

The imitation game

Apprendimento per rinforzo

di Antonio Celani

Un alieno sbarca sulla Terra. È arrivato in un bel posto di campagna, ma si sente a disagio. Sul suo pianeta tutto è leggero e immateriale. Il corpo umano che indossa per non dare nell'occhio invece è scomodo e ingombrante. Si sente anche spaesato: si era ripromesso di studiare tutto il possibile su questo pianeta durante il lungo viaggio, ma poi ha dormito tutto il tempo.

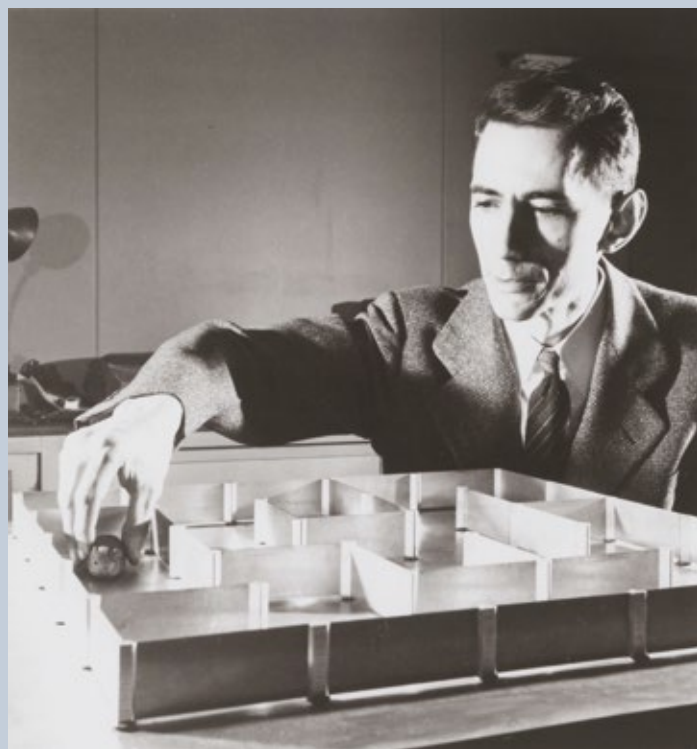
Ma ecco, un umano si avvicina, molto velocemente, in bicicletta. Poi si ferma, scende agilmente dalla bici che appoggia a un cancello, entra in una casa. Incuriosito l'alieno si avvicina. Anche lui vorrebbe muoversi con quella stessa destrezza. Ispeziona la bici e monta in sella. Mentre mette i piedi sui pedali, inizia a inclinarsi su un lato e cade rovinosamente a terra. Dopo averci pensato un po', decide di consultare il *database* collegato alla sua mente e in pochi secondi visualizza un milione di video di persone che vanno in bicicletta. "Ora so perfettamente cosa fare", si dice. Imitando i gesti che ormai gli sono familiari, si rimbecca i pantaloni, risale in sella e inizia a pedalare con entusiasmo. Ma dopo poche vigorose pedalate, purtroppo in senso sbagliato, la bici rimane sul posto e lui cade di nuovo.

"Devo chiedere aiuto a un esperto", pensa. Allora guarda un milione di *tutorial* su come si va in bicicletta, impara tutti i termini tecnici e studia l'equivalente di due dottorati in ingegneria e in fisica. Pensa: "Questa volta non posso sbagliare". E infatti sembra funzionare, la bici si muove, accelera, ma dopo pochi metri il novello ciclista perde l'equilibrio e finisce in un fosso.

Allertata dai rumori, la proprietaria della bicicletta esce di casa e si scaglia contro l'alieno urlando: "Ma cosa sta facendo?". Poi vedendo la sua espressione confusa e imbarazzata, si rende conto che non si tratta di un ladro. "Cosa pensava di fare?" chiede, più incuriosita che arrabbiata adesso. "Imparare ad andare in bicicletta", risponde. "Ho letto tutto quello che c'era da leggere, guardato tutto quello che c'era da guardare e studiato tutto quello che c'era da sapere, ma non riesco". Lei ride di cuore e gli spiega: "Ma signore, non lo sa che certe cose si imparano solo facendole?".

