

Computer intelligenti: un percorso tra informatica e filosofia

Gaetano Francesco Anastasi

ITET G. Caruso, Alcamo (TP), Italy
gaetanofrancesco.anastasi@istruzione.it

Sommario Se ci riferiamo all'intelligenza come alla capacità di risolvere problemi complessi mettendo in relazione elementi diversi tra loro, i computer che oggi utilizziamo possono già essere considerati intelligenti. La nostra interazione con essi è diventata talmente complessa che interrogarci sulla natura del nostro rapporto reciproco non è più solamente un problema speculativo ma un argomento di attualità immediata, che rivendica il suo spazio anche nell'ambito della formazione scolastica. In questo articolo si propone un percorso interdisciplinare che possa guidare le nuove generazioni ad una riflessione sui possibili scenari futuri.

1 Introduzione

Negli anni '80 del secolo appena trascorso i computer sono entrati timidamente nella nostra quotidianità, per poi reclamare uno spazio che adesso è preponderante. I computer sono ora pervasivamente immersi nel nostro spazio e nel nostro tempo e l'interazione con essi è diventata estremamente complessa. Interrogarci dunque sulla natura del nostro rapporto con i computer non è più solamente un problema speculativo ma un argomento di attualità immediata e di necessario interesse. Come tale, questo argomento rivendica il suo spazio anche nell'ambito della formazione scolastica: la metacognizione sulle esperienze scolastiche dovrebbe anche significare uscire da una dimensione di studio del passato per riflettere sul presente e provare ad immaginare il futuro.

Scopo di questo articolo è appunto quello di analizzare il rapporto tra intelligenza artificiale e coscienza artificiale, proponendo un percorso formativo interdisciplinare che possa guidare gli allievi nel tentativo di trovare risposte a questo tipo di domande. L'articolo è organizzato come segue: nella Sezione 2 ci si concentra sulle reti neurali come approccio all'intelligenza artificiale. La Sezione 3 accenna al tema della coscienza artificiale, ponendosi a cavallo tra il piano filosofico e quello scientifico-informatico. Nella Sezione 4 viene proposto un progetto didattico che possa implementare il percorso elaborato nelle sezioni precedenti. La Sezione 5 conclude l'articolo.

2 Intelligenza artificiale

Quando si discute informalmente di intelligenza artificiale, il pensiero corre soprattutto al cinema di fantascienza, che, più di tutte le altre arti, ha contribuito

nelle scorse decadi a creare una visione di come sarebbe stato il mondo nel momento in cui gli umani avessero integrato le “macchine pensanti” nella loro società. Se prendiamo ad esempio pellicole come *Blade Runner* [1] – ne potremmo citare diverse altre – ci si accorge che il termine intelligenza artificiale rimanda a scenari apocalittici, in cui generazioni di robot si ribellano ai loro creatori cercando di distruggerli. Tali robot hanno sembianze umane e si muovono come gli umani, agiscono come gli umani ma non provano emozioni: sentimenti come disgusto, rabbia, amore, compassione non alterano il loro comportamento, né influenzano in alcun modo il loro agire. Dal punto di vista ingegneristico, invece, diversi sono gli approcci che si possono ascrivere all’intelligenza artificiale. Tra questi, le *reti neurali* utilizzano l’esperienza di un medesimo fatto allo scopo di apprendere e suscitano particolare interesse nell’ambito dell’intelligenza artificiale, poiché costituiscono un valido strumento per tentare di replicare il cervello umano e studiarne i processi cognitivi.

Le reti neurali artificiali sono nate con l’obiettivo di riprodurre il modello di funzionamento del cervello umano. Un neurone artificiale presenta lo stesso modello computazionale di quello biologico, ha più ingressi ed una sola uscita. Ad ogni canale di ingresso viene associato un peso, che indica la forza della connessione e specifica se il canale è eccitatore o inibitore. L’uscita del neurone viene calcolata dalla sua funzione di attivazione $y = f(z)$, in cui la variabile z è proprio la somma pesata degli ingressi. In letteratura sono state proposte varie funzioni di attivazione, che contraddistinguono la rete neurale artificiale e spesso ne determinano la maggiore o minore adattabilità a svolgere un determinato compito. Uno dei modelli più conosciuti è quello del *perceptron* [2], in cui l’uscita della rete è binaria, determinata da una funzione di attivazione a soglia. In altri termini, l’uscita della rete corrisponde al valore 1 se la somma pesata degli ingressi è maggiore o uguale alla soglia, altrimenti l’uscita corrisponde a 0. Questa particolare funzione di attivazione rende il perceptron adatto a svolgere solamente i cosiddetti compiti di classificazione lineare, come determinare se degli oggetti si trovino a sinistra o a destra rispetto ad una linea tracciata sul terreno. Ad esempio, se l’uscita corrisponde ad 1, la rete classifica l’oggetto a destra della linea, se corrisponde a 0 l’oggetto è a sinistra. Il semplice modello del perceptron è stato migliorato nel corso degli anni, connettendo i neuroni in un’architettura che utilizza più livelli (o strati). E’ solitamente sufficiente utilizzare anche un solo strato nascosto per poter sfruttare una rete neurale artificiale anche per effettuare classificazioni non lineari, come ad esempio riconoscere una forma complessa all’interno di una fotografia. Tali compiti, pur essendo banali da eseguire per un essere umano, sono particolare difficili da progettare e risolvere con i classici algoritmi, costituiti da una serie predefinita di passi che si devono compiere per poter giungere alla risoluzione del problema. Le reti neurali artificiali, invece, sono caratterizzate dalla possibilità di apprendere e devono essere addestrate. Una rete neurale impara da successi ed insuccessi, che sono legati a come la rete risponde agli stimoli. Imparando, la rete si addestra appunto a svolgere al meglio il proprio compito.

