Sull'uso di operatori equivarianti non espansivi per l'apprendimento automatico

Patrizio Frosini

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

Seminari su Intelligenza Artificiale e Machine Learning
Gruppo UMI "AI&ML&MAT", 4 marzo 2022

Sommario

Il ruolo chiave degli osservatori nell'interpretazione dei dati

Introduciamo la formalizzazione matematica

Definizioni formali e proprietà geometriche dei GENEO

Metodi per costruire GENEO

Come possiamo usare i GENEO in ambito applicativo?

Il ruolo chiave degli osservatori nell'interpretazione dei dati

Introduciamo la formalizzazione matematica

Definizioni formali e proprietà geometriche dei GENEO

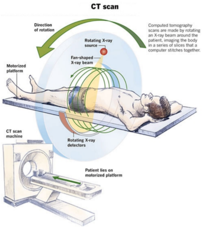
Metodi per costruire GENEO

Come possiamo usare i GENEO in ambito applicativo?

I dati sono spesso rappresentabili come funzioni

Alcuni esempi di dati esprimibili come funzioni:

- Un elettrocardiogramma (una funzione da $\mathbb{R}$ a $\mathbb{R}$);
- Un'immagine a livelli di grigio (una funzione da $\mathbb{R}^2$ a $\mathbb{R}$);
- Una TAC (una funzione da un'elica a $\mathbb{R}$).



I dati nel nostro modello

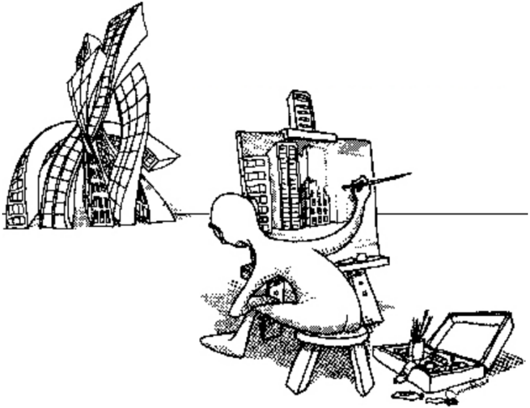
Nel nostro modello i dati sono descritti da funzioni a valori reali o vettoriali.

Indichiamo con Φ l'insieme dei dati ammissibili, cioè delle funzioni che possono essere interpretate come segnali da elaborare.

È importante osservare che solo alcune funzioni descrivono dati ammissibili: ad esempio, una funzione dal piano reale ai numeri reali che rappresenti un'immagine a livelli di grigio dovrà assumere valori compresi in un intervallo delimitato. Se questo non accade, la funzione non viene interpretata come la rappresentazione di un'immagine.

I dati vengono elaborati dagli osservatori

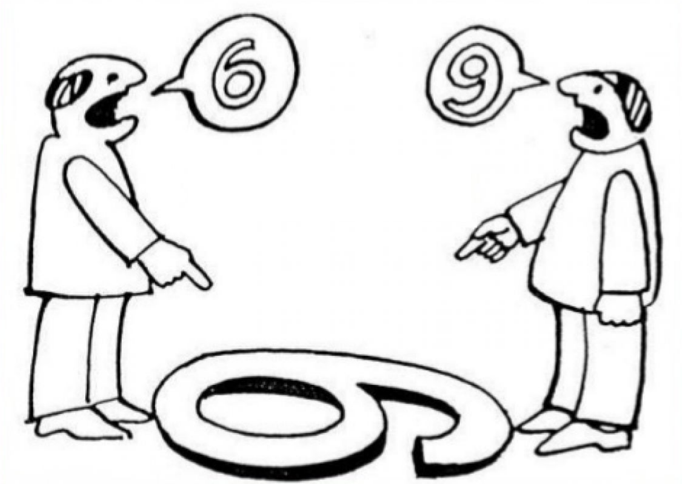
I dati di per sé non hanno significato. Occorre un osservatore che li elabori.



Un osservatore è un agente che trasforma dati in altri dati.

Gli osservatori sono delle variabili nell'analisi dei dati

Spesso l'interpretazione dei dati dipende fortemente dagli osservatori:



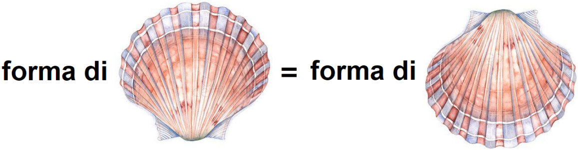
La coppia (dato, osservatore)

In realtà non siamo quasi mai interessati direttamente ai dati, ma alla relazione fra i dati e l'osservatore. Quel che ci interessa è soprattutto il comportamento dell'osservatore.