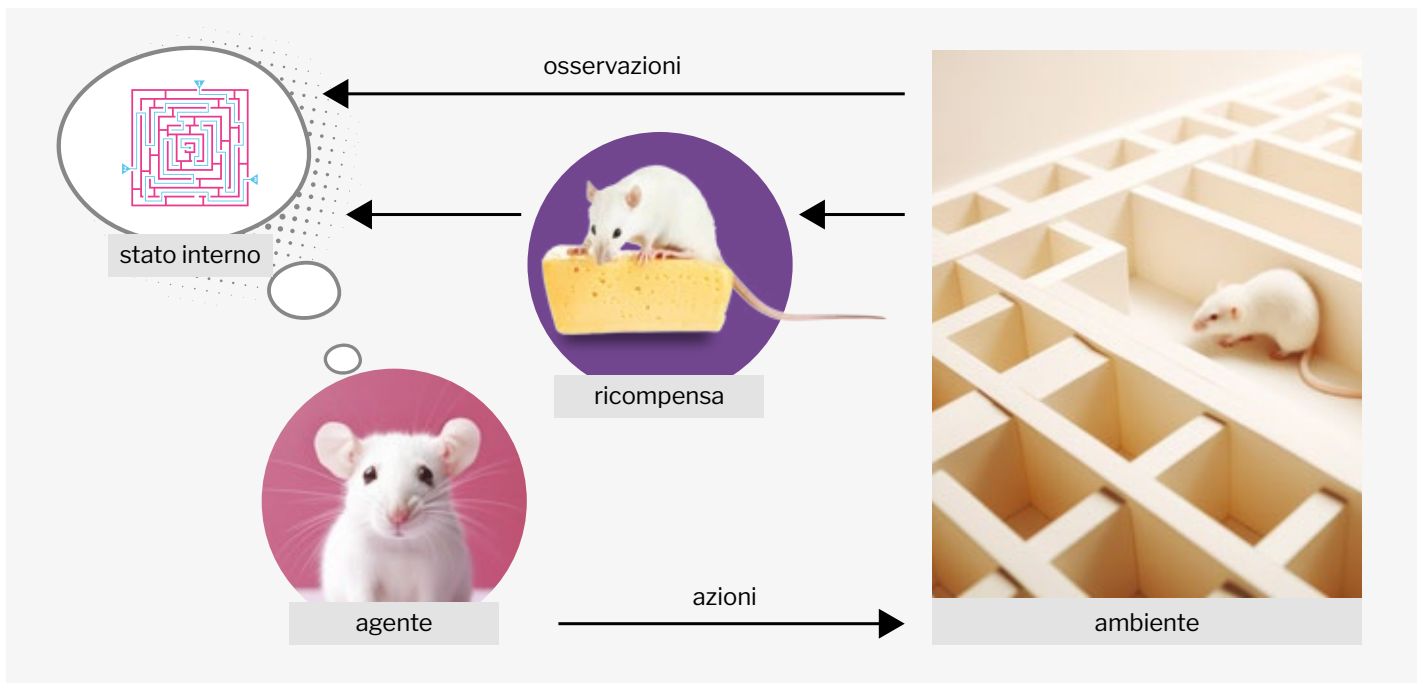
Quello che all'alieno sfugge è l'importanza dell'"apprendimento attivo".

Per noi umani, è un'esperienza talmente naturale che la diamo spesso per scontata. In pedagogia, l'apprendimento attivo considera i bambini protagonisti del proprio apprendimento,



a.

Claude Shannon, il fondatore della teoria dell'informazione con il topo meccanico "Theseus" in un labirinto, uno dei primi esperimenti di apprendimento in robotica. Shannon era anche noto per essere un abile giocoliere e ciclista di monocicli, quindi lui stesso dotato di una notevole "intelligenza incarnata".



b.

