**Goal**

Our goal is to explain, through examples of time series analysis and speech signal processing:

Topology of artificial neural network architectures (Explainable AI);

Topological inputs for designing neural networks (Interpretable AI).

**Topological Time Series**

**Theoretical Basis**

Theorem (Takens 1981). Let M be a compact manifold of dimension n. Given pairs (A,y) with A:M→M a smooth diffeomorphism and y:M→R a smooth function, it is a generic property that the map A\_\_:M→R2n:

is an embedding.

**Methods**

Let us assume that sampled signals are distributed over a manifold. To topologically analyze time series, we then proceed as follows (Fig. 1):

**Classification of Speech Signals**

**Inspiration**

Using persistent homology, Carlsson et al. qualitatively analyzed approximately 4.5x10⁶ high-contrast local patches of natural images obtained by van Hateren and van der Schaaf and previously studied by Lee et al. In their 2008 article (Fig. 3), 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 et al. have used 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 accuracy with smaller training sets (Fig. 4).

As a warm-up, our research group (Zhiwang Yu, Haiyu Zhang) has reproduced some of their results (Fig. 5).

**Results**

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 into Euclidean spaces.

For phonetic data, linguists created a charted “distribution space” of vowels (Fig. 6).

Using speech files from SpeechBox, our topological approach achieved an average accuracy exceeding 96% in classifying voiced and voiceless consonants via machine learning (Fig. 7). 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.

**Method**

As a demonstration of their effectiveness, in both accuracy and efficiency, our streamlined algorithm TopCap significantly outperformed traditional deep learning neural networks for the classification of voiced and voiceless consonants from real human speech data (Fig. 8).

In view of the capability of topological methods to discern vibration patterns in time series, we apply them to classify consonant signals into voiced and voiceless categories (Fig. 9).

**Two Novel Approaches in Nonlinear Time Series Analysis**

**Beyond Speech Signals**

A new, apparently paradoxical parameter selection scheme for choosing high Euclidean embedded dimension for time series data whose intrinsic dimensionality is low (Fig. 10);

Venturing beyond periodicity, an implementation with TDA to detect what we termed the “three fundamental variations” as finer structures inherent in time series data, namely, variabilities of frequency, of amplitude, and of average line, enabled by a scrutiny of learnable TDA descriptors (Fig. 11).

**Related Work**

We have been experimenting with more extensive datasets, including LJSpeech, LibriSpeech, TIMIT, as well as extending comparison of our approach to state-of-the-art methods to demonstrate its advantages. It will be useful to design and fine-tune them topologically (Fig. 12).

Joint with Huan Li of optical science and engineering at Zhejiang University and Xinxiang Niu of Huawei.