La demenza dell'Intelligenza Artificiale – Dall'Analisi dei Dati all'IA qualche elemento di riflessione

written by Renato Miceli | 29 Gennaio 2024

1. Introduzione

Possono esservi efficaci applicazioni informatiche che si avvalgono delle tecnologie cosiddette di Intelligenza Artificiale; mi pare tuttavia, vi sia anche della demenza nel loro utilizzo pervasivo.

In questo scritto mi occupo di Intelligenza Artificiale (IA) e Machine Learning (MaL) ambiti diversi da quello dell'Analisi dei Dati in Psicologia e Sociologia cui ho dedicato gran parte dei miei studi e della mia vita. Fra i diversi settori esistono, tuttavia, intense relazioni, constatato che molti degli algoritmi impiegati nelle nuove tecnologie informatiche utilizzano (spesso con forzature) metodi, modelli e tecniche tradizionali dell'Analisi dei Dati.

Agli esperti di IA chiedo comprensione se qualche mia argomentazione può risultare troppo superficiale o grossolana; lo scopo di queste pagine è prevalentemente didattico, volto ad evidenziare qualche punto di connessione e su questo riflettere. Per raggiungere lo scopo, ho realizzato (a fini didattici, appunto) un'applicazione che emula un Classificatore Bayesiano Naif (CBN)[1]; un CBN in miniatura che, mantenendo integro l'algoritmo bayesiano su cui si fonda, permette però, tramite l'utilizzo di pochi dati di fantasia e di un contesto originale, di assolvere a due propositi. Da un lato illustrare, negli opportuni ambienti di apprendimento, i legami con l'Analisi dei Dati (frequenze, tabelle di

contingenza, probabilità, modello logit, etc.); dall'altro, per un pubblico più ampio, conoscere – almeno per sommi capi – come può essere realizzato un CBN, consentendo di riflettere sui pregi, i limiti e i pericoli insiti nell'utilizzo di tecnologie di questo tipo.

Le pagine che seguono sono dedicate a questo secondo obiettivo.

2. Uno sguardo su Machine Learning e Intelligenza Artificiale

Sempre più sovente ci viene detto che, anche inconsapevolmente, stiamo interagendo con strumenti dell'IA e del MaL. <<Forse non lo sapete, ma il machine learning vi ha circondato. […] Ogni volta che usate un computer, ci sono buone probabilità che da qualche parte entri in gioco il machine learning>> (P. Domingos, 2016, p. 11)[2]. Il MaL è un termine generico per indicare una branca dell'IA che sviluppa algoritmi per risolvere problemi, generalmente di classificazione, con metodi in prevalenza di tipo matematico-statistico. Così, non è sorprendente che la comunicazione di massa usi la più generica e allusiva espressione di IA per farci sentire non solo a nostro agio, ma anche orgogliosi di poter utilizzare queste nuove tecnologie.

Stupisce, invece, la scarsità di informazioni comunemente accessibili e non specialistiche che illustrino il modo di operare di questi strumenti informatici. Accanto alla celebrazione dei loro pregi, è possibile sapere se e quanto sbagliano? Quali limiti di utilizzo è ragionevole attendersi? Presentano qualche difetto conosciuto? Viene comunemente detto che l'utilizzo diffuso di queste nuove tecnologie modifica, e modificherà ancor di più in futuro, l'ambiente in cui viviamo e le nostre vite. Di quali modifiche si tratta? Siamo sicuri che siano rivolte al meglio? Nell'uso pervasivo di queste tecnologie si possono annidare pericoli di tipo sociale? E, nel caso, quali?

Qualcuna di queste domande, le relative risposte e le questioni su cui vertono sono sicuramente e ampiamente discusse nella ristretta cerchia ingegneristico-informatica degli addetti ai lavori, ma – mi pare – non ne giunga eco all'esterno. Eppure l'arco delle applicazioni è già molto vasto e destinato ad estendersi ancor più in futuro; solo per fornire una breve e certamente non esaustiva panoramica, prodotti dell'IA sono coinvolti:

– nei motori di ricerca sul web, per decidere quali risultati (e anche quale pubblicità) mostrare a ciascun utente;

– nei filtri anti-spam delle mail;

– nell'uso del traduttore automatico di Google;

– nei siti di Amazon, Netfix, Facebook, Twitter (ora X), etc. per consigliare cosa può piacerci o quali aggiornamenti mostrare;

– nell'interpretazione dei comandi vocali e nei correttori degli errori di battitura;

– nell'identificazione facciale;

– nella diagnosi di malattie;

– nel determinare l'affidabilità creditizia;

– nell'assistere i giudici nelle loro decisioni (almeno negli USA);

– etc. etc.