L'apprendimento per rinforzo descrive il processo di continua interazione tra agente e ambiente, qui esemplificata da un topolino in un labirinto. L'agente riceve informazioni sotto forma di osservazioni che possono essere in questo caso visive od olfattive. Le osservazioni contribuiscono alla formazione di una rappresentazione interna dell'ambiente, sotto forma di una mappa o anche più astratta. Lo stato interno guida il processo di decisione e determina le azioni, cioè i movimenti del topolino. Durante la navigazione, o anche solo al momento del raggiungimento del traguardo, l'agente riceve ricompense che rinforzano le decisioni favorevoli e scoraggiano quelle sfavorevoli. Dopo molte ripetizioni del compito, il topolino impara la strada più breve che gli permette di raccogliere la maggior quantità di formaggio.

non destinarli passivi di informazioni. La conoscenza viene costruita attivamente tramite l'esperienza diretta e pratica e non semplicemente assorbita. L'esplorazione, la manipolazione e la sperimentazione sono quindi fondamentali. I bambini utilizzano i sensi e le azioni materiali per comprendere il mondo: si tratta di una comprensione intuitiva e non formale della fisica e delle sue leggi, che possiamo chiamare "intelligenza incarnata".

L'apprendimento attraverso l'interazione fisica con l'ambiente non si limita ai bambini ma riveste un ruolo cruciale per tutti gli animali e in ogni stadio della loro vita. Per esempio, è fondamentale per imparare a ottimizzare l'acquisizione di risorse vitali come cibo e acqua, esplorando attivamente nuove fonti e apprendendo le strategie più efficaci per sfruttarle. Così anche per imparare a navigare con efficacia in spazi anche sconosciuti, creando mappe cognitive e scegliendo percorsi ottimali. In questo apprendimento per tentativi ed errori, un ruolo fondamentale è giocato dalla gratificazione ricevuta in risposta a un certo comportamento, che può consistere in una ricompensa materiale o anche in una soddisfazione intrinseca. Il processo attraverso il quale un animale impara comportamenti sempre più efficienti nel raccogliere delle ricompense si chiama "apprendimento per rinforzo" (*reinforcement learning*).

A differenza di altri tipi di apprendimento, i segnali di rinforzo non danno istruzioni esplicite su quale comportamento adottare, ma esprimono una forma di incoraggiamento o di dissuasione piuttosto generica. Inoltre, la gratificazione può arrivare solo in seguito a una lunga e complessa catena di decisioni, spesso caratterizzate dalla mancanza di ricompense intermedie, se non addirittura marcate da esperienze spiacevoli, proprio come quando si impara ad andare in bicicletta. Il *feedback* limitato e la distanza temporale tra le azioni prese e le loro conseguenze a lungo termine rendono difficile identificare le cause prime di successi e fallimenti, rallentando e frustrando il processo di apprendimento.

C'è anche un altro problema. Da una parte c'è la necessità di cercare attivamente informazioni che accrescono la conoscenza dell'ambiente, cioè il processo di esplorazione caratteristico dell'apprendimento attivo. Ma dall'altra parte, l'obiettivo è sempre quello di massimizzare la ricompensa che l'ambiente offre, cioè sfruttare le conoscenze che si sono già acquisite per il proprio scopo. Queste due richieste sono quasi sempre in tensione tra di loro. Se l'alieno si limitasse ad agire unicamente in base alla gratificazione ottenuta, probabilmente desisterebbe molto presto, visti i



risultati deludenti, se non proprio punitivi, dei primi tentativi. Viceversa, se una volta imparato a pedalare con una certa destrezza si abbandonasse al gusto per le più audaci sperimentazioni, magari guidare la bici con il manubrio invertito, la gratificazione ne risentirebbe, per usare un eufemismo. Trovare il giusto equilibrio tra curiosità e produttività è la chiave dell'apprendimento per rinforzo. Quindi, se vogliamo insegnare a una macchina a svolgere ogni tipo di compito che richiede una ripetuta interazione con un mondo esterno, come andare in bici, giocare a ping pong o a un videogioco, dobbiamo affrontare queste difficoltà in modo sistematico.

