Topological combined machine learning for consonant recognition

Yifei Zhu

Southern University of Science and Technology

2023.11.26

In 1966, Mark Kac asked the famous question:

In 1966, Mark Kac asked the famous question:

Can you hear the shape of a drum?

In 1966, Mark Kac asked the famous question:

Can you hear the shape of a drum?

To hear the shape of a drum is to infer information about the shape of the drumhead from the sound it makes, using mathematical theory.

In 1966, Mark Kac asked the famous question:

Can you hear the shape of a drum?

To hear the shape of a drum is to infer information about the shape of the drumhead from the sound it makes, using mathematical theory.

In this talk, we mirror the question across senses and address instead:

In 1966, Mark Kac asked the famous question:

Can you hear the shape of a drum?

To hear the shape of a drum is to infer information about the shape of the drumhead from the sound it makes, using mathematical theory.

In this talk, we mirror the question across senses and address instead:

Can we see the sound of a human speech?

Topological combined machine learning for consonant recognition

Yifei Zhu

Southern University of Science and Technology

2023.11.26

Periodic phenomena: a motivating example

Let $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ be the 2D torus. Consider the dynamical system given by

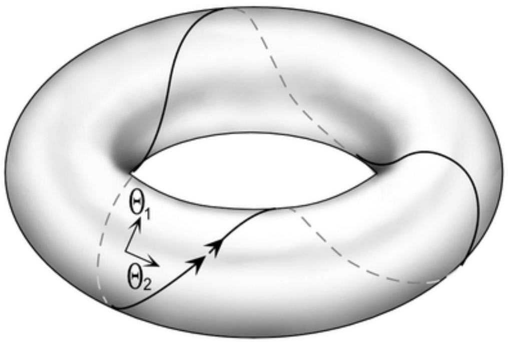
$$\begin{aligned} \Phi_\sigma \colon \mathbb{T}^2 \times \mathbb{R} &\to \mathbb{T}^2 \\ \big((a,b),t\big) &\mapsto (a+t, b+\sigma t) \end{aligned}$$

Periodic phenomena: a motivating example

Let $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ be the 2D torus. Consider the dynamical system given by

$$\begin{aligned}\Phi_\sigma \colon \mathbb{T}^2 \times \mathbb{R} &\to \mathbb{T}^2 \\ \big((a,b),t\big) &\mapsto (a+t, b+\sigma t)\end{aligned}$$

If σ is rational, then every orbit is periodic.

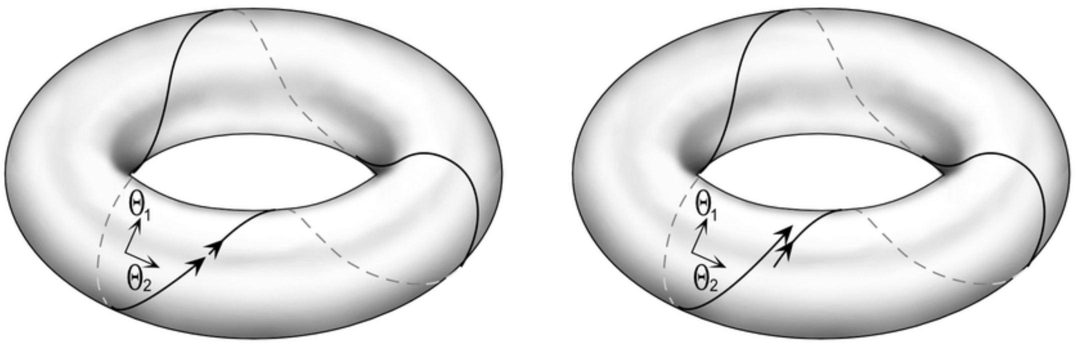


Periodic phenomena: a motivating example

Let $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ be the 2D torus. Consider the dynamical system given by

$$\Phi_\sigma \colon \mathbb{T}^2 \times \mathbb{R} \to \mathbb{T}^2$$
$$\big((a, b), t\big) \mapsto (a + t, b + \sigma t)$$

If σ is rational, then every orbit is periodic. Otherwise every orbit is dense in $\mathbb{T}^2$.



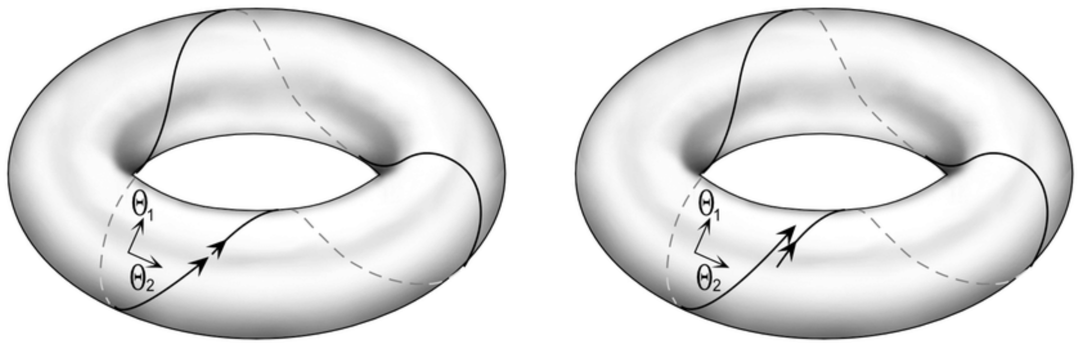
(a) (b)

Periodic phenomena: a motivating example

Let $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ be the 2D torus. Consider the dynamical system given by

$$\begin{aligned}\Phi_\sigma\colon \mathbb{T}^2 \times \mathbb{R} &\to \mathbb{T}^2 \\ \big((a,b),t\big) &\mapsto (a+t, b+\sigma t)\end{aligned}$$

If σ is rational, then every orbit is periodic. Otherwise every orbit is dense in $\mathbb{T}^2$.



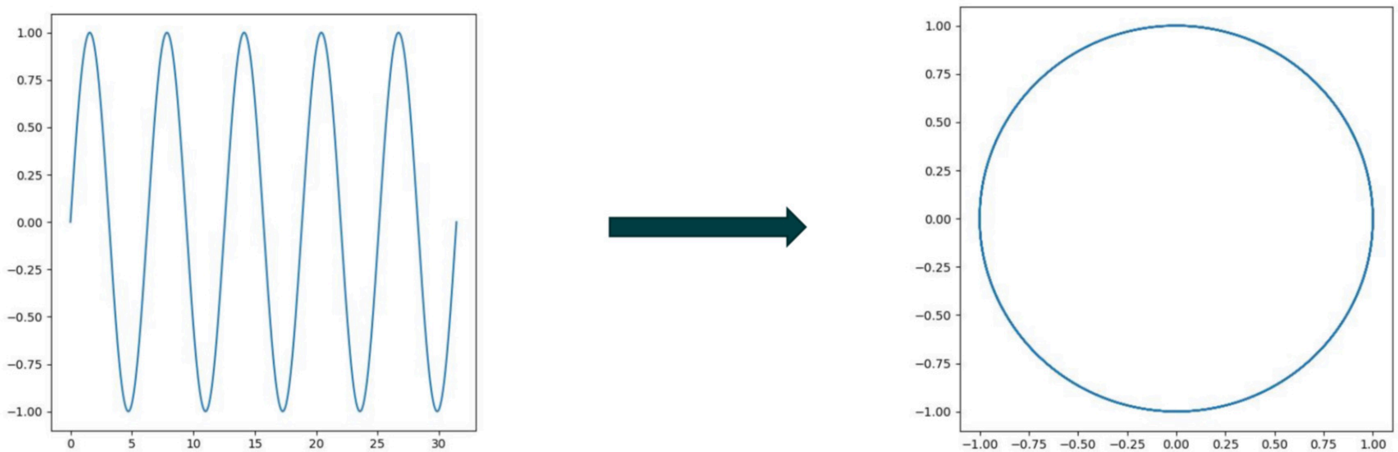
(a) (b)

From time series to topological shapes

Most periodic time series can be realized by a topological circle S^1 embedded in a Euclidean space of higher dimension.

Ideas of topological data analysis (TDA)

The topological type (more precisely, homotopy type) is robust against perturbations.



$y = \sin x$

Ideas of topological data analysis (TDA)

The topological type (more precisely, homotopy type) is robust against perturbations.

$y = \sin x$

$y = \sin x + \epsilon(x)$

Ideas of topological data analysis (TDA)

Features of topological shapes, such as the number of holes, can be captured by algebraic invariants that are computable.

Ideas of topological data analysis (TDA)

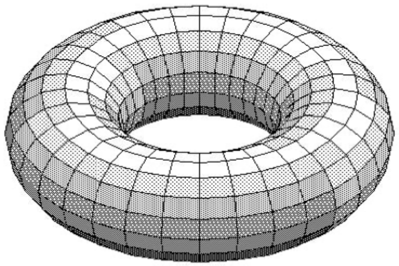
Features of topological shapes, such as the number of holes, can be captured by algebraic invariants that are computable.

Comparing these invariants effectively distinguishes the topological types of shapes.

Ideas of topological data analysis (TDA)

Features of topological shapes, such as the number of holes, can be captured by algebraic invariants that are computable.

Comparing these invariants effectively distinguishes the topological types of shapes.

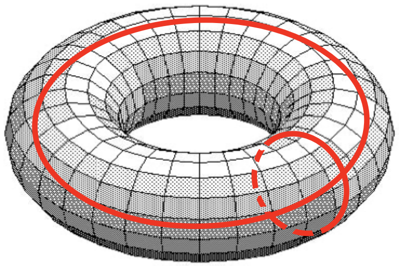


$$H_k(\mathbb{T}^2) = \begin{cases} \mathbb{Z} & k = 0 \\ & \\ & \end{cases}$$

Ideas of topological data analysis (TDA)

Features of topological shapes, such as the number of holes, can be captured by algebraic invariants that are computable.

Comparing these invariants effectively distinguishes the topological types of shapes.

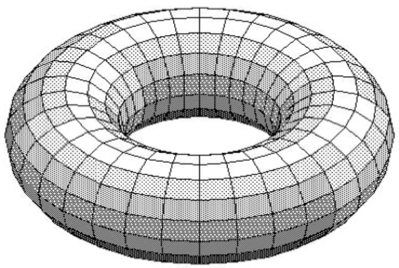


$$H_k(\mathbb{T}^2) = \begin{cases} \mathbb{Z} & k = 0 \\ \mathbb{Z} \oplus \mathbb{Z} & k = 1 \\ & \end{cases}$$

Ideas of topological data analysis (TDA)

Features of topological shapes, such as the number of holes, can be captured by algebraic invariants that are computable.

Comparing these invariants effectively distinguishes the topological types of shapes.

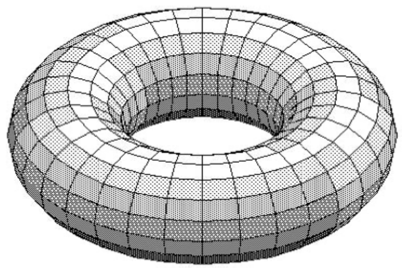


$$H_k(\mathbb{T}^2) = \begin{cases} \mathbb{Z} & k = 0 \\ \mathbb{Z} \oplus \mathbb{Z} & k = 1 \\ \mathbb{Z} & k = 2 \end{cases}$$

Ideas of topological data analysis (TDA)

Features of topological shapes, such as the number of holes, can be captured by algebraic invariants that are computable.

Comparing these invariants effectively distinguishes the topological types of shapes.

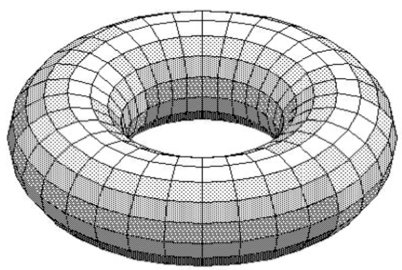


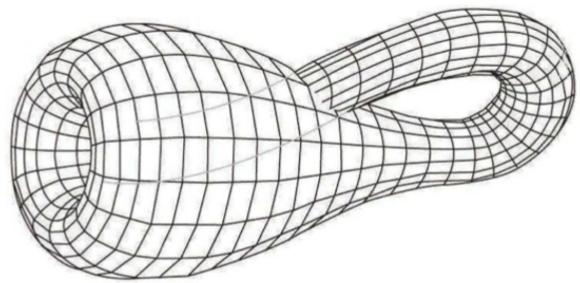
$$H_k(\mathbb{T}^2) = \begin{cases} \mathbb{Z} & k = 0 \\ \mathbb{Z} \oplus \mathbb{Z} & k = 1 \\ \mathbb{Z} & k = 2 \\ 0 & k > 2 \end{cases}$$

Ideas of topological data analysis (TDA)

Features of topological shapes, such as the number of holes, can be captured by algebraic invariants that are computable.

Comparing these invariants effectively distinguishes the topological types of shapes.





$$H_k(\mathbb{T}^2) = \begin{cases} \mathbb{Z} & k = 0 \\ \mathbb{Z} \oplus \mathbb{Z} & k = 1 \\ \mathbb{Z} & k = 2 \\ 0 & k > 2 \end{cases}$$

$$H_k(\text{Klein bottle}) = \begin{cases} \mathbb{Z} & k = 0 \\ \mathbb{Z} \oplus \mathbb{Z}/2 & k = 1 \\ 0 & k = 2 \\ 0 & k > 2 \end{cases}$$

Topological time series analysis

Let us make the assumption that sampled signals are distributed over a manifold.

Topological time series analysis

Let us make the assumption that sampled signals are distributed over a manifold. To topologically analyze time series, we then proceed as follows:

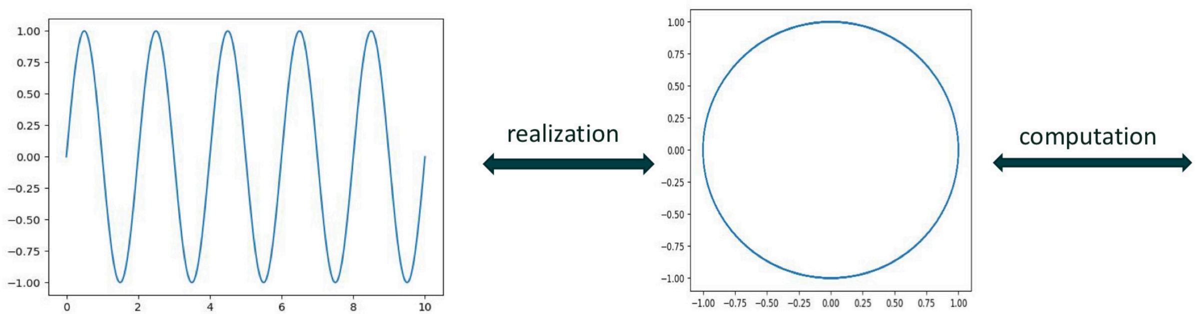
<u>Step 1</u> Embed the data into a Euclidean space of suitable dimension;

Topological time series analysis

Let us make the assumption that sampled signals are distributed over a manifold. To topologically analyze time series, we then proceed as follows:

Step 1 Embed the data into a Euclidean space of suitable dimension;

Step 2 Compute the algebraic invariants for statistical inference.



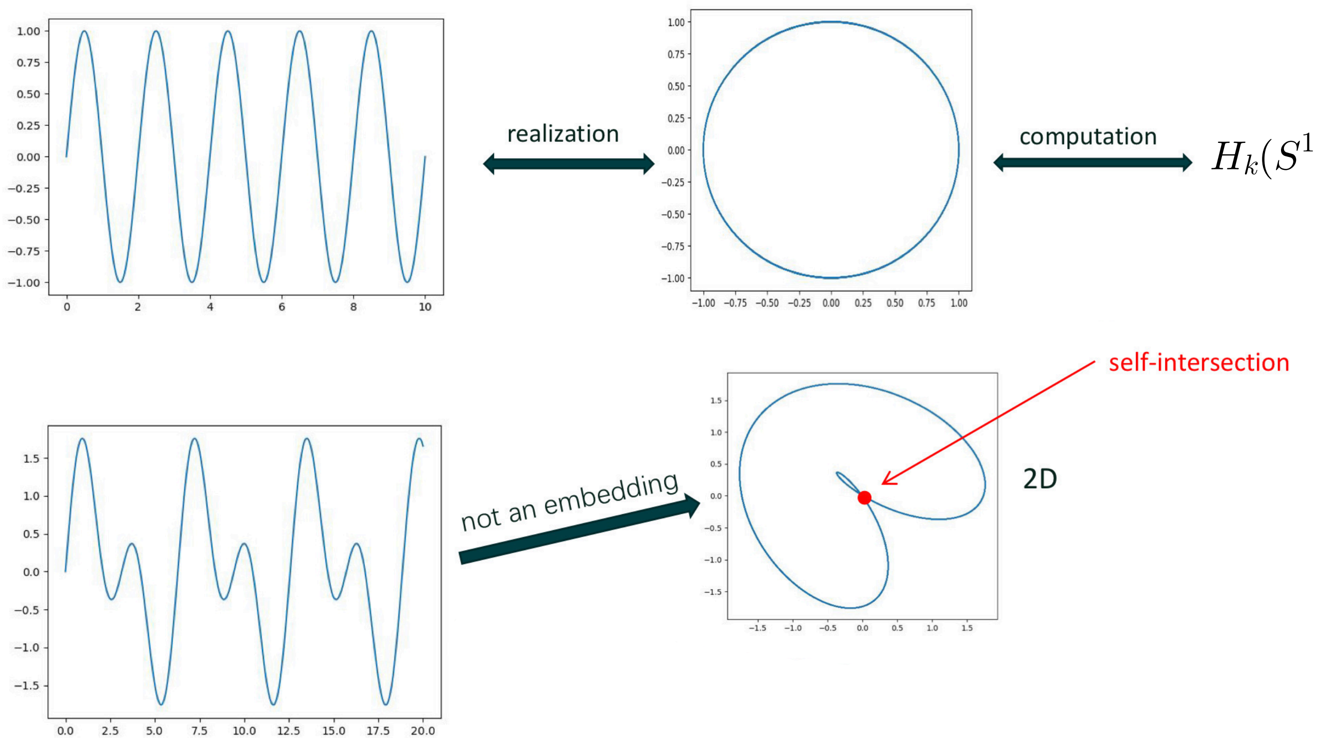
$$H_k(S^1) = \begin{cases} \mathbb{Z} & k = 0 \\ \mathbb{Z} & k = 1 \\ 0 & k > 1 \end{cases}$$

Topological time series analysis

Let us make the assumption that sampled signals are distributed over a manifold. To topologically analyze time series, we then proceed as follows:

<u>Step 1</u> Embed the data into a Euclidean space of suitable dimension;

<u>Step 2</u> Compute the algebraic invariants for statistical inference.