Come peraltro avviene in tutti gli ambiti, gli specialisti del campo si preoccupano di far funzionare al meglio il loro prodotto per risolvere il problema che è stato loro posto (quale algoritmo o quale combinazione di algoritmi è più idonea, quali e quanti dati sono necessari per istruire l'algoritmo, quanto rapidamente risponde alla chiamata, e così

via). Ragionevolmente, non è compito loro informare il largo pubblico sul funzionamento del prodotto; noi fruitori, spesso inconsapevoli, non siamo nemmeno i committenti di quella tecnologia. Questi ultimi sono aziende interessate a migliorare i loro servizi e a massimizzare il loro profitto; a loro interessa che il prodotto funzioni per i loro fini; ad esse compete la diffusione e la pubblicità del loro operato, non può certo essere demandato neppure a loro il compito di informare adeguatamente il pubblico.

Se a tutto questo si aggiunge anche l'alone mitico, di mistero e di inaccessibilità che accompagna la propaganda intorno all'IA, sembra proprio che a noi, comuni mortali, non resti altro da fare che cliccare.

Così, però, sul tema dell'IA si rischia di alimentare false credenze e il formarsi di faziosità spesso dominate dall'ignoranza: contro chi crede al mito "tecno-digitale" delle macchine, che fanno e faranno tutto e meglio degli umani, si schiera una professione di fede opposta, pregiudicando lo sviluppo di una visione critica che necessita comprendere, almeno approssimativamente, come funziona questa tecnologia.

D'altra parte, è proprio la rilevanza degli scenari aperti dalla diffusione dell'IA che impone lo sviluppo di momenti informativi accessibili a tutti e capaci di alimentare un sano ed equilibrato dibattito pubblico di cui si sente urgente bisogno, ma di cui – almeno qui in Italia – non v'è traccia.

Attingendo ancora dal già citato testo di P. Domingos, è possibile trarre qualche elemento di conoscenza capace, almeno un po', di dissipare la nebbia che avvolge l'argomento. Facendo riferimento al MaL, due sono gli aspetti sui quali insistono i vari algoritmi: apprendere e prevedere; come fare affinché l'algoritmo apprenda, reagendo adeguatamente all'acquisizione di nuovi dati e come utilizzare le conoscenze acquisite per prevedere cosa accadrà nel futuro.

Pur nella varietà delle risposte che possono essere date a questi quesiti, è confortante sapere che <<le centinaia di nuovi algoritmi di apprendimento inventati ogni anno si basano su un numero ristretto di idee fondamentali>> (ibid., p. 11) che, a loro volta, permettono di articolare il campo in cinque scuole di pensiero che dispongono di un proprio algoritmo di riferimento e che si ispirano a diversi ambiti del sapere. Sinteticamente (ibid., p. 18) le cinque scuole di pensiero sono:

1) i simbolisti che si ispirano alla filosofia, alla psicologia e alla logica, per i quali l'apprendimento è l'inverso della deduzione;

2) i connessionisti che prendono spunto dalle neuroscienze e dalla fisica, che considerano l'apprendimento un'operazione di reverse engineering;

3) gli evoluzionisti che fanno riferimento alla genetica e alla biologia evolutiva, per i quali l'apprendimento si realizza tramite simulazioni numeriche dell'evoluzione;

4) i bayesiani che poggiano sulla statistica e vedono l'apprendimento come una forma di inferenza probabilistica;

5) gli analogisti vicini alla psicologia e all'ottimizzazione matematica, che ritengono l'apprendimento un processo fondato su estrapolazioni basate su criteri di somiglianza.

