Un'escursione dai numeri di Betti al deep learning geometrico passando per l'omologia persistente e l'analisi topologica dei dati

Patrizio Frosini

Dipartimento di Matematica, ARCES, Alma Human-AI, AM^2 - Università di Bologna
patrizio.frosini@unibo.it

Scuola Normale Superiore, Pisa, 20 ottobre 2023

Sommario

Scopo del seminario

Lo scopo di questo seminario è illustrare, piuttosto informalmente, il lungo percorso che ha portato dalle idee di Betti alla Topological Data Analysis e all'uso di quest'ultima nel Machine Learning.

Sommario

I numeri di Betti

Nell'ottobre 1863 Enrico Betti scriveva al suo collega Placido Tardy:

"Ho nuovamente parlato con Riemann della connessione degli spazi, e me ne sono fatta un'idea esatta. (...) La nozione delle sezioni è venuta in mente a Riemann per una definizione che gliene ha dato Gauss in un colloquio familiare, parlando di un altro soggetto."

Betti pubblicò le sue idee nell'articolo *Sopra gli spazi di un numero qualunque di dimensioni* (Annali di Matematica Pura ed Applicata, vol. 4, 140–158 (1870)), descrivendo delle quantità che verranno chiamate *numeri di Betti* da Henri Poincaré (1854-1912) nella celebre memoria *Analysis situs* (H. Poincaré, Analysis situs, Journal de l'École polytechnique, (II) 1, 1-121 (1895)).

I numeri di Betti

Sopra gli spazi di un numero qualunque di dimensioni.

(*del prof.* ENRICO BETTI, *a Pisa.*)

1.

Siano $z_1, z_2, \dots z_n$ n variabili che possono prendere tutti i valori reali da $-\infty$ a $+\infty$. Il campo n volte infinito dei sistemi di valori di queste variabili lo diremo uno spazio di n dimensioni e lo dinoteremo con S_n. Un sistema $(z_1^0, z_2^0, \dots, z_n^0)$ determinerà un punto L_0 di questo spazio, e $z_1^0, z_2^0, \dots z_n^0$ si diranno le coordinate di questo punto.

Un sistema di m equazioni determinerà un campo dei sistemi di valori di $n-m$ variabili indipendenti che sarà uno spazio S_{n-m} di altrettante dimensioni, contenuto in S_n. Uno spazio di una sola dimensione che forma una semplice continuità lo chiameremo una linea.

Sia:

$$F(z_1, z_2, \dots, z_n) = 0 \tag{1}$$

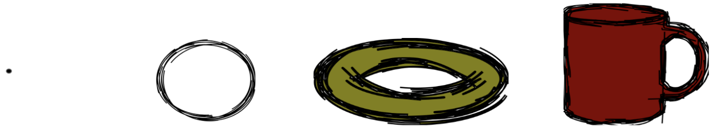
la equazione di uno spazio S_{n-1} di $n-1$ dimensioni. Se la funzione F è continua e ad un sol valore per tutti i valori reali delle coordinate, lo spazio S_{n-1} in generale separerà S_n in due regioni, in una delle quali sarà $F<0$, e nell'altra $F>0$; e non si potrà con variazioni continue dal sistema di valori delle coordinate di un punto della prima regione passare al sistema di valori delle coordinate di un punto dell'altra regione senza passare per un sistema di valori che sodisfaccia alla equazione (1). Le due regioni saranno due spazi di n dimensioni limitati dallo spazio S_{n-1}. Se dal sistema di valori delle coordinate di un punto qualunque di una delle due regioni si potrà sempre con variazione continua passare al sistema di valori delle coordinate di un altro punto qualunque della medesima regione senza passare per i valori delle coordinate di un punto di S_{n-1}, si dirà che questa regione è uno spazio connesso.

I numeri di Betti

Oltre la connessione lineare che si presenta sola nelle superficie, io ho osservato che negli spazi di un numero di dimensioni maggiore di due si possono considerare altre specie di connessioni.

Se in uno spazio R di n dimensioni limitato da uno o più spazi di $n-1$ dimensioni, ogni spazio chiuso di m dimensioni, essendo $m<n$, è il contorno di una parte di uno spazio linearmente connesso di $m+1$ dimensioni, tutta quanta contenuta in R, avremo una connessione secondo $m+1$ dimensioni, e diremo che R ha semplice la connessione di m^{esima} specie. Se uno spazio R ha semplici tutte le connessioni, diremo che è semplicemente connesso. Se invece in R si può imaginare un numero p_m di spazi chiusi di m dimensioni che non possano formare il contorno di una parte linearmente connessa di uno spazio di $m+1$ dimensioni, tutta quanta contenuta in R, e tali che ogni altro spazio chiuso di m dimensioni formi solo o con una parte di essi o con tutti il contorno di una parte linearmente connessa di uno spazio di $m+1$ dimensioni tutta quanta contenuta in R, diremo che R ha di $(p_m+1)^{\text{esimo}}$ ordine la connessione di m^{esima} specie.

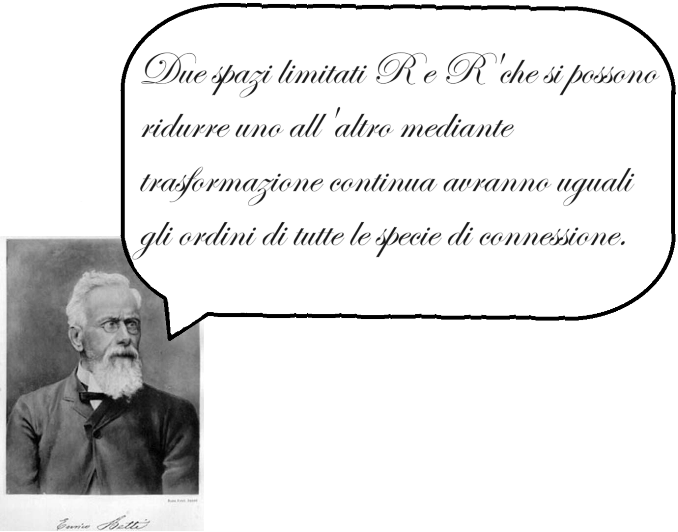
I numeri di Betti



$\beta_0 = 1$	$\beta_0 = 1$	$\beta_0 = 1$	$\beta_0 = 1$
$\beta_1 = 0$	$\beta_1 = 1$	$\beta_1 = 2$	$\beta_1 = 2$
$\beta_2 = 0$	$\beta_2 = 0$	$\beta_2 = 1$	$\beta_2 = 1$

Informalmente, il k-esimo numero di Betti conta (in modo opportuno) il numero di buchi k-dimensionali. Un "buco k-dimensionale" è un "ciclo k-dimensionale" che non è bordo di un oggetto $(k+1)$-dimensionale. Formalmente, il k-esimo numero di Betti è il rango del k-esimo gruppo di omologia.

