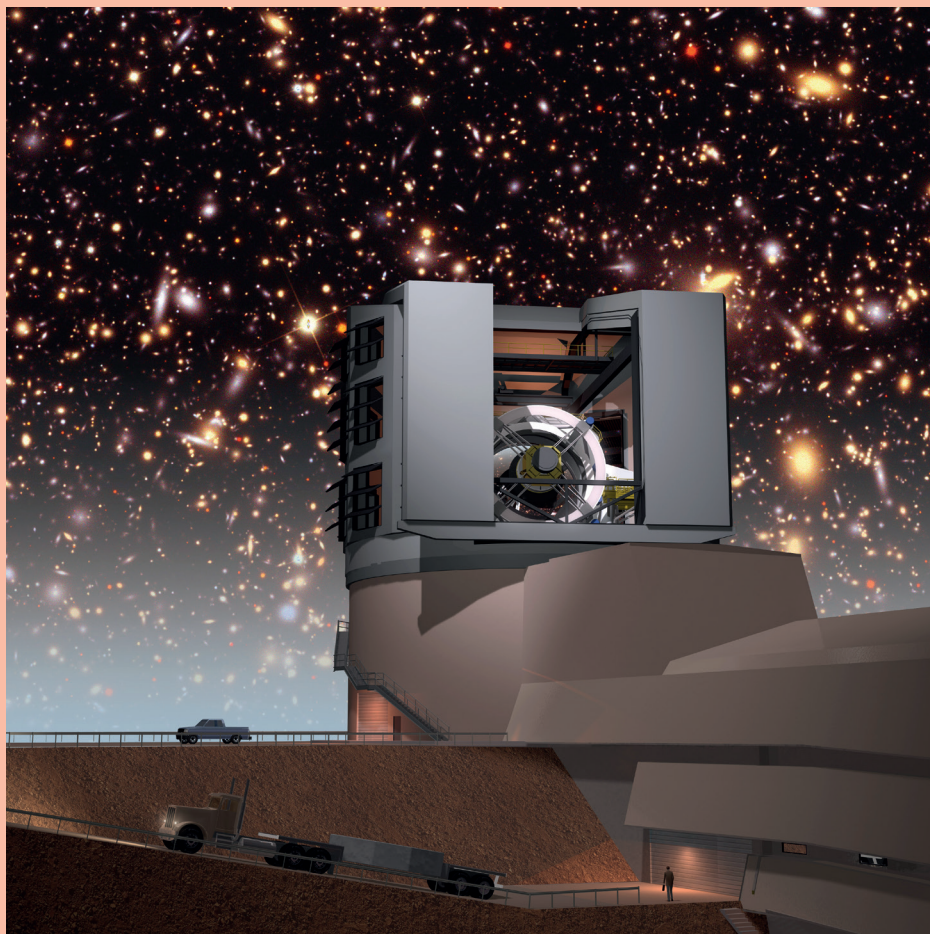


Nello zoo galattico

L'apprendimento profondo dell'universo

di Viviana Acquaviva



I termini “intelligenza artificiale” e “*machine learning*” sono divenuti ormai ubiqui in qualunque campo scientifico, dalla biologia alla chimica alla medicina, passando, ovviamente, per la fisica. Ma è difficile immaginare una disciplina che abbia subito una rivoluzione metodologica così rapida e sostanziale come l'astrofisica.

Le ragioni di questo cambiamento sono in parte da ricercarsi nella natura stessa dell'indagine sull'universo. I

metodi dell'intelligenza artificiale, dal *machine learning* al *data mining*, sono ideali per descrivere e investigare fenomeni i cui modelli fisici non sono sufficientemente sofisticati per catturarne la complessità. Tipicamente, questo accade quando abbiamo a che fare con oggetti estremamente complicati, che sono descritti da un numero di variabili molto elevato, o quando la nostra comprensione e interpretazione dei fenomeni fisici è limitata. E questa

a.