Cercando di riassumere le principali caratteristiche comuni a tutti questi algoritmi, si può dire che:

(a) producono prevalentemente classificazioni;

(b) possono essere molto semplici, realizzabili – dal punto di vista informatico – con poche righe di codice;

(c) si nutrono di grandi quantità di dati, i cosiddetti "BigData";

(d) modificano il loro comportamento in funzione dei dati usati in addestramento;

(e) rispondono anche in presenza di situazioni non presenti nei dati di addestramento;

(f) spesso sono secretati, brevettati da aziende private o sottoposti a “proprietà intellettuale” o a “segreto industriale”.

Come detto, fra le diverse scuole vi è quella bayesiana, che produce algoritmi basati sul teorema di Bayes ed è a questa che farò riferimento nel seguito.

3. Il Classificatore Bayesiano Naif

Il teorema di Bayes viene spesso presentato come uno strumento utile per passare “agevolmente” dalle probabilità degli effetti, alle probabilità delle cause. Più in generale e con riferimento al linguaggio in uso con le tabelle di contingenza, il teorema consente di ottenere le proporzioni o probabilità condizionali di riga (data la colonna), se si dispone delle probabilità condizionali di colonna e viceversa. Nella terminologia Bayesiana, assumendo di conoscere la probabilità di riga, questa viene detta “a priori”, mentre la probabilità condizionale di riga data la colonna si chiama “verosimiglianza”; la probabilità di colonna prende il nome di “evidenza” e la probabilità condizionale di colonna data la riga, che si può calcolare grazie al teorema, è detta “a posteriori”. Così il teorema di Bayes può essere espresso in modo conciso da questa semplice equazione:

Per esempio, possiamo immaginare di sapere che fra tutti i pazienti la probabilità di avere la febbre è 0.3 (a priori) e che la probabilità di avere l’influenza è 0.4 (evidenza); conoscendo anche che, fra i pazienti con la febbre, la probabilità condizionale di essere affetti dall’influenza è

0.6 (verosimiglianza), è possibile ottenere la probabilità condizionale (a posteriori) di essere affetti dall'influenza dato che si ha la febbre:

Chiaramente, nell'esempio, l'influenza è una delle possibili cause dell'effetto o sintomo: avere la febbre; d'altra parte è altrettanto evidente che, per poter stabilire di avere l'influenza, vorremmo e dovremmo disporre della presenza o assenza di altri sintomi come brividi, mal di gola, etc.

Così facendo, però, la questione diventa più complicata dato che raramente, o mai, i vari effetti – proprio come i sintomi dell'esempio – sono fra loro (probabilisticamente) indipendenti.

Generalizzando l'utilizzo del teorema a situazioni in cui è presente più di un effetto, si assiste ad una esplosione combinatoria[3] che rende tecnicamente impossibile considerare (nella formula del teorema) le probabilità condizionali (degli effetti data ciascuna causa), anche quando il numero degli effetti è limitato (per esempio: considerando solo 40 effetti di tipo booleano, le combinazioni da gestire sono già un numero di 12 cifre: $1.0995 \cdot 10^{12}$).

Lo stratagemma adottato per superare l'ostacolo consiste nel compiere un'assunzione (piuttosto ardita): si assume che tutti gli effetti considerati siano mutuamente indipendenti, data la causa. Da qui deriva l'attributo "naif" conferito all'algoritmo: <<… si tratta proprio di un'assunzione ingenua. […] Il machine learning, però, è l'arte del fare assunzioni false e cavarsela lo stesso>> (P. Domingos, 2016, p. 182).

Sotto questa assunzione e qualche ulteriore accorgimento tecnico, la realizzazione pratica di un algoritmo basato sul teorema di Bayes diventa gestibile. Accompagnando poi l'output dell'algoritmo (che è un vettore contenente valori di

probabilità) ad una opportuna regola di decisione (che può essere anche molto semplice, come l'estrazione del valore massimo), si ottiene un Classificatore Bayesiano Naif (Naïve Bayes Classifier).

