Anche il Deep Learning può sbagliare- seconda parte-

Data 28 novembre 2021
Categoria Medicinadigitale

Nella precedente pillola abbiamo illustrato gli algoritmi genetici che hanno aperto nuovi orizzonti alla ricerca medica basata sulla intelligenza artificiale: in particolare abbiamo sottolineato come questo tipo di algoritmi procedano attraverso un numero elevatissimo di “prove - errori- correzione degli errori -nuove prove”, fino ad ottenere risultati soddisfacenti.
Tuttavia, come abbiamo ricordato nella conclusione della precedente pillola, le procedure di Deep Learning consentono solo un controllo parziale sulla “struttura” ed un controllo sui “risultati”, ma non consentono controlli dettagliati sul complesso “processo” che porta, attraverso combinazioni casuali ignote ai ricercatori, a conseguire quei risultati.
La osservazione critica è importante perchè secoli di studi sulle leggi del caso e sui processi casuali hanno dimostrato che solo controllando il più accuratamente possibile i dati a nostra disposizione e le eventuali loro correlazioni ed interazioni possiamo minimizzare l'errore che è comunque una costante possibilità in ogni attività di ricerca, specie le più avanzate e sofisticate.

In questa pillola prendiamo in esame una importante ricerca pubblicata oltre un anno fa, che individua seri limiti del Deep Learning, specie se non utilizzato con procedure rigorose(1).

Un gruppo di esperti di ricerca clinica e di intelligenza artificiale dell’ Imperial College di Londra, delle Università di Oxford, di Stanford e della California, ha confrontato le ricerche sulla diagnostica per immagini effettuate mediante dispositivi di intelligenza artificiale con quelle effettuate da medici e ricercatori con le metodiche tradizionali.
La qualità degli studi è stata esaminata secondo gli standard previsti da Consort (2) per gli studi randomizzati, e da Tripod (3)per i non randomizzati, mentre la probabilità di bias veniva valutata mediante le metodiche Cochrane(4) per gli studi randomizzati e Probast(5) per quelli non randomizzati.

[b]Cosa valutare in uno studio[/b]
(Valutazione minimale ma utile: per valutazioni più approfondite si rinvia ai siti sopra citati e ad un ottimo manuale(6)
1) Finalità ed Obiettivi: debbono essere chiari e ben descritti
2) Modalità di reclutamento degli " oggetti di studio"
3) Criteri di inclusione ed esclusione degli "oggetti"
4) Numerosità del campione: deve essere coerente con gli obiettivi
5) Durata del follow-up: adeguata alle finalità
6) Procedure di confronto e modalità di confronto

I risultati furono sorprendenti: solo 10 su 91 ricerche effettuate mediante procedure di Deep Learning erano randomizzate, e tra le 81 ricerche non randomizzate solo nove erano prospettiche.
Solo in due studi dei 91 esaminati non sono stati rilevati severi bias ma purtroppo, malgrado la buona qualità di questi studi, la numerosità dei campioni esaminati non era tanto rilevante da poter trarre conclusioni generalizzabili a contesti molto più ampi…
Gli autori dello studio esprimevano l'auspicio che per il futuro si programmino studi di numerosità adeguata, con corretta randomizzazione, con criteri di inclusione ed esclusione ben definiti e con programmi e procedure controllati mediante valutazioni di qualità e di minimizzazione dei bias secondo quanto proposto dalle organizzazioni sopra citate(2-5).

Commento
Negli ultimi anni i mezzi di comunicazione ed anche parte delle riviste mediche hanno esaltato le performance dei dispositivi di intelligenza artificiale in tutti gli ambiti di ricerca ed in particolare in ambito medico. L'entusiasmo per gli indiscutibili risultati ottenuti in alcuni settori, ad esempio la radiologia, ha diffuso una forte propensione all'ottimismo che, come ci insegnano gli esperti di psicologia cognitiva, ci porta inconsapevolmente a sopravvalutare le potenzialità di questistrumenti.
In realtà decenni di ricerche sull'Errore in medicina e secoli di ricerca sul “Caso” ci hanno lasciato in eredità preziosi insegnamenti che non dovremmo assolutamente scordare.
Per ciò che riguarda all'errore ricordiamo l'insegnamento fondamentale di generazioni di filosofi, bene ripreso dal grande Reason(7): l'errore è una eventualità di ogni attività umana.; esso si può prevenire ma non eliminare e quindi bisogna sistematicamente ricercarlo, evidenziarlo e prevederne le conseguenze.
Una percentuale significativa degli errori segue le leggi del caso: utilizzare algoritmi avanzati come quelli genetici produce nuove affascinanti realtà ma inevitabilmente conduce prima o poi ad errori che solo un piccolo animale, chissa perchè chiamato uomo può correggere…



Continua nella prossima pillola…

Riccardo De Gobbi e Giampaolo Collecchia

Bibliografia

1) Myura Nagendran, Yang Chen et Al.: Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies BMJ 2020;368:m689 http://dx.doi.org/10.1136/bmj.m689
2)http://www.consort-statement.org/checklists/view/32--consort-2010/66-title
3)https://www.tripod-statement.org/
4)https://get.covidence.org/cochrane-
5) http://www.probast.org/wp-content/uploads/2020/02/PROBAST_20190515.pdf
6) **Rossi Renato Luigi: Come leggere uno studio clinico. Il Pensiero Scientifico Editore: Roma 2021.**
pensiero.it/catalogo/libri/professionisti/come-leggere-uno-studio-clinico
7) Reason James: Human error: models and management BMJ VOLUME 320 18 MARCH 2000 www.bmj.com

Perapprofondire:
Collecchia G. De Gobbi R.: Intelligenza Artificiale e Medicina Digitale. Una guida critica. Il Pensiero Scientifico Ed. Roma 2020
pensiero.it/catalogo/libri/pubblico/intelligenza-artificiale-e-medicina-digitale