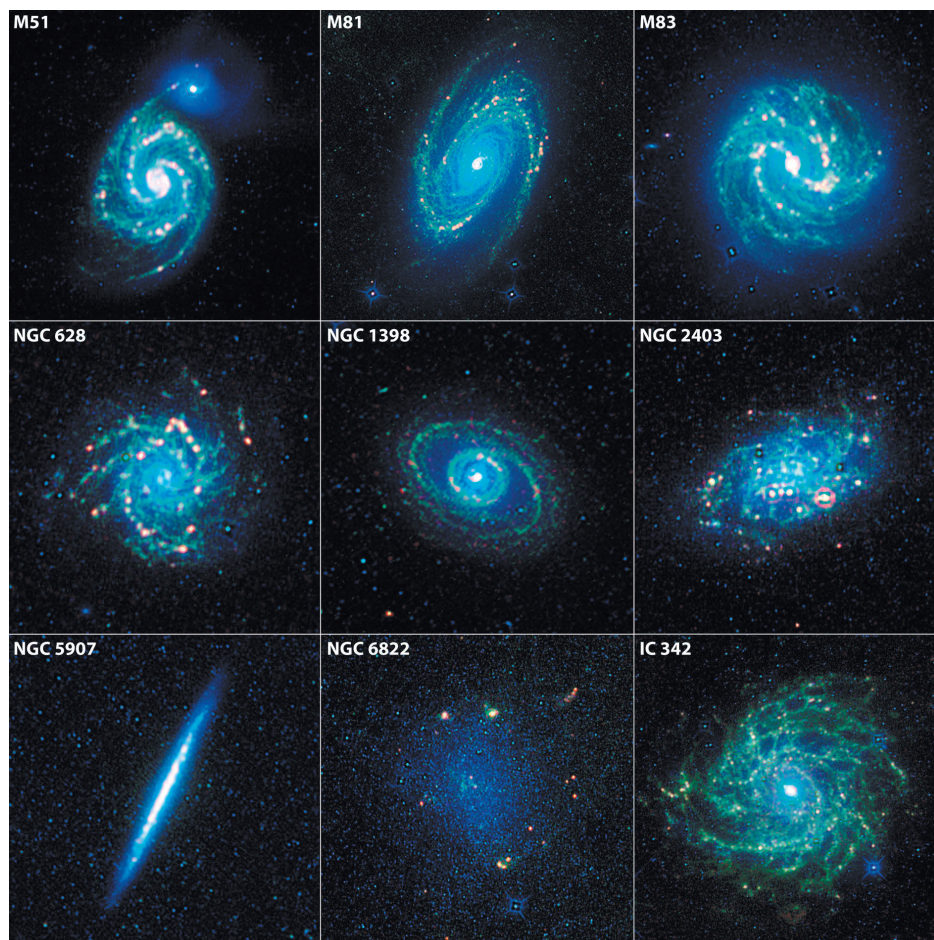
Il telescopio Lsst, attivo a partire dal 2023, produrrà immagini per miliardi di galassie, con un flusso di dati di decine di terabyte per notte.

descrizione calza a pennello a tantissimi problemi in ambito astrofisico, dalla formazione delle galassie alle dinamiche stellari, dalle simulazioni cosmologiche che mostrano la distribuzione della materia oscura nell'universo, all'individuazione di segnali variabili in un flusso di dati. Considerando anche il fatto che missioni come il Large Synoptic Survey Telescope (Lsst) genereranno nel giro di pochi anni una quantità di dati centuplicata rispetto agli osservatori delle generazioni precedenti, rendendo alcuni dei metodi correnti, come ad esempio l'ispezione manuale, completamente obsoleti, non c'è da stupirsi che l'astrofisica sia uno dei settori che ha adottato le nuove tecniche in modo rapido e probabilmente irreversibile. Ma quali sono le applicazioni del *machine learning* che hanno avuto più successo in astrofisica e cosmologia? La lista è certamente troppo lunga per essere completa. Un grandissimo punto di forza di queste tecniche, e specialmente del *deep learning*, è la loro efficacia nell'analizzare dati multidimensionali, come ad esempio le immagini, che contengono informazioni spesso difficili da descrivere in modo

analitico. Per esempio, identificare una galassia a spirale, vista sotto diverse possibili angolazioni, è un compito relativamente semplice per l'occhio (e il cervello) umano, ma scrivere una formula matematica - dunque, un modello - che possa "cercare" questa galassia nell'immagine e tenere conto di tutte le possibili prospettive è estremamente complicato. Qui le reti neurali vengono in nostro soccorso, perché la loro architettura permette facilmente di catturare correlazioni spaziali in due o tre dimensioni, purché il sistema abbia a disposizione una serie di esempi rappresentativi da cui imparare, un po' come un bambino impara facilmente a riconoscere gli animali dopo averli visti alcune volte in fotografia. Non a caso, una delle prime applicazioni del *machine learning* ai dati astronomici è stata quella di classificare le galassie in base alla loro morfologia, in collaborazione con il team di Galaxy Zoo, in cui il pubblico è stato

chiamato in causa per formare, sulla base del riconoscimento "umano", una collezione di elementi rappresentativi che potessero guidare l'algoritmo. Più in generale, l'abilità delle reti neurali nell'analizzare immagini permette al *machine learning* di migliorare la qualità dell'analisi dati rispetto ai metodi tradizionali, per cui le medesime immagini di stelle e galassie appaiono più nitide e ricche di informazioni se analizzate con i nuovi metodi. Inoltre, permette di catturare informazioni "nascoste" in strutture complicate, come i dettagli delle proprietà statistiche delle mappe della radiazione del fondo cosmico a microonde o del "lensing gravitazionale debole" (cioè la distorsione subita dalle immagini delle galassie lontane dovuta alla deflessione della luce da parte della materia incontrata nel percorso tra la sorgente e la Terra). Questo apre anche una nuova promettente direzione nel campo delle simulazioni cosmologiche, che

