Geometric observations for machine learning and artificial intelligence

Patrizio Frosini

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

Bologna, 3 May, 2024

Outline

The key role of observers in data analysis

Topological and metric basics for the theory of GENEOs

Building linear and nonlinear GENEOs

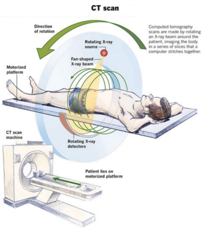
Intelligence and contradiction

How can we use GENEOs in applications?

Data can be often regarded as functions

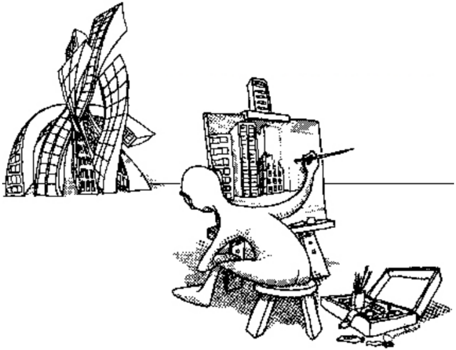
Some examples of data that can be seen as functions:

- An electrocardiogram (a function from $\mathbb{R}$ to $\mathbb{R}$);
- A gray-level image (a function from $\mathbb{R}^2$ to $\mathbb{R}$);
- A computerized tomography scan (a function from a helix to $\mathbb{R}$).



Data are processed by observers

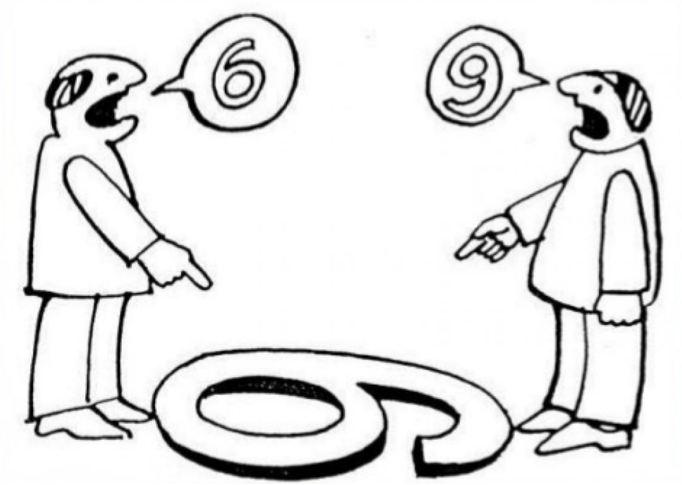
Data have no meaning if no observer elaborates them.



An observer is an agent that transforms data while respecting its symmetries.

Observers are variables in data analysis

Data interpretation strongly depends on the chosen observer:

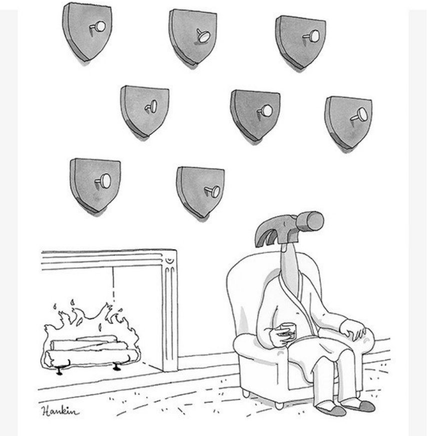


Our interest in data is greatly overrated

We are rarely directly interested in the data, but rather in how observers react to their presence.

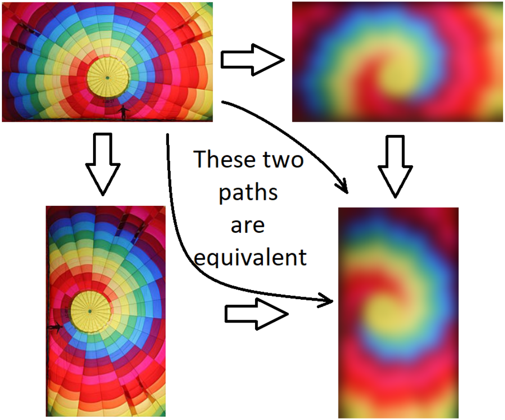
No data structure

Generally speaking, there is no structure in data. The structure of data is a projection of the structure of the observer.



