Topological Machine Learning class

Aina Ferrà Marcús

30/11/2023

Universitat de Barcelona

Topological classifier

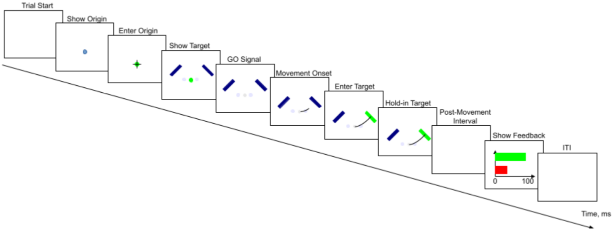
Topological classifier

Algorithm

1. Separate the training set according to the different classes.
2. For each class, compute its homological silhouette in dimension 0.
3. Given an input test, add it in parallel to each of the training subsets.
4. For every class, recompute its homological silhouette.
5. Compute the euclidean distance between the training silhouettes and the newly obtained silhouettes. .
6. Assign the label of the class whose new silhouette was closer to the trained one.

A neuroimaging problem

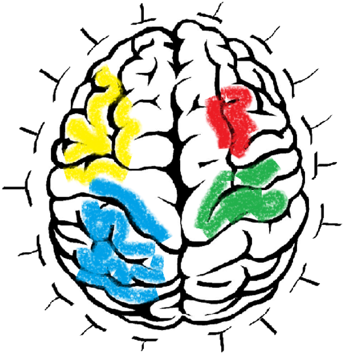
We study data coming from a an experiment designed to identify **how the expression of motivation is distributed across the brain**.



Participants in this study are asked to perform a precision task together with a partner (simulated). Depending on the skill of the partner, the participants are induced **different motivational states** modulated by social pressure.

A neuroimaging problem

Participants wear a 64-electrode cap, so a 64-channel encephalogram is recorded while they perform the task.



Time series are collapsed computing the mean of intensities. Therefore, for each task that a participant is doing we obtain a point in $\mathbb{R}^{64}$.

A neuroimaging problem

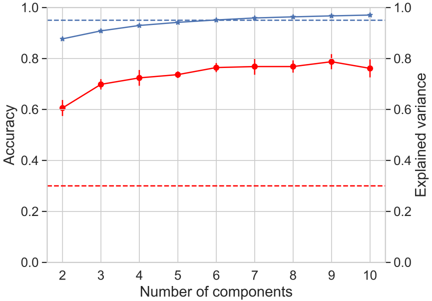
The following **classification problem** is proposed: given a point in $\mathbb{R}^{64}$ that represents a task by a participant, can we tell to which motivational state it belongs?

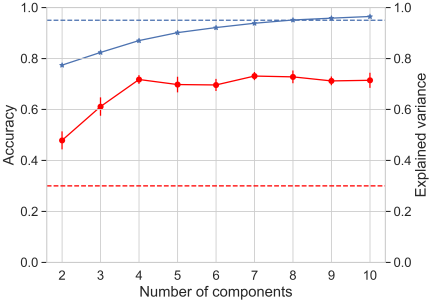
This same problem has been solved using other classifiers. The topological classifier works similar to them, but what extra information can we get?

The interesting thing came when we applied Principal Component Analysis, a dimensionality reduction technique, before classifying.

A neuroimaging problem

The accuracy of the classifier does not increase proportionally to the explained variance.





It appears that there is a peak around dimension 4, which furthers discussion about the latent dimension of the problem.

A neuroimaging problem

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

A topological classifier to characterize brain states: When shape matters more than variance

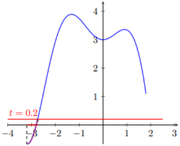
Aina Ferrà, Gloria Cecchini, Fritz-Pere Nobbe Fisas, Carles Casacuberta, Ignasi Cos

Published: October 2, 2023 • https://doi.org/10.1371/journal.pone.0292049

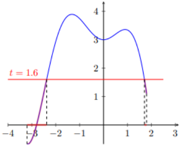
More details of this study can be found here.

TDA as a support system for deep learning

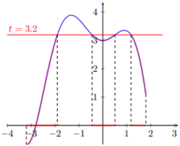
Persistent homology of a function



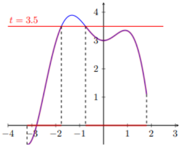
(A) A component is born at $t = -0.674$.



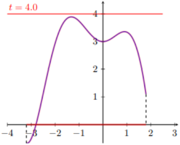
(B) Another component is born at $t = 1.097$. At the moment, we have two connected components.



(C) A third component is born at $t = 3$.



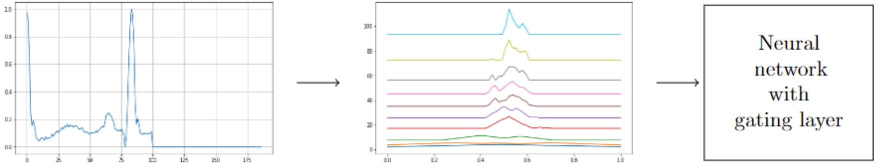
(D) Two components are merged at $t = 3.358$. We have two connected components again.



(E) Finally, the two components merge at $t = 3.892$. From now to infinitum, there will only remain one connected component.

Time series classification

The idea is to **decompose the signal into its landscape levels**, stack them together and feed it to a neural network.



Input

Persistence landscape

In a very natural way, we achieve a **hierarchical representation** such that the network can choose which information is necessary to perform the classification.