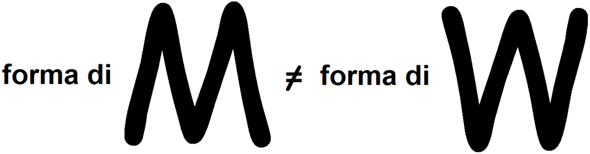
I casi in cui l'interesse per i dati sembra centrale sono quelli in cui la modalità di reazione degli osservatori è tacitamente condivisa.
Tutto ciò induce a privilegiare lo studio della "forma degli osservatori" allo studio della "forma dei dati".

Osservatori e gruppi di invarianza

Spesso gli osservatori giudicano alcuni dati equivalenti ad altri rispetto a opportuni gruppi di invarianza.



Il gruppo G è una variabile del problema e cambia al cambiare delle applicazioni.



Coppie di percezione

Chiameremo *coppia di percezione* ogni coppia (Φ, G) dove Φ sia uno spazio arbitrario di funzioni da un insieme X a $\mathbb{R}$ (o, più in generale, $\mathbb{R}^k$) e G è un gruppo di corrispondenze biunivoche che preservano Φ mediante l'azione di composizione a destra.

Ogni coppia di percezione rappresenta i dati che l'osservatore può elaborare e descrive l'equivalenza fra dati.

Per proseguire, è importante scegliere delle topologie adeguate su X e su G.

Una pseudo-metrica su X

Definiamo su X la pseudometrica

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

(Una pseudometrica è semplicemente una metrica d senza la proprietà $d(x_1, x_2) = 0 \implies x_1 = x_2$.)

L'uso di D_X implica che possiamo distinguere due punti solo se esiste una misurazione che assume due differenti valori su tali punti.

Teorema

Ogni funzione in Φ è continua rispetto a D_X.

Teorema

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

Un tocco di magia: ogni biiezione è un'isometria

Sia $\mathrm{Bij}(X)$ il gruppo di tutte le biiezioni da X a X, e indichiamo con $\mathrm{Bij}_\Phi(X)$ il sottogruppo di tutti i $g \in \mathrm{Bij}(X)$ tali che $\varphi \circ g \in \Phi$ e $\varphi \circ g^{-1} \in \Phi$ per ogni $\varphi \in \Phi$. Sia $\mathrm{Homeo}(X)$ il gruppo di tutti gli omeomorfismi da X a X rispetto a D_X, e indichiamo con $\mathrm{Homeo}_\Phi(X)$ il sottogruppo di tutti i $g \in \mathrm{Homeo}(X)$ tali che $\varphi \circ g \in \Phi$ e $\varphi \circ g^{-1} \in \Phi$ per ogni $\varphi \in \Phi$. Sia $\mathrm{Iso}(X)$ il gruppo di tutte le isometrie da X a X, e indichiamo con $\mathrm{Iso}_\Phi(X)$ il sottogruppo di tutti i $g \in \mathrm{Iso}(X)$ tali che $\varphi \circ g \in \Phi$ e $\varphi \circ g^{-1} \in \Phi$ per ogni $\varphi \in \Phi$.

Proposizione

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

Una pseudo-metrica su G

Scegliamo ora un sottogruppo G di $\mathrm{Homeo}_\Phi(X)$.
Possiamo definire una pseudometrica D_G su G ponendo

$$D_G(g_1, g_2) := \sup_{\varphi \in \Phi} \|\varphi \circ g_1 - \varphi \circ g_2\|_\infty .$$

Teorema

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

Teorema

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

Coppie di percezione compatte

La scelta delle pseudometriche D_X e D_G ci permette di ottenere il seguente teorema (provato da Faraz Ahmad):

Teorema

Se (Φ, G) è una coppia di percezione e lo spazio Φ è totalmente limitato, esiste un completamento compatto $(\bar{\Phi}, \bar{G})$ di (Φ, G).

In altri termini, la totale limitatezza dello spazio Φ delle funzioni ammissibili implica che, mediante una procedura di completamento e l'uso di opportune topologie, ci si può sempre ricondurre al caso il cui lo spazio delle funzioni ammissibili, il loro dominio e il gruppo di equivarianza sono compatti.