Representing observers as equivariant operators

Observers are structures able to change data into other data, and usually do that by respecting some data symmetries, i.e., by commuting with some transformations (**equivariance**).



Representing observers as equivariant operators

As a first approximation, observers can be represented as **Group Equivariant Operators** (**GEOs**).
In this talk we will illustrate some results on the theory of **Group Equivariant Non-Expansive Operators** (**GENEOs**).

Why "non-expansive"?
Because

1. observers are often assumed to simplify the metric structure of data in order to produce meaningful interpretations;
2. non-expansiveness guarantees good topological properties.

How could we represent observers?

nature machine intelligence

ARTICLES

https://doi.org/10.1038/s42256-019-0087-3

Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning

Mattia G. Bergomi[1], Patrizio Frosini[2,3]*, Daniela Giorgi[4] and Nicola Quercioli[2,3]

We provide a general mathematical framework for group and set equivariance in machine learning. We define group equivariant non-expansive operators (GENEOs) as maps between function spaces associated with groups of transformations. We study the topological and metric properties of the space of GENEOs to evaluate their approximating power and set the basis for general strategies to initialize and compose operators. We define suitable pseudo-metrics for the function spaces, the equivariance groups and the set of non-expansive operators. We prove that, under suitable assumptions, the space of GENEOs is compact and convex. These results provide fundamental guarantees in a machine learning perspective. By considering isometry-equivariant non-expansive operators, we describe a simple strategy to select and sample operators. Thereafter, we show how selected and sampled operators can be used both to perform classical metric learning and to inject knowledge in artificial neural networks.

https://rdcu.be/bP6HV

All begins with the space of admissible functions

Let X be a nonempty set. Let Φ be a topological subspace of the set $\mathbb{R}_b^X$ of all bounded functions φ from X to $\mathbb{R}$, endowed with the topology induced by the metric

$$D_\Phi(\varphi_1, \varphi_2) := \|\varphi_1 - \varphi_2\|_\infty .$$

We can see X as the space where we can make our measurements, and Φ as the space of all possible measurements. We will say that Φ is the *set of admissible functions*. In other words, Φ **is the set of all functions from X to $\mathbb{R}$ that can be produced by our measuring instruments (or by other observers)**. For example, a gray-level image can be represented as a function from the real plane to the interval $[0,1]$ (in this case $X = \mathbb{R}^2$).

Perception pairs

Let us consider a group G of bijections $g : X \to X$ such that $\varphi \in \Phi \implies \varphi \circ g \in \Phi$ for every $\varphi \in \Phi$. We say that (Φ, G) is a perception pair.

The choice of a perception pair states which data can be considered as legitimate measurements (the functions in Φ) and which group represents the symmetries between data (the group G).

To proceed, we need to introduce suitable topologies on X and G. Before doing that, we recall that the initial topology τ_{in} on X with respect to Φ is the coarsest topology on X such that every function φ in Φ is continuous.

A pseudo-metric on X

Let us define on X the pseudo-metric

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

Recall that a pseudo-metric is just a distance d *without* the property $d(x_1, x_2) = 0 \implies x_1 = x_2$.

D_X induces a topology τ_{D_X} on X.
The use of D_X implies that we can distinguish two points only if a measurement exists, taking those points to different values.

Proposition

The topology τ_{D_X} is finer than the initial topology τ_{in} on X with respect to Φ. If Φ is totally bounded, then τ_{D_X} coincides with τ_{in}.

A pseudo-metric on X

The following properties are of use in our model.

Proposition

Every function in Φ is non-expansive, and hence continuous.

Proposition

If Φ is compact and X is complete, then X is compact.