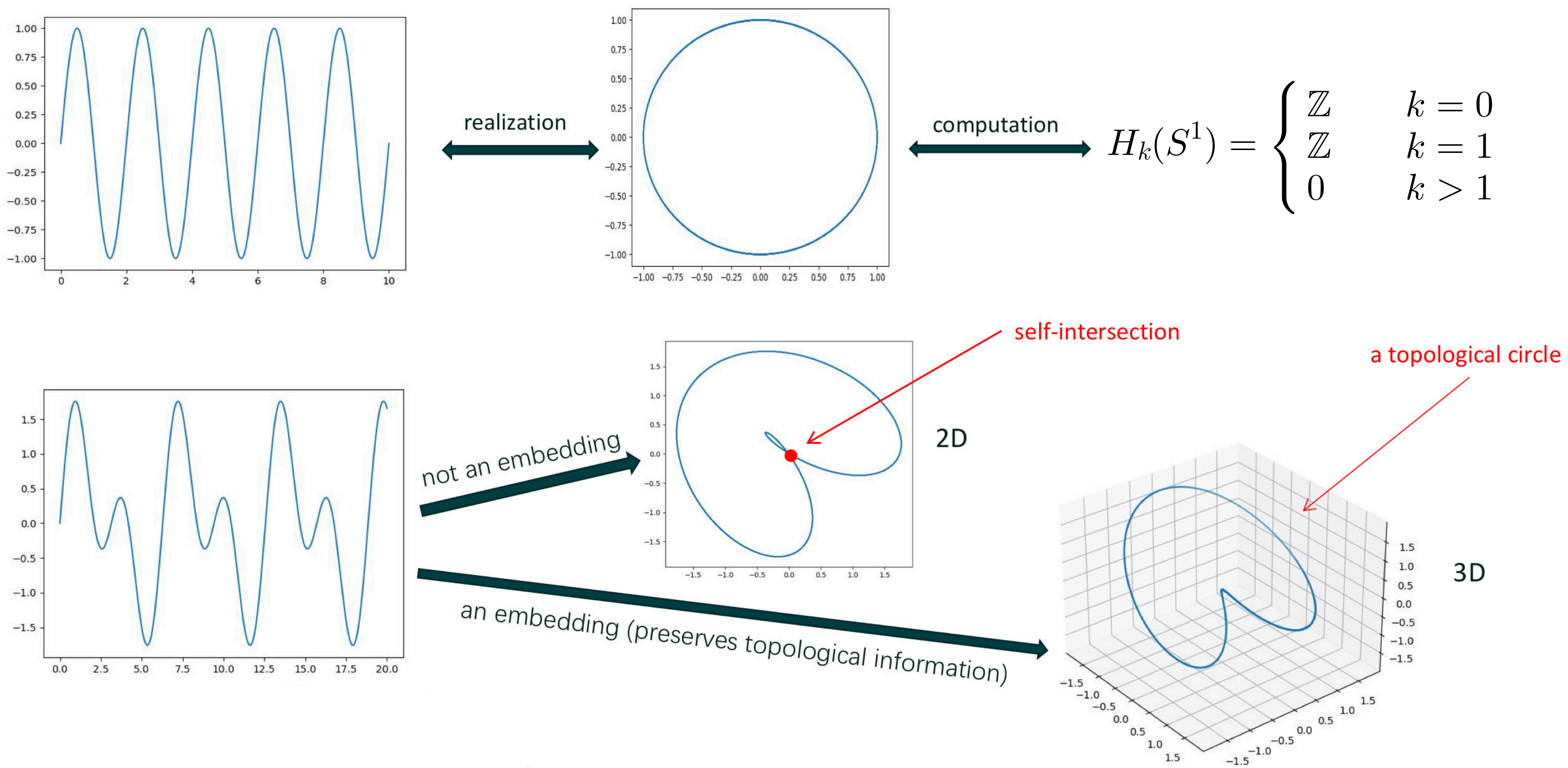
$$H_k(S^1) = \begin{cases} \mathbb{Z} & k = 0 \\ \mathbb{Z} & k = 1 \\ 0 & k > 1 \end{cases}$$

Topological time series analysis

Let us make the assumption that sampled signals are distributed over a manifold. To topologically analyze time series, we then proceed as follows:

<u>Step 1</u> Embed the data into a Euclidean space of suitable dimension;

<u>Step 2</u> Compute the algebraic invariants for statistical inference.



An application: detection of wheeze in medical science (pulmonology)

Wheezes are abnormal lung sounds and usually imply obstructive airway diseases.

An application: detection of wheeze in medical science (pulmonology)

Wheezes are abnormal lung sounds and usually imply obstructive airway diseases.

The most important characteristic of wheeze signals is their periodic patterns.

.

An application: detection of wheeze in medical science (pulmonology)

Wheezes are abnormal lung sounds and usually imply obstructive airway diseases.

The most important characteristic of wheeze signals is their periodic patterns.

The accuracy of topological periodicity detection is 98.39% (Emrani et al., IEEE Signal Processing Letters, 2014), while in two earlier papers with different methods they are 86.2% and 95.5%.

An application: detection of wheeze in medical science (pulmonology)

Wheezes are abnormal lung sounds and usually imply obstructive airway diseases.

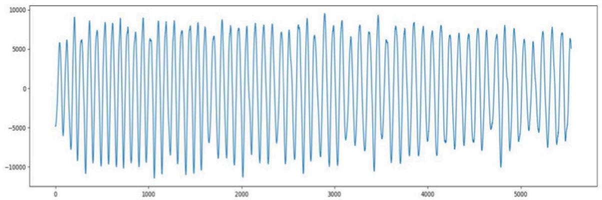
The most important characteristic of wheeze signals is their periodic patterns.

The accuracy of topological periodicity detection is 98.39% (Emrani et al., IEEE Signal Processing Letters, 2014), while in two earlier papers with different methods they are 86.2% and 95.5%.

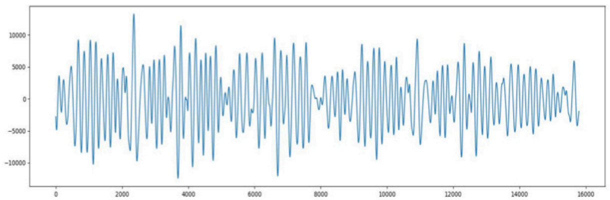
As a warm-up, our research group (Siheng Yi) has reproduced their results using the original data and open-source TDA programming package.

An application: detection of wheeze in medical science (pulmonology)

wheeze



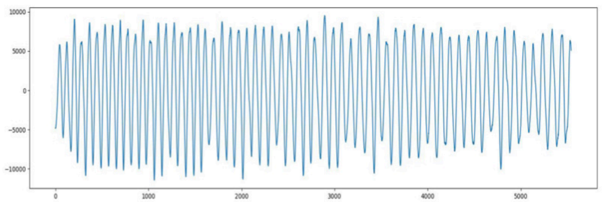
normal

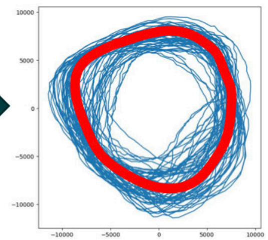


Original sound signals

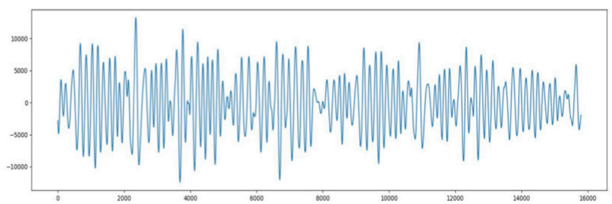
An application: detection of wheeze in medical science (pulmonology)

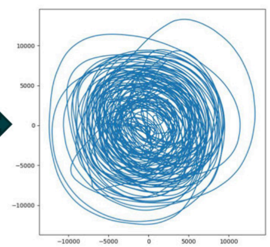
wheeze





normal





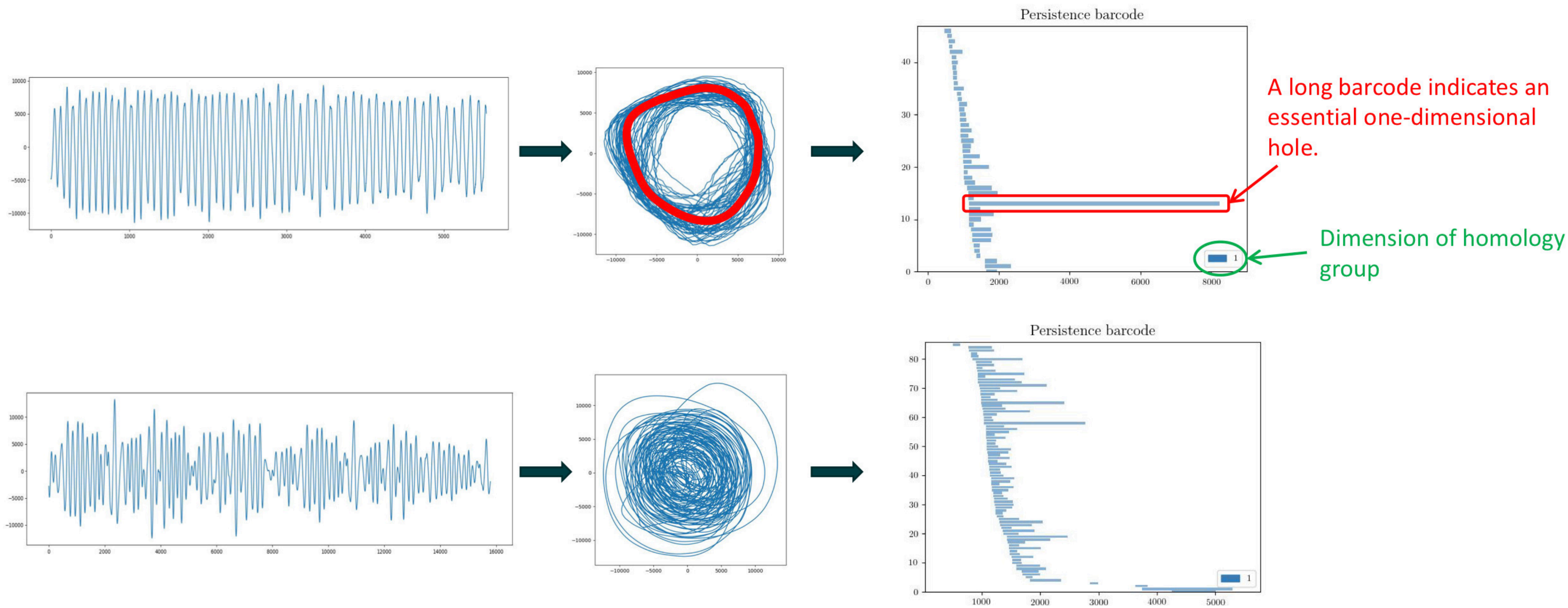
Original sound signals

Realized topological shapes embedded in 2D Euclidean space

An application: detection of wheeze in medical science (pulmonology)

wheeze

normal



Original sound signals

Realized topological shapes embedded in 2D Euclidean space

“Persistence barcodes” as representations of the algebraic invariant (1D homology group)

A pipeline for topological time series analysis

Time series data: $x_1, x_2, x_3, x_4, \ldots$

A pipeline for topological time series analysis

Time series data: $x_1, x_2, x_3, x_4, \ldots$

Preprocessing → Sliding window embedding

2D Euclidean shape: point cloud $(x_1, x_{1+\tau}), (x_2, x_{2+\tau}), (x_3, x_{3+\tau}), \ldots$

delay

A pipeline for topological time series analysis

Time series data: $x_1, x_2, x_3, x_4, \ldots$

Preprocessing → Sliding window embedding

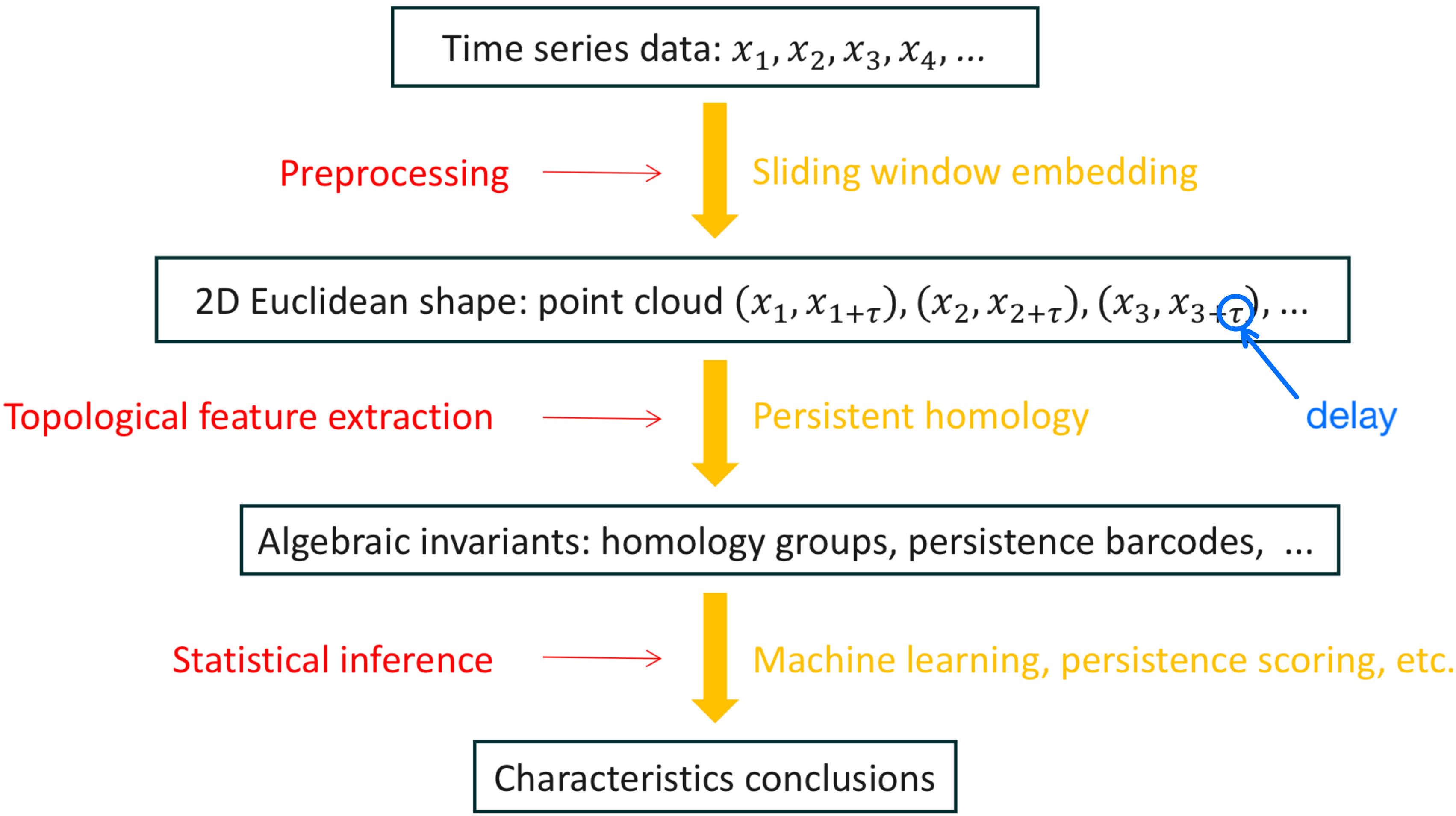
2D Euclidean shape: point cloud $(x_1, x_{1+\tau}), (x_2, x_{2+\tau}), (x_3, x_{3+\tau}), \ldots$

delay

Topological feature extraction → Persistent homology

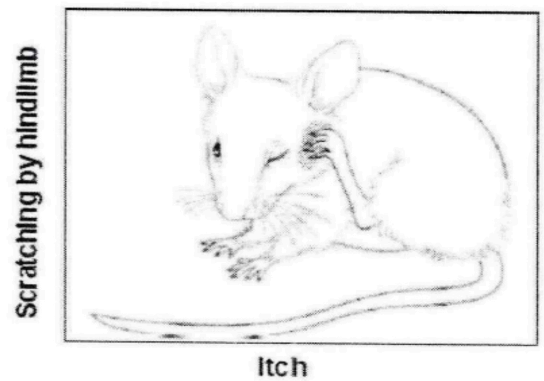
Algebraic invariants: homology groups, persistence barcodes, ...

A pipeline for topological time series analysis



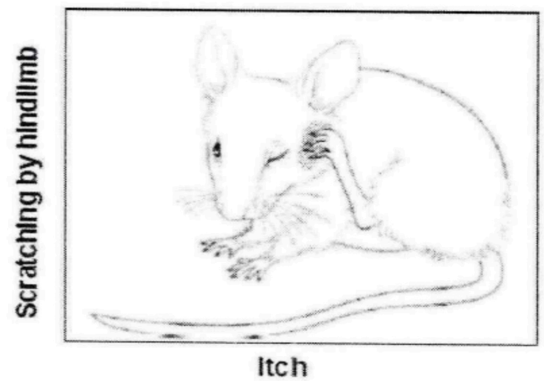
Application I: detection of mouse scratching behavior

Joint with the biomedical engineering group led by Fangyi Chen and the data science group led by Zhen Zhang, both at SUSTech, we applied topological methods to the problem of automated and real-time detection of mouse scratching behavior, with motivations from pharmacology.

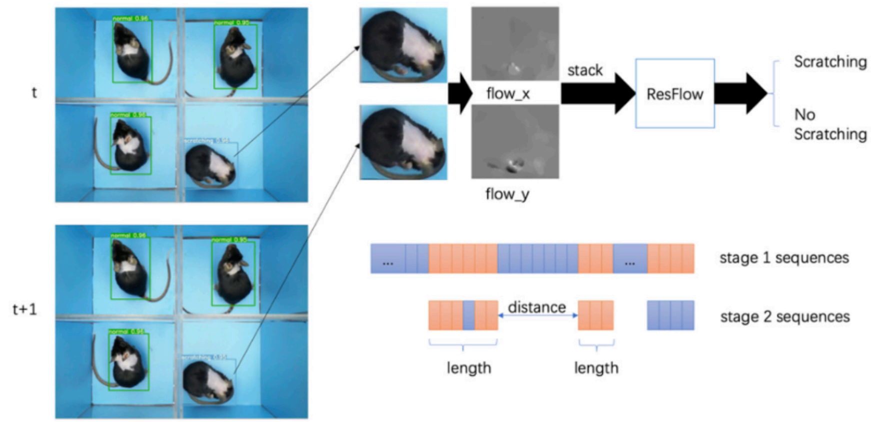


Application I: detection of mouse scratching behavior

Joint with the biomedical engineering group led by Fangyi Chen and the data science group led by Zhen Zhang, both at SUSTech, we applied topological methods to the problem of automated and real-time detection of mouse scratching behavior, with motivations from pharmacology.

