The theory of group equivariant non-expansive operators as a possible bridge between TDA and Deep Learning

Patrizio Frosini
(Joint work with Mattia Bergomi (Fundação Champalimaud, PT), Daniela Giorgi (ISTI-CNR, IT) and Nicola Quercioli (University of Bologna, IT))

Department of Mathematics and ARCES, University of Bologna
patrizio.frosini@unibo.it

Valencia, ICIAM, July 15, 2019

Outline

A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

A recent paper

In this talk we will illustrate the content of this paper:

M. G. Bergomi, P. Frosini, D. Giorgi, N. Quercioli,
Towards a topological-geometrical theory of group equivariant non-expansive operators for data analysis and machine learning,
https://arxiv.org/pdf/1812.11832.pdf

The role of equivariant operators in machine learning

- As pointed out by several authors (Mallat, Poggio, Rosasco...) the role of equivariant operators in machine learning is getting more and more important.
- The comparison of DATA is always a process depending on an agent/observer. We could say that data comparison consists in the study of the relationship between an agent or observer and the reality he/she can MEASURE. In our setting, data coincide with measurements and agents/observers are represented by equivariant operators.

What does MEASUREMENT mean?

Before proceeding, we have to determine what measurements are in our mathematical model.

Measurement is the assignment of a number to a characteristic of an object or event, which can be compared with other objects or events.

WIKIPEDIA

According to this definition, measurements (and hence data) can be seen as functions φ associating a real number $\varphi(x)$ with each point x of a set X of characteristics. (This definition admits a natural extension to vector-valued functions but, for the sake of simplicity, we will treat here the case of scalar-valued functions).

Assumptions in our model

- Data are represented as functions defined on topological spaces, since only data that are stable w.r.t. a certain criterion (e.g., with respect to some kind of measurement) can be considered for applications, and stability requires a topological structure.
- Data cannot be studied in a direct and absolute way. They are only knowable through acts of transformation made by an agent/observer. From the point of view of data analysis, only the pair (data, agent) matters. In general terms, agents are not endowed with purposes or goals: they are just ways and methods to transform data. Acts of measurement are a particular class of acts of transformation.

Assumptions in our model

- Agents are described by the way they transform data while respecting some kind of invariance. In other words, any agent can be seen as a group equivariant operator acting on a function space.
- Data similarity depends on the output of the considered agent.

A topology on the space X of characteristics

Since we wish to develop a theory that can be applied in real situations, we need stability with respect to noise. This naturally leads us to use a topology on the set Φ of possible measurements on a set X. A natural topology on the set Φ of possible measurements is the one induced by the L^∞ metric $D_\Phi(\varphi_1, \varphi_2) := \|\varphi_1 - \varphi_2\|_\infty$.

Since measurements are the central concept in our approach, the topology on X is derived from D_Φ.
We define this pseudometric D_X on X by setting

$$D_X(x_1, x_2) := \sup_{\varphi \in \Phi} |\varphi(x_1) - \varphi(x_2)|.$$

In plain words: Two points $x_1, x_2 \in X$ are close to each other if and only if every measurement in Φ takes similar values at those points.

Every function in Φ is continuous

In this talk we will assume that the topological space Φ is compact.

EXAMPLE 1. $X := S^1 = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, $\Phi =$ set of all 1-Lipschitzian functions from S^1 to $[0,1]$.

EXAMPLE 2. $X := [-1,1] \times [-1,1]$, $\Phi =$ set of all functions from X to $[0,1]$ that are 1-Lipschitzian both in $X_1 := [-1,0] \times [-1,1]$ and in $X_2 := (0,1] \times [-1,1]$. Please observe that the functions in Φ can be discontinuous at the points (x,y) with $x = 0$, with respect to the Euclidean topology on X. However, every function in Φ is continuous with respect to the topology induced by D_X.

Theorem

The topology induced by D_X is the initial topology on X, i.e. the coarsest topology on X such that each function $\varphi \in \Phi$ is continuous.

Homeomorphisms with respect to D_X

The next step consists in understanding what a Φ-preserving homeomorphism with respect to D_X is (a bijection $g : X \to X$ is called Φ-preserving if $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$).

Theorem

The Φ-preserving homeomorphisms with respect to D_X are exactly the Φ-preserving bijections from X to X.

Let us now consider a group G of homeomorphisms from X to X, whose elements preserves Φ by right composition.

We will say that (Φ, G) is a PERCEPTION PAIR.