In the following, we will usually assume that Φ is compact and X is complete (and hence compact).

Some magic happens: each bijection is an isometry

- $\mathrm{Bij}_\Phi(X) = \{$**bijections** $g: X \to X$ s.t. $\Phi \circ g, \Phi \circ g^{-1} \subseteq \Phi\}$;
- $\mathrm{Homeo}_\Phi(X) = \{$**homeomorphisms** $g: X \to X$ s.t. $\Phi \circ g, \Phi \circ g^{-1} \subseteq \Phi\}$;
- $\mathrm{Iso}_\Phi(X) = \{$**isometries** $g: X \to X$ s.t. $\Phi \circ g, \Phi \circ g^{-1} \subseteq \Phi\}$.

Proposition

$\mathrm{Bij}_\Phi(X) = \mathrm{Homeo}_\Phi(X) = \mathrm{Iso}_\Phi(X)$.

A pseudo-metric on G

Let us now focus our attention on a subgroup G of $\mathrm{Homeo}_\Phi(X)$.
We can define a pseudo-metric D_G on G by setting

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

Theorem

G is a topological group with respect to D_G and the action of G on Φ by right composition is continuous.

Theorem

If Φ is compact and G is complete, then G is compact.

GEOs and GENEOs

Each pair (Φ, G) with $G \subseteq \mathrm{Homeo}_\Phi(X)$ is called a *perception pair*.

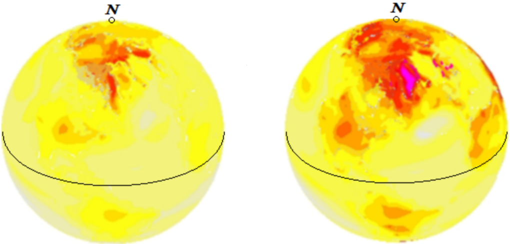
Let us assume that two perception pairs (Φ, G), (Ψ, H) are given, and fix a group homomorphism $T : G \to H$.

Each function $F : \Phi \to \Psi$ such that $\boxed{F(\varphi \circ g) = F(\varphi) \circ T(g)}$ for every $\varphi \in \Phi, g \in G$ is called a *Group Equivariant Operator (GEO)* associated with the homomorphism T.

If F is also non-expansive (i.e., $\boxed{D_\Psi(F(\varphi_1), F(\varphi_2)) \leq D_\Phi(\varphi_1, \varphi_2)}$ for every $\varphi_1, \varphi_2 \in \Phi$), then F is called a *Group Equivariant Non-Expansive Operator (GENEO)* associated with the homomorphism T.

An example of GENEO

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 assume that only two opposite points N, S can be localized on the sphere.

An example of GENEO

Let us introduce two perception pairs $(\Phi, G), (\Psi, H)$ by setting

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

and

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

An example of GENEO

This is a simple example of GENEO from (Φ, G) to (Ψ, H):

- $T(g)$ is the rotation $h \in H$ of the equator S^1 that is induced by the rotation g of S^2, for every $g \in G$.
- $F(\varphi)$ is 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 isomorphism $T : G \to H$.

In plain words, our GENEO simplifies the data by transforming "temperature distributions on the earth" into "temperature distributions on the equator".

Two key results (and two good news for applications)

Let us assume that a homomorphism $T : G \to H$ has been fixed.
Let us define a metric D_{GENEO} on $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ by setting

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

Theorem

If Φ *and* Ψ *are* **compact**, *then* $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ *is* **compact** *with respect to* D_{GENEO}.

Theorem

If Ψ *is* **convex**, *then* $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ *is* **convex**.

Two key observations (1)

- While the space of data is often non-convex (and hence averaging data does not make sense), the assumption of convexity of Ψ implies the convexity of the space of observers and allows us to consider the "average of observers".