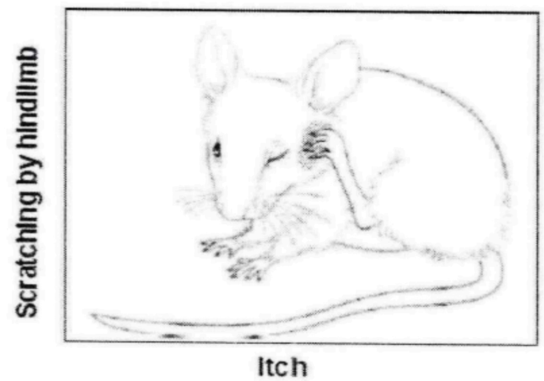


Prior to our group's involvement, machine learning via neural networks was applied with satisfactory accuracy (https://yifeizhu.github.io/scratch.mp4).

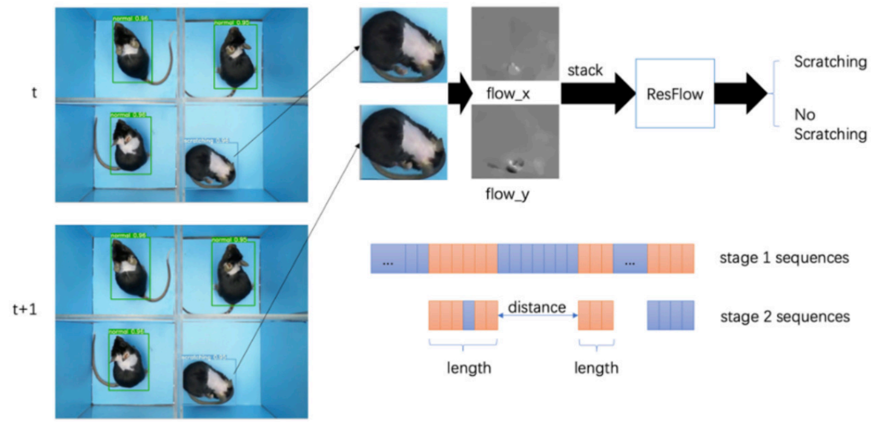


Application I: detection of mouse scratching behavior

Joint with the biomedical engineering group led by Fangyi Chen and the data science group led by Zhen Zhang, both at SUSTech, we applied topological methods to the problem of automated and real-time detection of mouse scratching behavior, with motivations from pharmacology.



Prior to our group's involvement, machine learning via neural networks was applied with satisfactory accuracy (https://yifeizhu.github.io/scratch.mp4).



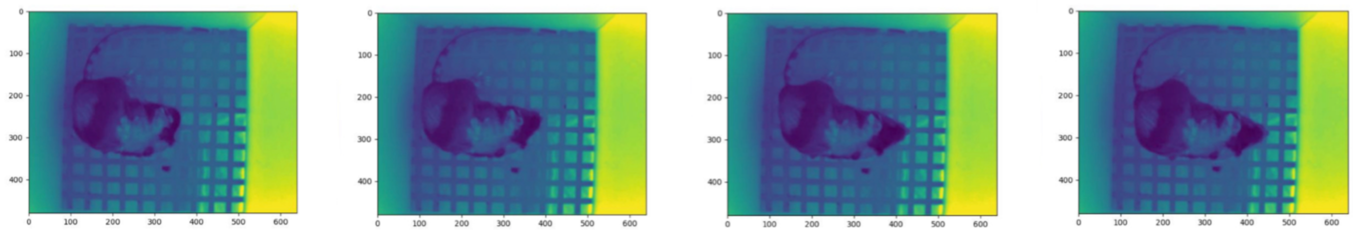
However, the learning process was time consuming, which is impractical for time-sensitive purposes and lab efficiency.

Application I: detection of mouse scratching behavior

We observed that the scratching behavior exhibits periodicity.

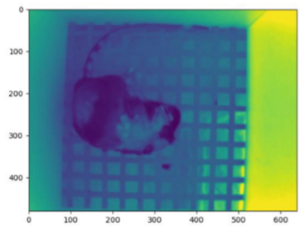
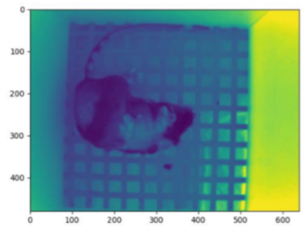
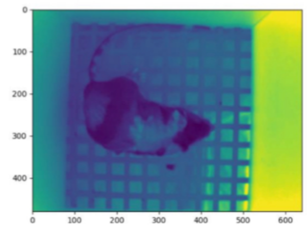
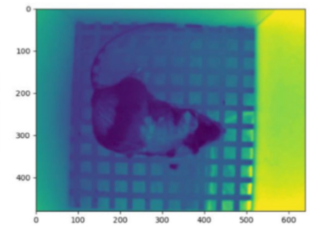
Application I: detection of mouse scratching behavior

We observed that the scratching behavior exhibits periodicity. In the meantime, however, global movements of a mouse may significantly reduce the pattern.



Application I: detection of mouse scratching behavior

We observed that the scratching behavior exhibits periodicity. In the meantime, however, global movements of a mouse may significantly reduce the pattern.

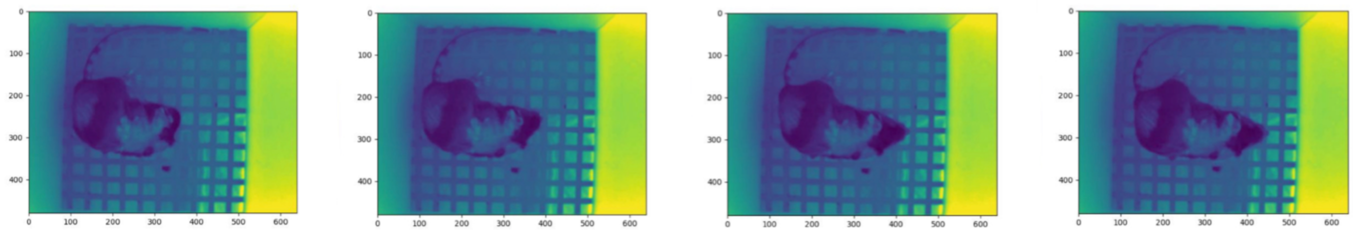





To resolve this issue, we adopted the following approaches:

Approach 1 Sum up all 460 x 640 pixels to extract a series of 1D data which ignores differences caused by global movements. Too coarse?

Application I: detection of mouse scratching behavior

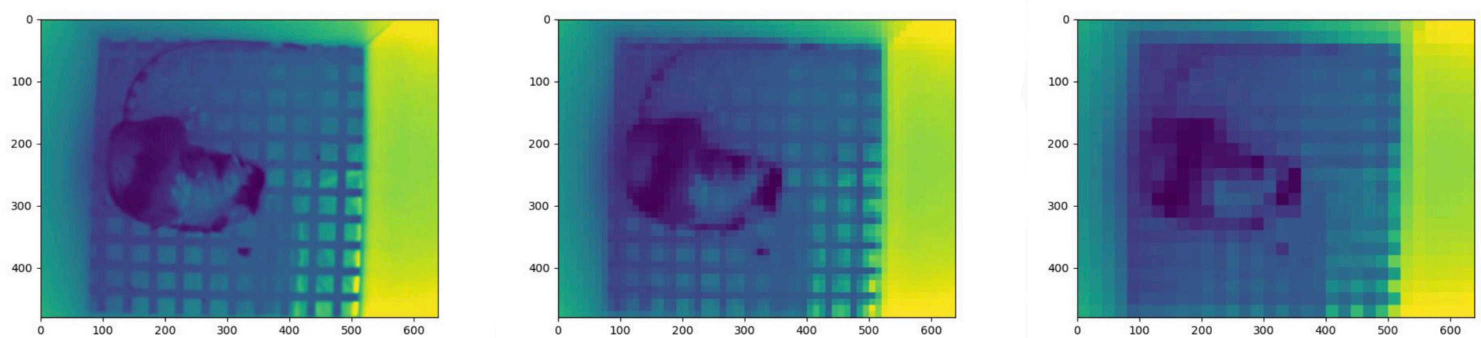
We observed that the scratching behavior exhibits periodicity. In the meantime, however, global movements of a mouse may significantly reduce the pattern.



To resolve this issue, we adopted the following approaches:

Approach 1 Sum up all 460 x 640 pixels to extract a series of 1D data which ignores differences caused by global movements. Too coarse?

Approach 2 Blur the images by pooling, and feed the topological pipeline with reduced 100-dimensional data. Still too refined?

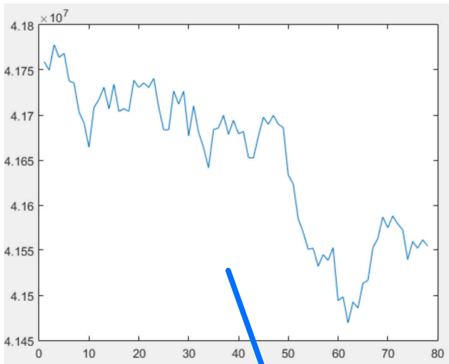


Application I: detection of mouse scratching behavior

Approach 1 (1D data, Qingrui Qu), combined with carefully designed filtration for wave signals + suitably chosen geometric statistics, yielded a close-to-real-time, decently accurate detection performance.

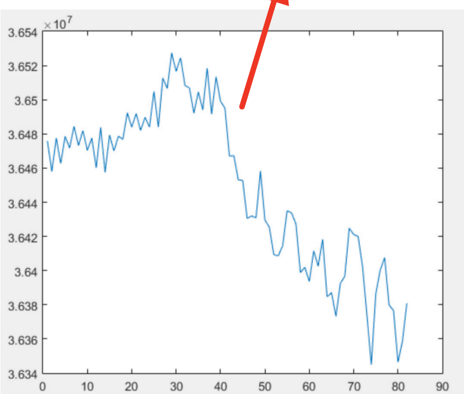
Application I: detection of mouse scratching behavior

Approach 1 (1D data, Qingrui Qu), combined with carefully designed filtration for wave signals + suitably chosen geometric statistics, yielded a close-to-real-time, decently accurate detection performance.



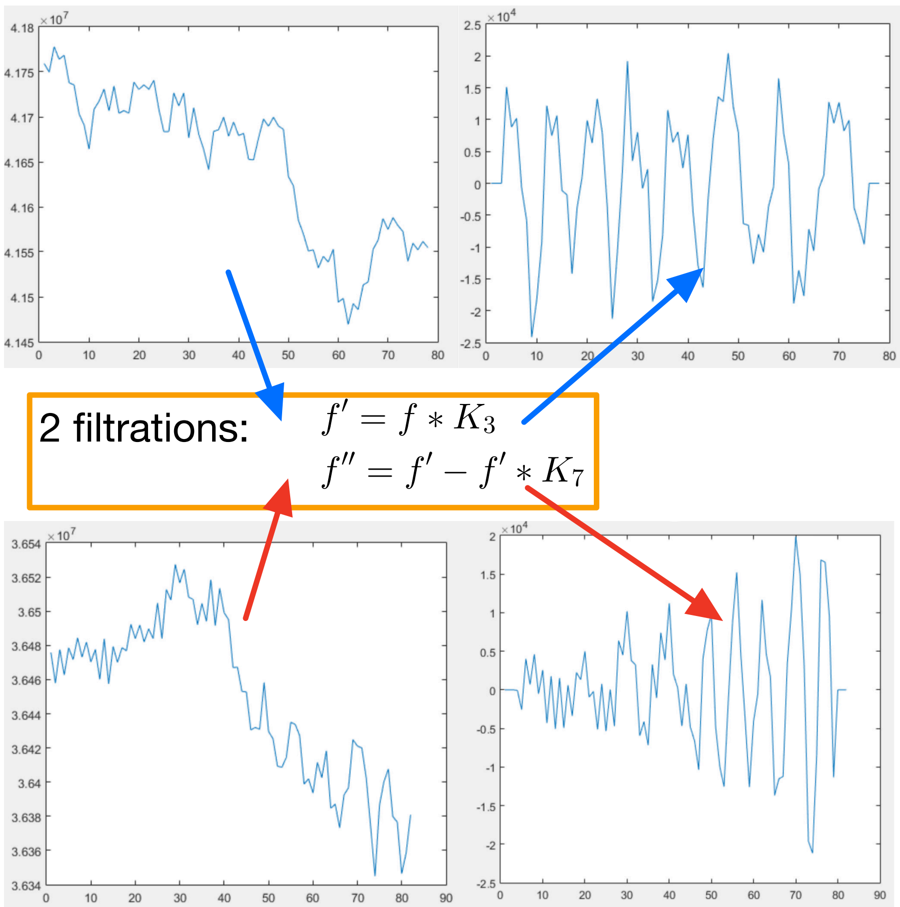
2 filtrations:

$$f' = f * K_3$$
$$f'' = f' - f' * K_7$$



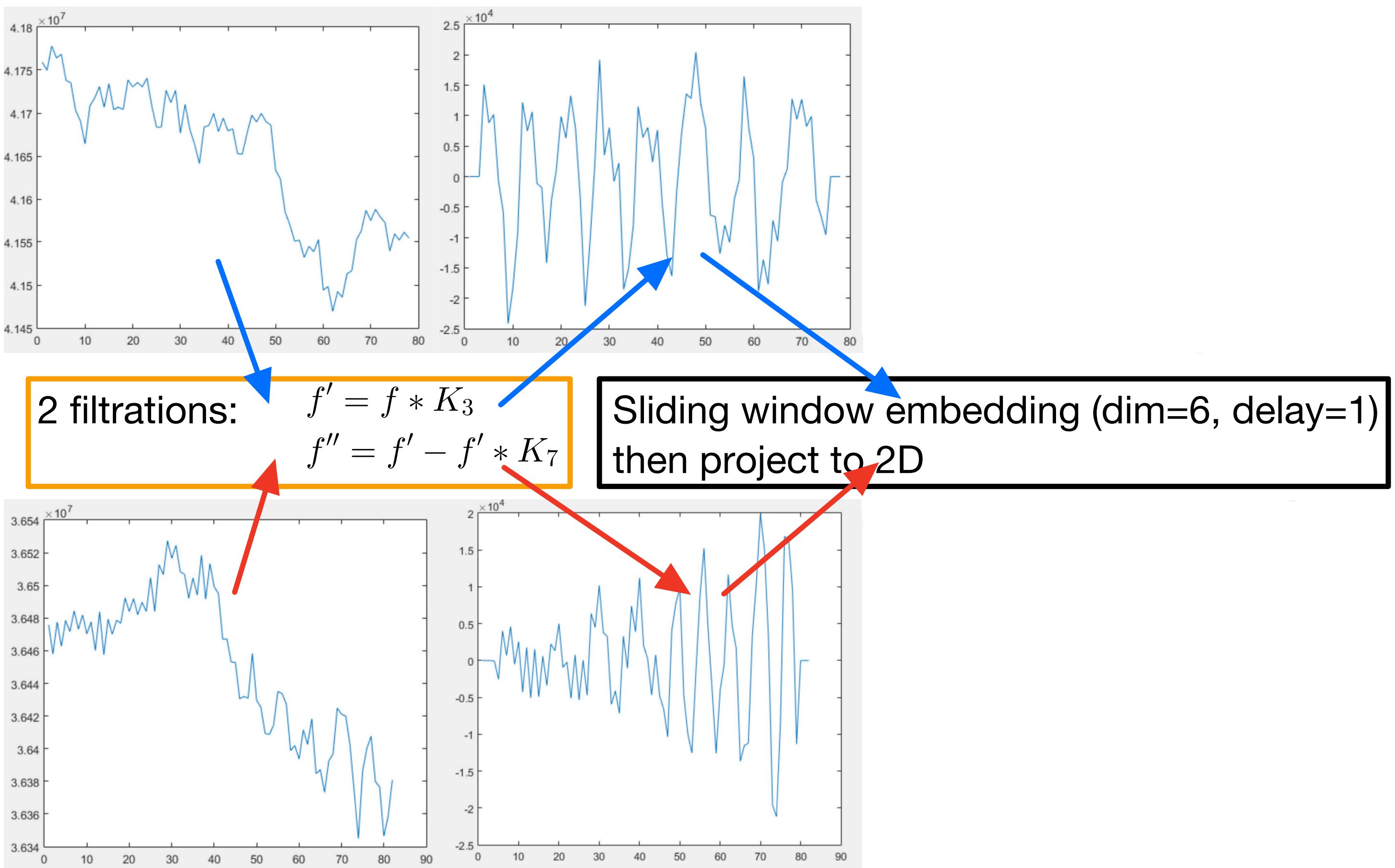
Application I: detection of mouse scratching behavior

Approach 1 (1D data, Qingrui Qu), combined with carefully designed filtration for wave signals + suitably chosen geometric statistics, yielded a close-to-real-time, decently accurate detection performance.



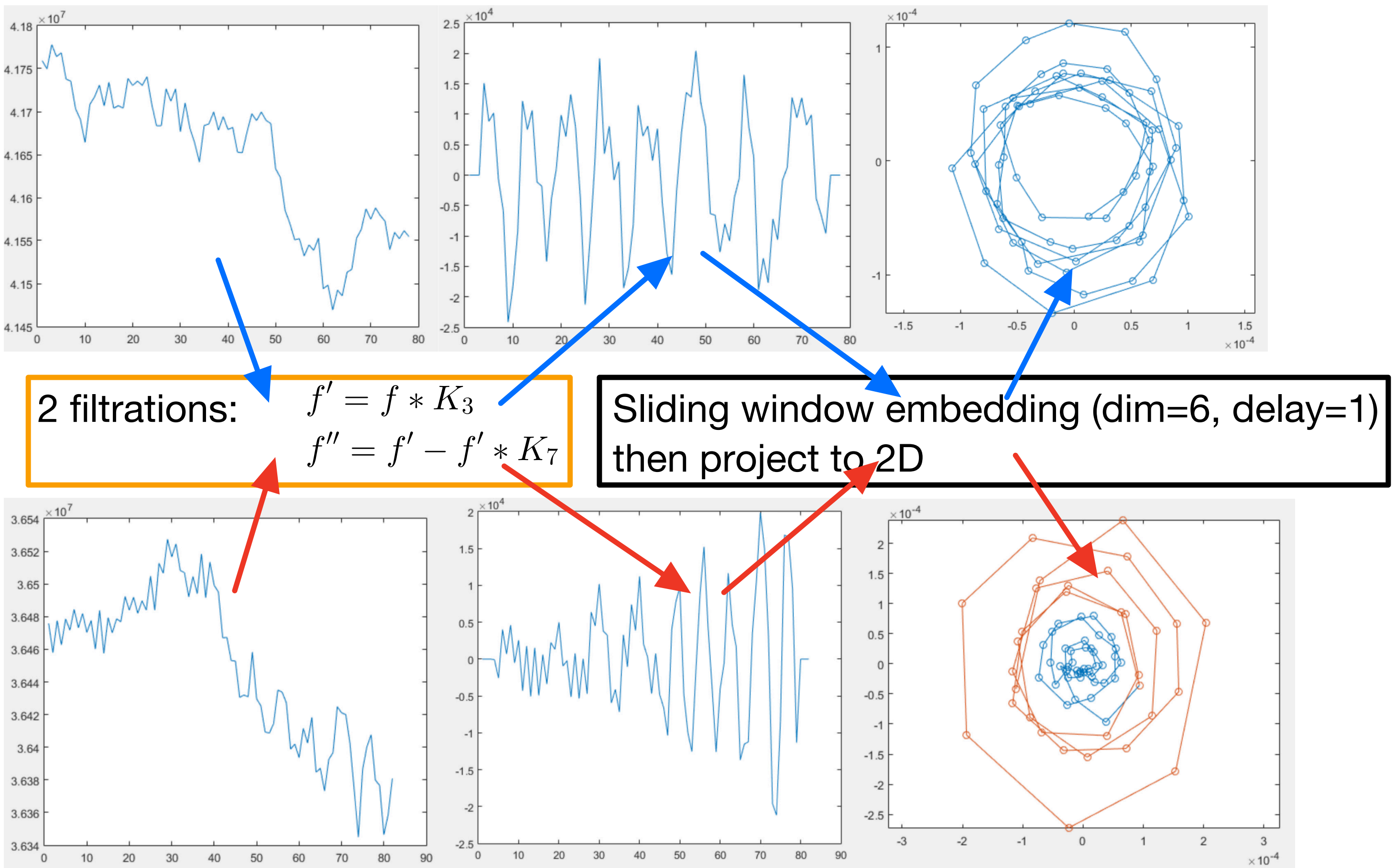
Application I: detection of mouse scratching behavior

Approach 1 (1D data, Qingrui Qu), combined with carefully designed filtration for wave signals + suitably chosen geometric statistics, yielded a close-to-real-time, decently accurate detection performance.