A che serve tutto questo nell'analisi dei dati?



A che serve tutto questo nell'analisi dei dati?

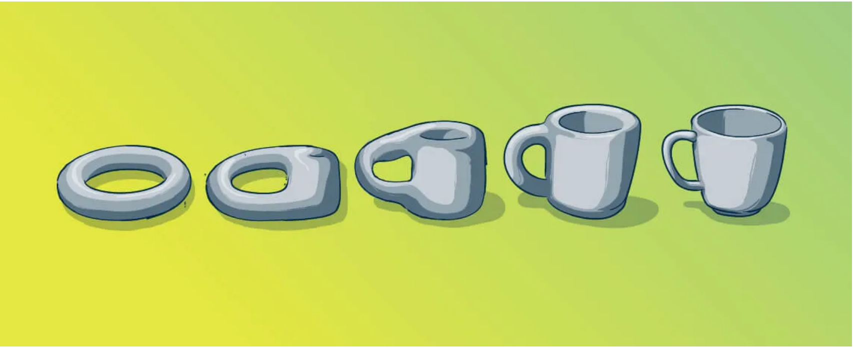
L'analisi dei dati riguarda lo *"studio della loro forma"*, cioè lo studio delle proprietà dei dati che sono invarianti in presenza di deformazioni prefissate e ammissibili.

Confrontare i numeri di Betti ci permette di distinguere oggetti che non si possono ottenere uno dall'altro mediante deformazioni continue (omeomorfismi).

Sommario

Andare oltre l'omologia: la Topological Data Analysis

Contare i buchi degli oggetti non è sempre sufficiente per distinguerli.



Andare oltre l'omologia: la Topological Data Analysis

Purtroppo, confrontare le forme non vuol dire semplicemente contare il numero di buchi presenti negli oggetti. Il numero di buchi è solo una delle tante proprietà che contribuiscono a definire il concetto di forma. Si tratta inoltre di una proprietà che non è stabile in presenza di rumore.

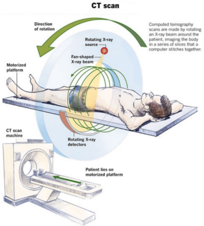
Un approccio probabilmente più evoluto è quello in cui la similitudine di forma viene espressa in termini metrici, tenendo conto che spesso consideriamo come uguali due oggetti se e solo se si possono ottenere uno dall'altro applicando una trasformazione appartenente a un gruppo fissato.

Ciò ci conduce al concetto di pseudo-distanza naturale *rispetto a un gruppo di omeomorfismi*. La differenza di forma è espressa e quantificata da tale distanza.

I dati possono essere spesso visti come funzioni

Alcuni esempi di dati che possono essere visti come funzioni:

- Un elettocardiogramma è una funzione da $\mathbb{R}$ a $\mathbb{R}$;
- Un'immagine a livelli di grigio è una funzione da $\mathbb{R}^2$ a $\mathbb{R}$;
- Una tomografia computerizzata è una funzione da un'elica a $\mathbb{R}$.



(... ma anche le "nubi di punti" e i grafi possono essere visti come funzioni).

La definizione della pseudo-distanza naturale d_G

Siano X e G un **spazio metrico compatto** e un **sottogruppo del gruppo** $\mathrm{Homeo}(X)$ di tutti gli omeomorfismi da X a X, rispettivamente. Se φ_1, φ_2 sono due funzioni continue e limitate da X a $\mathbb{R}$ possiamo considerare il valore $\boxed{\inf_{g \in G} \|\varphi_1 - \varphi_2 \circ g\|_\infty}$. Questo valore è chiamato **pseudo-distanza naturale** $d_G(\varphi_1, \varphi_2)$ fra φ_1 e φ_2 rispetto al gruppo G.

Ricordo che una pseudo-distanza è semplicemente una distanza d priva della proprietà $d(x,y) = 0 \implies x = y$.

Dotando $C^0(X, \mathbb{R})$ e G della topologia della convergenza uniforme, si ha che G diviene **un gruppo topologico che agisce con continuità su $C^0(X, \mathbb{R})$ per composizione a destra**.

La definizione della pseudo-distanza naturale d_G

Se G è il gruppo banale Id, allora d_G è la distanza indotta dalla norma del massimo $\|\varphi_1 - \varphi_2\|_\infty$. Inoltre, se G_1 e G_2 sono sottogruppi di $\mathrm{Homeo}(X)$ e $G_1 \subseteq G_2$, allora

$$d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2) \leq d_{G_2}(\varphi_1, \varphi_2) \leq d_{G_1}(\varphi_1, \varphi_2) \leq \|\varphi_1 - \varphi_2\|_\infty$$

per ogni $\varphi_1, \varphi_2 \in C^0(X, \mathbb{R})$.

La pseudo-metrica d_G viene usualmente ristretta a $\Phi \times \Phi$, dove Φ è un sottoinsieme limitato di $C^0(X, \mathbb{R})$, detto **insieme dei dati ammissibili**. Tale pseudo-distanza naturale d_G può essere vista come "verità di riferimento" e descrive le differenze che l'osservatore può percepire fra le misurazioni appartenenti a Φ, rispetto all'equivalenza espressa dal gruppo G.

Vediamo ora alcune interessanti proprietà di d_G.