Forse anche per la sua semplicità (una sola equazione), l'algoritmo è ampiamente utilizzato nei prodotti informatici che sfruttano le tecniche dell'IA. L'algoritmo, senza paternità, circola dagli anni '70 del secolo scorso, ma <<cominciò ad affermarsi seriamente negli anni novanta, quando i ricercatori si accorsero con stupore che spesso era più accurato di learner molto più sofisticati […]. Oggi Naïve Bayes è molto diffuso. Ad esempio, è alla base di molti filtri antispam […]. Anche i motori di ricerca più semplici utilizzano un algoritmo molto simile a Naïve Bayes per decidere che pagine web mostrarvi […]. La lista dei problemi di previsione cui si è applicato Naïve Bayes è praticamente infinita>> (P. Domingos, 2016, p.183). <<Partendo da un database di cartelle cliniche contenenti i sintomi dei pazienti, i risultati delle analisi e la presenza di patologie pregresse, Naïve Bayes può imparare a diagnosticare una malattia in una frazione di secondo, con un'accuratezza che spesso supera quella di medici con anni di studio alle spalle>> (P. Domingos, 2016, p. 47).

3.1 Un'applicazione

Questo stesso algoritmo è alla base del funzionamento di un'applicazione, realizzata a scopo didattico, che consente di "giocare" con il CBN per monitorarne il funzionamento e per ragionare su quanto gli ruota attorno.

L'applicazione volutamente fantasiosa e bizzarra per quanto attiene al contesto che però riguarda l'esperienza di ogni studente, è stata pensata cercando di ridurre al minimo indispensabile i dati da utilizzare, sia per la fase di apprendimento, sia per l'interrogazione (anche a scapito della ragionevolezza).

Lo scopo ipotizzato di questa applicazione consiste nel disporre di uno strumento che, come una magica sfera di cristallo, consenta ad un ipotetico studente di prevedere l'esito di un esame universitario (in una qualche disciplina), sfruttando esclusivamente dati comportamentali o di atteggiamento, senza avvalersi di alcuna informazione nel merito della disciplina stessa.

Così, l'applicazione fornisce il suo responso, stimando la probabilità fra due esiti (Classi) alternativi: "Promosso" / "Respinto", indicati rispettivamente come "S" e "N", utilizzando le informazioni desumibili da quattro Indicatori (Caratteristiche o Features) relativi a ciascun studente; due Indicatori sono di tipo categoriale con tre Categorie ciascuno, mentre gli altri due sono booleani; l'insieme degli Indicatori costituisce così un Profilo (Pattern o Configurazione) relativo ad ogni studente. Il significato degli Indicatori e delle relative categorie è riportato in Figura1 (fra parentesi quadre è indicato il codice attribuito a ciascun Indicatore e a ogni Categoria).

Figura1

Si può ipotizzare che ogni studente risponda a quattro domande del tipo: "Lei ha seguito le lezioni in maniera: Continuativa / Saltuaria / Mai ?"; etc.

Senza che i professori possano accedere a queste informazioni, viene poi svolto l'esame tradizionale, così da poter confrontare l'esito empirico (reale) con quello pronosticato dal CBN. Oppure, in maniera "avveniristica", si può pensare che, svolgendo lo studente tutte le sue attività sul web, altri classificatori siano incaricati di definire la categoria che gli compete, per ogni Indicatore; un opportuno sistema di IA potrebbe allora sostituire i professori (divenuti così obsoleti e inutilmente costosi), fornendo il suo intelligente

responso (forse e ahimè, vi è più di qualcuno che auspica questa seconda ipotesi).

Comunque sia, il nostro Classificatore necessita di dati per essere istruito. Grandi quantità di dati (BigData) sono certamente utili e indispensabili per addestrare un CBN a svolgere compiti importanti come la diagnostica di malattie o il riconoscimento di volti; l'addestramento con grandi quantità di dati richiede, tuttavia, che questi siano disponibili, oltre a tempi di elaborazione molto lunghi e computer dedicati. Su questo aspetto si ritornerà nel seguito, intanto però, è bene precisare che non è questa la strada utile per la nostra applicazione.

Per la fase di apprendimento iniziale del nostro CBN ci si è avvalsi di pochi dati (20 osservazioni) frutto di fantasia. Più precisamente, i dati sono stati generati tramite simulazione, facendo in modo però che alcune relazioni (fra Classe e Indicatore) fossero più forti di altre; così da rendere la matrice dei dati in input aderente allo stereotipo che vuole una maggiore probabilità di promozione per chi, convenzionalmente, fa le cose per bene (segue le lezioni, svolge gli esercizi, studia intensamente, etc.): una matrice dati, insomma, benpensante.

La matrice dati utilizzata per la fase di apprendimento iniziale è riportata in Figura2.