in generale richiedono l'uso di risorse computazionali molto elevate. Alcuni tipi di reti neurali, come le Gan, nate solo nel 2014 come algoritmi per riconoscere i falsi, sono state usate con successo per generare nuovi dati e di fatto ampliare il volume delle simulazioni, con un costo computazionale assai ridotto. Perfezionare questi metodi consentirà, in futuro, di ottenere simulazioni cosmologiche ad alta risoluzione su grandissimi volumi, che oggi risultano proibitive per via degli eccessivi tempi richiesti. Le Rnn si prestano a un'analisi molto efficace di dati in sequenza temporale (*time series*), che sono fondamentali per l'individuazione tempestiva di sorgenti variabili, come supernovae o raggi cosmici, e saranno certamente fondamentali sia nell'analisi dei dati di Lsst che nel campo nascente dell'astronomia multimessaggera, l'analisi combinata di segnali di origine elettromagnetica e gravitazionale. Ma non ci sono solo le reti neurali. Da un lato, l'astrofisica non è fatta solo di immagini. Dall'altro, tutti i metodi di *deep learning* tendono a sacrificare l'interpretabilità in favore della precisione. In altre parole, è difficile capire perché una rete neurale prende una certa decisione e - di conseguenza - imparare nuova fisica. Per questo motivo, mentre lo sviluppo di strumenti che possano decodificare l'*output* delle reti neurali diventa un argomento



b.
Forme e colori delle galassie nell'universo svelano la storia dell'evoluzione cosmica: le tecniche di *machine learning* consentono di classificarle correttamente in tempi molto rapidi.



c. La figura del *support scientist*, a metà tra il ricercatore e il programmatore, diventerà sempre più cruciale nelle scienze ad alto contenuto di dati, come l'astrofisica.

sempre più discusso, altri metodi più semplici e trasparenti hanno potenti applicazioni in campi in cui i modelli lasciano a desiderare. Un buon esempio è il caso del calcolo delle distanze delle galassie basato sulla fotometria (*photometric redshifts*). In questo caso, metodi di *machine learning* come Random Forest (una particolare combinazione di Bdt) o Principal Component Analysis Decomposition, o anche soltanto combinazioni lineari di spettri empirici, risultano spesso più precisi dei metodi basati sui modelli dell'emissione energetica delle galassie.

Per continuare, non possiamo non citare l'importanza di tecniche di *machine learning* nell'alleviare il problema sempre più pressante della rappresentazione e compressione delle informazioni contenute nei dati. Il volume dei dati prodotti dalle nuove generazioni di telescopi (per ritornare al caso di Lsst, dell'ordine di decine di terabytes per notte) è impossibile da maneggiare per un utente che lavora sul suo computer personale, e ovviamente la situazione peggiora quando si considerano i prodotti derivati come cataloghi, distribuzioni di probabilità di parametri, e così via. La splendida e doverosa pratica della condivisione pubblica dei dati, originali e derivati, tipica dell'astrofisica e pilastro della riproducibilità dei risultati richiesta dal metodo scientifico, rischia di arenarsi per un problema tecnico di accesso ai dati stessi. Mentre alcuni di questi problemi potranno essere risolti solo con l'allocatione di nuove risorse, il *machine learning* può aiutare tramite l'uso di tecniche specializzate, note come "riduzione della dimensionalità", che consentono di individuare le proiezioni più interessanti nello spazio dei dati e ridurne di fatto la dimensione, con un sacrificio minimo del contenuto di informazione.

Da ultimo, è bene ricordare che la rivoluzione creata

dal diffondersi dei metodi del *machine learning* ha delle implicazioni profonde a livello umano. Cambia la figura del ricercatore, che diventa forse meno specializzato e deve saper parlare il linguaggio della programmazione, della visualizzazione dei dati e diventare, ora più che mai, un "narratore" di storie complesse e affascinanti. Aumentano le possibilità di ricerca interdisciplinare, in cui la *data science* fornisce quel linguaggio comune che forse finora è un po' mancato a statistici, informatici e astronomi. Poi si aprono nuove possibilità di carriera, a cavallo tra l'astrofisica e l'informatica, come dimostrato dalle nuove figure professionali come il "*support scientist*", sempre più comuni negli annunci di lavoro. E più di ogni altra cosa, è importante tener presente questi cambiamenti nella formazione dei fisici di domani, così che possano sfruttare al meglio questa eccezionale occasione di analizzare dati in maniere finora impossibili o sconosciute, e scoprire nuove verità sull'universo che ci circonda.

Biografia

Viviana Acquaviva, astrofisica, utilizza i metodi dell'intelligenza artificiale nelle sue attività di ricerca sull'evoluzione dell'universo. È professore presso la City University of New York, membro del Centro di Astrofisica Computazionale del Flatiron Institute, *visiting scientist* presso il Museo di Storia Naturale di New York e ambasciatrice (Harlow Shapley Lecturer) per la American Astronomical Society. Nel 2018 è stata premiata come una delle 50 donne italiane più influenti nel campo della tecnologia da "InspiringFifty".

DOI: 10.23801/asimmetrie.2019.27.7