Pseudo-distanza naturale nel caso di varietà

Teorema

Sia $\mathscr{M}$ una varietà chiusa di classe C^1 e supponiamo che $\varphi_1, \varphi_2 : \mathscr{M} \to \mathbb{R}$ siano funzioni C^1. Poniamo $d := d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$. Allora esiste un intero positivo k tale che valga una delle seguenti proprietà:

1) *k è dispari e kd è la distanza fra un valore critico di φ_1 e un valore critico di φ_2;*
2) *k è pari e kd è o la distanza fra due valori critici di φ_1 o la distanza fra due valori critici di φ_2.*

- P. Donatini, P. Frosini, *Natural pseudodistances between closed manifolds*, Forum Mathematicum, vol. 16 (2004), n. 5, 695-715.

Pseudo-distanza naturale nel caso di superfici

Teorema

Sia $\mathscr{S}$ *una superficie chiusa di classe* C^1 *e supponiamo che* $\varphi_1, \varphi_2 : \mathscr{S} \to \mathbb{R}$ *siano funzioni* C^1*. Poniamo* $d := d_{\mathrm{Homeo}(\mathscr{S})}(\varphi_1, \varphi_2)$*. Allora vale almeno una delle seguenti proprietà:*

1) *d è la distanza fra un valore critico di φ_1 e un valore critico di φ_2;*
2) *d è la metà della distanza fra due valori critici di φ_1;*
3) *d è la metà della distanza fra due valori critici di φ_2;*
4) *d è un terzo della distanza fra un valore critico di φ_1 e un valore critico di φ_2.*

- P. Donatini, P. Frosini, *Natural pseudodistances between closed surfaces*, Journal of the European Mathematical Society, vol. 9 (2007), n. 2, 331–353.

Pseudo-distanza naturale nel caso di curve

Teorema

Sia $\mathscr{C}$ una curva chiusa di classe C^1 e supponiamo che $\varphi_1, \varphi_2 : \mathscr{C} \to \mathbb{R}$ siano funzioni C^1. Poniamo $d := d_{\mathrm{Homeo}(\mathscr{C})}(\varphi_1, \varphi_2)$. Allora vale almeno una delle seguenti proprietà:

a) *d è la distanza fra un valore critico di φ_1 e un valore critico di φ_2;*

b) *d è metà della distanza fra due valori critici di φ_1;*

c) *d è metà della distanza fra due valori critici di φ_2.*

- P. Donatini, P. Frosini, *Natural pseudo-distances between closed curves*, Forum Mathematicum, vol. 21 (2009), Issue 6, 981–999.

Un'osservazione importante

La pseudo-distanza naturale può essere vista come la pseudo-metrica di riferimento nel nostro modello ma, sfortunatamente, calcolarla è difficile, poiché il gruppo G è spesso troppo grande.

Come possiamo ottenere informazioni sulla pseudo-distanza naturale?

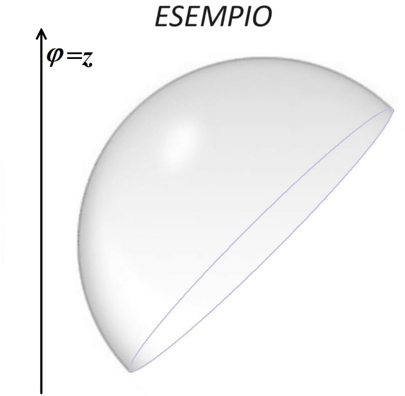
Abbiamo bisogno di un nuovo approccio, che trae origine dalle idee proposte da Betti e costituisce il principale strumento matematico nella Topological Data Analysis: **l'omologia persistente**.

In termini molto imprecisi, l'omologia persistente studia i buchi di uno spazio topologico rispetto a funzioni filtranti, intese come variabili. In un certo senso l'omologia persistente aggiunge all'omologia classica la scelta di un "punto di vista".

Sommario

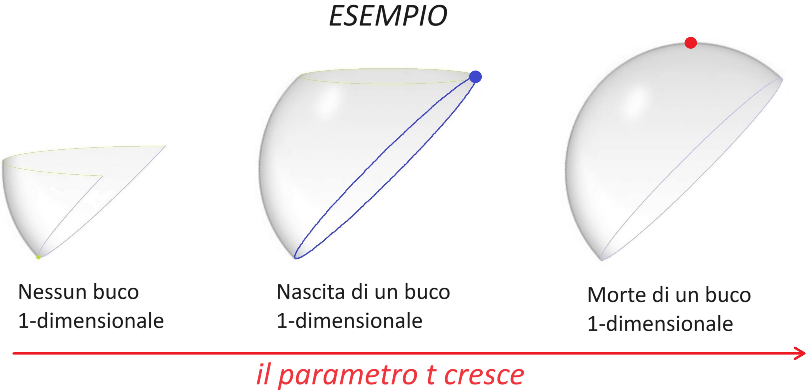
Cos'è l'omologia persistente?

Se $\varphi : X \to \mathbb{R}$ è una funzione continua, possiamo considerare gli insiemi di sottolivello $X_t := \{x \in X : \varphi(x) \leq t\}$. Quando t varia vediamo la nascita e la morte di buchi k-dimensionali.



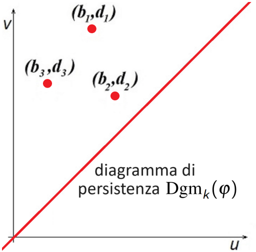
Cos'è l'omologia persistente?

Se $\varphi : X \to \mathbb{R}$ è una funzione continua, possiamo considerare gli insiemi di sottolivello $X_t := \{x \in X : \varphi(x) \leq t\}$. Quando t varia vediamo la nascita e la morte di buchi k-dimensionali.



Cos'è l'omologia persistente?

In parole povere, il **diagramma di persistenza** $\mathrm{Dgm}_k(\varphi)$ di φ in grado k è la collezione delle coppie (b_i, d_i) dove b_i e d_i sono i tempi di nascita e di morte dell'i-esimo buco k-dimensionale.



I punti del diagramma di persistenza sono dotati di molteplicità. Si assume, per definizione, che ogni punto della diagonale $u = v$ faccia parte del diagramma di persistenza e che abbia molteplicità infinita.

A cosa serve l'omologia persistente?

