as k-Nearest Neighbours and Support Vector Machine. Notably, in classifying voiced and voiceless consonants, TopCap achieves an accuracy exceeding 96% and is geared towards designing topologically enhanced convolutional layers for deep learning of speech and audio signals.

Architecture



Flow chart of experiment. Here $|S|$ denotes the number of samples in a time series, $|P|$ denotes the number of points in the point cloud, and T denotes the (minimal) period of the time series computed by the ACL function.

Results

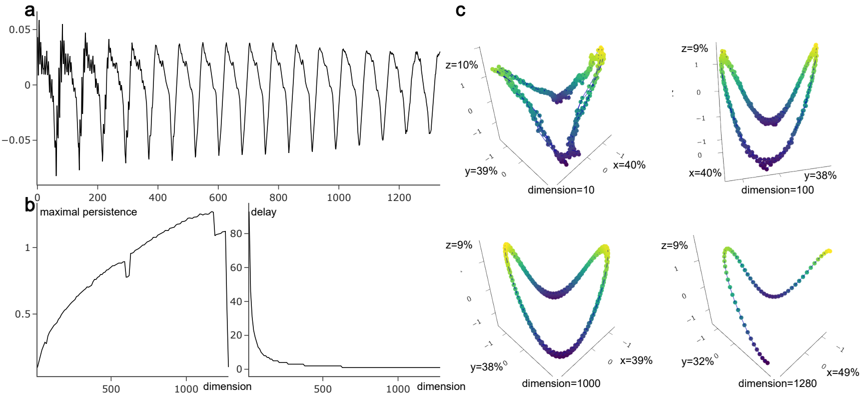

a
True Positive Rate
False Positive Rate
Tree
Discriminant
LogReg
NaiveBayes
SVM
KNN
Kernel
Ensemble
b
AUC
Accuracy
96.45
96.71
98.2
96.71
98.21
96.38
98.26
96.05
98.42
97.37
98.19
97.04
98.59
97.04
98.47
96.38
98.86
97.37
Tree
Discriminant
LogReg
NaiveBayes
SVM
KNN
Kernel
Ensemble
NN
c
d
voicedless
voiced

Machine learning results with topological features. **a**, ROCs of traditional machine learning algorithms. **b**, Accuracy and AUC of each of these algorithms. **c**, Diagrams of records represented as (birth time, lifetime) for voiced consonants (left) and voiceless consonants (right), where voiced consonants exhibit relatively higher birth time and lifetime. The colour represents the density of points in each unit grid box. **d**, Histograms of records represented by their lifetime for voiced and voiceless consonants, together with kernel density estimation and rug plot. The distributions of maximal persistence can distinguish voiced and voiceless consonants.

Remarks

- Parameters in time-delay embedding: dimension, delay, skip; and their influences on persistent homology.
- Using topological methods to identify trend and variation: what are the obstacles?
- There is evidence showing that topological methods make AI more interpretable and efficient.
- Topological data analysis: current development and related works.

Dimension


a
0.05
0
−0.05
0
200
400
600
800
1000
1200
b
maximal persistence
1
0.5
500
1000
dimension
delay
80
60
40
20
0
500
1000
dimension
c
z=10%
y=39%
x=40%
dimension=10
z=9%
x=40%
y=38%
dimension=100
z=9%
y=38%
x=39%
dimension=1000
z=9%
y=32%
x=49%
dimension=1280

Point-cloud behaviour with increasing embedded dimension. **a**, Original .wav file of a record of [ŋ] (voiced consonant). **b**, maximal persistence of the series after TDE as dimension increases (left) and the corresponding delay that ensures the time series to reach theoretically optimal maximal persistence (right). Skip equals 5 when computing PD. **c**, Visualisation of the embedded point clouds, which shows principal component analysis (PCA) of the embedded point clouds in 3D as projected from various dimensions. Skip equals 1 when performing PCA. The percentage along each axis indicates the PCA explained variance ratio.

Skip

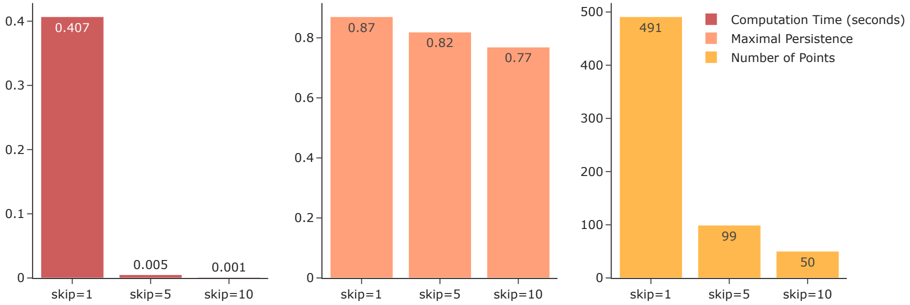


Figure 7: Given a sound record of the voiced consonant [m], computation time, maximal persistence, and the size of point clouds as skip increases. An increase in skip can lead to a significant reduction in computation time, owing to the reduced size of the point cloud. However, maximal persistence remains resilient to an increase in the skip parameter.

Multiple dependency

dimension = 10	desired delay = 40		dimension = 50	desired delay = 8		dimension = 100	desired delay = 4	
delay	skip	maximal persistence	delay	skip	maximal persistence	delay	skip	maximal persistence
1	1	0.0610	1	1	0.2834	1	1	0.4270
10	1	0.1299	3	1	0.3021	2	1	0.4337
20	1	0.1312	4	1	0.3054	2	5	0.4146
30	1	0.1281	5	1	0.3058	3	1	0.4357
39	1	0.1229	6	1	0.3042	3	5	0.4120
39	5	0.1134	7	1	0.3052	4	1	0.4381
40	1	0.1290	7	5	0.2886	4	5	0.4139
40	5	0.1195	8	1	0.3093	5	1	0.4375
41	1	0.1200	8	5	0.2928	5	5	0.4105
41	5	0.1153	9	1	0.3091	6	1	0.4347
45	1	0.0940	9	5	0.2913	6	5	0.4114
50	1	0.1226	10	1	0.3069	7	1	0.4380
60	1	0.1315	15	1	0.3070	8	1	0.4378
94	1	empty	18	1	empty	9	1	empty

Table 1: maximal persistence for choices of dimension, delay, and skip in TDE. Empty in maximal persistence means the delay is too large to obtain point-cloud data.

The End