Le reti neurali vengono oramai realizzate tramite software. La tecnologia ci

permette, allo stato dell'arte, di riprodurre solamente su bassa scala la struttura del cervello umano. Secondo una stima che considera in media un milione di miliardi di sinapsi all'interno del cervello umano – ciascuna delle quali richiede almeno 4 byte per essere rappresentata nella memoria di un computer – e considerando che negli ultimi 20 anni la capacità delle memorie volatili (RAM) è aumentata di un fattore 10 ogni 4 anni, si dovrà aspettare almeno l'anno 2029 per avere una tecnologia che permetta di riprodurre la struttura del cervello umano con lo stesso ordine di grandezza [3].

Ci si può aspettare che le macchine, una volta acquisito un cervello artificiale paragonabile al nostro, possano diventare consapevoli, coscienti, in grado di identificare sé stesse come entità pensanti. Tuttavia, l'equazione tra intelligenza e consapevolezza non è proprio immediata. Possiamo affermare con molta probabilità che le macchine saranno intelligenti come noi – dato che per certi aspetti lo sono già – ma esse potranno mai essere consapevoli? Esisterà mai una coscienza artificiale?

3 Coscienza artificiale

Uno dei principali problemi da considerare quando si parla di coscienza, intesa come consapevolezza, è che abbiamo (e possiamo avere) esperienza diretta solo della nostra coscienza, non di quella degli altri. Siamo naturalmente portati ad attribuirle agli altri esseri umani per similitudine.

Ma cosa è dunque la coscienza? Seguendo la tradizione del dualismo cartesiano [4], la coscienza può essere considerata come un'entità immateriale, che non dipende né è in alcun modo legata all'attività cerebrale. Tale teoria, che concretamente identifica la coscienza con l'anima, trova riscontro anche da un punto di vista religioso, poiché si crede che l'anima venga creata dall'afflato divino e sia separata dal corpo. Tuttavia, il dualismo cartesiano non riesce a spiegare alcuni fenomeni neuronali osservati sperimentalmente, come ad esempio il fatto che un cervello stimolato con degli elettrodi possa generare emozioni e pensieri di ordine superiore che solitamente associamo alle nostre virtù [5]. Per questo motivo, ultimamente presso la comunità scientifica sta ricevendo credito la cosiddetta teoria olistica, secondo cui la coscienza è un prodotto cerebrale che emerge solamente considerando il cervello nella sua globalità.

Proprio per evitare di cadere in gineprai filosofico-teologici, il matematico Alan Turing propose nella metà del Novecento un semplice test che potesse permettere di stabilire se un computer fosse in grado di pensare o meno. Tale test, che lo stesso Turing ha denominato come “il gioco dell'imitazione” [6], ci consente di affermare che una macchina pensa quando è in grado di chattare con un umano su un qualsiasi argomento, senza che l'umano si accorga che il suo interlocutore è per l'appunto una macchina. Turing prevedeva che nell'arco di mezzo secolo ci sarebbero stati computer in grado di superare sistematicamente tale test ma le sue previsioni si sono rivelate oltremodo rosee. Infatti, soprattutto per quanto concerne le conversazioni su temi generali, i computer odierni dimostrano ancora scarsa capacità espressiva e linguistica.

Nonostante le difficoltà che i computer tutt'ora incontrano nel superare il test di Turing, c'è chi comunque ne critica radicalmente la fondatezza. La critica più celebre al test di Turing è opera del filosofo americano John Searle ed è nota come "l'obiezione della stanza cinese" [7]. Searle immagina di chiudere in una stanza un uomo che non conosce il cinese, rendendolo invisibile dall'esterno. L'uomo all'interno della stanza è dotato di un certo numero di libroni, che costituiscono un manuale tramite il quale egli può associare a determinati ideogrammi forniti in ingresso degli altri ideogrammi da fornire in uscita. Dall'esterno vengono dunque inseriti dei fogli con delle domande scritte in cinese e l'uomo nella stanza, dopo aver consultato il manuale, fornisce delle risposte sempre in cinese. Se il manuale è stato redatto in maniera particolarmente accurata, dall'esterno della stanza si riterrà che chi è dentro conosca il cinese, cosa che in realtà è non vera. Chi è dentro ha solo manipolato dei simboli, ma non ne conosceva il significato intrinseco. Inutile dire che l'obiezione della stanza cinese ha innescato un aspro dibattito, cui nemmeno lo stesso Searle si è sottratto, intervenendo più volte nel corso degli anni per difendere la propria posizione filosofica. Tra le innumerevoli critiche fatte all'obiezione della stanza cinese, citiamo come esempio significativo la contro-obiezione del matematico Martin Davis, il quale, nella sua opera *Il calcolatore universale* [8], osserva che Searle effettua un'arbitraria separazione tra hardware (l'uomo nella stanza) e software (il manuale che l'uomo utilizza).

