

Padova, 30 luglio 2021

INTELLIGENZA ARTIFICIALE IN BIOLOGIA: UNIPD COORDINA IL GRUPPO DI LAVORO INTERNAZIONALE

La pubblicazione di DOME rappresenta un passo fondamentale per migliorare la produttività dell'applicazione di metodi basati sul machine learning in campo biologico

Data, Optimization, Models and Evaluation ([DOME](#)) è un set di linee guida per la validazione dei



metodi di intelligenza artificiale applicati all'ambito biologico sviluppato da un gruppo di ricercatori internazionali, nell'ambito dell'infrastruttura europea [ELIXIR](#) per i dati biologici, **coordinato dal Professor Silvio Tosatto del Dipartimento di Scienze Biomediche dell'Università di Padova**. Il lavoro ambisce a creare un modello che permetta di valutare la qualità dei metodi applicati al settore della biologia. L'analisi si concentra sulla qualità dei dati utilizzati (*Data*), la strategia scelta in fase di training per

ottimizzare il modello da generare (*Optimization*), la capacità di riprodurre il modello generato (*Models*) e la valutazione della qualità del modello stesso (*Evaluation*).

Ma perché sono necessari degli standard? Il Machine Learning è un ambito dell'intelligenza artificiale (abbreviato AI, dall'inglese *artificial intelligence*) che si occupa di creare sistemi che apprendono o migliorano le proprie performance in base ai dati che gli vengono sottoposti ed è alla base delle maggiori applicazioni degli ultimi anni. Le potenzialità del machine learning attirano ormai da tempo l'interesse degli scienziati di ogni campo del sapere e, non da ultimo, del campo biologico e biomedico, in cui l'applicazione di questa tecnologia può davvero determinare benefici concreti per la salute di tutti.

«Dalla metà degli anni '90 ad oggi - **dice il Professor Silvio Tosatto** - è stata registrata una crescita esponenziale dei lavori scientifici con metodi di machine learning in campo biologico: dalle poche decine di pubblicazioni del 1996, l'anno scorso abbiamo raggiunto le oltre nove mila. Tuttavia il solo fatto che queste ricerche siano state realizzate con una nuova tecnologia non garantisce che si tratti sempre di lavori all'avanguardia, soprattutto in ambito biologico. Per questo è importante migliorare la trasparenza e riproducibilità dei metodi pubblicati».

Se in biologia vi sono molti tipi di data-set che si possono facilmente "dare in pasto" ad un algoritmo di machine learning, quali ad esempio immagini al microscopio e sequenze proteiche, buona parte dei dati biologici sono dati privi di struttura, variabili e difficilmente riproducibili, il che li rende inadatti ad essere analizzati da metodi di machine learning. A causa della loro natura variabile, sono infatti maggiormente soggetti a casi di *data leakage*.

Il problema della selezione dei dati da sottoporre ad un software in fase di training e poi in fase di valutazione finale del modello generato è centrale affinché il modello realizzato sia affidabile e privo di distorsioni. In fase di *training* (allenamento) gli algoritmi di machine learning sono in grado di imparare in un modo in certa misura analogo a ciò che accade nel cervello umano, analizzando i dati che gli vengono forniti (*training set*). Man mano che si estende la quantità di dati sottoposta ad analisi, il software accumula di volta



Professor Silvio Tosatto

in volta maggiori informazioni sull'attività che deve svolgere e impara per esperienza, più o meno come fa un bambino, a tentativi, imparando dagli errori e osservando chi lo circonda. Si creano delle vere e proprie reti neurali artificiali in grado di raggiungere performance notevoli.

La pubblicazione di DOME rappresenta un passo fondamentale per migliorare la produttività dell'applicazione di metodi basati sul machine learning in campo biologico e ci si augura che questo porterà a notevoli avanzamenti in tutte le aree di questo ambito scientifico.

Walsh, I., Fishman, D., Garcia-Gasulla, D. et al. DOME: recommendations for supervised machine learning validation in biology. *Nat Methods* (2021). <https://doi.org/10.1038/s41592-021-01205-4>
<https://dome-ml.org/>
<https://elixir-europe.org/focus-groups/machine-learning>
[DOME logo](#)