A pseudo-metric on our Φ-preserving group G

If a perception pair (Φ, G) is given, we can define the function

$$D_G(g_1, g_2) = \sup_{\varphi \in \Phi} D_\Phi(\varphi \circ g_1, \varphi \circ g_2) \tag{0.1}$$

from $G \times G$ to $\mathbb{R}$.
The function D_G is a pseudo-metric on G.
Please note that also the definition of D_G is inherited from the definition of D_Φ.

Theorem

G is a topological group with respect to the pseudo-metric topology and the action of G on Φ through right composition is continuous.

Compactness of X and G

We recall that we are assuming Φ compact.

Theorem

If X is complete then it is also compact with respect to D_X.

Theorem

If G is complete then it is also compact with respect to D_G.

In this talk we will assume that X and G are complete, and hence compact.

A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

Our ground truth: the natural pseudo-distance d_G

Definition

The pseudo-distance $d_G : \Phi \times \Phi \to \mathbb{R}$ is defined by setting

$$d_G(\varphi_1, \varphi_2) = \inf_{g \in G} D_\Phi(\varphi_1, \varphi_2 \circ g).$$

It is called the natural pseudo-distance associated with the group G.

If $G = \{Id : x \mapsto x\}$, then d_G equals the sup-norm distance D_Φ on Φ.
If G_1 and G_2 are groups of Φ-preserving self-homeomorphisms of X and $G_1 \subseteq G_2$, then the definition of d_G implies that

$$d_{G_2}(\varphi_1, \varphi_2) \leq d_{G_1}(\varphi_1, \varphi_2) \leq D_\Phi(\varphi_1, \varphi_2)$$

for every $\varphi_1,\ \varphi_2 \in \Phi$.

Our ground truth: the natural pseudo-distance d_G

The natural pseudo-distance d_G is our ground truth: it describes the differences that the agent/observer can perceive between the measurements in Φ with respect to the equivalence expressed by the group G.

A possible objection: *"The use of the concept of homeomorphism makes the natural pseudo-distance d_G difficult to apply. For example, in shape comparison two similar objects can be non-homeomorphic, hence this pseudo-metric cannot be applied to real problems."*

A possible objection

Answer: **the homeomorphisms do not concern the "objects" but the space X where the measurements are made**.

- For example, if we are interested in grey level images, the domain of our measurements can be modelled as the real plane and each image can be represented as a function from $\mathbb{R}^2$ to $\mathbb{R}$. Therefore, the space X is not given by the (possibly non-homeomorphic) objects displayed in the pictures, but by the topological space $\mathbb{R}^2$.
- If we make two CAT scans, the topological space X is always given by an helix turning many times around a body, and no requirement is made about the topology of such a body.

In other words, it is usually legitimate to assume that the topological space X is determined only by the measuring instrument we are using to get our measurements.

A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

Group equivariant non-expansive operators

The natural pseudo-distance d_G represents our ground truth.

Unfortunately, d_G is difficult to compute. This is also a consequence of the fact that we can easily find subgroups G of $\mathrm{Homeo}(X)$ that cannot be approximated with arbitrary precision by smaller finite subgroups of G (i.e. $G =$ group of rigid motions of $X = \mathbb{R}^3$).

Nevertheless, in this talk we will show that d_G can be approximated with arbitrary precision by means of a **DUAL** approach based on persistent homology and group equivariant non-expansive operators (GENEOs).

The space of GENEOs

Definition

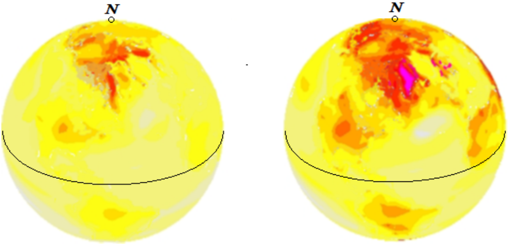
Assume that (Φ, G), (Ψ, H) are two perception pairs and that a homomorphism $T : G \to H$ has been fixed. A *Group Equivariant Non-Expansive Operator (GENEO) from* (Φ, G) *to* (Ψ, H) is a map $F : \Phi \to \Psi$ such that the following properties hold for every $\varphi_1, \varphi_2 \in \Phi$:

1. $F(\varphi \circ g) = F(\varphi) \circ T(g)$ for every $g \in G$;
2. $D_\Psi(F(\varphi_1), F(\varphi_2)) \leq D_\Phi(\varphi_1, \varphi_2)$.

We will use the symbol $\mathscr{F}^{\mathrm{all}}$ to denote the set of all GENEOs from (Φ, G) to (Ψ, H) with respect to T.

An example of GENEO

We give an example of the use of the definition of GENEO between two different perception pairs (Φ, G), (Ψ, H).
Let us assume to be interested in the comparison of the distributions of temperatures on a sphere, taken at two different times:



Let us also imagine that only two opposite points N, S can be localized on the sphere.

