Le Scienze
EDIZIONE ITALIANA DI SCIENTIFIC AMERICAN

Se l'intelligenza artificiale impara da sola come vincere

NUOVI ALGORITMI DI "APPRENDIMENTO PROFONDO" SONO IN GRADO DI IMPARARE DA SOLI A GIOCARE A UNA MOLTEPLICITÀ DI VIDEOGIOCHI E DI SCOPRIRE AUTONOMAMENTE LE STRATEGIE MIGLIORI PER VINCERE. LE IMPRESSIONANTI CAPACITÀ DI QUESTE FORME DI INTELLIGENZA ARTIFICIALE METTONO PERÒ IN EVIDENZA PER LA PRIMA VOLTA NELLA STORIA DELL'UMANITÀ UNA TOTALE DISSOCIAZIONE FRA INTELLIGENZA E CONSAPEVOLEZZA.

Nel corso della storia umana, l'intelligenza e la consapevolezza sono stati due concetti strettamente legati. Se avete un sacco della prima, si presume che, in qualche modo più o meno mal definito, siate più consapevoli del tizio dall'aria ottusa in fondo alla strada. Una persona intelligente sarà anche molto consapevole, in grado di parlarvi dettagliatamente delle sue esperienze (perché la coscienza è questo: la capacità di sperimentare qualcosa, qualsiasi cosa, che si tratti di un mal di denti, della visione di una casa giallo canarino o di una rabbia bruciante).

Ma questo stretto rapporto potrebbe sgretolarsi.

Prendiamo le ultime vicende di DeepMind, una piccola azienda di Londra di cui è stato cofondatore, nel 2011, il britannico Demis Hassabis, bambino prodigio degli scacchi, designer di videogame e neuroscienziato computazionale. Lo scorso anno DeepMind è stata acquistata per centinaia di milioni di dollari da Google. Il suo nuovo codice fa cose mozzafiato: impara da solo a giocare ai videogiochi, e spesso molto meglio dei giocatori umani. La svolta tecnica è descritta in uno studio pubblicato a febbraio su *Nature*.

Per farvi un'idea, cercate su YouTube il video intitolato *DeepMind Artificial Intelligence @ FDOT14*. Si tratta di un breve estratto, ripreso con uno smartphone del discorso di Hassabis alla conferenza Tech 2014, dedicato a un algoritmo informatico che impara a giocare il classico arcade Breakout. Lo scopo del gioco, una variante di Pong, è rompere dei mattoncini allineati in righe sulla parte superiore dello schermo usando una palla che rimbalza in alto e sulle pareti laterali. Se la palla tocca la parte inferiore dello schermo, il giocatore perde.

Co-creato da Steve Wozniak nel 1976, per gli standard odierni il gioco è primitivo ma comunque avvincente. Hassabis lo ha usato per spiegare il problema al pubblico. All'inizio l'algoritmo non sapeva nulla

e muoveva la barra in modo casuale e senza molta coordinazione, colpendo la palla solo occasionalmente. Dopo un'ora di allenamento, giocando più e più volte, le prestazioni sono migliorate, riuscendo spesso a rinviare la palla e a rompere i mattoni. Dopo due ore di allenamento, è diventato più bravo della maggior parte degli esseri umani, respingendo palle veloci e ad angoli stretti.

I programmatori hanno lasciato che l'algoritmo continuasse a giocare da solo, e a migliorare. Dopo quattro ore di gioco, l'algoritmo ha scoperto una strategia innovativa per Breakout, che ha fatto schizzare le sue prestazioni ben oltre quelle mai ottenute da qualsiasi essere umano. L'algoritmo ha compiuto l'impresa imparando a scavare un tunnel attraverso la parete di mattoncini a partire da un lato, in modo che la palla distruggesse rapidamente un gran numero di mattoni da dietro. Molto intelligente. Il successo è stato così impressionante che tutti gli esperti presenti sono esplosi in uno scrosciante applauso spontaneo (un evento raro in una conferenza scientifica).