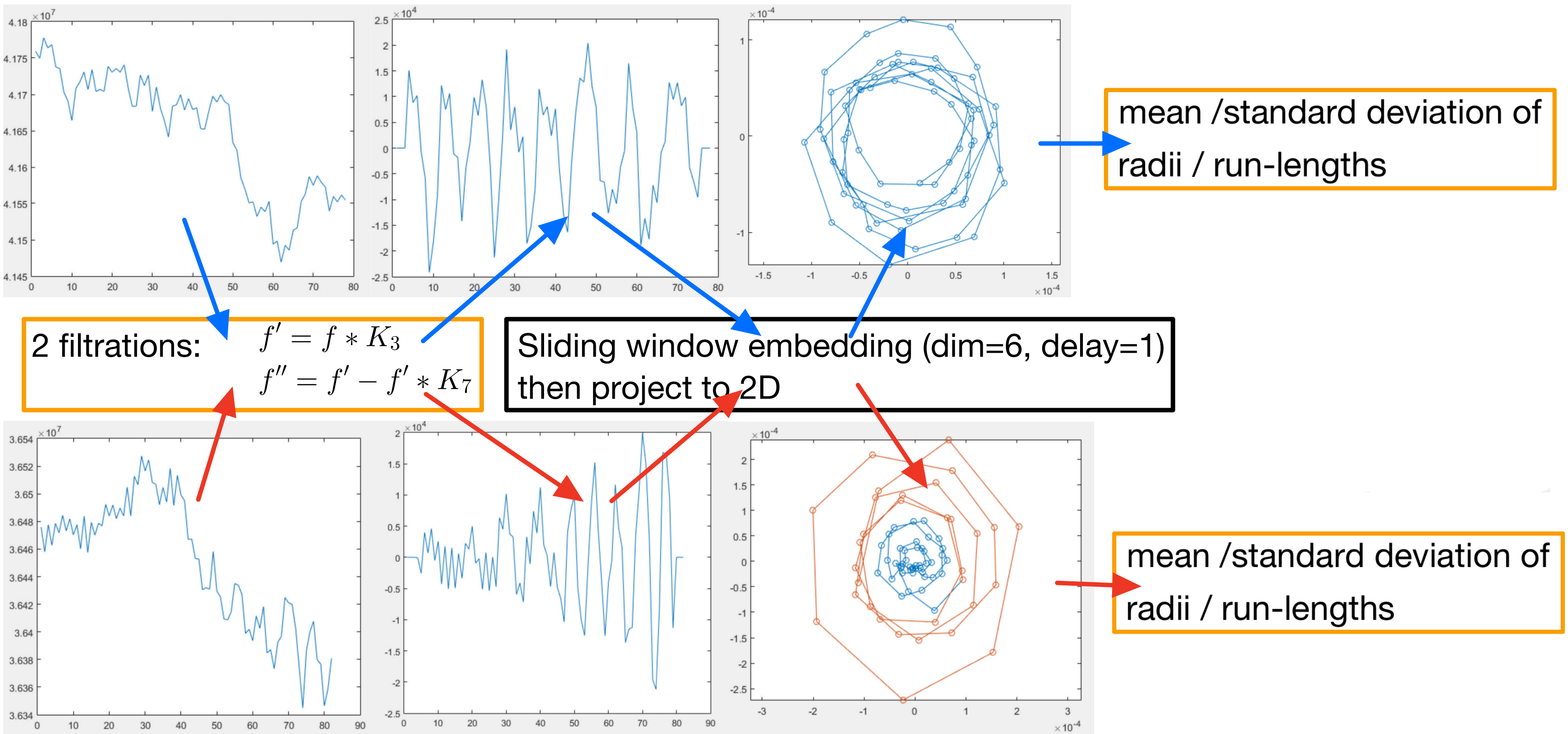
Application I: detection of mouse scratching behavior

Approach 1 (1D data, Qingrui Qu), combined with carefully designed filtration for wave signals + suitably chosen geometric statistics, yielded a close-to-real-time, decently accurate detection performance.



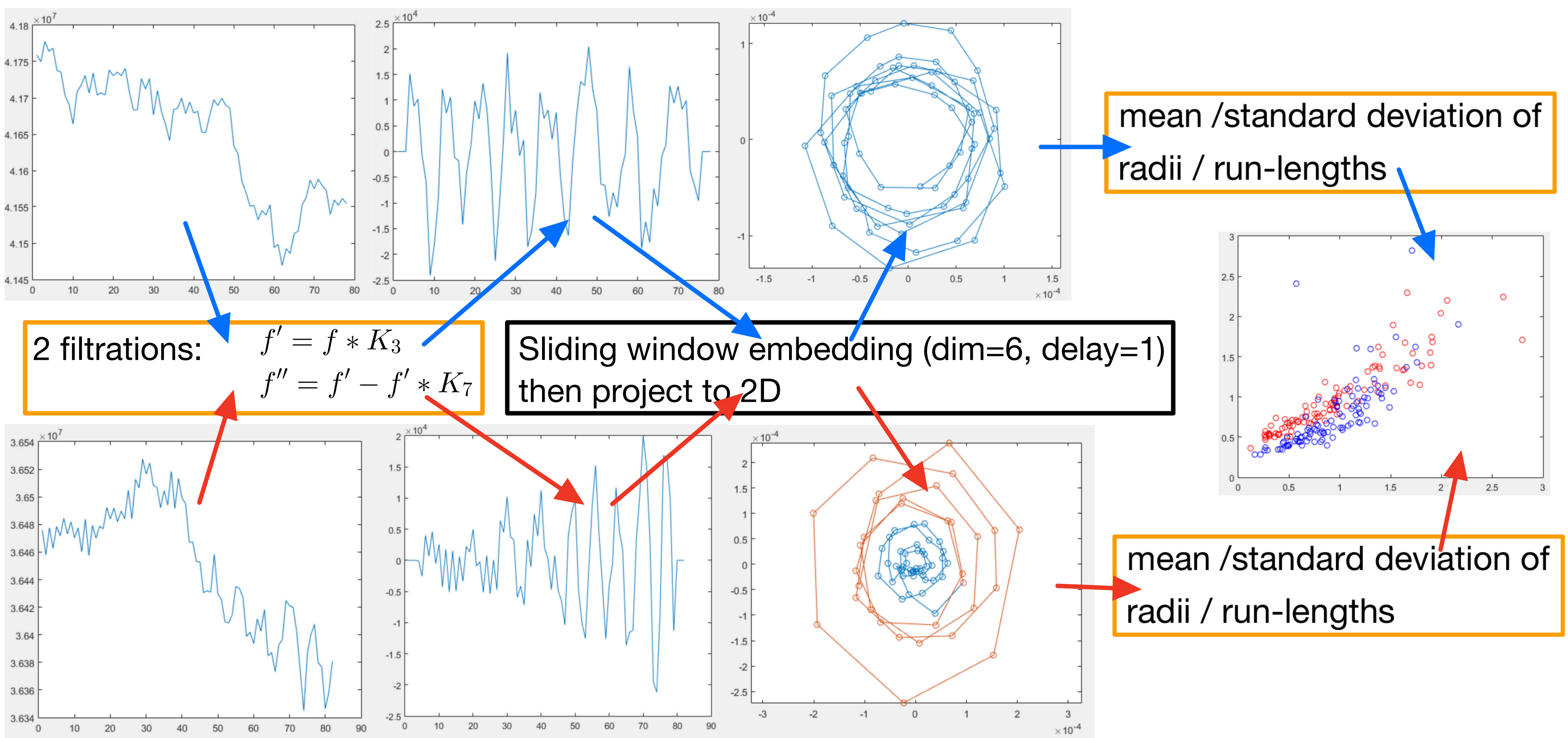
Application I: detection of mouse scratching behavior

Approach 1 (1D data, Qingrui Qu), combined with carefully designed filtration for wave signals + suitably chosen geometric statistics, yielded a close-to-real-time, decently accurate detection performance.



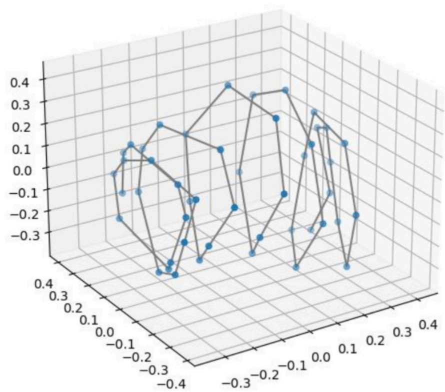
Application I: detection of mouse scratching behavior

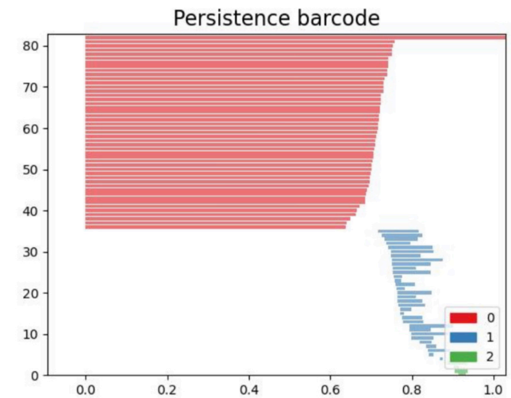
Approach 1 (1D data, Qingrui Qu), combined with carefully designed filtration for wave signals + suitably chosen geometric statistics, yielded a close-to-real-time, decently accurate detection performance.

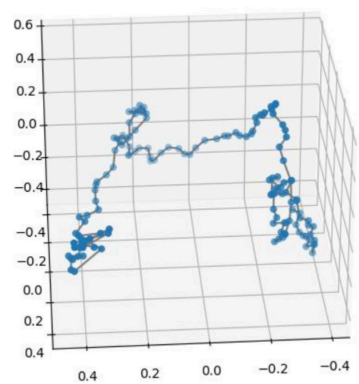


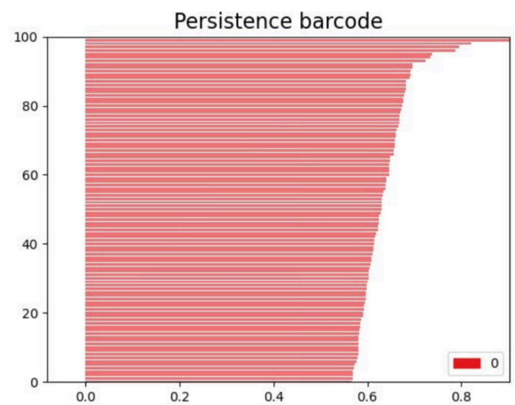
Application I: detection of mouse scratching behavior

Approach 2 (multi-dimensional data, Siheng Yi), combined with persistent homology and its representations, yielded recognizable characteristics but required considerable computational time.









Application II: classification of speech signals

Joint with Meng Yu of Tencent AI Lab, we applied topological methods to classify voiced/voiceless and vowel/consonant speech data, with motivations from industrial applications.

Application II: classification of speech signals

Joint with Meng Yu of Tencent AI Lab, we applied topological methods to classify voiced/voiceless and vowel/consonant speech data, with motivations from industrial applications.

We were inspired by Carlsson et al.'s discovery of the Klein-bottle distribution of local natural images, as well as their subsequent recent work of topological convolutional neural networks learning video data.

Application II: classification of speech signals

Joint with Meng Yu of Tencent AI Lab, we applied topological methods to classify voiced/voiceless and vowel/consonant speech data, with motivations from industrial applications.

We were inspired by Carlsson et al.'s discovery of the Klein-bottle distribution of local natural images, as well as their subsequent recent work of topological convolutional neural networks learning video data. We would like to understand an analogous distribution space for speech data and how its input may enable smarter learning.

Application II: classification of speech signals

Joint with Meng Yu of Tencent AI Lab, we applied topological methods to classify voiced/voiceless and vowel/consonant speech data, with motivations from industrial applications.

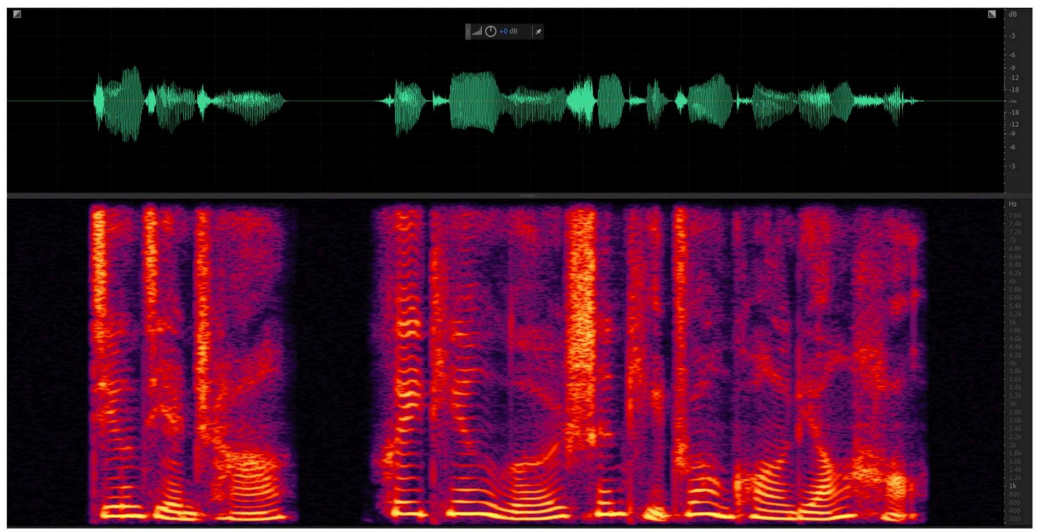
We were inspired by Carlsson et al.'s discovery of the Klein-bottle distribution of local natural images, as well as their subsequent recent work of topological convolutional neural networks learning video data. We would like to understand an analogous distribution space for speech data and how its input may enable smarter learning.

Display of speech signals

There are speech signal processing softwares for professional use.

Time

Time-Frequency



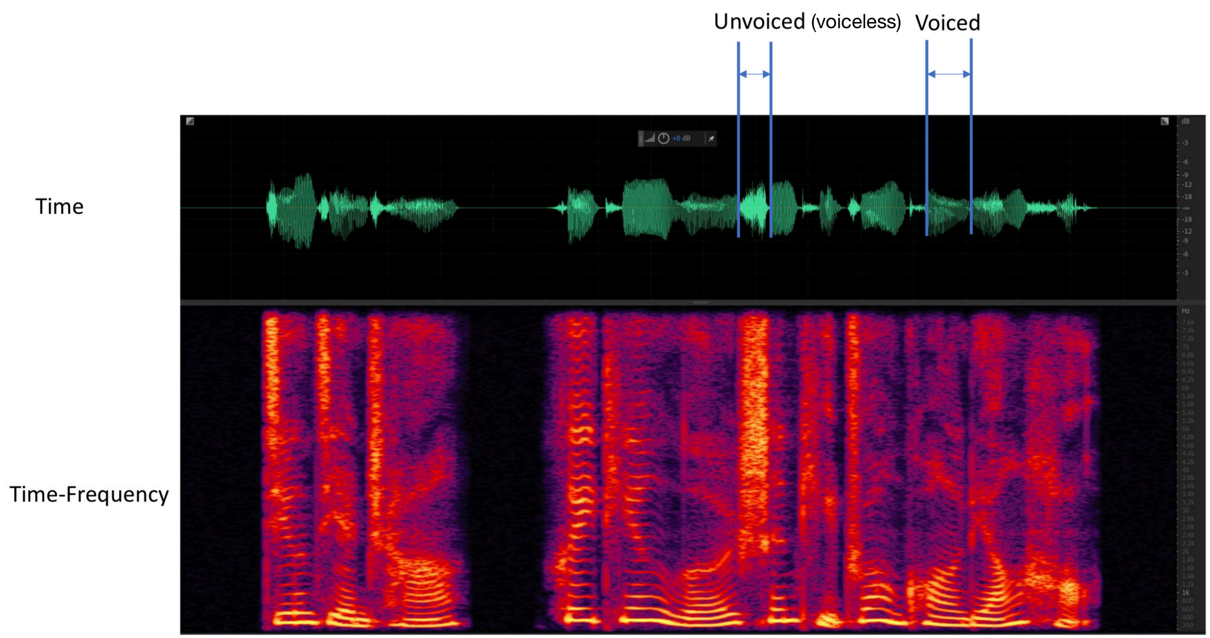
Application II: classification of speech signals

Joint with Meng Yu of Tencent AI Lab, we applied topological methods to classify voiced/voiceless and vowel/consonant speech data, with motivations from industrial applications.

We were inspired by Carlsson et al.'s discovery of the Klein-bottle distribution of local natural images, as well as their subsequent recent work of topological convolutional neural networks learning video data. We would like to understand an analogous distribution space for speech data and how its input may enable smarter learning.

Display of speech signals

There are speech signal processing softwares for professional use.