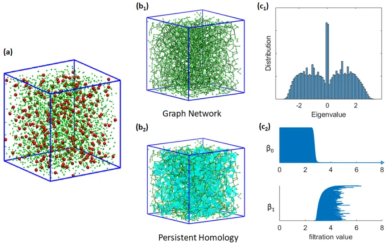
I diagrammi di persistenza sono il principale strumento matematico nella Topological Data Analysis (TDA). Permettono di descrivere e confrontare metricamente le "forme" degli spazi topologici e delle varietà rispetto alle funzioni filtranti considerate. Sono largamente utilizzati per studiare le informazioni in maniera resistente al rumore.

scientific reports

Weighted persistent homology for osmolyte molecular aggregation and hydrogen-bonding network analysis

D. Vijay Anand, Zhenyu Meng, Kelin Xia & Yuguang Mu

Scientific Reports **10**, Article number: 9685 (2020) | Cite this article



Cosa sono le funzioni dei numeri di Betti persistenti?

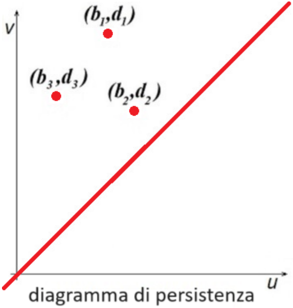
I diagrammi di persistenza non sono adatti per finalità statistiche, perché non esiste una **buona** definizione di **media** di diagrammi di persistenza.

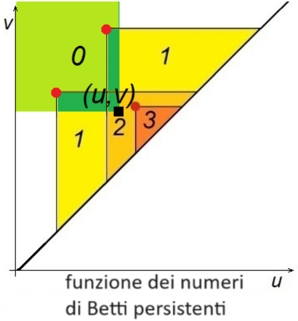
Le funzioni dei numeri di Betti persistenti (FNBP) descrivono essenzialmente la stessa informazione rappresentata dai diagrammi di persistenza ma permettono di introdurre il concetto di media e di fare elaborazioni di tipo statistico.

Definizione

La k-esima **funzione dei numeri di Betti persistenti** $\beta_k(u, v)$ è il numero di buchi k-dimensionali il cui tempo di nascita è più piccolo di u e il cui tempo di morte è più grande di v.

Cosa sono le funzioni dei numeri di Betti persistenti?





La funzione dei numeri di Betti persistenti calcolata in (u, v) è il numero dei punti del diagramma di persistenza che si trovano nella regione verde. Se usiamo l'omologia di Čech, i diagrammi di persistenza sono equivalenti alle funzioni dei numeri di Betti persistenti.

Cosa sono le funzioni dei numeri di Betti persistenti?

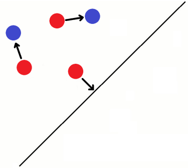
Formalmente:

Definizione

Sia $\varphi : X \to \mathbb{R}$ una funzione continua. Se $u, v \in \mathbb{R}$ e $u < v$, possiamo considerare l'inclusione i di X_u in X_v. Tale inclusione induce un omomorfismo $i^* : H_k(X_u) \to H_k(X_v)$ fra i gruppi di omologia di X_u e X_v in grado k. Il gruppo $PH_k^{\varphi}(u, v) := i^*(H_k(X_u))$ viene detto il ***k*-esimo gruppo di omologia persistente** *rispetto alla funzione* $\varphi : X \to \mathbb{R}$*, calcolato nel punto* (u, v). Il rango $r_k(\varphi)(u, v)$ di questo gruppo viene detto *la* ***k*-esima funzione dei numeri di Betti persistenti** *rispetto alla funzione* $\varphi : X \to \mathbb{R}$*, calcolata in* (u, v).

La media di funzioni di numeri di Betti persistenti viene definita come la media usuale di funzioni a valori reali.

Una metrica per i diagrammi di persistenza



$\boxed{d_{\mathrm{match}}(D_1, D_2)}$

Un modo per trasformare D_1 in D_2

I diagrammi di persistenza (e quindi le funzioni dei numeri di Betti persistenti) possono essere confrontati per mezzo della **distanza di bottleneck** d_{match}. La distanza di bottleneck $\boxed{d_{\mathrm{match}}(D_1, D_2)}$ fra due diagrammi di persistenza D_1, D_2 è il minimo dei costi dei matching fra i due diagrammi. Il costo di un matching è il massimo degli spostamenti dei punti di D_1 nei punti di D_2 definiti da tale matching, dove ogni spostamento è dato dalla **distanza in norma del massimo** in $\mathbb{R}^2$ fra i punti e le loro immagini. Spostare un punto sulla diagonale significa "cancellarlo".

I diagrammi di persistenza danno informazioni su d_G

Una proprietà fondamentale dei diagrammi di persistenza è la loro **stabilità** rispetto a perturbazioni delle funzioni filtranti:

Teorema

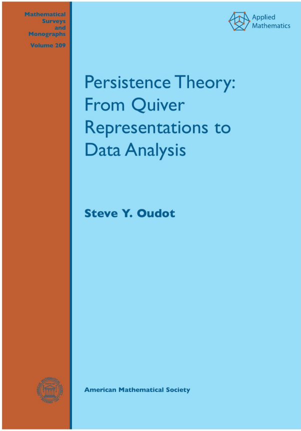
Se k è un numero naturale e $\varphi_1, \varphi_2 \in C^0(X, \mathbb{R})$, allora

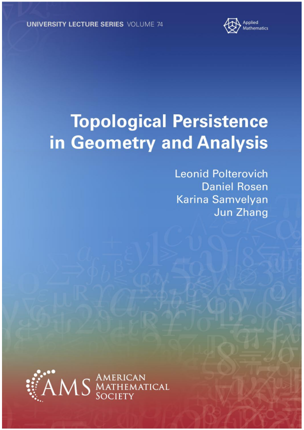
$$d_{\mathrm{match}}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2)) \leq d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2) \leq \|\varphi_1 - \varphi_2\|_\infty .$$

- M. d'Amico, P. Frosini and C. Landi (2005), *Natural pseudo-distance and optimal matching between reduced size functions*, Technical Report no. 66, DISMI, University of Modena and Reggio Emilia, Italy. (limitato al caso $k = 0$)
- D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, *Stability of persistence diagrams*, *Discr. Comput. Geom.*, 37:103–120, 2007.