Il ruolo chiave degli osservatori nell'interpretazione dei dati

Introduciamo la formalizzazione matematica

Definizioni formali e proprietà geometriche dei GENEO

Metodi per costruire GENEO

Come possiamo usare i GENEO in ambito applicativo?

Lo spazio dei GENEO in termini formali

Definizione

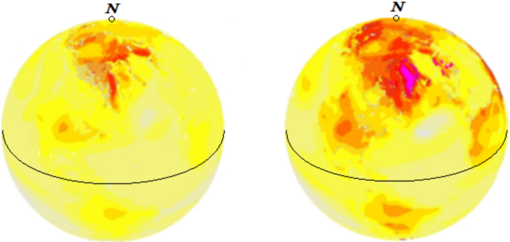
Supponiamo che (Φ, G) e (Ψ, H) siano due coppie di percezione e che sia stato fissato un omomorfismo $T : G \to H$. Un Group Equivariant Non-Expansive Operator (GENEO) da (Φ, G) a (Ψ, H) rispetto a T è una mappa $F : \Phi \to \Psi$ tale che valgano le seguenti proprietà per ogni $\varphi, \varphi_1, \varphi_2 \in \Phi$ e ogni $g \in G$:

1. $F(\varphi \circ g) = F(\varphi) \circ T(g)$;
2. $\|F(\varphi_1) - F(\varphi_2)\|_\infty \leq \|\varphi_1 - \varphi_2\|_\infty$.

Se vale solo la proprietà 1) parleremo di *Group Equivariant Operator (GEO)*.

Un esempio di GENEO fra due coppie di percezione

Diamo un esempio di GENEO fra due coppie di percezione (Φ, G), (Ψ, H) fra loro diverse.
Supponiamo di essere interessati al confronto di distribuzioni di temperature su una sfera:



Supponiamo anche che solo due punti opposti N, S possano essere localizzati sulla sfera.

Un esempio di GENEO fra due coppie di percezione

In questo caso possiamo porre

- $X = S^2$
- $\Phi =$ insieme delle funzioni 1-lipschitziane da S^2 a un intervallo fissato $[a, b]$ (ogni $\varphi \in \Phi$ descrive una possibile distribuzione di temperature sulla sfera)
- $G =$ gruppo delle rotazioni di S^2 attorno all'asse $N - S$

Possiamo anche considerare l'"equatore" della nostra sfera, rappresentato dallo spazio S^1.

Quindi possiamo porre

- $Y =$ equatore S^1 di S^2
- $\Psi =$ insieme delle funzioni 1-lipschitziane da S^1 a $[a, b]$
- $H =$ gruppo delle rotazioni di S^1

Un esempio di GENEO fra due coppie di percezione

In questo caso possiamo costruire un semplice esempio di GENEO da (Φ, G) a (Ψ, H) ponendo

- $T(g)$ uguale alla rotazione $h \in H$ dell'equatore S^1 che è indotta dalla rotazione g di S^2, per ogni $g \in G$.
- $F(\varphi)$ uguale alla funzione ψ che porta ogni punto y dell'equatore S^1 nella media della funzione φ (che descrive una possibile distribuzione di temperature) lungo il meridiano contenente y, per ogni $\varphi \in \Phi$;

Possiamo controllare facilmente che F verifica le proprietà che definiscono il concetto di operatore equivariante non espansivo rispetto all'omomorfismo $T : G \to H$.

In parole povere, il nostro GENEO trasforma "distribuzioni di temperature sulla sfera" in "distribuzioni di temperature sull'equatore".

Una metrica per lo spazio $\mathscr{F}^{\mathrm{all}}$

Sullo spazio $\mathscr{F}^{\mathrm{all}}$ di tutti i GENEO da (Φ, G) a (Ψ, H) rispetto all'omomorfismo $T : G \to H$ si può considerare la seguente metrica:

Definizione

Se $F_1, F_2 \in \mathscr{F}^{\mathrm{all}}$ poniamo

$$D_{\mathrm{GENEO}}(F_1, F_2) := \sup_{\varphi \in \Phi} \| F_1(\varphi) - F_2(\varphi) \|_\infty .$$

Alcune proprietà molto utili

Sia $\mathscr{F}^{\mathrm{all}}$ l'insieme di tutti i GENEO da (Φ, G) a (Ψ, H) rispetto a un omomorfismo fissato $T : G \to H$.