Per capire che cosa sta succedendo e perché è qualcosa di notevole, andiamo a dare un'occhiata più approfondita. L'algoritmo ha tre caratteristiche, tutte riprese dalla neurobiologia: apprendimento con rinforzo, reti neurali a strati di convoluzione (*convolutional neural networks*) e *loop* sulla memoria selettiva.

Un'eredità duratura del comportamentismo, l'indirizzo che ha dominato lo studio del comportamento umano e animale nella prima parte del XX secolo, è l'idea che gli organismi imparano il comportamento ottimale mettendo in relazione la conseguenza di una particolare azione con uno stimolo specifico che l'ha preceduta. Questo stimolo è detto rinforzo del comportamento.

Prendiamo il caso del mio cane Ruby, quando, da cucciolo, l'ho dovuto educare. Subito dopo avergli dato da bere, a intervalli prestabiliti, lo portavo in un punto particolare del giardino e aspettavo. A un certo punto, avrebbe fatto pipì spontaneamente,e io l'avrei riempito di lodi. Se capitava un "incidente" in casa, lo sgridavo severamente. I cani rispondono bene a questi segnali sociali positivi e negativi. Nel giro di un mese o due Ruby aveva imparato che se a uno stimolo interno, la vescica piena, seguiva un certo comportamento – fare pipì nel solito posto – poteva aspettarsi un premio ed evitare una punizione.

L'apprendimento per rinforzo è stato formalizzato e implementato nelle reti neurali per insegnare ai computer come giocare vari giochi. Gerald Tesauro della IBM ha usato una versione particolare di apprendimento per rinforzo – il cosiddetto apprendimento per differenze temporali - per la progettazione di una rete che gioca a backgammon. Il programma analizza la tavola di gioco ed esamina tutte le possibili mosse lecite e le risposte del giocatore avversario a queste mosse. Tutte le posizioni di gioco che ne risultano vanno ad alimentare il cuore del programma, la sua funzione di valore.

La mossa scelta dal programma è quella che porta alla posizione sulla tavola con il punteggio più alto. Dopo una mano, la rete è leggermente ottimizzata, così che il programma prevede che cosa succederà un po' meglio di quanto poteva fare dopo la sua mossa precedente. Partendo da zero, continua a migliorare per tentativi ed errori. Ciò che rende complicato l'apprendimento per rinforzo è che di solito c'è un notevole ritardo tra una mossa e il suo esito utile o dannoso. Il superamento di questo handicap richiede addestramento, addestramento e ancora addestramento: arrivare a battere a backgammon un giocatore umano esperto richiedeva al programma di Tesauro di giocare 200.000 partite contro se stesso.

Il secondo ingrediente del successo di DeepMind si chiama rete neurale a strati di convoluzione. Si basa su un modello dei circuiti cerebrali scoperto nel sistema visivo dei mammiferi da Torsten Wiesel e David H. Hubel fra la fine degli anni cinquanta e i primi anni sessanta. (Per questo lavoro Hubel e Wiesel furono poi insigniti del premio Nobel.) Il modello postula uno strato di elementi, o unità, di elaborazione, che calcolano una somma ponderata dei segnali in ingresso. Se la somma è sufficientemente grande, il modello manda un segnale in uscita, altrimenti rimane "spento".

Alcuni teorici considerano il sistema visivo null'altro che una cascata di strati simili di elaborazione, indicata con il nome di rete *feedforward*. Ogni strato riceve un input da un livello precedente e invia un output al livello successivo. Il primo strato è la retina che intercetta la pioggia di fotoni in arrivo, registra le variazioni di luminosità dell'immagine e passa questi dati alla fase di elaborazione successiva. L'ultimo strato è costituito da un gruppo di unità che segnalano se alcuni elementi di alto livello, per esempio vostra nonna o Jennifer Aniston, sono presenti in quell'immagine.