Interesse per la Topological Data Analysis

Negli ultimi anni sono comparsi molti articoli e libri sulla TDA.





Utilità della Topological Data Analysis

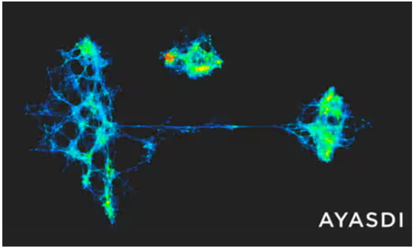
La TDA ha avuto grande successo nell'analisi dei dati.

New big data firm to pioneer topological data analysis

The Guardian

Stanford University project goes commercial following groundbreaking research into cancer therapy and counter-terrorism strategy

- More from the Guardian on big data
- More data journalism and data visualisations from the Guardian



Ayasdi launched today in Palo Alto, California, having secured $10.25m from investors. The funds will be used to build on its Insight Discovery platform.

A US big data firm is set to establish algebraic topology as the gold standard of data science with the launch of the world's leading topological data analysis (TDA) platform.

Alcuni riferimenti bibliografici

- P. Frosini, *A distance for similarity classes of submanifolds of a Euclidean space*, Bulletin of the Australian Mathematical Society, 42, 3 (**1990**), 407-416.
- P. Frosini, *Measuring shapes by size functions*, Proc. of SPIE, Intelligent Robots and Computer Vision X: Algorithms and Techniques, Boston, MA 1607 (**1992**), 122-133.
- S. Barannikov, *Framed Morse complex and its invariants*, Advances in Soviet Mathematics, ADVSOV. 21 (**1994**) 93–115.
- V. Robins, *Towards computing homology from finite approximations*, Topology proceedings, 24(1) (**1999**), 503-532.
- H. Edelsbrunner, D. Letscher, A. Zomorodian, *Topological Persistence and Simplification*. Discrete & Computational Geometry. 28 (4) (**2002**): 511–533.
- A. Zomorodian, G. Carlsson, *Computing Persistent Homology*, Discrete & Computational Geometry. 33 (2) (**2004**) 249–274.

Cose che la TDA non sa fare (1)

Come abbiamo visto, la teoria della TDA **ci dà una limitazione inferiore per la pseudo-distanza naturale** d_G quando $G = \mathrm{Homeo}(X)$:

$$d_{\mathrm{match}}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2)) \leq d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2).$$

Per definizione, questa è anche una limitazione inferiore per d_G quando $G \subset \mathrm{Homeo}(X)$, **ma in generale non è una "buona" limitazione inferiore.**

Come ottenere buone limitazioni inferiori per d_G nel caso generale ($G \neq \mathrm{Homeo}(X)$)?

Cose che la TDA non sa fare (2)

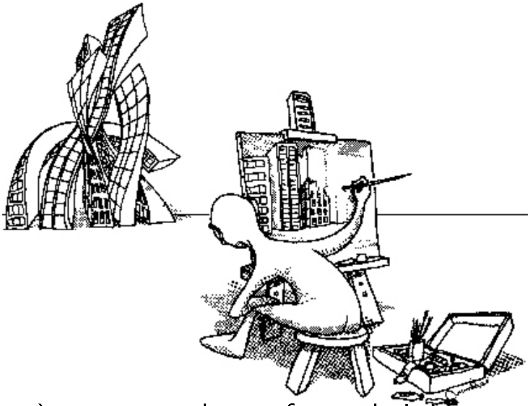
Quando la TDA è nata era focalizzata sui dati, ma in molti casi l'analisi dei dati dipende fortemente dall'osservatore.
Come possiamo inserire nella TDA il ruolo dell'osservatore?

GUIDO MORETTI CON UNA SUA SCULTURA (LA STESSA, DA TRE PUNTI DI VISTA DIVERSI)

Occorre un nuovo paradigma, basato su altri princìpi. Vediamone alcuni.

1. I dati sono sempre acquisiti da osservatori

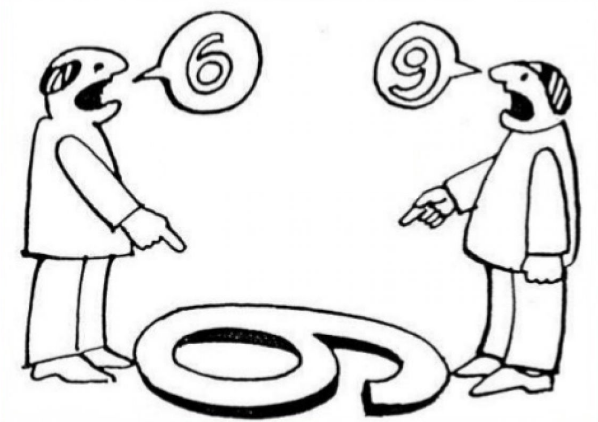
I dati non hanno significato senza un osservatore che li elabori.



Un osservatore è un agente che trasforma dati rispettandone le simmetrie.

2. Gli osservatori sono delle variabili del problema

L'interpretazione dei dati dipende fortemente dall'osservatore prescelto. L'osservatore può variare.

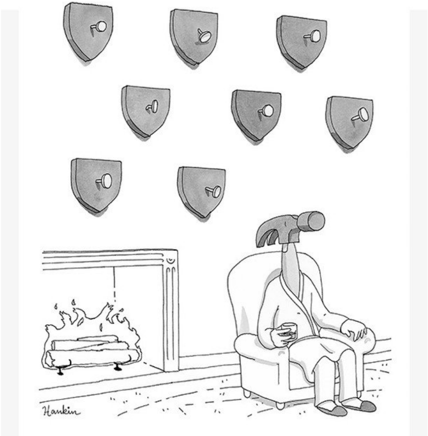


3. L'interesse per i dati è spesso sovrastimato

Non ci interessano quasi mai direttamente i dati, ma piuttosto il modo in cui gli osservatori reagiscono alla loro presenza.