Qui entra in gioco la teoria dell'apprendimento per rinforzo. Si tratta di un insieme di metodi in cui confluiscono diverse discipline della matematica come la teoria dei processi di decisione, l'ottimizzazione, la statistica e la teoria dell'informazione, nutriti da una corrente sotterranea che attinge dalla psicologia, dalle neuroscienze e dall'economia, che sfociano in algoritmi messi in pratica dagli esperti di informatica e robotica, da ingegneri e da fisici. Questo approccio interdisciplinare permette di descrivere una grande varietà di situazioni di interesse pratico con un unico

linguaggio astratto, in cui nozioni come agente, ambiente, ricompense, osservazioni, strategie hanno un significato ben preciso che si declina poi caso per caso.

L'apprendimento per rinforzo ha diversi punti di contatto con la fisica. Innanzitutto, l'ottimizzazione dei processi di decisione è essenzialmente un problema di controllo di un sistema dinamico. Ad esempio, possiamo pensare alla robotica come una versione (molto) più complicata del problema di scoprire quando e quanta forza applicare a un sistema di punti materiali affinché raggiungano una certa configurazione desiderata in un dato tempo. Anche se non conosciamo tutti i parametri che definiscono questo sistema (masse, forze interne, attriti, perturbazioni, ecc.) possiamo comunque imparare il controllo ottimale per tentativi ed errori grazie all'apprendimento per rinforzo.

A un livello ancora maggiore di astrazione, la tensione tra ricompensa ed esplorazione ricorda il concetto di bilanciamento tra energia ed entropia in meccanica statistica. Questa analogia ha ispirato molti algoritmi di uso corrente in cui il processo di apprendimento è interpretato come un lento "raffreddamento" verso le strategie di decisione ottimali. Inoltre, dal punto di vista teorico, questa

c.
Un singolo algoritmo di apprendimento per rinforzo ha imparato a giocare a livelli sovrumani tre giochi diversi (scacchi, go e shōgi, quest'ultimo visibile in foto), battendo programmi specializzati e campioni del mondo.



connessione fornisce la chiave per comprendere la relazione tra apprendimento e teoria dell'informazione.

L'apprendimento per rinforzo ha già raggiunto traguardi importanti nelle applicazioni dell'intelligenza artificiale. Ha dimostrato abilità superiori a quelle umane in giochi complessi come *go* e *shōgi* e si è distinto nel controllo di robot per manipolazioni precise e compiti complessi, come risolvere il cubo di Rubik. Tra le maggiori aspettative per il futuro c'è sicuramente il suo uso in medicina per la personalizzazione delle cure e l'assistenza alla

chirurgia robotica minimamente invasiva. L'importanza dell'apprendimento per rinforzo nell'ambito dell'intelligenza artificiale è stata recentemente riconosciuta con il conferimento del "Turing award" a Richard Sutton e Andrew Barto, che hanno introdotto a partire dagli anni '80 la maggior parte dei concetti e degli algoritmi alla base di questa disciplina. Grazie ai recenti sviluppi, possiamo dire che la predizione di "macchine che imparano dall'esperienza", che fece Alan Turing nel 1947, sia adesso una realtà concreta. Per gli alieni in bicicletta invece, dovremo ancora aspettare.

d. Richard Sutton (a sinistra) e Andrew Barto (a destra), i pionieri della teoria dell'apprendimento per rinforzo. In mezzo a loro il collega John Moore.

Biografia

Antonio Celani è *senior research scientist* presso l'International Center for Theoretical Physics di Trieste, dove coordina le attività della sezione di Quantitative Life Sciences. Si occupa di apprendimento per rinforzo e delle sue applicazioni al comportamento animale e artificiale.