Figura2

Sulla base di queste conoscenze acquisite il CBN fornisce prestazioni discretamente buone con un tasso di Classificazioni corrette pari a 0.8 (80%) e una percentuale di falsi positivi (previsti come promossi e invece respinti all'esame) pari a circa il 22% (2 falsi positivi, su 9 empiricamente respinti).

Stante la struttura definita di questo CBN (2 Classi, 2 Indicatori con 3 Categorie e 2 Indicatori con 2 Categorie) le Configurazioni teoriche di cui può disporre sono 72 ($3^2 \cdot 2^2 = 36$ per 2 Classi); senza considerare le Classi, i Profili disponibili sono 36, e i dati utilizzati ne coprono 11 (30.56%), quindi vi sono 25 Profili totalmente sconosciuti al CBN.

Provando ad interrogare il CBN, fornendogli qualcuno di questi Profili (sconosciuti ai dati di apprendimento), si ottengono comunque delle risposte sensate; per esempio: uno studente con il Profilo “1010” (frequenza alle lezioni continuativa; svolgimento esercizi mai; studio profondo; interesse basso) ottiene la previsione di essere promosso con una probabilità di 0.7056; un altro studente con il Profilo “0200” (frequenza alle lezioni mai; svolgimento esercizi raro; studio superficiale; interesse basso) ottiene la previsione di essere respinto con una probabilità di 0.8274.

Al di là quindi dei pochi dati conosciuti, questo Classificatore sembra rappresentare adeguatamente il “mondo” da cui quei dati provengono; un mondo in cui per superare l’esame serve soprattutto studiare in maniera approfondita. Chiamiamo questo mondo “A” e indichiamo come: “A1”, “A2” e “A3” tre studenti che presentano altrettanti Profili; interrogando il CBN otteniamo i responsi riportati in Figura3.

Figura3

Studente	Profilo	Responso (Prob.)
A1	1000	RESPINTO (0.7188)
A2	1001	RESPINTO (0.8099)
A3	0001	RESPINTO (0.9697)

Accompagnati dalla loro triste profezia, i tre studenti si avviano ad affrontare l'esame e, nell'incertezza che accompagna sempre la vita, otterranno il loro esito empirico che potrà risultare: coerente ("accidenti, il CBN aveva visto giusto") o incoerente ("tiè, mi è andata bene lo stesso").

Ma il mondo non è solo incerto, è anche vario e instabile: senza disturbare Eraclito, sappiamo che le cose cambiano.

Possiamo immaginare che in un luogo diverso da quello di mondo "A" (per condizione spaziale e/o temporale), diciamo in mondo "B", vengano raccolti nuovi dati che possono alimentare e accrescere la base di conoscenze del CBN. Si tratta di ben 100 osservazioni (sempre di fantasia), provenienti da mondo "B", in cui domina uno stereotipo diverso da quello di mondo "A"; qui si assiste ad una maggiore probabilità di promozione per chi ha (o manifesta) un atteggiamento di alto interesse per la disciplina. Anche da una superficiale analisi è possibile ricavare alcuni indizi che mostrano la contraddittorietà dei nuovi dati rispetto a quelli iniziali, ma è proprio dinnanzi ad una tale situazione che si evidenzia ciò che qui più interessa.

Nella circostanza ipotizzata, nuovi dati disponibili, una qualche "intelligenza" umana o artificiale dovrà necessariamente decidere il da farsi: includere nella base di conoscenze del CBN i nuovi dati SÌ o NO?

Se si opta per il NO (NON recepire i nuovi dati) il CBN resterà ancorato alla sua conoscenza del mondo che riflette solo quella di mondo "A". Se, viceversa, si decide per il SÌ (recepire i nuovi dati), facendo in modo che il CBN aggiorni la sua conoscenza, considerando anche quanto accade in mondo "B", possiamo interrogare nuovamente il CBN, utilizzando gli stessi Profili dei tre studenti visti in precedenza ("A1", "A2", "A3") e confrontare i responsi già visti con quelli che ora ottengono tre diversi studenti ("B1", "B2", "B3"). Si può osservare così che il CBN è effettivamente in grado di

apprendere dai nuovi dati; l'esito del confronto è riportato in Figura4.