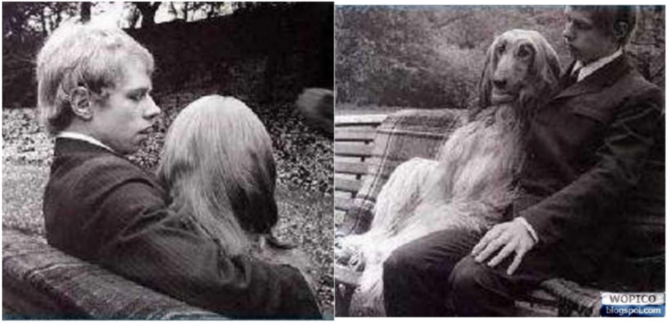
4. In generale, non c'è alcuna struttura nei dati

Parlando in termini generali, i dati non hanno alcuna struttura. La struttura dei dati è una proiezione della struttura dell'osservatore.



Una teoria topologica per gli spazi di osservatori

Dobbiamo dunque sviluppare una teoria **topologica** per gli spazi di **osservatori**. La domanda principale non è *"Qual è la forma dei dati?"* ma *"Qual è la forma degli osservatori?"*



Sommario

Osservatori come operatori equivarianti

Gli osservatori sono strutture in grado di trasformare dati in altri dati, e solitamente lo fanno rispettando alcune equivalenze fra dati, cioè commutando con alcune trasformazioni.

In prima approssimazione, gli osservatori possono essere rappresentati come operatori equivarianti rispetto a un gruppo (Group Equivariant Operators - GEOs), ma in questo seminario illustreremo alcuni risultati sulla teoria degli **operatori equivarianti non espansivi** (**Group Equivariant Non-Expansive Operators - GENEOs**).

Perché "non espansivi"? Perché

1. **spesso si presume che gli osservatori semplifichino la struttura metrica dei dati al fine di produrre interpretazioni concise e significative;**
2. **la non espansività garantisce buone proprietà topologiche.**

Come possiamo rappresentare gli osservatori?

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

Tutto inizia con lo spazio delle funzioni ammissibili

Sia X un insieme non vuoto. Sia Φ un sottospazio topologico dell'insieme $\mathbb{R}_b^X$ di tutte le funzioni φ da X a $\mathbb{R}$ limitate, dotato della topologia indotta dalla metrica

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

Possiamo vedere X come lo spazio dove effettuiamo le nostre misurazioni, e Φ come lo spazio di tutte le possibili misurazioni. Diremo che Φ è l'**insieme delle funzioni ammissibili**. In altre parole, **Φ è l'insieme di tutte le funzioni da X a $\mathbb{R}$ che possono essere prodotte dai nostri strumenti di misurazione (o da altri osservatori)**. Per esempio, un'immagine a livelli di grigio può essere rappresentata come una funzione dal piano reale nell'intervallo $[0,1]$ (in questo caso $X = \mathbb{R}^2$).

Coppie di percezione

Consideriamo un gruppo G di biiezioni $g : X \to X$ tale che $\varphi \in \Phi \implies \varphi \circ g \in \Phi$ per ogni $\varphi \in \Phi$. Diciamo che (Φ, G) è una **coppia di percezione**.

La scelta di una coppia di percezione stabilisce quali dati possono essere considerati misurazioni legittime (le funzioni in Φ) e quale gruppo rappresenta l'equivalenza tra i dati (il gruppo G).

Per procedere è necessario introdurre opportune topologie su X e G. Prima di farlo, ricordiamo che la topologia iniziale τ_{in} su X rispetto a Φ è la topologia meno fine su X tra quelle che rendono continue tutte le funzioni φ in Φ.

Una pseudo-metrica su X

Definiamo su X la pseudo-metrica

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

D_X induce una topologia τ_{D_X} su X.

L'uso di D_X implica che possiamo distinguere due punti soltanto se esiste una misurazione che porta quei punti in valori diversi fra loro.

Proposizione

La topologia τ_{D_X} è più fine della topologia iniziale τ_{in} su X rispetto a Φ. Se Φ è totalmente limitato, allora τ_{D_X} coincide con τ_{in}.

Una pseudo-metrica su X

Le seguenti proprietà sono utili nel nostro modello.

Proposizione

Ogni funzione in Φ è non espansiva e quindi continua.

Proposizione

Se Φ è compatto e X è completo, allora X è compatto.

Nel seguito, assumeremo solitamente che Φ sia compatto e X sia completo (e quindi compatto).

Un'utile conseguenza: ogni biiezione è un'isometria

- $\mathrm{Bij}_\Phi(X) = \{\textbf{biiezioni } g : X \to X \text{ s.t. } \Phi \circ g, \Phi \circ g^{-1} \subseteq \Phi\}$;
- $\mathrm{Homeo}_\Phi(X) = \{\textbf{omeomorfismi } g : X \to X \text{ s.t. } \Phi \circ g, \Phi \circ g^{-1} \subseteq \Phi\}$;
- $\mathrm{Iso}_\Phi(X) = \{\textbf{isometrie } g : X \to X \text{ s.t. } \Phi \circ g, \Phi \circ g^{-1} \subseteq \Phi\}$.

Proposizione

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

Una pseudo-metrica su G

Concentriamo ora la nostra attenzione su un sottogruppo G di $\mathrm{Homeo}_\Phi(X)$. Possiamo definire una pseudo-metrica D_G su G ponendo

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

Proposizione

G è un gruppo topologico rispetto a D_G e l'azione di G su Φ per composizione a destra è continua.

Proposizione

Se Φ è compatto e G è completo, allora G è compatto.

I GEO e i GENEO

Ricordo che ogni coppia (Φ, G) con $G \subseteq \mathrm{Homeo}_\Phi(X)$ è detta **coppia di percezione**.

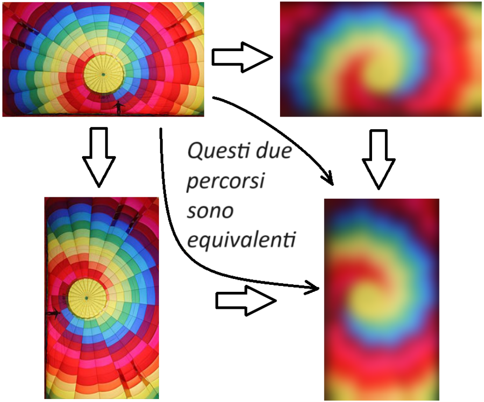
Supponiamo che siano date due coppie di percezione (Φ, G), (Ψ, H) e fissiamo un omomorfismo di gruppi $T : G \to H$.