Two key observations (2)

- Our main goal is to develop a good geometric and compositional theory to approximate an ideal observer. In our model, "to approximate an observer" means to look for a GENEO F that minimizes a suitable "cost function" $c(F)$. The cost function quantifies the error that is committed by taking the GENEO F instead of the ideal observer. Since the space of GENEOs is compact and convex (under the assumption that the data spaces are compact and convex), if the cost function $c(F)$ is strictly convex we have that there is one and only one GENEO that best approximates the ideal observer.

Elementary methods to build GENEOs

In order to use our model profitably we need constructive methods to produce GENEOs in the presence of pre-established data and equivariance groups.

Without going into technical details, here we simply observe that under reasonable assumptions

- the composition of GENEOs is still a GENEO;
- the maximum and the minimum of GENEOs are still GENEOs;
- the translation of a GENEO is still a GENEO;
- the convex combination of GENEOs is still a GENEO.

(But there's much more than that...)

Permutant measures

Let us consider the set $\Phi = \mathbb{R}^X \cong \mathbb{R}^n$ of all functions from a finite set $X = \{x_1, \ldots, x_n\}$ to $\mathbb{R}$, and a subgroup G of the group $\mathrm{Bij}(X)$ of all permutations of X.

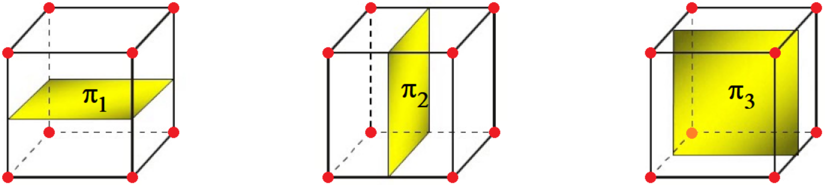
Definition

A finite (signed) measure μ on $\mathrm{Bij}(X)$ is called a *permutant measure* with respect to G if every <u>subset</u> H of $\mathrm{Bij}(X)$ is measurable and μ is invariant under the conjugation action of G (i.e., $\mu(H) = \mu(gHg^{-1})$ for every $g \in G$).

An example of permutant measure

Let us consider the set X of the vertices of a cube in $\mathbb{R}^3$, and the group G of the orientation-preserving isometries of $\mathbb{R}^3$ that take X to X. Let π_1, π_2, π_3 be the three planes that contain the center of mass of X and are parallel to a face of the cube. Let $h_i : X \to X$ be the orthogonal symmetry with respect to π_i, for $i \in \{1,2,3\}$.

We can now define a permutant measure μ on the group $\mathrm{Bij}(X)$ by setting $\mu(h_1) = \mu(h_2) = \mu(h_3) = c$, where c is a positive real number, and $\mu(h) = 0$ for any $h \in \mathrm{Bij}(X)$ with $h \notin \{h_1, h_2, h_3\}$.



Representation Theorem for linear GENEOs

Theorem (Representation Theorem for linear GENEOs)

Let us assume that $G \subseteq \mathrm{Bij}(X)$ *transitively acts on the finite set* X *and that* F *is a map from* $\mathbb{R}^X$ *to* $\mathbb{R}^X$. *The map* F *is a linear GENEO from* $\mathbb{R}^X$ *to* $\mathbb{R}^X$ *with respect to the identical homomorphism* $\mathrm{id}_G : g \mapsto g$ **if and only if** *a permutant measure* μ *with respect to* G *exists, such that* $F(\varphi) = \sum_{h \in \mathrm{Bij}(X)} \varphi h^{-1} \; \mu(h)$ *for every* $\varphi \in \mathbb{R}^X$, *and* $\sum_{h \in \mathrm{Bij}(X)} |\mu(h)| \leq 1$.

Intelligence and contradiction

The representation of observers as functional operators has another important consequence: a sort of "principle of contradiction".

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 10 (2009) 297–315

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Does intelligence imply contradiction?