Figura4

		Solo mondo "A"	…		mondo "A" con mondo "B"
Profilo	**Studente**	**Responso (Prob.)**	**…**	**Studente**	**Responso (Prob.)**
1000	**A1**	**RESPINTO (0.7188)**	**…**	**B1**	**PROMOSSO (0.6492)**
1001	**A2**	**RESPINTO (0.8099)**	**…**	**B2**	**PROMOSSO (0.7215)**
0001	**A3**	**RESPINTO (0.9697)**	**…**	**B3**	**PROMOSSO (0.8981)**

Come in precedenza, i tre nuovi studenti ("B1", "B2", "B3") otterranno l'esito empirico (in un contesto incerto) che potrà risultare: coerente con quanto previsto dal CBN ("bene, il CBN aveva visto giusto") o incoerente ("perdinci, è andata male").

Il CBN che ha assimilato, nella sua base di conoscenza, i nuovi dati provenienti da mondo "B" fornisce ora prestazioni un poco inferiori a quelle precedenti (tasso di Classificazioni corrette pari a 0.7) e una percentuale di falsi positivi (previsti come promossi e invece respinti all'esame) pari a poco meno del 27% (17 falsi positivi, su 64 empiricamente respinti).

Ora però, più di uno studente su quattro prova una forte delusione dovuta all'incoerenza fra il pronostico del CBN (favorevole) e l'esito empirico (infausto); ci si può chiedere pertanto cosa potrebbe accadere nel mondo reale.

Considerando la retorica che ruota intorno ai temi dell'IA, dei BigData e degli algoritmi intelligenti, pare realistico ipotizzare che la richiesta sarebbe quella di adeguare la realtà (il contesto empirico) ai responsi dell'algoritmo.

Professori troppo severi, affetti da favoritismo e parzialità, istituzioni inadeguate, etc. potrebbero essere argomentazioni utilizzate a tale scopo. Soprattutto se si pensa alla potenza mediatica che può esercitare chi ha investito ingenti risorse nella realizzazione di quel prodotto hi-tech, si può ipotizzare che non indietreggerebbe facilmente; la propaganda sarebbe diffusa e continua alimentando fazioni contrapposte, stimolate – in modo demenziale – a schierarsi pro o contro il progresso, la tecnologia o la scienza(!).

D'altra parte, lo stesso scenario potrebbe presentarsi anche qualora si optasse (come è stato detto in precedenza) per la scelta di "NON recepire i nuovi dati", lasciando il CBN ancorato a quanto sapeva dal solo mondo "A". Quando la conoscenza dell'applicazione verteva sulle prime 20 osservazioni, la percentuale di falsi positivi (previsti come promossi e invece respinti all'esame) era inferiore a quella successiva (circa uno su cinque); anche in quel caso, però, si sarebbe andati incontro alla delusione di alcuni studenti e la richiesta di adeguamento della realtà all'algoritmo potrebbe essere avanzata in modo altrettanto pressante. Considerato che allora il CBN operava in modo più aderente ad uno stereotipo sensato (maggiore probabilità di promozione per chi studia in modo approfondito), anche la richiesta potrebbe apparire ragionevole, ma il problema, la sua gravità, così come la follia innescata dalla semplice presenza dello strumento, restano inalterati.

Nelle applicazioni dell'IA, chi governa il processo inerente i dati, più che la realizzazione degli algoritmi, ha enormi responsabilità sociali, può manipolare – come mai prima d'ora – l'ideologia degli individui. Disgraziatamente l'assenza di informazioni su come vengono gestiti i dati sembra essere la norma, ancora più stringente della segretezza imposta su alcuni algoritmi.

4. Qualche ulteriore considerazione (BigData e dintorni)

Si potrebbe pensare che la situazione descritta in precedenza e artatamente generata con pochi dati di fantasia sia dovuta alla scelta (qui obbligata) di non fare ricorso ai cosiddetti BigData. D'altra parte, come già accennato, grandi quantità di dati sono necessari in alcuni casi e il loro utilizzo esige strumenti dedicati allo scopo, comunemente accessibili solo a chi dispone di adeguate risorse.