Teorema

Se Φ e Ψ sono compatti, allora $\mathscr{F}^{\mathrm{all}}$ è compatto rispetto a D_{GENEO}.

Corollario

$\mathscr{F}^{\mathrm{all}}$ può essere ε-approssimato da un insieme finito per ogni $\varepsilon > 0$.

Teorema

Se Ψ è convesso, allora $\mathscr{F}^{\mathrm{all}}$ è convesso.

Alcune proprietà molto utili

Se gli spazi dei dati sono compatti, allora anche lo spazio degli osservatori è compatto.

QUINDI

Se gli spazi dei dati sono compatti (e dunque approssimabili con un errore arbitrariamente piccolo tramite un insieme finito), allora lo spazio degli osservatori è approssimabile con un errore arbitrariamente piccolo tramite un insieme finito di osservatori.

Due proprietà molto utili

Se lo spazio dei dati in uscita è convesso, allora anche lo spazio degli osservatori è convesso.

QUINDI

(per le topologie che abbiamo scelto)

Se gli spazi dei dati sono compatti e convessi, allora anche lo spazio degli osservatori è compatto e convesso e dunque ogni funzione di costo strettamente convessa ammette un solo minimo sullo spazio degli osservatori.

Quindi lo spazio dei GENEO è particolarmente adatto ai processi di approssimazione e minimizzazione.

Approssimazione di osservatori

IL NOSTRO OBIETTIVO PRINCIPALE: APPROSSIMARE UN OSSERVATORE IDEALE

Nel nostro modello "approssimare un osservatore" significa cercare un GENEO F che minimizzi un'opportuna "funzione costo" $c(F)$.

La funzione costo quantifica l'errore che si commette prendendo il GENEO F al posto dell'osservatore ideale.

Dato che lo spazio dei GENEO è compatto e convesso (sotto l'ipotesi che gli spazi dei dati siano compatti e convessi), se la funzione costo $c(F)$ è strettamente convessa abbiamo che esiste uno e un solo GENEO che approssima al meglio l'osservatore ideale.

Un'osservazione importante

Mentre molto spesso non ha senso considerare le medie dei dati originali (perché l'insieme dei dati ammissibili è frequentemente non convesso), l'assunzione che il codominio dei GENEO sia convesso permette di considerare la media degli osservatori.

Il ruolo chiave degli osservatori nell'interpretazione dei dati

Introduciamo la formalizzazione matematica

Definizioni formali e proprietà geometriche dei GENEO

Metodi per costruire GENEO

Come possiamo usare i GENEO in ambito applicativo?

Metodi elementari per costruire 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* $L^*(F_1, \ldots, F_n)(\Phi) \subseteq \Phi$ *(dove* L^* *è la mappa indotta da* L*), allora* $L^*(F_1, \ldots, F_n) \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto a* T*.*

Quest'ultima proposizione implica i seguenti tre risultati.

Metodi elementari per costruire 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* $\max(F_1, \dots, F_n)(\Phi), \min(F_1, \dots, F_n) \subseteq \Phi$*, allora* $\max(F_1, \dots, F_n)$, $\min(F_1, \dots, F_n) \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto a* T.

Proposizione (Traslazione)

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

Proposizione (Combinazione convessa)

Se $F_1, \dots, F_n \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto a* $T : G \to H$, $\sum_{i=1}^{n} |a_i| \leq 1$ *e* $F_\Sigma(\Phi) \subseteq \Phi$ *per* $F_\Sigma(\varphi) := \sum_{i=1}^{n} a_i F_i(\varphi)$*, allora* $F_\Sigma \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *rispetto 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 permutante rispetto a G è una misura finita (con segno) μ su $\mathrm{Bij}(X)$ tale che ogni sottoinsieme H di $\mathrm{Bij}(X)$ è misurabile e μ è invariante per l'azione di coniugio di G (cioè, $\mu(H) = \mu(gHg^{-1})$ per ogni $g \in G$).

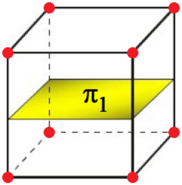
Proposizione

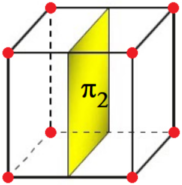
Se μ è una misura permutante rispetto a G, allora la mappa $F_\mu : \mathbb{R}^X \to \mathbb{R}^X$ definita ponendo $F_\mu(\varphi) := \sum_{h \in \mathrm{Bij}(X)} \varphi h^{-1} \mu(h)$ è un GEO lineare.