I teorici dell'apprendimento hanno sviluppato efficaci metodi matematici per regolare i "pesi" di queste unità - ossia quanto debba essere influente un input rispetto a un altro - ottenendo reti feedforward che imparano a svolgere specifici compiti di rilevazione.

Per esempio, una rete è esposta a decine di migliaia di immagini prese da Internet, ciascuna delle quali è classificata in un certo modo a seconda che includa un gatto o no. Dopo ogni esposizione, tutti i pesi sono leggermente modificati. Se l'addestramento è abbastanza lungo (anche in questo caso l'addestramento deve essere davvero intensivo) e le immagini sono elaborate da reti abbastanza profonde, ossia con molti strati di elementi di elaborazione, la rete neurale riesce a fare una generalizzazione ed è in grado di riconoscere con precisione se una nuova fotografia contiene un felino. La rete ha appreso, in modo supervisionato, a distinguere le immagini dei gatti da quelle di cani, persone, automobili e così via.

La situazione non è dissimile da quella di una madre che sfogliando un libro illustrato con il suo bambino, gli indica tutti i gatti. Le reti a strati di convoluzione profondi sono di gran moda fra Google, Facebook, Apple e altre aziende della Silicon Valley che cercano di etichettare automaticamente le immagini, tradurre il parlato in testo, trovare i passanti in un video e identificare i tumori nelle scansioni del seno.

L'apprendimento supervisionato è differente dall'apprendimento con rinforzo. Nel primo, ogni immagine in ingresso è accoppiata a un'etichetta (un'immagine contiene un gatto); nel secondo, no. Nell'apprendimento per rinforzo, l'effetto di ogni mossa sul punteggio di gioco si sviluppa nel tempo, le azioni possono produrre dei benefici (un punteggio migliore) anche molte mosse più tardi.

Hassabis e i suoi collaboratori hanno usato una variante dell'apprendimento per rinforzo detta *Q-learning*, che fa da supervisore alla rete di apprendimento profondo. L'input della rete consiste in una versione sfocata dello schermo colorato di gioco, che include il punteggio – uguale a quello che vede un giocatore umano – ma anche le schermate associate alle ultime tre mosse. L'output della rete è un comando al joystick di spostarsi in una delle otto direzioni cardinali, con o senza l'azionamento del pulsante "Fuoco". Partendo da un'impostazione casuale dei suoi pesi, la proverbiale tabula rasa, l'algoritmo è arrivato a capire quali azioni portano a un punteggio finale più alto, quando la barra ha maggiori probabilità di intercettare la palla sul fondo in modo da respingerla con una traiettoria tale da rompere i mattoni. In questo modo, la rete ha imparato - attraverso la ripetizione e l'apprendimento rinforzato - i metodi più efficaci per giocare a Breakout, superando di uno sconvolgente 1327 per cento il punteggio di un testatore professionista del gioco.

Il terzo componente critico dell'algoritmo è il loop della memoria selettiva, simile a quello che si pensa si verifichi nell'ippocampo, una regione del cervello associata alla memoria. Nell'ippocampo, i modelli di attività delle cellule nervose associate a una particolare esperienza, per esempio quella di percorrere un labirinto, vengono "rivisti", ma a un ritmo più veloce. L'algoritmo, cioè, può richiamare dalla sua memoria, a caso, un particolare episodio di gioco (comprese le proprie mosse) incontrato in precedenza, adeguare la propria azione sulla base dell'esperienza precedente e aggiornare di conseguenza la sua funzione di valutazione.

Ai progettisti di DeepMind però non bastava che il loro algoritmo imparasse un solo gioco, e lo hanno allenato a 49 diversi giochi per Atari 2600, tutti sviluppati per generazioni di adolescenti. Fra questi vi

erano Pinball, StarGunner, Robot Tank, Road Runner, Pong, Space Invaders, Ms. Pac-Man, Alien e la vendetta di Montezuma. In tutti i casi è stato usato sempre lo stesso algoritmo, con le stesse impostazioni. L'unica cosa che cambiava era l'output, calibrato sulle esigenze specifiche di ciascun gioco. I risultati hanno sbaragliato quelli di tutti gli altri algoritmi "giocatori". Non solo, in 29 di questi giochi l'algoritmo ha superato del 75 per cento o più un testatore professionista umano, battendolo a volte con un margine molto ampio.