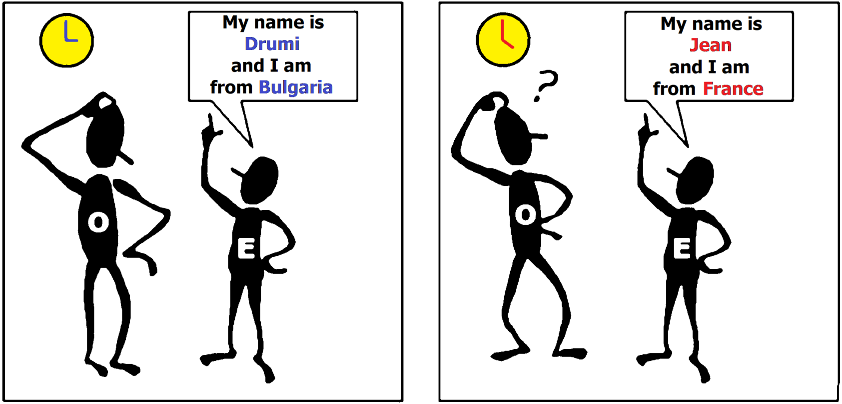
Action editor: Vasant Honavar

P. Frosini *

Department of Mathematics and ARCES, University of Bologna, I-40126 Bologna, Italy

Intelligence and contradiction

What do we mean by contradiction?



Intelligence and contradiction

Every sufficiently intelligent entity is contradictory

Ludwig Josef Johann Wittgenstein

Tractatus
Logico-Philosophicus

By
LUDWIG WITTGENSTEIN

With an Introduction by
BERTRAND RUSSELL, F.R.S.

NEW YORK
HARCOURT, BRACE & COMPANY, INC.
LONDON: KEGAN PAUL, TRENCH, TRUBNER & CO., LTD.
1922

ı proposizione.

Intelligence and contradiction

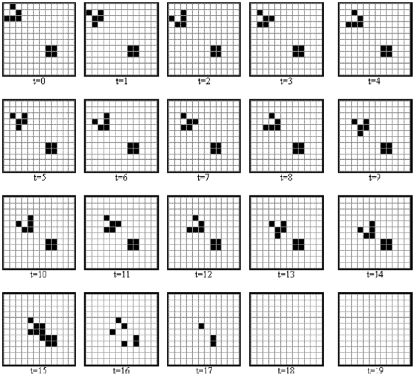
Equivalently, we can say that

The behavior of any sufficiently intelligent entity is unpredictable.

Intelligence and contradiction

How can we prove that?

We can use an approach based on cellular automata.



Intelligence and contradiction

Sketch of proof:

- An observer is identified (understood as an operator who transforms the functions that represent the states of the cellular automaton into functions that describe the perceived entity and the environment that surrounds it).
- The intelligence of an entity is defined as its ability to survive in the environment according to the judgment of the observer.
- It is proved that there is a threshold for intelligence (dependent on the number of states that the observer can associate with the entity and the environment), beyond which the observed entity appears necessarily contradictory to the chosen observer.

In this model, contradictoriness and non-predictability do not appear as limitations of intelligent structures but as necessary conditions for the development of complex intellectual behaviors.

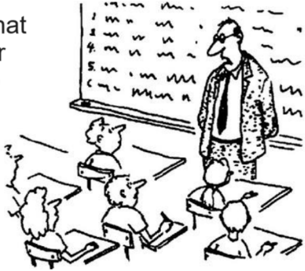
Intelligence and contradiction

Theorem. Let E be an entity with a finite lifespan and assume that its environment is deterministic. If the intelligence of E is greater than the product of the cardinalities of the sets P_{ent} and P_{ENV} the entity must necessarily be contradictory.

ATTENTION! The theorem does not assert that intelligent entities must change their behavior (this fact is obvious) but that they must do so without the observer understanding why.

P_{ent} = set of states of E recognized by the observer.

P_{ENV} = set of environmental states recognized by the observer.



"I expect you all to be independent, innovative, critical thinkers who will do exactly as I say!"