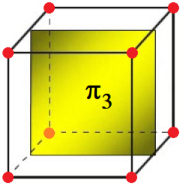
Un esempio di misura permutante

Consideriamo l'insieme X dei vertici di un cubo in $\mathbb{R}^3$ e il gruppo G delle isometrie di $\mathbb{R}^3$ che portano X in X e conservano l'orientazione. Siano π_1, π_2, π_3 i tre piani contenenti il centro di massa di X e paralleli a qualche faccia del cubo. Sia $h_i : X \to X$ la simmetria ortogonale rispetto a π_i, per $i \in \{1, 2, 3\}$.

Possiamo ora definire una misura permutante μ sul gruppo $\mathrm{Bij}(X)$ ponendo $\mu(h_1) = \mu(h_2) = \mu(h_3) = c$, dove c è un numero reale positivo, e $\mu(h) = 0$ per ogni $h \in \mathrm{Bij}(X)$ con $h \notin \{h_1, h_2, h_3\}$.







Costruire GENEO per mezzo di misure permutanti

Vale il seguente teorema di rappresentazione.

Teorema

Assumiamo 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 $(\mathbb{R}^X, G)$ a $(\mathbb{R}^X, G)$ (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 $\sum_{h \in \mathrm{Bij}(X)} |\mu(h)| \leq 1$.

N.B.: In molti casi $\mathrm{card}\ \mathrm{supp}\ \mu \ll \mathrm{card}\ G$.

Ulteriori dettagli sono disponibili in questo preprint:
S. Botteghi, M. Brasini, P. Frosini and N. Quercioli, On the finite representation of group equivariant operators via permutant measures
https://arxiv.org/pdf/2008.06340.pdf

Il ruolo chiave degli osservatori nell'interpretazione dei dati

Introduciamo la formalizzazione matematica

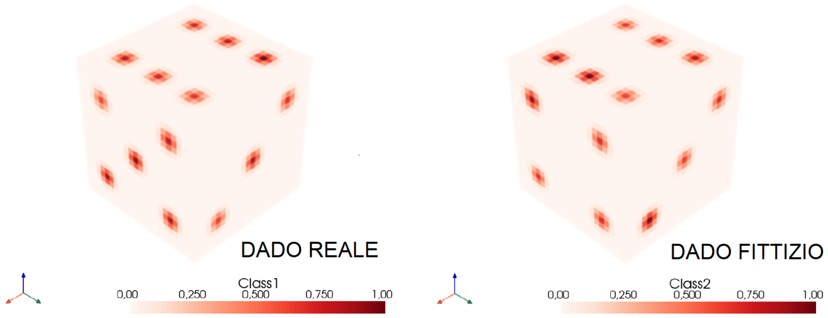
Definizioni formali e proprietà geometriche dei GENEO

Metodi per costruire GENEO

Come possiamo usare i GENEO in ambito applicativo?

Che succede quando applichiamo dei GENEO ai dati?

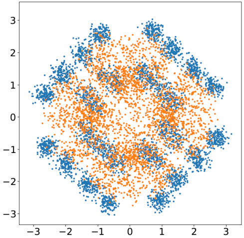
Un esempio: confronto fra dadi reali e dadi fittizi.



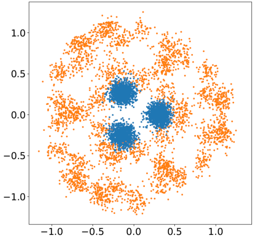
(Esperimenti e calcoli di Giovanni Bocchi)

Che succede quando applichiamo dei GENEO ai dati?

Abbiamo simulato 10000 dadi (7000 per il training set e 3000 per il test set), poi abbiamo applicato la PCA al test set e al test set trasformato da un opportuno GENEO ottimizzato sul training set:



PCs DEI DATI ORIGINALI



PCs DEI DATI TRASFORMATI

Per ogni dado viene riportato il punto del piano dato dalle prime due componenti principali. I **punti blu** si riferiscono ai dadi reali, i **punti arancioni** ai dadi fittizi. Il GENEO usato è una combinazione convessa di 3 GENEO associati a misure permutanti.