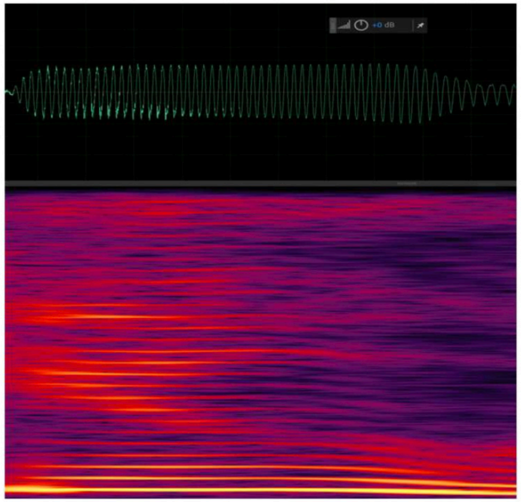


Application II: classification of speech signals

Voiced

Sinusoid in time domain

Harmonics in frequency domain



Time and Time-Frequency domain

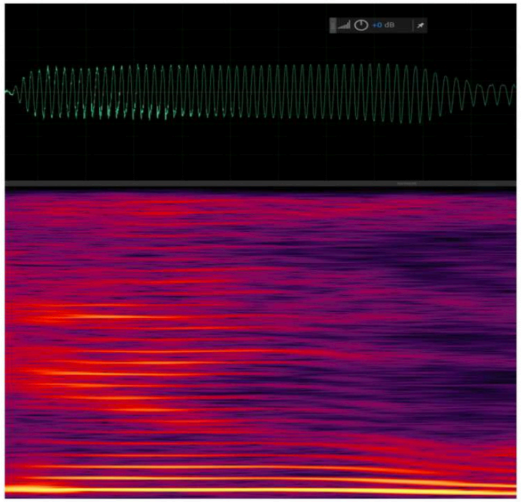
Frequency response

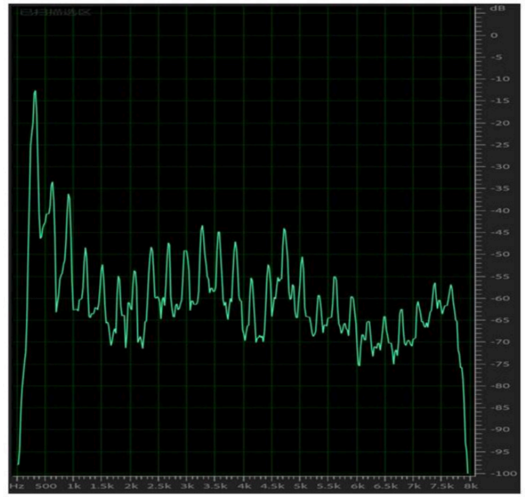
Application II: classification of speech signals

Voiced

Sinusoid in time domain

Harmonics in frequency domain



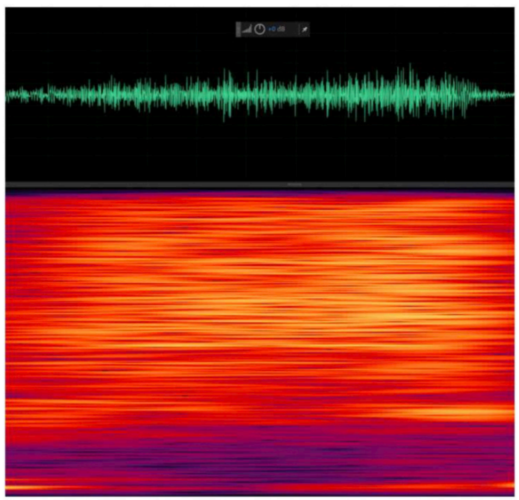


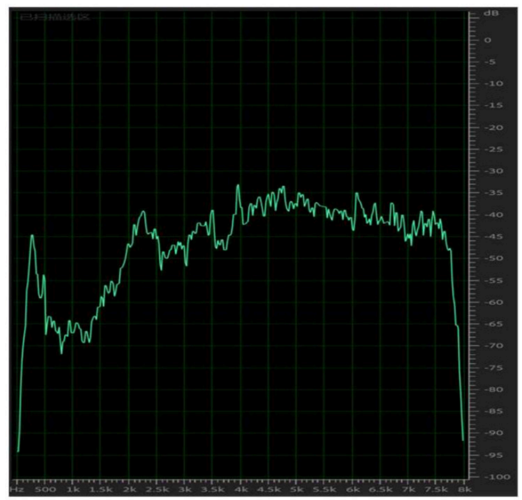
Time and Time-Frequency domain

Frequency response

Voiceless

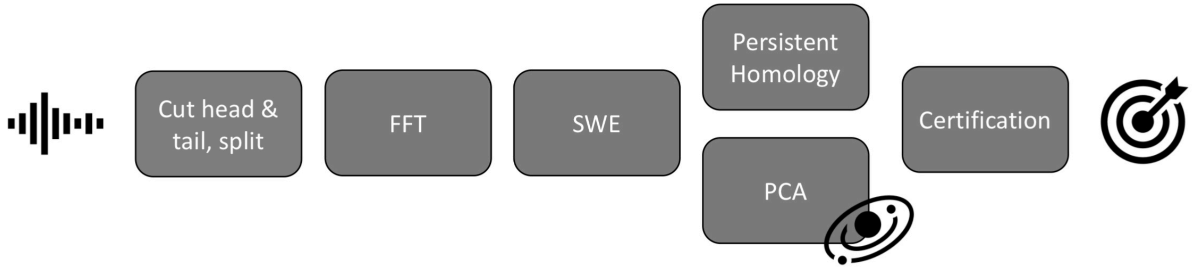
Like a white noise





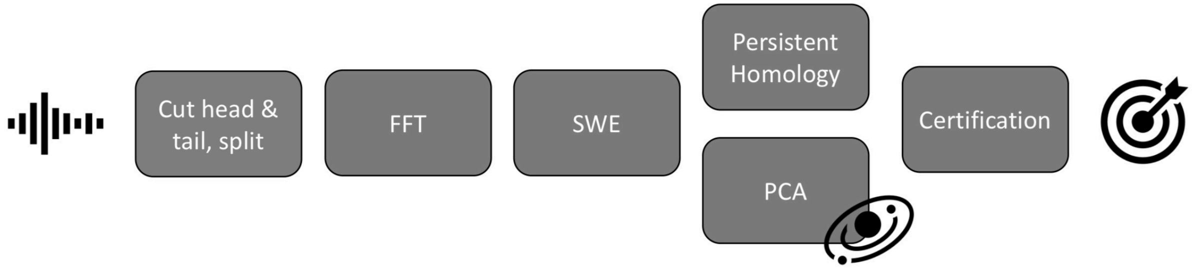
Application II: classification of speech signals

Here is a flowchart for our topological approach:

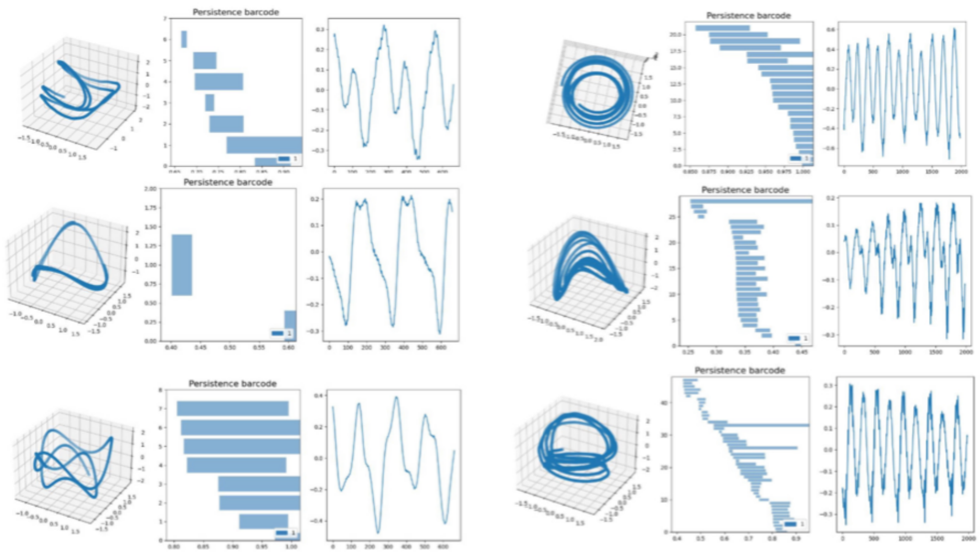


Application II: classification of speech signals

Here is a flowchart for our topological approach:



Topological profiles for vowels and consonants (Pingyao Feng)

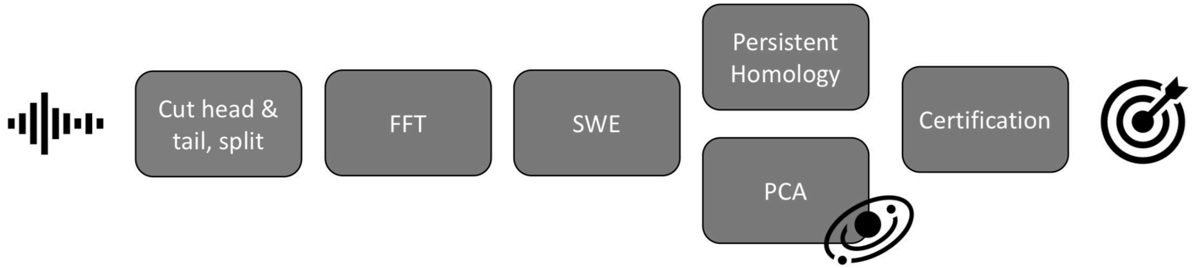


Features for vowels

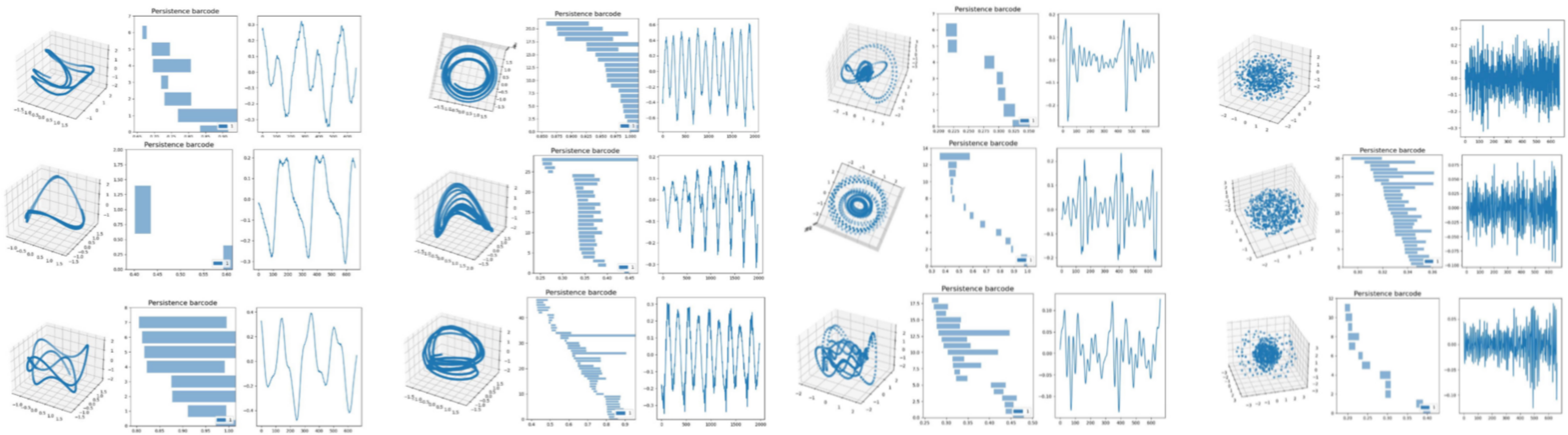
Left: frame size: 15ms, frame shift: 5ms; Right: frame size: 45ms, frame shift: 22.5ms

Application II: classification of speech signals

Here is a flowchart for our topological approach:



Topological profiles for vowels and consonants (Pingyao Feng)



Features for vowels

Features for consonants

Left: frame size: 15ms, frame shift: 5ms; Right: frame size: 45ms, frame shift: 22.5ms

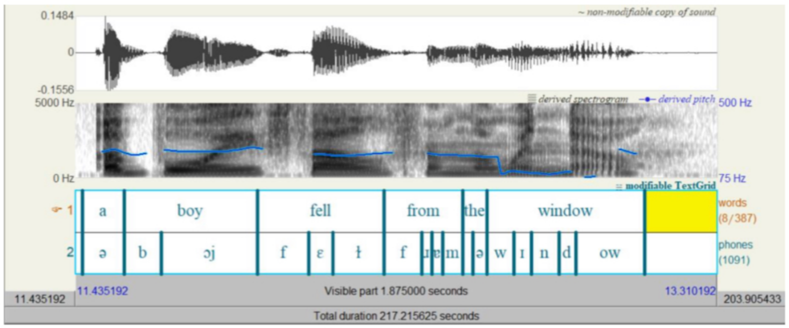
Left: pulmonic consonant; Right: non-pulmonic consonant

Application II: classification of speech signals

Using real-world speech data from the MFA aligner, our research group (Feng) further fed the topological features for machine learning, and obtained positive preliminary results for classification.

Application II: classification of speech signals

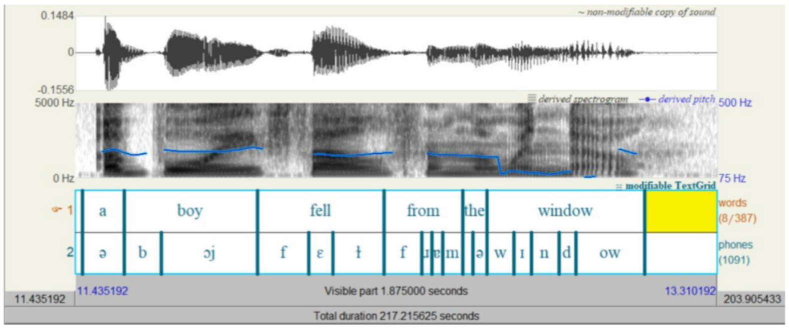
Using real-world speech data from the MFA aligner, our research group (Feng) further fed the topological features for machine learning, and obtained positive preliminary results for classification.



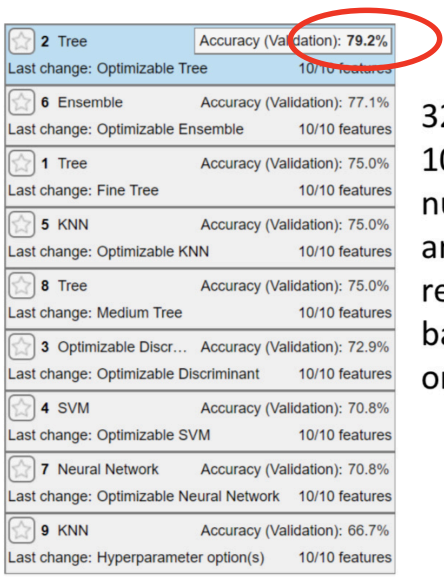

```
vowel_phones=['ɔj','ɛ','ə','ɪ','aj','
ɑ','æ','i','o','ʊ','aw','e','u','a']
consonant_phones=['b','f','m','ɹ','ð'
,'w','h','p','t','z','n','g','dʒ','s'
,'ʃ','v','l','ŋ','k','θ','j','tʃ','ʒ'
,'d']
```

Application II: classification of speech signals

Using real-world speech data from the MFA aligner, our research group (Feng) further fed the topological features for machine learning, and obtained positive preliminary results for classification.



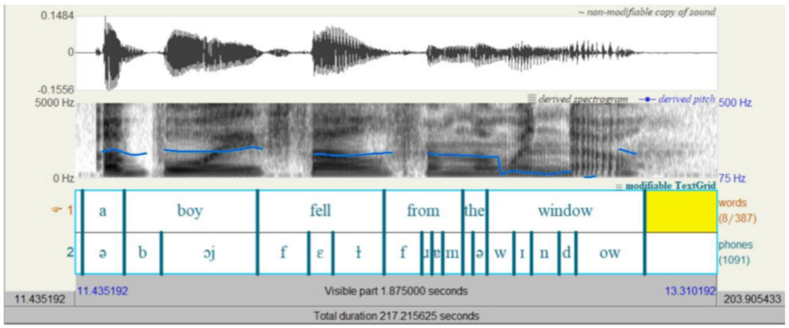

```
vowel_phones=['ɔj','ɛ','ə','ɪ','aj','
ɑ','æ','i','o','ʊ','aw','e','u','a']
consonant_phones=['b','f','m','ɹ','ð'
,'w','h','p','t','z','n','g','dʒ','s'
,'ʃ','v','l','ŋ','k','θ','j','tʃ','ʒ'
,'d']
```



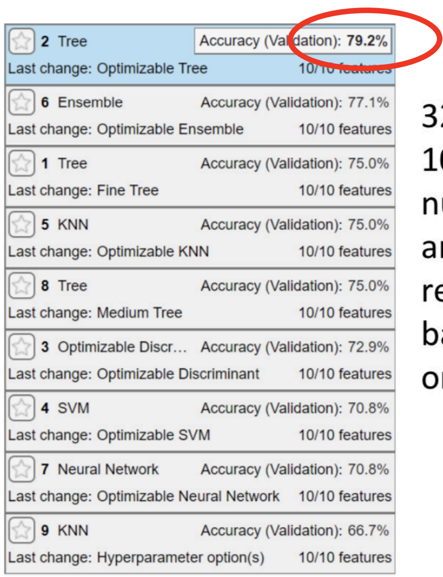
32 vowels, 16 consonants.
10 features: 5 are barcodes number of 5 diag, other 5 are number of barcodes that reaches inf(both consider barcode of 1 dimension for only)

Application II: classification of speech signals

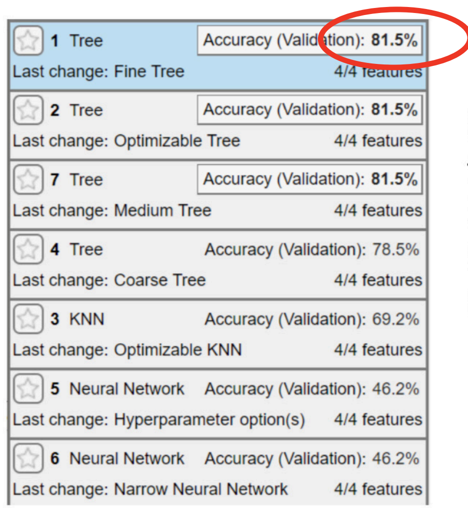
Using real-world speech data from the MFA aligner, our research group (Feng) further fed the topological features for machine learning, and obtained positive preliminary results for classification.




```
vowel_phones=['ɔj','ɛ','ə','ɪ','aj','ɑ','æ','i','o','ʊ','aw','e','u','a']
consonant_phones=['b','f','m','ɹ','ð','w','h','p','t','z','n','g','dʒ','s','ʃ','v','l','ŋ','k','θ','j','tʃ','ʒ','d']
```



32 vowels, 16 consonants. 10 features: 5 are barcodes number of 5 diag, other 5 are number of barcodes that reaches inf(both consider barcode of 1 dimension for only)

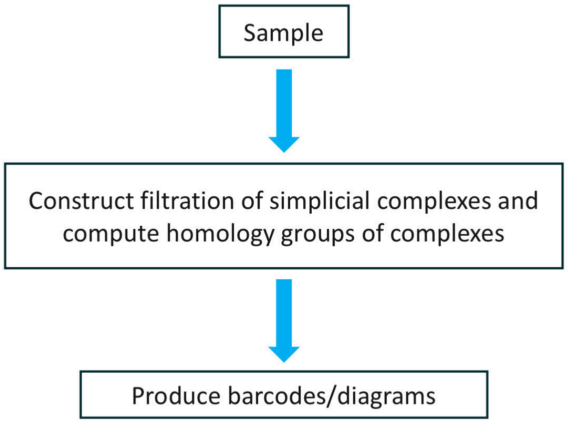


32 vowels, 33 consonants. 4 features: bottleneck distance between neighborhood barcode(currently the best result)