L'algoritmo ha i suoi limiti. Le sue prestazioni migliorano sempre più lentamente via via che i giochi richiedono una pianificazione progressivamente più a lungo termine. Per esempio, le prestazioni dell'algoritmo per MS Pac-Man sono abbastanza modeste perché il gioco richiede di scegliere il percorso da seguire nel labirinto per evitare di essere mangiato da un fantasma anche a 10 o più mosse di distanza.

Il programma, tuttavia, preannuncia un nuovo livello di sofisticazione nell'intelligenza artificiale. Deep Blue, il programma IBM che nel 1997 sfidò a scacchi il grande maestro Garry Kasparov, e Watson, il sistema IBM che ha battuto Ken Jennings e Brad Rutter nel quiz Jeopardy, erano raggruppamenti di algoritmi altamente specializzati messi a punto con cura artigianale per affrontare un tipo particolare di problema.

Il segno distintivo della nuova generazione di algoritmi è che, come le persone, imparano dai propri successi e dai propri fallimenti. Partendo esclusivamente dalla sfilza di pixel della schermata di gioco, alla fine gareggiano in giochi sparatutto, di boxe, di corse automobilistiche.

Naturalmente, i mondi in cui operano sono fisicamente molto semplificati e ubbidiscono a regole molto rigide, e le loro azioni sono molto limitate. Non vi è alcun segno di sensibilità in questi algoritmi. Non hanno alcuno dei comportamenti che associamo con la coscienza. Secondo gli attuali modelli teorici della coscienza le reti convoluzionali profonde non sono consapevoli.

Sono degli zombie che agiscono nel mondo, ma lo fanno senza alcun sentimento, mostrando una forma di intelligenza fredda, limitata e aliena: un algoritmo «sfrutta spietatamente le debolezze del sistema che trova. In modo del tutto automatico», ha detto Hassabis nel suo discorso del 2014. Questi algoritmi, inclusi quelli che controllano le auto a guida autonoma di Google o quelli che eseguono gli scambi sui mercati finanziari, dimostrano che per la prima volta nella storia del pianeta, l'intelligenza è completamente dissociata dalla sensibilità, dalla coscienza.

Sono intelligenti, nel senso che possono imparare ad adattarsi a nuovi mondi, motivati unicamente dalla massimizzazione dei premi quali sono definiti dal punteggio di gioco. Non ho alcun dubbio che i progettisti di DeepMind siano impegnati a lavorare su motori di apprendimento più sofisticati, per insegnare ai loro algoritmi a dominare in prima persona giochi sparatutto come Doom o Halo, o giochi di strategia, come StarCraft. Questi algoritmi riusciranno a eseguire sempre meglio compiti specifici in nicchie molto specifiche, che nel mondo moderno abbondano. Ma non creeranno né apprezzeranno l'arte, né si meraviglieranno di fronte a un bellissimo tramonto.

Se questa sarà una buona cosa per l'umanità, lo si vedrà a lungo termine. La ragione per cui dominiamo il mondo naturale non è perché siamo più veloci o più forti, e men che meno più saggi di altri animali, ma perché siamo più intelligenti. Forse questi algoritmi di apprendimento sono nuvole scure all'orizzonte dell'umanità. Forse saranno la nostra ultima invenzione.

CHRISTOF KOCH

Christof Koch è presidente dell'Allen Institute for Brain Science, editorialista di *Scientific American Mind* (la versione statunitense di *Mente e Cervello*) e membro del consiglio di amministrazione di *Scientific American*.

La versione originale di questo articolo è apparsa sul numero di luglio/agosto 2015 (n.4, vol. 26) di *Scientific American Mind*.