Un'applicazione reale: trovare "tasche" nelle proteine

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

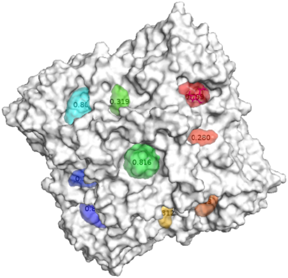
[3] Department of Pharmaceutical Sciences, Università degli Studi di Milano

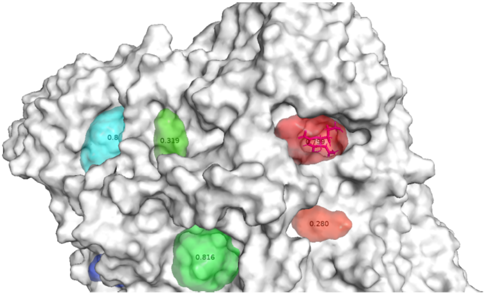
[4] Dipartimento di Scienze della Salute, Università degli Studi "Magna Græcia di Catanzaro"

[5] Dompé Farmaceutici SpA

https://arxiv.org/abs/2202.00451

Un'applicazione reale: trovare tasche nelle proteine

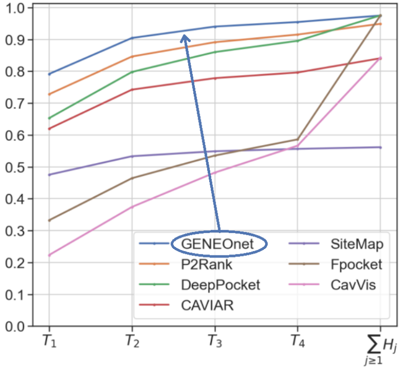




La ricerca di tasche è stata effettuata identificando un GENEO ottimale nell'inviluppo convesso di 8 GENEO (ciascuno di essi sensibile a una particolare proprietà fisica o geometrica delle tasche).

Un'applicazione reale: trovare tasche nelle proteine

Ecco i risultati dei nostri esperimenti:

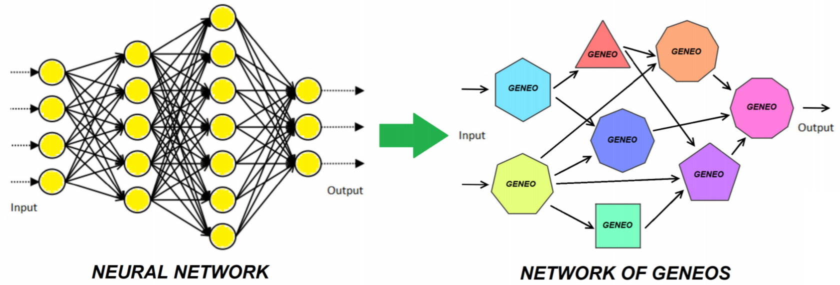


Sottolineiamo che GENEOnet usa 17 parametri, mentre una rete neurale convoluzionale come DeepPocket richiede 665122 parametri.

Il punto chiave nel nostro approccio

In prospettiva, stiamo cercando di sviluppare una buona teoria composizionale che permetta la costruzione di reti di GENEO efficienti e trasparenti.

Gli esperimenti che stiamo effettuando suggeriscono che rimpiazzare neuroni con GENEO potrebbe rendere il deep learning più facilmente interpretabile e velocizzare il processo di apprendimento.



Questioni aperte

- Come possiamo approssimare un osservatore reale (diciamo, p.e., un medico) per mezzo di GENEO, al fine di emulare il comportamento dell'osservatore in rapporto ai dati?
- Possiamo individuare procedure costruttive che ci permettano di realizzare/approssimare ogni possibile GENEO rispetto a un dato gruppo di equivarianza?
- Qual è il modo più corretto di confrontare GENEO in un setting di tipo topologico-statistico?
- Come possiamo selezionare insiemi rappresentativi in uno spazio di probabilità di GENEO?
- Come possiamo prevedere e controllare il comportamento di reti di GENEO?
- Come possiamo valutare i vantaggi e i limiti di un approccio all'analisi dei dati basato sull'interazione fra GENEO e TDA?

SOME REFERENCES

- 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*, **Nature Machine Intelligence**, vol. 1, n. 9, 423–433 (2 September 2019)
- 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, P. Frosini, A. Micheletti, A. Pedretti, C. Gratteri, F. Lunghini, A. R. Beccari, C. Talarico, *GENEOnet: A new machine learning paradigm based on Group Equivariant Non-Expansive Operators. An application to protein pocket detection*, 2022 https://arxiv.org/abs/2202.00451


Grazie per
l'attenzione!