L'algoritmo qui utilizzato, come molti altri, può richiedere l'utilizzo di grandi quantità di dati in fase di addestramento; ma, raggiunta una prestazione adeguata del sistema, i dati effettivamente utilizzati si riducono ad una manciata (le probabilità condizionali e poco più) e questi valori (anche se centinaia o migliaia) possono essere agevolmente incorporati in un prodotto commerciale, il cui contenuto può essere secretato e sottoposto a Copy Right.

Nell'applicazione descritta in precedenza, se fossero state utilizzate grandi quantità di dati, la scelta relativa a includere o meno i dati relativi a mondo "B" nella base di conoscenze del CBN sarebbe stata comunque da compiere; forse relegata nella sola fase iniziale di addestramento, ma comunque presente.

Il ricorso ai BigData non esime dal considerare il portato informativo dei dati utilizzati che, anche se "big", sono sempre soltanto dati parziali e generati in condizioni specifiche.

Non è realistico ipotizzare l'utilizzo di dati (in quantità grande o piccola) che sia scevro da difficoltà e scelte che inevitabilmente condizionano il loro significato: la fonte o le fonti, i tempi e luoghi di acquisizione, le modalità che portano un fatto ad essere trasformato in un dato, le finalità per cui il dato viene generato e, come se non bastasse, le procedure messe in opera per collegare (link) fra loro insiemi di dati provenienti da situazioni diverse, la gestione dei dati mancanti, e così via.

Tutto ciò non può essere soddisfatto evocando esclusivamente i BigData, come se la quantità potesse sopperire la qualità. I dati, a dispetto del loro nome, non sono semplicemente dati (participio passato del verbo dare). Senza considerare ciò, si corre il rischio di replicare la vecchia stupidaggine di chi asserisce che "i dati parlano da soli".

Gran parte dei cosiddetti BigData proviene dal web e viene generata dai click di milioni o miliardi di utenti, ma la realtà è assai più vasta della rete di Internet e in essa non si esaurisce; si tratta sempre e comunque di dati raccolti in quel contesto e, per giunta, come tutti i dati, riguardano ciò che è già accaduto, sono dati del passato, dai quali può essere demenziale pretendere di leggere il futuro.

Nell'Analisi dei Dati[4], dove si usano modelli anche articolati e complessi, ci si preoccupa spesso e giustamente dell'adattamento del modello ai dati (fit o data fitting) e/o dell'adeguatezza dei dati al modello usato; ma adattamento e capacità previsionale restano sempre due aspetti chiaramente distinti.

Può essere utile rammentare il celebre aforisma attribuito a Niels Bohr (ma anche ad altri) che recita: "è difficile fare previsioni, soprattutto sul futuro".

In un suo scritto, G. Gigerenzer[5] coglie adeguatamente questo aspetto, ricordando che nella scienza si sa che per predire accuratamente il futuro è necessario disporre di una buona teoria, di dati affidabili e di un contesto (relativamente) stabile. Così <<l'IA è magnifica nel gestire simili situazioni stabili, dal riconoscimento facciale per sbloccare il proprio smartphone al cercare la strada più veloce per raggiungere la propria destinazione, all'analisi di grandi quantitativi di dati nell'amministrazione. Ma le aziende hi-tech spesso predicono il comportamento umano senza una buona teoria, dati attendibili o un mondo stabile>> (G. Gigerenzer, 2023, pp. 55-56).

Un altro aspetto preoccupante dei BigData è il loro connubio con le tecniche di ricerca automatica di correlazioni (data mining). Le trappole della "correlazione spuria" e di quella "soppressa", solo per fare due esempi, sono sempre in agguato[6]. Come ricorda ancora Gigerenzer <<… gli entusiasti dei big data hanno spinto all'estremo le tesi di Pearson e Hume, sostenendo che le cause non sono neppure necessarie; piuttosto [costoro sostengono che], avendo a disposizione petabyte, "basta la correlazione">> (G. Gigerenzer, 2023, p. 153).

Gli aspetti che ruotano intorno all'uso dell'IA qui richiamati sono preoccupanti, ma molti altri che meriterebbero considerazione, sono altrettanto inquietanti e, a me pare, contrari all'intelligenza (almeno a quella umana).