4 Percorso didattico

Il percorso proposto nelle precedenti pagine si muove tra informatica e filosofia, dunque un suo innesto nella formazione scolastica dovrebbe trovare naturale *humus* in un Liceo Scientifico opzione Scienze Applicate, ordinamento in cui queste materie fanno già parte del curriculum scolastico. Tuttavia, lo si potrebbe implementare anche in altri istituti, in tutto o in parte, come percorso extracurricolare o avvalendosi dei docenti di Italiano per la parte storico-filosofica. In particolare, ad avviso dell'autore, sarebbe un percorso particolarmente interessante da proporre in un Istituto Tecnico Tecnologico ad indirizzo Informatico, ordinamento in cui il *focus* è spesso unicamente sulla tecnologia, senza alcun cenno alla dimensione epistemologica. Considerando un progetto che si articoli su 30 ore, si può riassumere il percorso proposto secondo lo schema della Tabella 1.

Ovviamente i contenuti, così come i prerequisiti e gli obiettivi didattici potrebbero differire notevolmente nei due ordinamenti citati, ovvero Liceo Scientifico e Istituto Tecnico. Per brevità di trattazione, un approfondimento risulta particolarmente difficile in queste pagine e viene riservato ad un futuro lavoro. A titolo esemplificativo accenniamo al fatto che, dal punto di vista informatico, si può identificare il modello di una rete neurale con quello di un grafo orientato, in cui i pesi sugli archi rappresentano la capacità di eccitazione di un collegamento. In particolare, le reti neurali stratificate, come quelle cui si fa riferimento in questo articolo, possono essere rappresentate da un grafo diretto aciclico (DAG [9]). In un Istituto Tecnico, si può proporre agli studenti l'implementazione dei grafi alla fine del quarto anno o all'inizio del quinto, dopo aver trattato la program-

| Contenuti | Ore |
|--|-----|
| Teoria dei grafi | 5 |
| Reti neurali | 5 |
| Applicazioni di classificazione/riconoscimento | 5 |
| La coscienza: il dualismo cartesiano | 5 |
| Turing e il gioco dell'imitazione | 5 |
| Searle e l'obiezione della stanza cinese | 5 |

Tabella 1. Progetto didattico per macro-aree

mazione ad oggetti, utilizzando classiche strutture dati come le liste. In un Liceo Scientifico, invece, si potrebbero avere studenti con un solido background matematico ma con poca dimestichezza nella programmazione ad oggetti e dunque potrebbe essere più opportuno approcciarsi alle reti neurali utilizzando ambienti di calcolo scientifico come Matlab e Scilab, che offrono strumenti specifici come ad esempio il Neural Network Toolbox [10].

5 Conclusioni

Introdurre il tema dell'intelligenza artificiale e della coscienza artificiale nelle nostre scuole è necessario affinché le nuove generazioni possano cominciare a riflettere sugli scenari possibili che il futuro ci prospetta. Nell'articolo, tramite una trattazione teorica, si sono proposte delle macro-aree che costituiscono un possibile percorso didattico interdisciplinare sui computer intelligenti.

Riferimenti bibliografici

1. H. Fancher and D. Peoples, "Blade runner, movie script." http://www.scifiscripts.com/scripts/bladerunner_final.txt, 1981. Visto il 01/03/2018.
2. F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," pp. 386–408, 1958. *Psychological Review*, Vol 65(6).
3. G. Buttazzo, "Coscienza artificiale: Missione impossibile?," *Mondo Digitale*, vol. 1, 2002.
4. R. Descartes, *Meditationes De Prima Philosophia*. 1641.
5. J. Parvizi, V. Rangarajan, W. R. Shirer, N. Desai, and M. D. Greicius, "The will to persevere induced by electrical stimulation of the human cingulate gyrus," *Neuron*, vol. 80, Issue 6, 2013. Pagg 1359 – 1367.
6. A. Turing, "Computing machinery and intelligence," 1950. Pubblicato in *Mind Design II*, a cura di Haugeland J., MIT Press, 1997.
7. J. Searle, "La mente è un programma?," *Le Scienze*, 1990. n. 259, marzo.
8. M. Davis, *Il Calcolatore Universale. Da Leibiniz a Turing*. Adelphi editore, 2012.
9. K. Thulasiraman and M. N. Swamy, *Directed Graphs*, ch. 5, pp. 97–125. Wiley-Blackwell, 2011.
10. H. Demuth, M. Beale, and M. Hagan, "Neural network toolbox™ 6," *User's guide*, vol. 10, p. 11, 2008.