Ciascuna funzione $F : \Phi \to \Psi$ tale che $\boxed{F(\varphi \circ g) = F(\varphi) \circ T(g)}$ per ogni $\varphi \in \Phi$ e ogni $g \in G$, è detta un **operatore equivariante (Group Equivariant Operator - GEO)** associato all'omomorfismo T.

Se F è anche non espansivo (cioè $\boxed{D_\Psi(F(\varphi_1), F(\varphi_2)) \leq D_\Phi(\varphi_1, \varphi_2)}$ per ogni $\varphi_1, \varphi_2 \in \Phi$), allora F è chiamato **operatore equivariante non espansivo (Group Equivariant Non-Expansive Operator - GENEO)** associato all'omomorfismo T.

Un esempio di GENEO

Ogni blurring di immagini, ottenuto tramite una convoluzione con un nucleo rotazionalmente invariante e di norma inferiore a 1 in L^1, è un GENEO rispetto al gruppo delle isometrie del piano.



Due risultati (e due buone notizie per le applicazioni)

Supponiamo che sia stato fissato un omomorfismo $T : G \to H$.
Definiamo una metrica D_{GENEO} su $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ ponendo

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

Teorema

Se Φ *e* Ψ *sono* **compatti**, *allora* $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ *è* **compatto** *rispetto a* D_{GENEO}.

Teorema

Se Ψ *è* **convesso**, *allora* $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ *è* **convesso**.

Due osservazioni importanti (1)

- Mentre lo spazio dei dati è spesso non convesso (e quindi non possiamo calcolare la media dei dati), l'assunzione di convessità di Ψ implica la convessità dello spazio degli osservatori e ci consente di considerare la **"media degli osservatori"**.

Due osservazioni importanti (2)

- Il nostro obiettivo principale è sviluppare una buona teoria geometrica e compositiva per approssimare un osservatore ideale. Nel nostro modello, “approssimare un osservatore” significa cercare un GENEO F che minimizza un’opportuna “funzione di costo” $c(F)$. La funzione di costo quantifica l’errore commesso prendendo il GENEO F invece dell’osservatore ideale. Poiché lo spazio dei GENEO è compatto e convesso (sotto l’ipotesi che gli spazi dei dati siano compatti e convessi), se la funzione di costo $c(F)$ è strettamente convessa esiste uno e un solo GENEO che meglio approssima l’osservatore ideale.

Sommario

Metodi elementari per costruire dei GENEO

Proposizione (Composizione)

Se $F_1 \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto a* $T_1 : G \to H$ *e* $F_2 \in \mathrm{GENEO}((\Psi, H), (\chi, K))$ *rispetto a* $T_2 : H \to K$ *allora* $F_2 \circ F_1 \in \mathrm{GENEO}((\Phi, G), (\chi, K))$ *rispetto a* $T_2 \circ T_1 : G \to K$.

Proposizione (Applicazione di una funzione 1-Lipschitz)

Se $F_1, \ldots, F_n \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto a* $T : G \to H$, L *è una mappa 1-Lipschitz da* $\mathbb{R}^n$ *a* $\mathbb{R}$, *e* $L^*(F_1, \ldots, F_n)(\Phi) \subseteq \Psi$ *(con* L^* *mappa indotta da* L*), allora* $L^*(F_1, \ldots, F_n) \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto a* T.

Da quest'ultima proposizione seguono le seguenti tre affermazioni.

Metodi elementari per costruire dei GENEO

Proposizione (Reticolo dei GENEO)

Se $F_1, \dots, F_n \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto a* $T : G \to H$ *e* $\max(F_1, \dots, F_n)(\Phi), \min(F_1, \dots, F_n)(\Phi) \subseteq \Psi$*, allora* $\max(F_1, \dots, F_n), \min(F_1, \dots, F_n) \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *risp. a* T.

Proposizione (Traslazione)

Se $F \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto a* $T : G \to H$*, e* $F_b(\Phi) \subseteq \Psi$ *per* $F_b(\varphi) := F(\varphi) - b$*, allora* $F_b \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *risp. a* T.

Proposizione (Combinazione convessa)

Se $F_1, \dots, F_n \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto a* $T : G \to H$*,* $(a_1, \dots, a_n) \in \mathbb{R}^n$ *con* $\sum_{i=1}^n |a_i| \leq 1$ *e* $F_\Sigma(\Phi) \subseteq \Psi$ *per* $F_\Sigma(\varphi) := \sum_{i=1}^n a_i F_i(\varphi)$*, allora* $F_\Sigma \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *risp. a* T.

Misure permutanti

Consideriamo l'insieme $\Phi = \mathbb{R}^X \cong \mathbb{R}^n$ di tutte le funzioni da un insieme finito $X = \{x_1, \ldots, x_n\}$ a $\mathbb{R}$ e un sottogruppo G del gruppo $\mathrm{Bij}(X)$ di tutte le permutazioni di X.

Definizione

Una misura finita (con segno) μ su $\mathrm{Bij}(X)$ è detta *misura permutante* rispetto a G se ogni sottoinsieme H di $\mathrm{Bij}(X)$ è misurabile e μ è invariante rispetto all'azione di coniugio di G (cioè $\mu(H) = \mu(gHg^{-1})$ per ogni $g \in G$).

Teorema di rappresentazione per GENEO lineari

Teorema (Teorema di rappresentazione per GENEO lineari)

Supponiamo che $G \subseteq \mathrm{Bij}(X)$ *agisca transitivamente sull'insieme finito* X *e che* F *sia una mappa da* $\mathbb{R}^X$ *a* $\mathbb{R}^X$. *La mappa* F *è un GENEO lineare da* $\mathbb{R}^X$ *a* $\mathbb{R}^X$ *rispetto all'omomorfismo identico* $\mathrm{Id}_G : g \mapsto g$ **se e solo se** *esiste una misura permutante* μ *rispetto a* G, *tale che* $F(\varphi) = \sum_{h \in \mathrm{Bij}(X)} \varphi h^{-1}\, \mu(h)$ *per ogni* $\varphi \in \mathbb{R}^X$, *e* $\sum_{h \in \mathrm{Bij}(X)} |\mu(h)| \leq 1$.