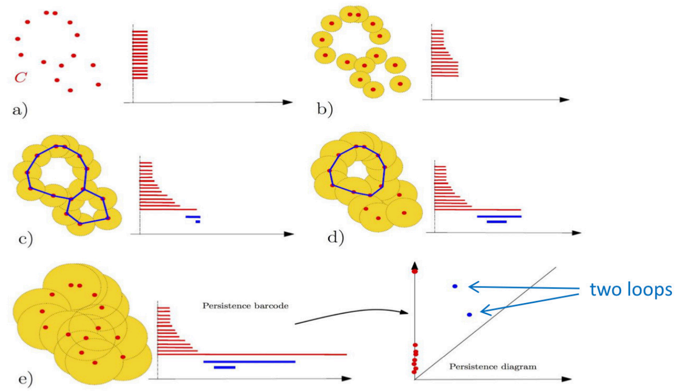
A formal recap of the topological methods applied

A formal recap of the topological methods applied

- Persistent homology

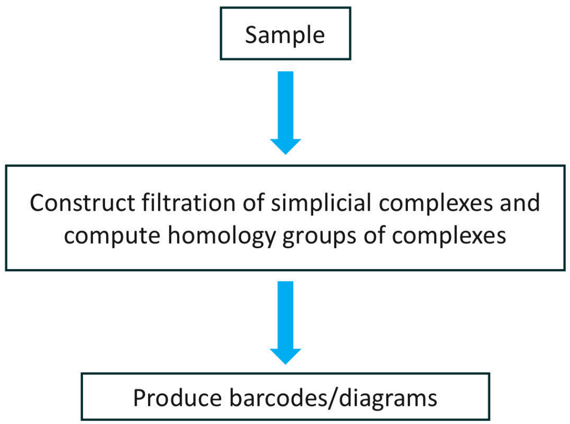


How filtration through varying distance measure reveals essential topological features

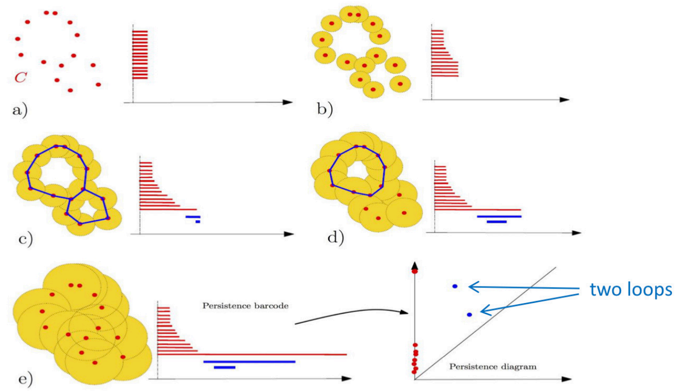


A formal recap of the topological methods applied

- Persistent homology



How filtration through varying distance measure reveals essential topological features

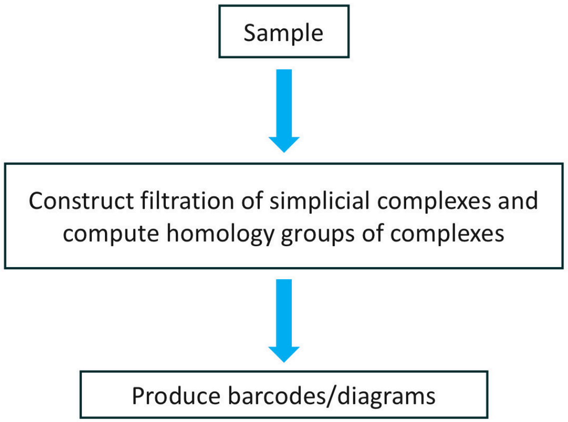


- Sliding window embedding (time-delay embedding)

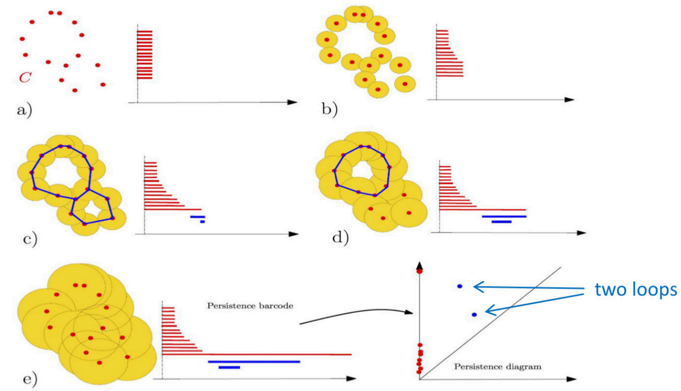
  Euclidean embedding of time series data dates back to Takens's work on fluid turbulence in the 1980s.

A formal recap of the topological methods applied

- Persistent homology



How filtration through varying distance measure reveals essential topological features



- Sliding window embedding (time-delay embedding)

Euclidean embedding of time series data dates back to Takens's work on fluid turbulence in the 1980s.

Theorem (Takens 1981). Let M be a compact manifold of dimension n. Given pairs (ϕ, y) with $\phi: M \to M$ a smooth diffeomorphism and $y: M \to \mathbb{R}$ a smooth function, it is a generic property that the map $\Phi_{(\phi,y)}: M \to \mathbb{R}^{2n+1}$ defined by

$$\Phi_{(\varphi,y)}(x) = \Big(y\,(x)\,, y\big(\varphi(x)\big), \ldots, y\big(\varphi^{2n}(x)\big)\Big)$$

is an embedding.

From topological data analysis to topological deep learning

From topological data analysis to topological deep learning

Using persistent homology, Carlsson, Ishkhanov, de Silva, and Zomorodian qualitatively analyzed approximately 4.5×10^6 high-contrast local patches of natural images

From topological data analysis to topological deep learning

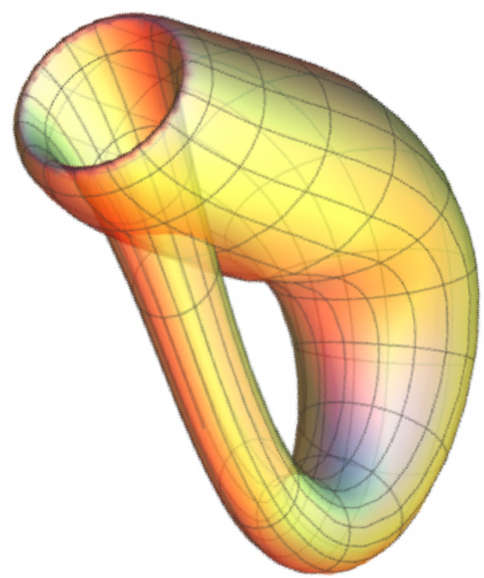
Using persistent homology, Carlsson, Ishkhanov, de Silva, and Zomorodian qualitatively analyzed approximately 4.5×10^6 high-contrast local patches of natural images obtained by van Hateren and van der Schaaf

From topological data analysis to topological deep learning

Using persistent homology, Carlsson, Ishkhanov, de Silva, and Zomorodian qualitatively analyzed approximately 4.5×10^6 high-contrast local patches of natural images obtained by van Hateren and van der Schaaf and previously studied by Lee, Mumford, and Petersen.

From topological data analysis to topological deep learning

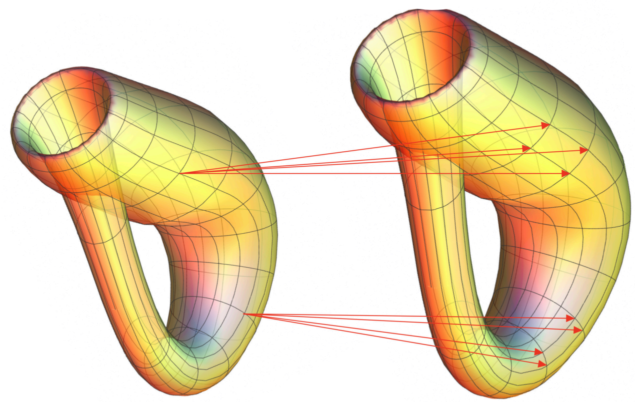
Using persistent homology, Carlsson, Ishkhanov, de Silva, and Zomorodian qualitatively analyzed approximately 4.5×10^6 high-contrast local patches of natural images obtained by van Hateren and van der Schaaf and previously studied by Lee, Mumford, and Petersen. In their 2008 article, they discovered that as vectors of pixels, the image data were unevenly distributed over a Klein bottle within the 7-dimensional Euclidean sphere!



From topological data analysis to topological deep learning

Using persistent homology, Carlsson, Ishkhanov, de Silva, and Zomorodian qualitatively analyzed approximately 4.5×10^6 high-contrast local patches of natural images obtained by van Hateren and van der Schaaf and previously studied by Lee, Mumford, and Petersen. In their 2008 article, they discovered that as vectors of pixels, the image data were unevenly distributed over a Klein bottle within the 7-dimensional Euclidean sphere!

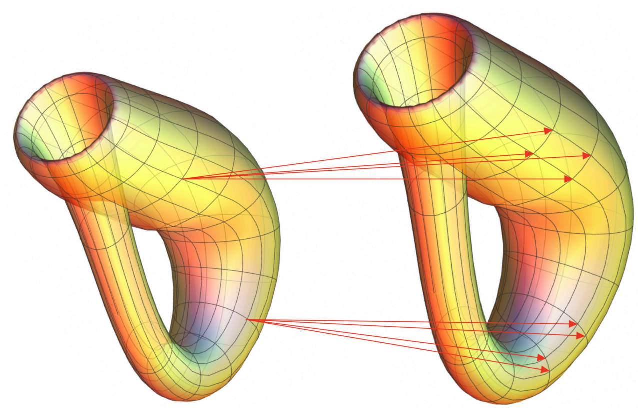
A decade later, Love, Filippenko, Maroulas, and Carlsson have made the Klein bottle as a topological input for designing convolutional layers in neural networks that learn image data.



From topological data analysis to topological deep learning

Using persistent homology, Carlsson, Ishkhanov, de Silva, and Zomorodian qualitatively analyzed approximately 4.5×10^6 high-contrast local patches of natural images obtained by van Hateren and van der Schaaf and previously studied by Lee, Mumford, and Petersen. In their 2008 article, they discovered that as vectors of pixels, the image data were unevenly distributed over a Klein bottle within the 7-dimensional Euclidean sphere!

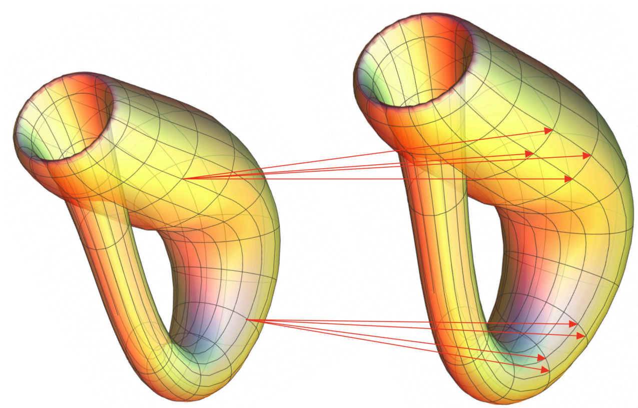
A decade later, Love, Filippenko, Maroulas, and Carlsson have made the Klein bottle as a topological input for designing convolutional layers in neural networks that learn image data. Moreover, they have incorporated the tangent bundle of a Klein bottle into TCNNs for learning video data.



From topological data analysis to topological deep learning

Using persistent homology, Carlsson, Ishkhanov, de Silva, and Zomorodian qualitatively analyzed approximately 4.5×10^6 high-contrast local patches of natural images obtained by van Hateren and van der Schaaf and previously studied by Lee, Mumford, and Petersen. In their 2008 article, they discovered that as vectors of pixels, the image data were unevenly distributed over a Klein bottle within the 7-dimensional Euclidean sphere!

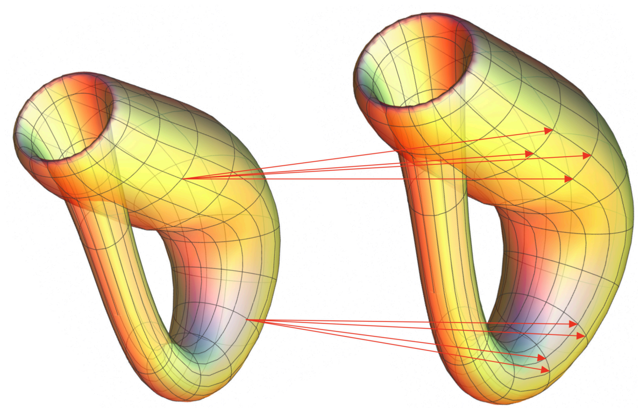
A decade later, Love, Filippenko, Maroulas, and Carlsson have made the Klein bottle as a topological input for designing convolutional layers in neural networks that learn image data. Moreover, they have incorporated the tangent bundle of a Klein bottle into TCNNs for learning video data. Both learnings achieved higher accuracies with smaller training sets.



From topological data analysis to topological deep learning

Using persistent homology, Carlsson, Ishkhanov, de Silva, and Zomorodian qualitatively analyzed approximately 4.5×10^6 high-contrast local patches of natural images obtained by van Hateren and van der Schaaf and previously studied by Lee, Mumford, and Petersen. In their 2008 article, they discovered that as vectors of pixels, the image data were unevenly distributed over a Klein bottle within the 7-dimensional Euclidean sphere!

A decade later, Love, Filippenko, Maroulas, and Carlsson have made the Klein bottle as a topological input for designing convolutional layers in neural networks that learn image data. Moreover, they have incorporated the tangent bundle of a Klein bottle into TCNNs for learning video data. Both learnings achieved higher accuracies with smaller training sets.



As a second warm-up, our research group (Zhiwang Yu) have reproduced some of their results.

From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators',

From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab,

From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals,

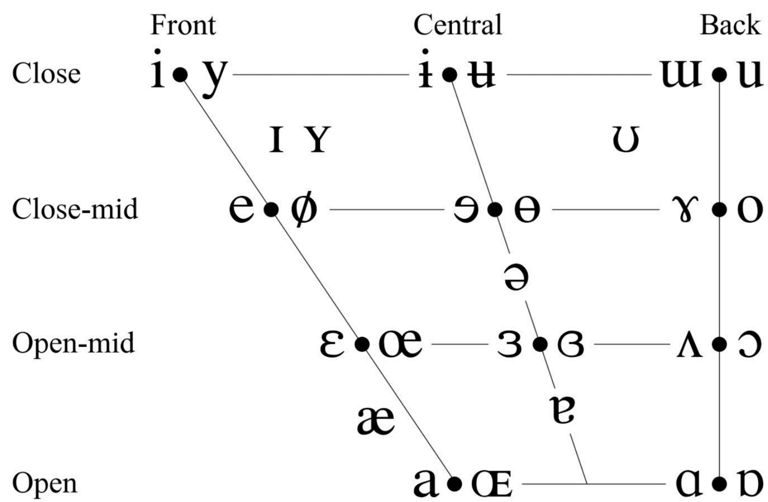
From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals, with the additional tool of time-delay embedding for turning time series data to point clouds in Euclidean spaces.

From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals, with the additional tool of time-delay embedding for turning time series data to point clouds in Euclidean spaces.

For phonetic data, linguists created a charted "distribution space" of vowels:

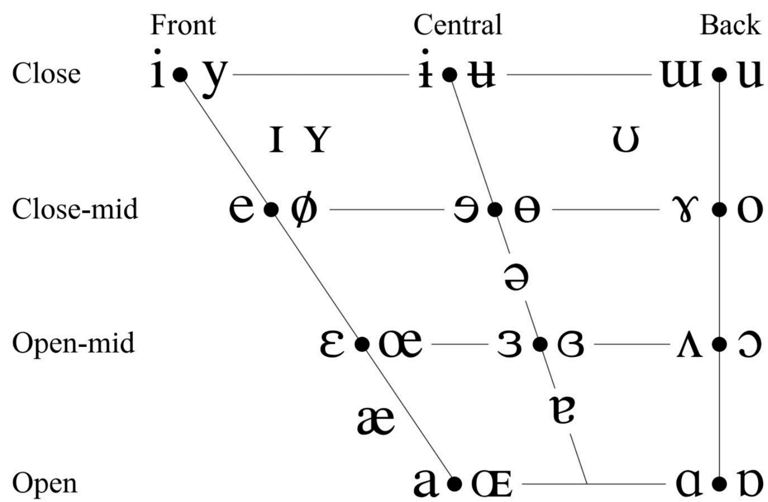


From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals, with the additional tool of time-delay embedding for turning time series data to point clouds in Euclidean spaces.

For phonetic data, linguists created a charted "distribution space" of vowels:

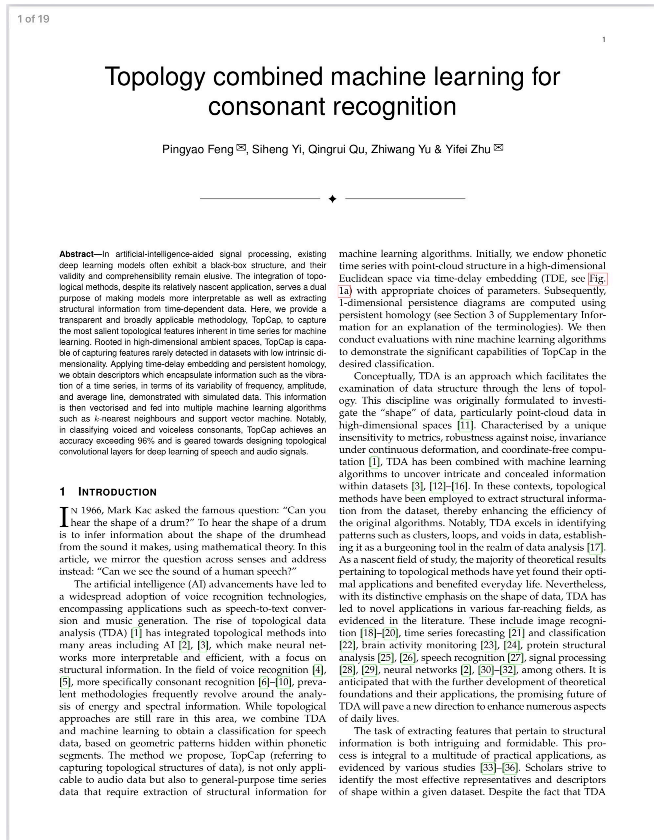
The vertical axis of the chart denotes vowel height. Vowels pronounced with the tongue lowered are at the bottom and raised are at the top. The horizontal axis of the chart denotes vowel backness. Vowels with the tongue moved towards the front of the mouth are in the left of the chart, while those with the tongue moved to the back are placed in right. The last parameter is whether the lips are rounded. At each given spot, vowels on the right and left are rounded and unrounded, respectively.