In conclusione, però, meritano una menzione particolare le applicazioni in ambito militare sulle quali ha recentemente posto l'attenzione Giorgio Parisi (Nobel per la fisica 2021), nella Prefazione al libro di F. Farruggia (2023), di cui riporto alcuni stralci (le sottolineature sono mie): <<L'intelligenza artificiale apre nuove possibilità per le applicazioni militari, in particolare per quanto riguarda i sistemi d'arma con significativa autonomia nelle funzioni critiche di selezione e attacco dei bersagli. Tali armi autonome potrebbero portare a una nuova corsa agli armamenti […]. Alcune organizzazioni chiedono un divieto sulle armi autonome, simile alle convenzioni in materia di armi chimiche o biologiche. […] In assenza di un divieto di sistemi di armi autonome letali (Lethal Autonomous Weapons Systems, LAWS), dovrebbe essere garantita la conformità di qualsiasi sistema d'arma al Diritto Internazionale Umanitario. Queste armi dovrebbero essere integrate nelle strutture di comando e controllo esistenti in modo tale che la responsabilità e la responsabilità legale rimangano associate a specifici attori umani. C'è una chiara necessità di trasparenza e di discussione pubblica delle questioni sollevate in questo

settore>>.

Certo il pericolo militare è soverchiante, ma – mi permetto di aggiungere – pure quello sociale è grave. Così, mi pare opportuno richiedere che anche l'attivazione e lo spegnimento dei prodotti dell'IA, in generale, venga consapevolmente ricondotto agli esseri umani.

Come minimo, andrebbe imposto l'obbligo di segnalare, con un avvertimento, che qualche prodotto dell'IA sta per essere utilizzato (simile a quello apposto su prodotti d'uso ritenuti pericolosi, come il tabacco). Poi sarebbe opportuno disporre sempre di un pulsante (anzi un Click) di "arresto" immediato, capace di inibire l'uso dell'IA in quella specifica applicazione.

Assai più delle illusorie speranze di riservatezza (privacy), solitamente elargite a piene mani, questi accorgimenti aiuterebbero la riflessione e la consapevolezza nell'utilizzo dell'IA da parte delle persone.

Così, si potrebbe forse evitare il rischio che, un domani, qualcuno possa trovarsi nella scomoda posizione di quell'astronauta, unico superstite della strage perpetrata dall'intelligenza del computer, nella lunga e indimenticabile sequenza di disattivazione di HALL9000 (il computer di bordo), che Stanley Kubrick ci ha lasciato nel suo film "2001: Odissea nello spazio".

Riferimenti bibliografici

Cardano, M., Miceli, R. (a cura di) (1991). *Il linguaggio delle variabili. Strumenti per la ricerca sociale.* Torino: Rosenberg & Sellier.

Domingos P., 2016, *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, Torino, Bollati Boringhieri.

Farruggia F., (a cura di), 2023, *Dai droni alle armi autonome. Lasciare l'Apocalisse alle macchine?*, Milano, Franco Angeli.

Gigerenzer G., 2023, *Perché l'intelligenza umana batte ancora gli algoritmi*, Milano, Raffaello Cortina.

Miceli R. (2001), *Percorsi di ricerca e analisi dei dati,* Torino, Bollati Boringhieri

Ricolfi L., (1993), *Tre variabili. Un'introduzione all'analisi multivariata*, Milano, Angeli Editore.

Ricolfi L. (a cura di), 1997, *La ricerca qualitativa*, Roma, La Nuova Italia Scientifica.

[1]L'applicazione informatica, realizzata artigianalmente con il linguaggio Java, consente di generare e gestire dei CBN che supportano fino a 10 Classi, 25 Indicatori Booleani o Categoriali (max 4 Categorie); chi, per ragioni didattiche, fosse interessato ad utilizzare questo software può richiederlo all'autore.

[2]Pedro Domingos è considerato uno degli scienziati di punta del machine learning a livello internazionale.

[3]Per un'argomentazione più rigorosa su questo aspetto si veda: Wikipedia, https://w.wiki/87oX

[4]Per approfondimenti, almeno nel campo delle Scienze Psicologiche e Sociali, si rimanda a: L. Ricolfi, 1997 e R. Miceli, 2001.

[5]Gerd Gigerenzer, scienziato cognitivo, lavora all'Istituto Max Planck di Berlino.

[6]Per approfondimenti si rimanda a: M. Cardano, R. Miceli, 1991 e L. Ricolfi, 1993.