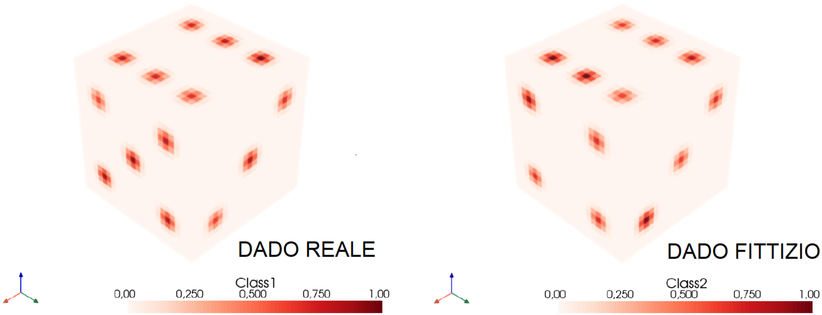
Riferimenti bibliografici sulla costruzione di GENEO

- F. Conti, P. Frosini, N. Quercioli, *On the construction of Group Equivariant Non-Expansive Operators via permutants and symmetric functions*, Frontiers in Artificial Intelligence, vol. 5 (2022), 1-11.
- G. Bocchi, S. Botteghi, M. Brasini, P. Frosini and N. Quercioli, *On the finite representation of linear group equivariant operators via permutant measures*, Annals of Mathematics and Artificial Intelligence, vol. 91 (2023), n. 4, 465–487.

Sommario

Cosa succede quando applichiamo GENEO ai dati?

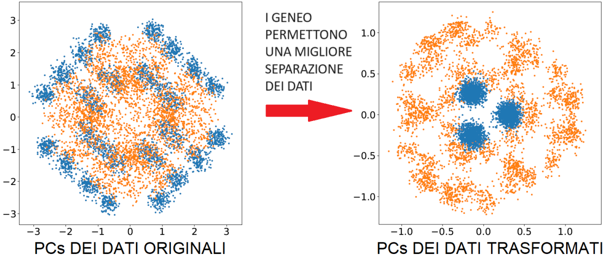
Un esempio di utilizzo: confronto tra dadi veri e dadi falsi.



(Esperimento e calcoli di Giovanni Bocchi)

Cosa succede quando applichiamo GENEO ai dati?

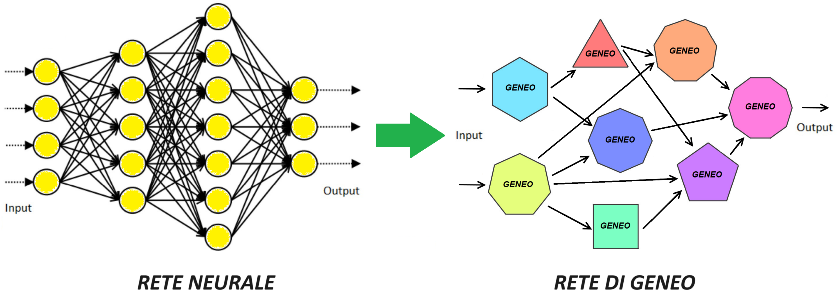
Abbiamo prodotto 10000 dadi (un training set di cardinalità 7000 e un test set di cardinalità 3000), abbiamo poi applicato la PCA al test set e al test set trasformato da un GENEO ottimizzato sul training set:



Per ogni matrice vengono riportate le prime due componenti principali. I punti blu sono associati ai **dadi reali**, mentre quelli arancioni a **dadi falsi**. Il GENEO che utilizziamo è una combinazione convessa di 3 GENEO definiti da misure permutanti.

Uso dei GENEO nel Machine Learning

In prospettiva, stiamo cercando di sviluppare una buona teoria composizionale per costruire reti efficienti di GENEO. Alcuni esperimenti preliminari suggeriscono che la sostituzione dei neuroni con GENEO potrebbero rendere il deep learning più trasparente e interpretabile e velocizzare il processo di apprendimento.



Uso dei GENEO nel Machine Learning

Per maggiori dettagli sull'uso di GENEO nel 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

Alcune linee di ricerca

Stiamo attualmente studiando questi problemi:

- Come possiamo estendere la teoria dei GENEO ai grafi? (F. Ahmad, M. Ferri, P. Frosini, *Generalized Permutants and Graph GENEOs*, `https://arxiv.org/pdf/2206.14798.pdf`.)
- Come possiamo estendere la teoria dei GENEO a spazi di probabilità di segnali? P. Cascarano, P. Frosini, N. Quercioli, A. Saki, *On the geometric and Riemannian structure of the spaces of group equivariant non-expansive operators*, `https://arxiv.org/pdf/2103.02543.pdf`

Alcune linee di ricerca

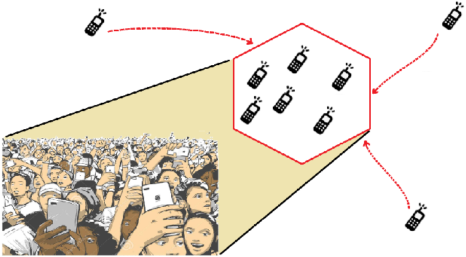
Un'altra linea di ricerca (appena avviata) è basata su questa osservazione: la dimostrazione del teorema di rappresentazione dei GENEO lineari è basata sullo studio di matrici bistocastiche (e quindi piani di trasporto su insiemi finiti) e, inoltre, molte applicazioni richiedono di lavorare non con gruppi di trasformazioni di un dominio ma con piani di trasporto da uno spazio di misura in sé (spesso, infatti, le trasformazioni considerate non sono deterministiche: non sappiamo in quale punto venga trasportato un dato punto x, ma solo la probabilità $p(x, y)$ che x venga portato in y).

Dunque: come possiamo estendere la teoria dei GENEO rimpiazzando i gruppi di trasformazioni (deterministiche) con monoidi di piani di trasporto (di carattere probabilistico)?

Un progetto di ricerca in corso

CNIT / WiLab - Huawei Joint Innovation Center (JIC)

Project on GENEOs for 6G



WILAB HUAWEI

GRAZIE PER L'ATTENZIONE