From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals, with the additional tool of time-delay embedding for turning time series data to point clouds in Euclidean spaces.

For phonetic data, linguists created a charted "distribution space" of vowels. Using speech files from SpeechBox, our topological approach achieved an average accuracy exceeding 96% in classifying voiced and voiceless consonants via machine learning.


1 of 19

1

Topology combined machine learning for consonant recognition

Pingyao Feng ✉, Siheng Yi, Qingrui Qu, Zhiwang Yu & Yifei Zhu ✉

Abstract—In artificial-intelligence-aided signal processing, existing deep learning models often exhibit a black-box structure, and their validity and comprehensibility remain elusive. The integration of topological methods, despite its relatively nascent application, serves a dual purpose of making models more interpretable as well as extracting structural information from time-dependent data. Here, we provide a transparent and broadly applicable methodology, TopCap, to capture the most salient topological features inherent in time series for machine learning. Rooted in high-dimensional ambient spaces, TopCap is capable of capturing features rarely detected in datasets with low intrinsic dimensionality. Applying time-delay embedding and persistent homology, we obtain descriptors which encapsulate information such as the vibration of a time series, in terms of its variability of frequency, amplitude, and average line, demonstrated with simulated data. This information is then vectorised and fed into 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 topological convolutional layers for deep learning of speech and audio signals.

1 INTRODUCTION

In 1966, Mark Kac asked the famous question: "Can you hear the shape of a drum?" To hear the shape of a drum is to infer information about the shape of the drumhead from the sound it makes, using mathematical theory. In this article, we mirror the question across senses and address instead: "Can we see the sound of a human speech?"

The artificial intelligence (AI) advancements have led to a widespread adoption of voice recognition technologies, encompassing applications such as speech-to-text conversion and music generation. The rise of topological data analysis (TDA) [1] has integrated topological methods into many areas including AI [2], [3], which make neural networks more interpretable and efficient, with a focus on structural information. In the field of voice recognition [4], [5], more specifically consonant recognition [6]–[10], prevalent methodologies frequently revolve around the analysis of energy and spectral information. While topological approaches are still rare in this area, we combine TDA and machine learning to obtain a classification for speech data, based on geometric patterns hidden within phonetic segments. The method we propose, TopCap (referring to capturing topological structures of data), is not only applicable to audio data but also to general-purpose time series data that require extraction of structural information for machine learning algorithms. Initially, we endow phonetic time series with point-cloud structure in a high-dimensional Euclidean space via time-delay embedding (TDE, see Fig. 1a) with appropriate choices of parameters. Subsequently, 1-dimensional persistence diagrams are computed using persistent homology (see Section 3 of Supplementary Information for an explanation of the terminologies). We then conduct evaluations with nine machine learning algorithms to demonstrate the significant capabilities of TopCap in the desired classification.

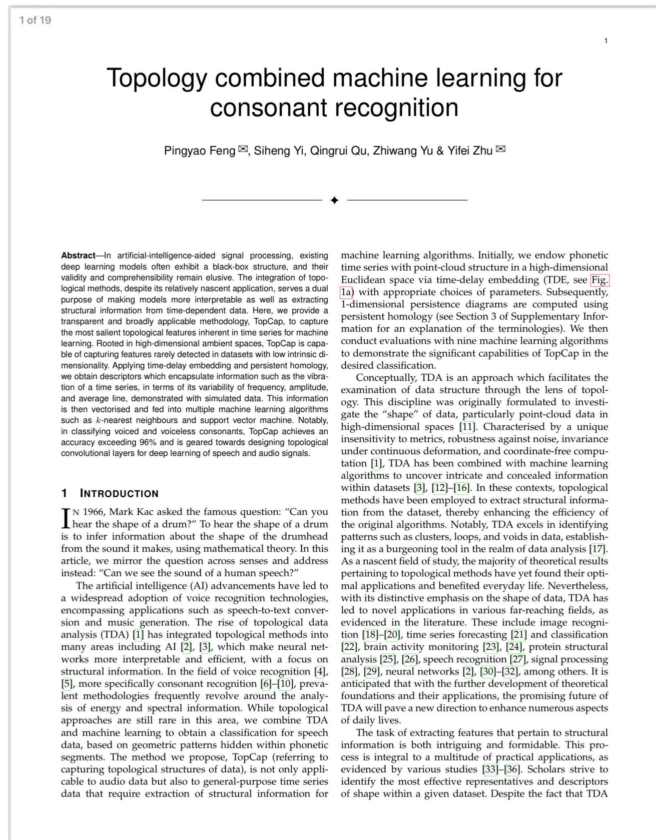
Conceptually, TDA is an approach which facilitates the examination of data structure through the lens of topology. This discipline was originally formulated to investigate the "shape" of data, particularly point-cloud data in high-dimensional spaces [11]. Characterised by a unique insensitivity to metrics, robustness against noise, invariance under continuous deformation, and coordinate-free computation [1], TDA has been combined with machine learning algorithms to uncover intricate and concealed information within datasets [3], [12]–[16]. In these contexts, topological methods have been employed to extract structural information from the dataset, thereby enhancing the efficiency of the original algorithms. Notably, TDA excels in identifying patterns such as clusters, loops, and voids in data, establishing it as a burgeoning tool in the realm of data analysis [17]. As a nascent field of study, the majority of theoretical results pertaining to topological methods have yet found their optimal applications and benefited everyday life. Nevertheless, with its distinctive emphasis on the shape of data, TDA has led to novel applications in various far-reaching fields, as evidenced in the literature. These include image recognition [18]–[20], time series forecasting [21] and classification [22], brain activity monitoring [23], [24], protein structural analysis [25], [26], speech recognition [27], signal processing [28], [29], neural networks [2], [30]–[32], among others. It is anticipated that with the further development of theoretical foundations and their applications, the promising future of TDA will pave a new direction to enhance numerous aspects of daily lives.

The task of extracting features that pertain to structural information is both intriguing and formidable. This process is integral to a multitude of practical applications, as evidenced by various studies [33]–[36]. Scholars strive to identify the most effective representatives and descriptors of shape within a given dataset. Despite the fact that TDA

From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals, with the additional tool of time-delay embedding for turning time series data to point clouds in Euclidean spaces.

For phonetic data, linguists created a charted "distribution space" of vowels. Using speech files from SpeechBox, our topological approach achieved an average accuracy exceeding 96% in classifying voiced and voiceless consonants via machine learning. A main goal remains to use topological methods to reveal a distribution space for speech data,



1 of 19

1

Topology combined machine learning for consonant recognition

Pingyao Feng ✉, Siheng Yi, Qingrui Qu, Zhiwang Yu & Yifei Zhu ✉

Abstract—In artificial-intelligence-aided signal processing, existing deep learning models often exhibit a black-box structure, and their validity and comprehensibility remain elusive. The integration of topological methods, despite its relatively nascent application, serves a dual purpose of making models more interpretable as well as extracting structural information from time-dependent data. Here, we provide a transparent and broadly applicable methodology, TopCap, to capture the most salient topological features inherent in time series for machine learning. Rooted in high-dimensional ambient spaces, TopCap is capable of capturing features rarely detected in datasets with low intrinsic dimensionality. Applying time-delay embedding and persistent homology, we obtain descriptors which encapsulate information such as the vibration of a time series, in terms of its variability of frequency, amplitude, and average line, demonstrated with simulated data. This information is then vectorised and fed into 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 topological convolutional layers for deep learning of speech and audio signals.

1 Introduction

In 1966, Mark Kac asked the famous question: "Can you hear the shape of a drum?" To hear the shape of a drum is to infer information about the shape of the drumhead from the sound it makes, using mathematical theory. In this article, we mirror the question across senses and address instead: "Can we see the sound of a human speech?"

The artificial intelligence (AI) advancements have led to a widespread adoption of voice recognition technologies, encompassing applications such as speech-to-text conversion and music generation. The rise of topological data analysis (TDA) [1] has integrated topological methods into many areas including AI [2], [3], which make neural networks more interpretable and efficient, with a focus on structural information. In the field of voice recognition [4], [5], more specifically consonant recognition [6]–[10], prevalent methodologies frequently revolve around the analysis of energy and spectral information. While topological approaches are still rare in this area, we combine TDA and machine learning to obtain a classification for speech data, based on geometric patterns hidden within phonetic segments. The method we propose, TopCap (referring to capturing topological structures of data), is not only applicable to audio data but also to general-purpose time series data that require extraction of structural information for machine learning algorithms. Initially, we endow phonetic time series with point-cloud structure in a high-dimensional Euclidean space via time-delay embedding (TDE, see Fig. 1a) with appropriate choices of parameters. Subsequently, 1-dimensional persistence diagrams are computed using persistent homology (see Section 3 of Supplementary Information for an explanation of the terminologies). We then conduct evaluations with nine machine learning algorithms to demonstrate the significant capabilities of TopCap in the desired classification.

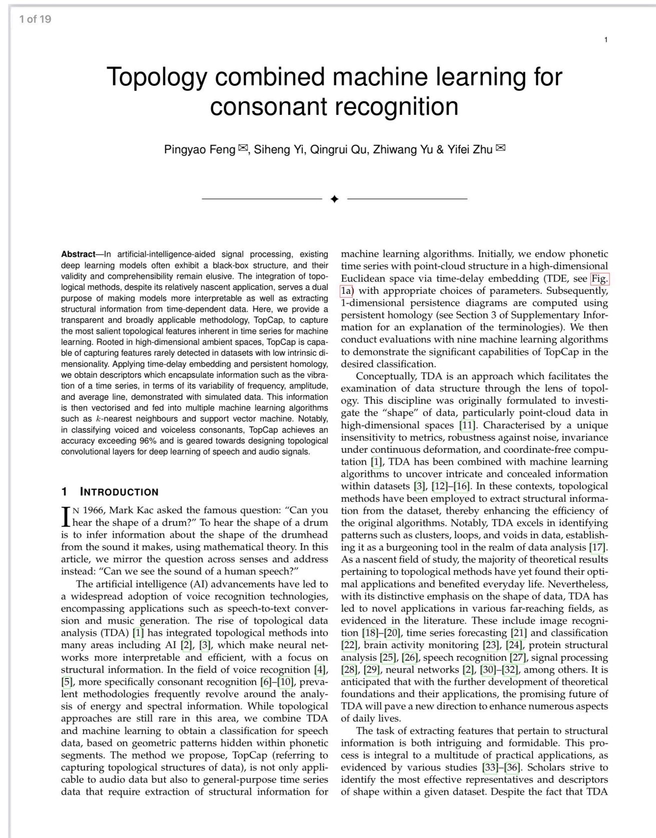
Conceptually, TDA is an approach which facilitates the examination of data structure through the lens of topology. This discipline was originally formulated to investigate the "shape" of data, particularly point-cloud data in high-dimensional spaces [11]. Characterised by a unique insensitivity to metrics, robustness against noise, invariance under continuous deformation, and coordinate-free computation [1], TDA has been combined with machine learning algorithms to uncover intricate and concealed information within datasets [3], [12]–[16]. In these contexts, topological methods have been employed to extract structural information from the dataset, thereby enhancing the efficiency of the original algorithms. Notably, TDA excels in identifying patterns such as clusters, loops, and voids in data, establishing it as a burgeoning tool in the realm of data analysis [17]. As a nascent field of study, the majority of theoretical results pertaining to topological methods have yet found their optimal applications and benefited everyday life. Nevertheless, with its distinctive emphasis on the shape of data, TDA has led to novel applications in various far-reaching fields, as evidenced in the literature. These include image recognition [18]–[20], time series forecasting [21] and classification [22], brain activity monitoring [23], [24], protein structural analysis [25], [26], speech recognition [27], signal processing [28], [29], neural networks [2], [30]–[32], among others. It is anticipated that with the further development of theoretical foundations and their applications, the promising future of TDA will pave a new direction to enhance numerous aspects of daily lives.

The task of extracting features that pertain to structural information is both intriguing and formidable. This process is integral to a multitude of practical applications, as evidenced by various studies [33]–[36]. Scholars strive to identify the most effective representatives and descriptors of shape within a given dataset. Despite the fact that TDA

From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals, with the additional tool of time-delay embedding for turning time series data to point clouds in Euclidean spaces.

For phonetic data, linguists created a charted "distribution space" of vowels. Using speech files from SpeechBox, our topological approach achieved an average accuracy exceeding 96% in classifying voiced and voiceless consonants via machine learning. A main goal remains to use topological methods to reveal a distribution space for speech data, even a digraph on it modeling the complex network of speech-signal sequences,



Topology combined machine learning for consonant recognition

Pingyao Feng ✉, Siheng Yi, Qingrui Qu, Zhiwang Yu & Yifei Zhu ✉

Abstract—In artificial-intelligence-aided signal processing, existing deep learning models often exhibit a black-box structure, and their validity and comprehensibility remain elusive. The integration of topological methods, despite its relatively nascent application, serves a dual purpose of making models more interpretable as well as extracting structural information from time-dependent data. Here, we provide a transparent and broadly applicable methodology, TopCap, to capture the most salient topological features inherent in time series for machine learning. Rooted in high-dimensional ambient spaces, TopCap is capable of capturing features rarely detected in datasets with low intrinsic dimensionality. Applying time-delay embedding and persistent homology, we obtain descriptors which encapsulate information such as the vibration of a time series, in terms of its variability of frequency, amplitude, and average line, demonstrated with simulated data. This information is then vectorised and fed into 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 topological convolutional layers for deep learning of speech and audio signals.

1 Introduction

In 1966, Mark Kac asked the famous question: "Can you hear the shape of a drum?" To hear the shape of a drum is to infer information about the shape of the drumhead from the sound it makes, using mathematical theory. In this article, we mirror the question across senses and address instead: "Can we see the sound of a human speech?"

The artificial intelligence (AI) advancements have led to a widespread adoption of voice recognition technologies, encompassing applications such as speech-to-text conversion and music generation. The rise of topological data analysis (TDA) [1] has integrated topological methods into many areas including AI [2], [3], which make neural networks more interpretable and efficient, with a focus on structural information. In the field of voice recognition [4], [5], more specifically consonant recognition [6]–[10], prevalent methodologies frequently revolve around the analysis of energy and spectral information. While topological approaches are still rare in this area, we combine TDA and machine learning to obtain a classification for speech data, based on geometric patterns hidden within phonetic segments. The method we propose, TopCap (referring to capturing topological structures of data), is not only applicable to audio data but also to general-purpose time series data that require extraction of structural information for machine learning algorithms. Initially, we endow phonetic time series with point-cloud structure in a high-dimensional Euclidean space via time-delay embedding (TDE, see Fig. 1a) with appropriate choices of parameters. Subsequently, 1-dimensional persistence diagrams are computed using persistent homology (see Section 3 of Supplementary Information for an explanation of the terminologies). We then conduct evaluations with nine machine learning algorithms to demonstrate the significant capabilities of TopCap in the desired classification.

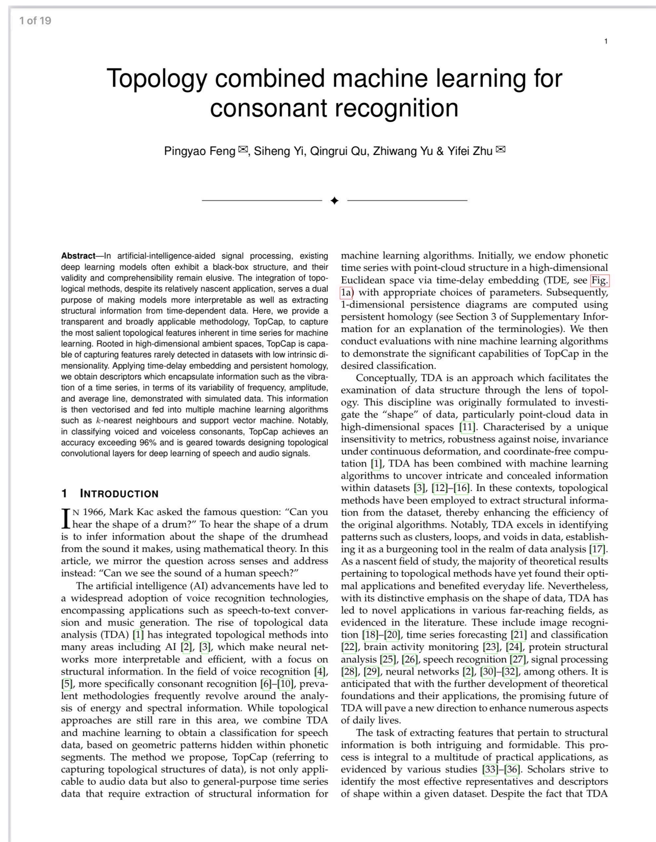
Conceptually, TDA is an approach which facilitates the examination of data structure through the lens of topology. This discipline was originally formulated to investigate the "shape" of data, particularly point-cloud data in high-dimensional spaces [11]. Characterised by a unique insensitivity to metrics, robustness against noise, invariance under continuous deformation, and coordinate-free computation [1], TDA has been combined with machine learning algorithms to uncover intricate and concealed information within datasets [3], [12]–[16]. In these contexts, topological methods have been employed to extract structural information from the dataset, thereby enhancing the efficiency of the original algorithms. Notably, TDA excels in identifying patterns such as clusters, loops, and voids in data, establishing it as a burgeoning tool in the realm of data analysis [17]. As a nascent field of study, the majority of theoretical results pertaining to topological methods have yet found their optimal applications and benefited everyday life. Nevertheless, with its distinctive emphasis on the shape of data, TDA has led to novel applications in various far-reaching fields, as evidenced in the literature. These include image recognition [18]–[20], time series forecasting [21] and classification [22], brain activity monitoring [23], [24], protein structural analysis [25], [26], speech recognition [27], signal processing [28], [29], neural networks [2], [30]–[32], among others. It is anticipated that with the further development of theoretical foundations and their applications, the promising future of TDA will pave a new direction to enhance numerous aspects of daily lives.

The task of extracting features that pertain to structural information is both intriguing and formidable. This process is integral to a multitude of practical applications, as evidenced by various studies [33]–[36]. Scholars strive to identify the most effective representatives and descriptors of shape within a given dataset. Despite the fact that TDA

From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators’, in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals, with the additional tool of time-delay embedding for turning time series data to point clouds in Euclidean spaces.