An example of GENEO

In this case we can set

- $X = S^2$
- Φ = set of 1-Lischitzian functions from S^2 to a fixed interval $[a, b]$
- G = group of rotations of S^2 around the axis $N - S$

We can also consider the "equator" of our sphere, represented as the space S^1.

Therefore, we can also set

- Y = the equator S^1 of S^2
- Ψ = set of 1-Lischitzian functions from S^1 to $[a, b]$
- H = group of rotations of S^1

An example of GENEO

In this case we can build a simple example of GENEO from (Φ, G) to (Ψ, H) by setting

- $T(g)$ equal to the rotation $h \in H$ of the equator S^1 that is induced by the rotation g of S^2, for every $g \in G$.
- $F(\varphi)$ equal to the function ψ that takes each point y belonging to the equator S^1 to the average of the temperatures along the meridian containing y, for every $\varphi \in \Phi$;

We can easily check that F verifies the properties defining the concept of group equivariant non-expansive operator with respect to the homomorphism $T : G \to H$.

A pseudo-metric for the space $\mathscr{F}^{\mathrm{all}}$

The following pseudo-metric is of use on $\mathscr{F}^{\mathrm{all}}$:

Definition

If $F_1, F_2 \in \mathscr{F}^{\mathrm{all}}$, we set

$$D_{\mathrm{GENEO}}(F_1, F_2) := \sup_{\varphi \in \Phi} D_{\Psi}(F_1(\varphi), F_2(\varphi)). \tag{0.2}$$

Some good news

Let $\mathscr{F}^{\mathrm{all}}$ be the set of all GENEOs from (Φ, G) to (Ψ, H) with respect to a fixed homomorphism $T : G \to H$.

Theorem

$\mathscr{F}^{\mathrm{all}}$ **is compact** *with respect to* D_{GENEO}.

Corollary

$\mathscr{F}^{\mathrm{all}}$ *can be* ε*-approximated by a finite subset for every* $\varepsilon > 0$.

Theorem

If Ψ *is convex, then* $\mathscr{F}^{\mathrm{all}}$ **is convex** .

A link with TDA - 1

Persistent homology enters this theoretical framework by means of an equality allowing us to approximate the natural pseudo-distance:

Theorem

If $(\Phi, G) = (\Psi, H)$, then

$$d_G(\varphi_1, \varphi_2) = \sup_{F \in \mathscr{F}^{\mathrm{all}}} d_{\mathrm{match}}\left(\mathrm{Dgm}(F(\varphi_1)), \mathrm{Dgm}(F(\varphi_2))\right)$$

where $\mathrm{Dgm}(F(\varphi))$ is the persistence diagram of the function $F(\varphi)$ and d_{match} is the usual bottleneck distance.

A link with TDA - 2

The computational machinery developed in persistent homology can be used in our mathematical approach as a proxy for the fast comparison of GENEOs, by replacing $D_{\mathrm{GENEO}}(F_1, F_2)$ with the pseudo-metric

$$\Delta_{\mathrm{GENEO}}(F_1, F_2) := \sup_{\varphi \in \Phi} d_{\mathrm{match}}\left(\mathrm{Dgm}(F_1(\varphi)), \mathrm{Dgm}(F_2(\varphi))\right).$$

The following result immediately follows from the stability of persistence diagrams:

Proposition

$$\Delta_{\mathrm{GENEO}}(F_1, F_2) \leq D_{\mathrm{GENEO}}(F_1, F_2).$$

Open questions

After defining an agent/observer as a collection of GENEOs, our purpose consists in looking for methods to approximate the agent/observer by a finite (and possible small) set of simple GENEOs. This leads us to the following open questions:

- How can we build a good library of GENEOs?
- How can we find a method to choose a finite set $\mathscr{F}^*$ of GENEOs that allows for both a good approximation of the natural pseudo-distance d_G and a fast computation?
- How can we provide a suitable statistical theory for group equivariant non-expansive operators?
- In which cases can the set of GENEOs be equipped with the structure of a Riemannian manifold?
- Could we compose operators to form networks, in the same way as computational units are connected in an artificial neural network?

Conclusions

- In our model, data comparison is based on measurements made by an agent/observer. Each measurement can be represented as a function defined on a topological space X.
- The agent/observer can be seen as and approximated by a collection of GENEOs, applied to the measurements. The operators are allowed to change both the space of measurements and the invariance group.
- Persistent homology provides an efficient way for the comparison of GENEOs.
- **The topological space of GENEOs deserves further research.**


Thanks for your attention!