Intelligence and contradiction

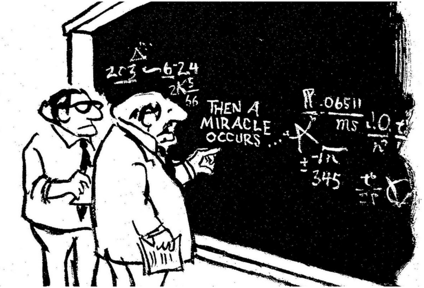
A precise formulation of this approach can be found here:

P. Frosini, Does intelligence imply contradiction?, Cognitive Systems Research, vol. 10 (2009), n. 4, 297-315.

(A synthetic and beautiful slideshow of this paper has been made by Mattia G. Bergomi. It is available at the link *https://mgbergomi.github.io/Contradiction/*.)

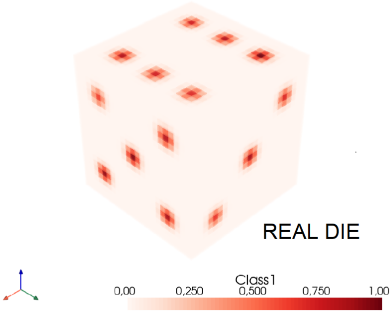
Intelligence and contradiction

We have seen that every agent A appears unpredictable in the eyes of a fixed observer if the *"intelligence"* of A exceeds a threshold expressed by the product of the number of states that the observer can perceive in the agent and in the environmental context. This implies that to have predictability of behavior it is necessary to choose models in which the aforementioned threshold is greater than the desired intelligence value.



What happens when we apply GENEOs to our data?

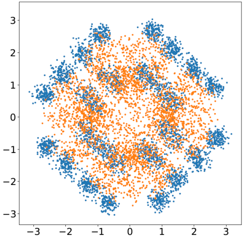
An example of use: comparison between real dice and fake dice.



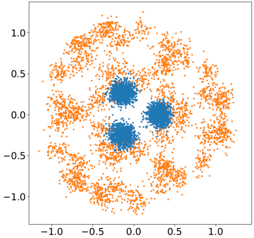
(Experiment and computations by Giovanni Bocchi)

What happens to data when we apply GENEOs?

We produced 10000 dice (a training set of size 7000 and a test set of size 3000), then we applied PCA to the test set and to the test set transformed by a suitable GENEO, optimized on the training set:



WITHOUT GENEOS



WITH GENEOS

For each die the first two principal components are plotted. Blue points are associated with **real dice**, while orange ones with **fake dice**. The GENEO we use was built by a convex combination of 3 GENEOs defined by permutant measures.

A real application: finding pockets in proteins

GENEOnet: A new machine learning paradigm based on Group Equivariant Non-Expansive Operators. An application to protein pocket detection.

Giovanni Bocchi [1], Patrizio Frosini [2], Alessandra Micheletti [1], Alessandro Pedretti [3] Carmen Gratteri [4], Filippo Lunghini [5] , Andrea Rosario Beccari [5] and Carmine Talarico [5]

[1] Department of Environmental Science and Policy, Università degli Studi di Milano

[2] Department of Mathematics, Università degli Studi di Bologna

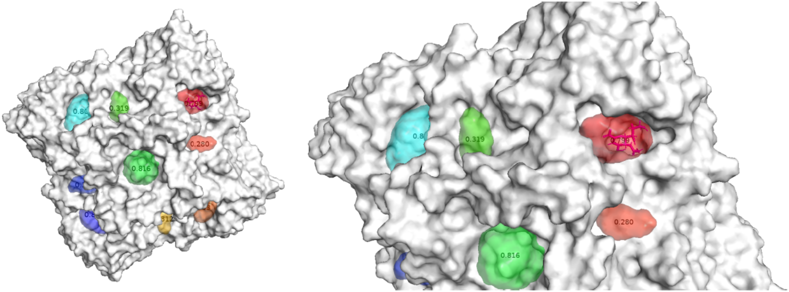
[3] Department of Pharmaceutical Sciences, Università degli Studi di Milano