For phonetic data, linguists created a charted “distribution space” of vowels. Using speech files from SpeechBox, our topological approach achieved an average accuracy exceeding 96% in classifying voiced and voiceless consonants via machine learning. A main goal remains to use topological methods to reveal a distribution space for speech data, even a digraph on it modeling the complex network of speech-signal sequences,



Topology combined machine learning for consonant recognition

Pingyao Feng ✉, Siheng Yi, Qingrui Qu, Zhiwang Yu & Yifei Zhu ✉

Abstract—In artificial-intelligence-aided signal processing, existing deep learning models often exhibit a black-box structure, and their validity and comprehensibility remain elusive. The integration of topological methods, despite its relatively nascent application, serves a dual purpose of making models more interpretable as well as extracting structural information from time-dependent data. Here, we provide a transparent and broadly applicable methodology, TopCap, to capture the most salient topological features inherent in time series for machine learning. Rooted in high-dimensional ambient spaces, TopCap is capable of capturing features rarely detected in datasets with low intrinsic dimensionality. Applying time-delay embedding and persistent homology, we obtain descriptors which encapsulate information such as the vibration of a time series, in terms of its variability of frequency, amplitude, and average line, demonstrated with simulated data. This information is then vectorised and fed into 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 topological convolutional layers for deep learning of speech and audio signals.

1 INTRODUCTION

In 1966, Mark Kac asked the famous question: “Can you hear the shape of a drum?” To hear the shape of a drum is to infer information about the shape of the drumhead from the sound it makes, using mathematical theory. In this article, we mirror the question across senses and address instead: “Can we see the sound of a human speech?”

The artificial intelligence (AI) advancements have led to a widespread adoption of voice recognition technologies, encompassing applications such as speech-to-text conversion and music generation. The rise of topological data analysis (TDA) [1] has integrated topological methods into many areas including AI [2], [3], which make neural networks more interpretable and efficient, with a focus on structural information. In the field of voice recognition [4], [5], more specifically consonant recognition [6]–[10], prevalent methodologies frequently revolve around the analysis of energy and spectral information. While topological approaches are still rare in this area, we combine TDA and machine learning to obtain a classification for speech data, based on geometric patterns hidden within phonetic segments. The method we propose, TopCap (referring to capturing topological structures of data), is not only applicable to audio data but also to general-purpose time series data that require extraction of structural information for machine learning algorithms. Initially, we endow phonetic time series with point-cloud structure in a high-dimensional Euclidean space via time-delay embedding (TDE, see Fig. 1a) with appropriate choices of parameters. Subsequently, 1-dimensional persistence diagrams are computed using persistent homology (see Section 3 of Supplementary Information for an explanation of the terminologies). We then conduct evaluations with nine machine learning algorithms to demonstrate the significant capabilities of TopCap in the desired classification.

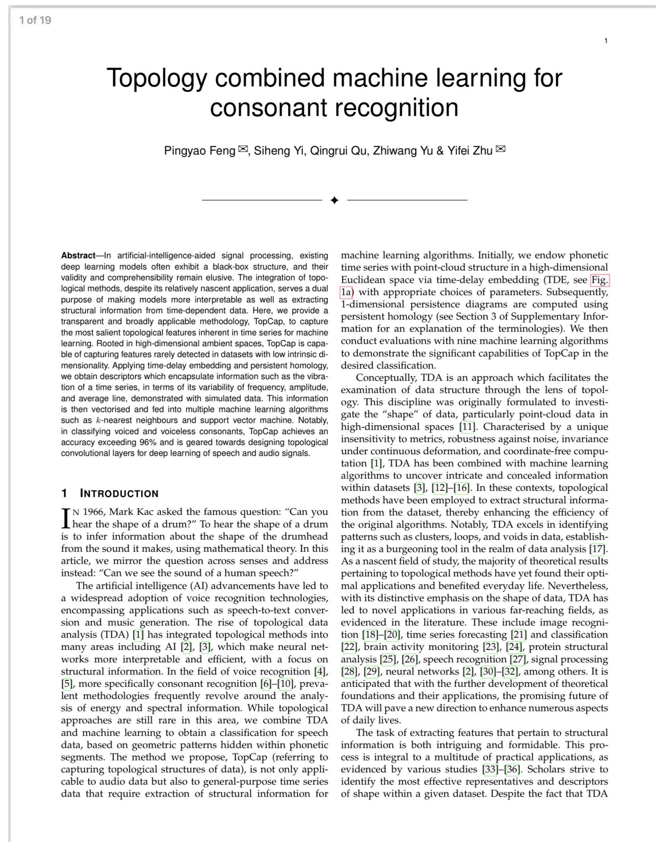
Conceptually, TDA is an approach which facilitates the examination of data structure through the lens of topology. This discipline was originally formulated to investigate the “shape” of data, particularly point-cloud data in high-dimensional spaces [11]. Characterised by a unique insensitivity to metrics, robustness against noise, invariance under continuous deformation, and coordinate-free computation [1], TDA has been combined with machine learning algorithms to uncover intricate and concealed information within datasets [3], [12]–[16]. In these contexts, topological methods have been employed to extract structural information from the dataset, thereby enhancing the efficiency of the original algorithms. Notably, TDA excels in identifying patterns such as clusters, loops, and voids in data, establishing it as a burgeoning tool in the realm of data analysis [17]. As a nascent field of study, the majority of theoretical results pertaining to topological methods have yet found their optimal applications and benefited everyday life. Nevertheless, with its distinctive emphasis on the shape of data, TDA has led to novel applications in various far-reaching fields, as evidenced in the literature. These include image recognition [18]–[20], time series forecasting [21] and classification [22], brain activity monitoring [23], [24], protein structural analysis [25], [26], speech recognition [27], signal processing [28], [29], neural networks [2], [30]–[32], among others. It is anticipated that with the further development of theoretical foundations and their applications, the promising future of TDA will pave a new direction to enhance numerous aspects of daily lives.

The task of extracting features that pertain to structural information is both intriguing and formidable. This process is integral to a multitude of practical applications, as evidenced by various studies [33]–[36]. Scholars strive to identify the most effective representatives and descriptors of shape within a given dataset. Despite the fact that TDA

From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals, with the additional tool of time-delay embedding for turning time series data to point clouds in Euclidean spaces.

For phonetic data, linguists created a charted "distribution space" of vowels. Using speech files from SpeechBox, our topological approach achieved an average accuracy exceeding 96% in classifying voiced and voiceless consonants via machine learning. A main goal remains to use topological methods to reveal a distribution space for speech data, even a digraph on it modeling the complex network of speech-signal sequences, and apply these topological inputs for smarter learning.


1 of 19

1

Topology combined machine learning for consonant recognition

Pingyao Feng ✉, Siheng Yi, Qingrui Qu, Zhiwang Yu & Yifei Zhu ✉

Abstract—In artificial-intelligence-aided signal processing, existing deep learning models often exhibit a black-box structure, and their validity and comprehensibility remain elusive. The integration of topological methods, despite its relatively nascent application, serves a dual purpose of making models more interpretable as well as extracting structural information from time-dependent data. Here, we provide a transparent and broadly applicable methodology, TopCap, to capture the most salient topological features inherent in time series for machine learning. Rooted in high-dimensional ambient spaces, TopCap is capable of capturing features rarely detected in datasets with low intrinsic dimensionality. Applying time-delay embedding and persistent homology, we obtain descriptors which encapsulate information such as the vibration of a time series, in terms of its variability of frequency, amplitude, and average line, demonstrated with simulated data. This information is then vectorised and fed into 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 topological convolutional layers for deep learning of speech and audio signals.

1 INTRODUCTION

In 1966, Mark Kac asked the famous question: "Can you hear the shape of a drum?" To hear the shape of a drum is to infer information about the shape of the drumhead from the sound it makes, using mathematical theory. In this article, we mirror the question across senses and address instead: "Can we see the sound of a human speech?"

The artificial intelligence (AI) advancements have led to a widespread adoption of voice recognition technologies, encompassing applications such as speech-to-text conversion and music generation. The rise of topological data analysis (TDA) [1] has integrated topological methods into many areas including AI [2], [3], which make neural networks more interpretable and efficient, with a focus on structural information. In the field of voice recognition [4], [5], more specifically consonant recognition [6]–[10], prevalent methodologies frequently revolve around the analysis of energy and spectral information. While topological approaches are still rare in this area, we combine TDA and machine learning to obtain a classification for speech data, based on geometric patterns hidden within phonetic segments. The method we propose, TopCap (referring to capturing topological structures of data), is not only applicable to audio data but also to general-purpose time series data that require extraction of structural information for machine learning algorithms. Initially, we endow phonetic time series with point-cloud structure in a high-dimensional Euclidean space via time-delay embedding (TDE, see Fig. 1a) with appropriate choices of parameters. Subsequently, 1-dimensional persistence diagrams are computed using persistent homology (see Section 3 of Supplementary Information for an explanation of the terminologies). We then conduct evaluations with nine machine learning algorithms to demonstrate the significant capabilities of TopCap in the desired classification.

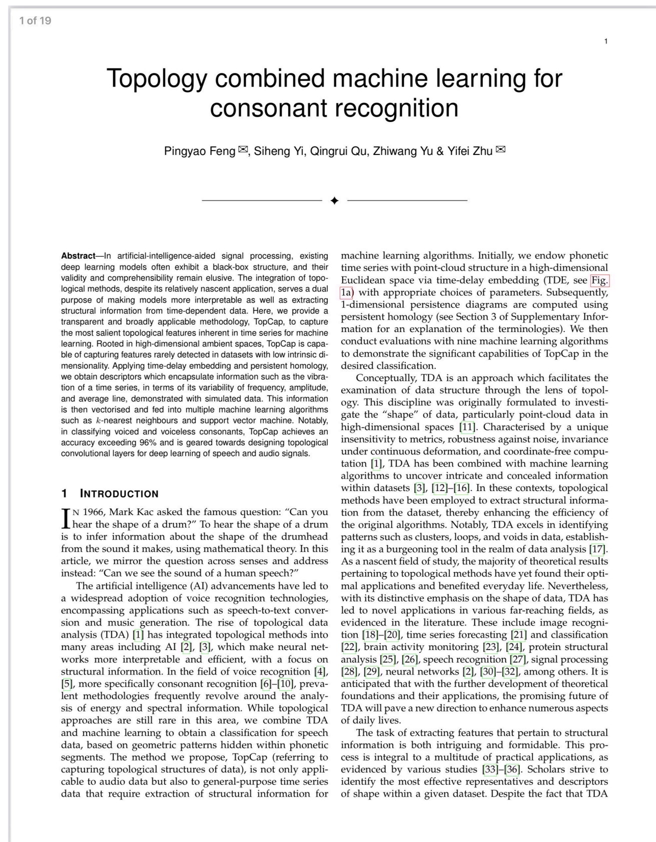
Conceptually, TDA is an approach which facilitates the examination of data structure through the lens of topology. This discipline was originally formulated to investigate the "shape" of data, particularly point-cloud data in high-dimensional spaces [11]. Characterised by a unique insensitivity to metrics, robustness against noise, invariance under continuous deformation, and coordinate-free computation [1], TDA has been combined with machine learning algorithms to uncover intricate and concealed information within datasets [3], [12]–[16]. In these contexts, topological methods have been employed to extract structural information from the dataset, thereby enhancing the efficiency of the original algorithms. Notably, TDA excels in identifying patterns such as clusters, loops, and voids in data, establishing it as a burgeoning tool in the realm of data analysis [17]. As a nascent field of study, the majority of theoretical results pertaining to topological methods have yet found their optimal applications and benefited everyday life. Nevertheless, with its distinctive emphasis on the shape of data, TDA has led to novel applications in various far-reaching fields, as evidenced in the literature. These include image recognition [18]–[20], time series forecasting [21] and classification [22], brain activity monitoring [23], [24], protein structural analysis [25], [26], speech recognition [27], signal processing [28], [29], neural networks [2], [30]–[32], among others. It is anticipated that with the further development of theoretical foundations and their applications, the promising future of TDA will pave a new direction to enhance numerous aspects of daily lives.

The task of extracting features that pertain to structural information is both intriguing and formidable. This process is integral to a multitude of practical applications, as evidenced by various studies [33]–[36]. Scholars strive to identify the most effective representatives and descriptors of shape within a given dataset. Despite the fact that TDA

From topological data analysis to topological deep learning

Motivated by the works of Carlsson and his collaborators', in consultation with Meng Yu of Tencent AI Lab, we have been investigating analogous questions for speech signals, with the additional tool of time-delay embedding for turning time series data to point clouds in Euclidean spaces.

For phonetic data, linguists created a charted "distribution space" of vowels. Using speech files from SpeechBox, our topological approach achieved an average accuracy exceeding 96% in classifying voiced and voiceless consonants via machine learning. A main goal remains to use topological methods to reveal a distribution space for speech data, even a digraph on it modeling the complex network of speech-signal sequences, and apply these topological inputs for smarter learning.



Topology combined machine learning for consonant recognition

Pingyao Feng ✉, Siheng Yi, Qingrui Qu, Zhiwang Yu & Yifei Zhu ✉

Abstract—In artificial-intelligence-aided signal processing, existing deep learning models often exhibit a black-box structure, and their validity and comprehensibility remain elusive. The integration of topological methods, despite its relatively nascent application, serves a dual purpose of making models more interpretable as well as extracting structural information from time-dependent data. Here, we provide a transparent and broadly applicable methodology, TopCap, to capture the most salient topological features inherent in time series for machine learning. Rooted in high-dimensional ambient spaces, TopCap is capable of capturing features rarely detected in datasets with low intrinsic dimensionality. Applying time-delay embedding and persistent homology, we obtain descriptors which encapsulate information such as the vibration of a time series, in terms of its variability of frequency, amplitude, and average line, demonstrated with simulated data. This information is then vectorised and fed into 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 topological convolutional layers for deep learning of speech and audio signals.

1 Introduction

In 1966, Mark Kac asked the famous question: "Can you hear the shape of a drum?" To hear the shape of a drum is to infer information about the shape of the drumhead from the sound it makes, using mathematical theory. In this article, we mirror the question across senses and address instead: "Can we see the sound of a human speech?"

The artificial intelligence (AI) advancements have led to a widespread adoption of voice recognition technologies, encompassing applications such as speech-to-text conversion and music generation. The rise of topological data analysis (TDA) [1] has integrated topological methods into many areas including AI [2], [3], which make neural networks more interpretable and efficient, with a focus on structural information. In the field of voice recognition [4], [5], more specifically consonant recognition [6]–[10], prevalent methodologies frequently revolve around the analysis of energy and spectral information. While topological approaches are still rare in this area, we combine TDA and machine learning to obtain a classification for speech data, based on geometric patterns hidden within phonetic segments. The method we propose, TopCap (referring to capturing topological structures of data), is not only applicable to audio data but also to general-purpose time series data that require extraction of structural information for machine learning algorithms. Initially, we endow phonetic time series with point-cloud structure in a high-dimensional Euclidean space via time-delay embedding (TDE, see Fig. 1a) with appropriate choices of parameters. Subsequently, 1-dimensional persistence diagrams are computed using persistent homology (see Section 3 of Supplementary Information for an explanation of the terminologies). We then conduct evaluations with nine machine learning algorithms to demonstrate the significant capabilities of TopCap in the desired classification.

Conceptually, TDA is an approach which facilitates the examination of data structure through the lens of topology. This discipline was originally formulated to investigate the "shape" of data, particularly point-cloud data in high-dimensional spaces [11]. Characterised by a unique insensitivity to metrics, robustness against noise, invariance under continuous deformation, and coordinate-free computation [1], TDA has been combined with machine learning algorithms to uncover intricate and concealed information within datasets [3], [12]–[16]. In these contexts, topological methods have been employed to extract structural information from the dataset, thereby enhancing the efficiency of the original algorithms. Notably, TDA excels in identifying patterns such as clusters, loops, and voids in data, establishing it as a burgeoning tool in the realm of data analysis [17]. As a nascent field of study, the majority of theoretical results pertaining to topological methods have yet found their optimal applications and benefited everyday life. Nevertheless, with its distinctive emphasis on the shape of data, TDA has led to novel applications in various far-reaching fields, as evidenced in the literature. These include image recognition [18]–[20], time series forecasting [21] and classification [22], brain activity monitoring [23], [24], protein structural analysis [25], [26], speech recognition [27], signal processing [28], [29], neural networks [2], [30]–[32], among others. It is anticipated that with the further development of theoretical foundations and their applications, the promising future of TDA will pave a new direction to enhance numerous aspects of daily lives.

The task of extracting features that pertain to structural information is both intriguing and formidable. This process is integral to a multitude of practical applications, as evidenced by various studies [33]–[36]. Scholars strive to identify the most effective representatives and descriptors of shape within a given dataset. Despite the fact that TDA

From topological data analysis to topological deep learning

Reservoir networks and photonic circuits have been applied to vowel recognition, too.

nature photonics

ARTICLES

PUBLISHED ONLINE: 12 JUNE 2017 | DOI: 10.1038/NPHOTON.2017.93

Deep learning with coherent nanophotonic circuits

Yichen Shen[1*†], Nicholas C. Harris[1*†], Scott Skirlo[1], Mihika Prabhu[1], Tom Baehr-Jones[2], Michael Hochberg[2], Xin Sun[3], Shijie Zhao[4], Hugo Larochelle[5], Dirk Englund[1] and Marin Soljačić[1]

Artificial neural networks are computational network models inspired by signal processing in the brain. These models have dramatically improved performance for many machine-learning tasks, including speech and image recognition. However, today's computing hardware is inefficient at implementing neural networks, in large part because much of it was designed for von Neumann computing schemes. Significant effort has been made towards developing electronic architectures tuned to implement artificial neural networks that exhibit improved computational speed and accuracy. Here, we propose a new architecture for a fully optical neural network that, in principle, could offer an enhancement in computational speed and power efficiency over state-of-the-art electronics for conventional inference tasks. We experimentally demonstrate the essential part of the concept using a programmable nanophotonic processor featuring a cascaded array of 56 programmable Mach–Zehnder interferometers in a silicon photonic integrated circuit and show its utility for vowel recognition.

From topological data analysis to topological deep learning

Reservoir networks and photonic circuits have been applied to vowel recognition, too.

nature photonics

ARTICLES

PUBLISHED ONLINE: 12 JUNE 2017 | DOI: 10.1038/NPHOTON.2017.93

Deep learning with coherent nanophotonic circuits

Yichen Shen[1*†], Nicholas C. Harris[1*†], Scott Skirlo[1], Mihika Prabhu[1], Tom Baehr-Jones[2], Michael Hochberg[2], Xin Sun[3], Shijie Zhao[4], Hugo Larochelle[5], Dirk Englund[1] and Marin Soljačić[1]

Artificial neural networks are computational network models inspired by signal processing in the brain. These models have dramatically improved performance for many machine-learning tasks, including speech and image recognition. However, today's computing hardware is inefficient at implementing neural networks, in large part because much of it was designed for von Neumann computing schemes. Significant effort has been made towards developing electronic architectures tuned to implement artificial neural networks that exhibit improved computational speed and accuracy. Here, we propose a new architecture for a fully optical neural network that, in principle, could offer an enhancement in computational speed and power efficiency over state-of-the-art electronics for conventional inference tasks. We experimentally demonstrate the essential part of the concept using a programmable nanophotonic processor featuring a cascaded array of 56 programmable Mach–Zehnder interferometers in a silicon photonic integrated circuit and show its utility for vowel recognition.

It will be useful to design and fine-tune them topologically (joint with Huan Li of optical science and engineering at Zhejiang University and Xinxiang Niu of Huawei).

Thank you.