Time series classification

The neural network processes each landscape level **separately and parallely**, and then we can choose which levels of landscapes were more useful to perform a classification.

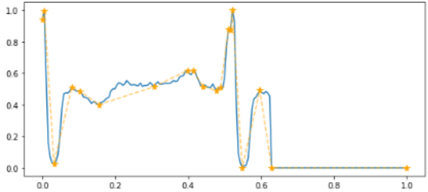
Dataset	Raw data	Unnormalized	Normalized	Selected	No.
ECG5000	94.72 ± 0.7	92.96 ± 0.5	93.12 ± 0.4	92.88 ± 0.3	3
FreezerRT	99.53 ± 0.4	63.70 ± 4.2	88.97 ± 0.3	88.70 ± 1.6	2
HandOutlines	89.20 ± 1.7	75.77 ± 4.6	85.26 ± 2.0	81.90 ± 2.0	5
ItalyPowerD	97.73 ± 1.1	87.18 ± 3.4	89.45 ± 1.7	90.36 ± 1.3	2
MoteStrain	90.98 ± 1.1	71.84 ± 1.8	77.33 ± 2.8	76.31 ± 1.7	3
PhalangesOC	64.02 ± 2.1	63.95 ± 4.2	68.95 ± 0.9	69.21 ± 0.8	4
StarlightCurves	95.70 ± 0.4	89.82 ± 2.3	94.92 ± 0.1	95.22 ± 0.5	3
Wafer	99.65 ± 0.2	90.86 ± 0.8	98.63 ± 0.3	98.79 ± 0.4	3
Yoga	82.48 ± 1.6	64.21 ± 4.1	75.36 ± 1.1	78.33 ± 1.0	4

We test it with many different time series datasets to assess its performance: in general, it works well.

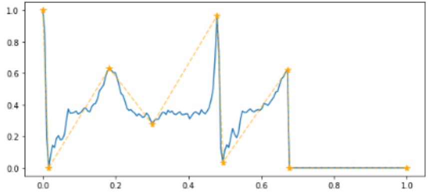
Arrhythmia identification

We need to make sure that the neural network is **focusing on the right information**.

Idea: feed a neural network with the heartbeat processed by the landscape levels, choose which landscape levels are more important and represent the points that originated such landscapes.

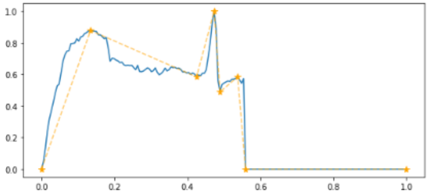


(a) Normal heartbeat

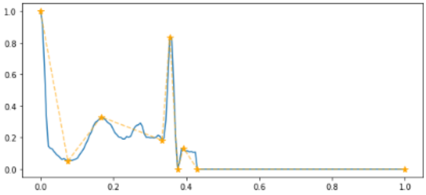


(b) Supraventricular premature beat

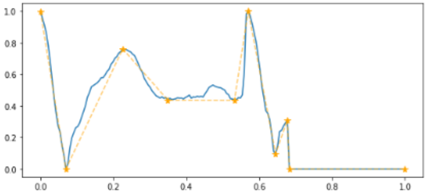
Arrhythmia identification



(c) Premature ventricular contraction



(d) Fusion of ventricular and normal

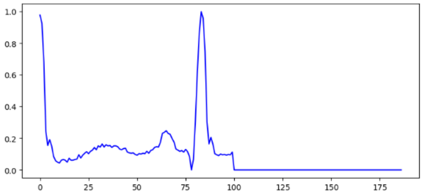


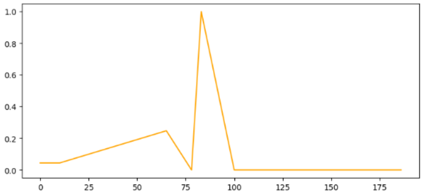
(e) Unclassifiable beat

Arrhythmia identification

But can we be really sure that this is the most important information for the network?

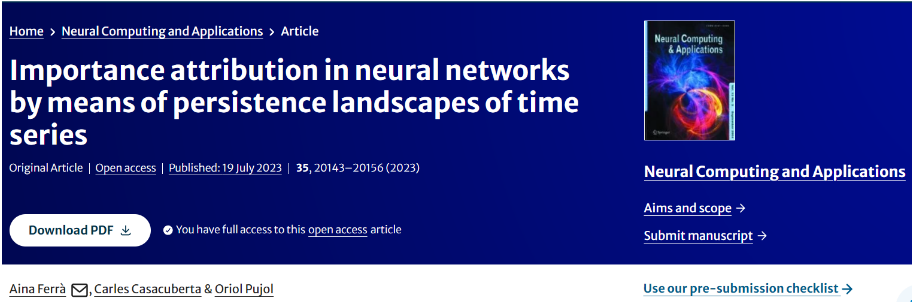
Actually yes! We can feed the neural network with only the **simplified version of the hearbeats** and see if it still works properly.





The network achieves a 96% of accuracy!

Time series classification


Home > Neural Computing and Applications > Article

Importance attribution in neural networks by means of persistence landscapes of time series

Original Article | Open access | Published: 19 July 2023 | **35**, 20143–20156 (2023)

Download PDF

You have full access to this open access article

Aina Ferrà, Carles Casacuberta & Oriol Pujol

Neural Computing and Applications

Aims and scope →

Submit manuscript →

Use our pre-submission checklist →

More details of this study can be found here.

A theoretical discussion about reconstruction can be found here.