[4] Dipartimento di Scienze della Salute, Università degli Studi "Magna Græcia di Catanzaro"

[5] Dompé Farmaceutici SpA

https://arxiv.org/ftp/arxiv/papers/2202/2202.00451.pdf

A real application: finding pockets in proteins



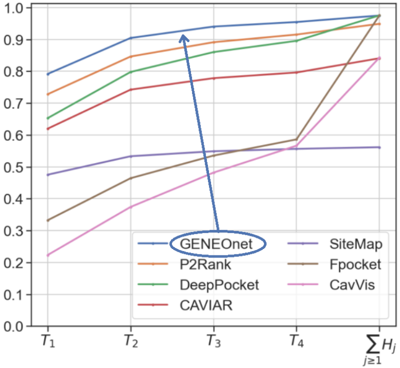
a) b)

Model predictions for protein 2QWE. In Figure a) the global view of the prediction is shown, where different pockets are depicted in different colors and are labelled with their scores. In Figure b) the zoomed of the pocket containing the ligand is shown.

The search for the pockets was carried out by identifying an optimal GENEO in the convex hull of 8 GENEOs (each focused on a particular property of the pockets).

A real application: finding pockets in proteins

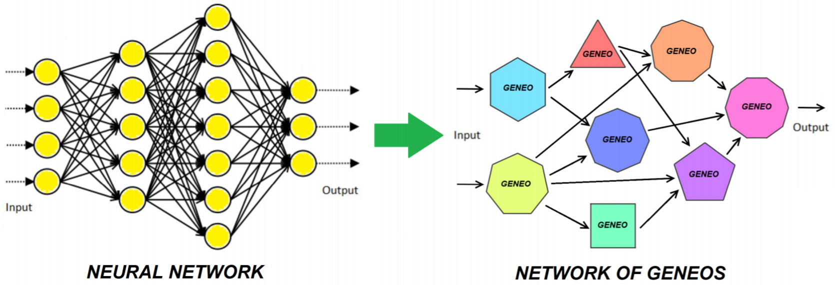
Here are the results of our experiments:



Please note that GENEOnet uses 17 parameters, while a CNN such as DeepPocket requires 665122 parameters.

The main point in our approach

In perspective, we are looking for a good compositional theory for building efficient and transparent networks of GENEOs.
Some preliminary experiments suggest that replacing neurons with GENEOs could make deep learning more transparent and interpretable and speed up the learning process.



GENEOs and Machine Learning

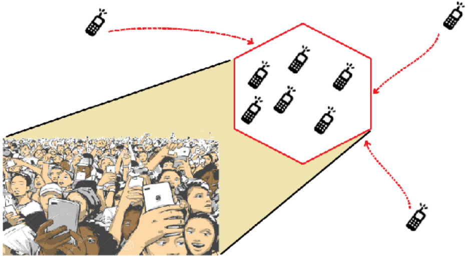
For more details about the use of GENEOs in Machine Learning:

- A. Micheletti, *A new paradigm for artificial intelligence based on group equivariant non-expansive operators*, In: *EMS Magazine*, Online First, 24 April 2023.
- `https://ems.press/content/serial-article-files/27673`

Current research projects

CNIT / WiLab - Huawei Joint Innovation Center (JIC)

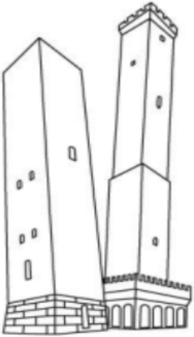
Project on GENEOs for 6G



Current research projects

Horizon Europe (HORIZON)
Call: HORIZON-CL4-2023-HUMAN-01-CNECT
Project: 101135775 — PANDORA
Funding: approximately 9 million euros.

Task 3.3 - Leveraging domain knowledge for explainable learning: This task aims to investigate the use of domain knowledge in the development of explainable AI models. Tools like GENEOs for applications in TDA and ML and new theoretical methods of GENEOs for explainable AI will be used.



Thanks for your attention!