

Direttore

Silvano TAGLIAGAMBE
Università degli Studi di Sassari

Comitato scientifico

Jesús Timoteo ÁLVAREZ
Universidad Complutense de Madrid

Dario ANTISERI
Libera Università Internazionale degli Studi Sociali “Guido Carli” (LUISS) di Roma

Gilberto CORBELLINI
Sapienza – Università di Roma

Roberto GIUNTINI
Università degli Studi di Cagliari

Amit HAGAR
Indiana University

FILOSOFIA DELLA SCIENZA

Il vero viaggio di scoperta
non consiste nel cercare nuovi orizzonti
ma nell'aver occhi nuovi.

— Marcel PROUST

Alla base di questa collana vi sono due idee guida. La prima è che i confini tra le discipline sussistano soprattutto per il piacere (e l'esigenza) di varcarli e che questa istanza sia più forte di qualsiasi implacabile "polizia di frontiera", tesa a impedire la libera interazione e lo scambio dialogico tra i diversi campi del sapere. Valeva ieri per la teoria di Copernico e per quella di Darwin, vale, a maggior ragione, oggi per le frontiere della cosmologia o per quelle della biologia e della fisica, per non parlare dell'informatica o dell'alta tecnologia. La seconda idea è che la filosofia più interessante, come amava ripetere Ludovico Geymonat, è quella che si annida nelle pieghe della scienza, per cui è a quest'ultima, nelle sue diverse articolazioni e nei suoi svariati indirizzi, che vanno al di là di ogni artificiosa barriera tra "scienze della natura" e "scienze umane", che bisogna guardare per dare una risposta seria e credibile ad alcune delle grandi domande che la filosofia si è posta nel corso del suo sviluppo storico.

In questo quadro generale i singoli contributi che vengono proposti sono tutti contrassegnati da frequenti segni d'interpunzione metaforici, per stimolare quel tipo di lettura di cui parla Wittgenstein nei suoi Pensieri diversi: «Con i miei numerosi segni d'interpunzione io vorrei rallentare il ritmo della lettura. Perché vorrei essere letto lentamente». Non sono libri "usa e getta", da affrontare in maniera fugace e sbrigativa. Sono opere che esigono di essere lette seguendo e facendo propria la bellissima (e sempre attuale) massima attribuita a Svetonio, che è un richiamo all'importanza della meditazione: «Festina lente».



Vai al contenuto multimediale

Nicola Sisti

**Sistemi automatici
e concetti mentali**





Aracne editrice

www.aracneeditrice.it
info@aracneeditrice.it

Copyright © MMXIX
Giacchino Onorati editore S.r.l. – unipersonale

www.giacchinoonoratieditore.it
info@giacchinoonoratieditore.it

via Vittorio Veneto, 20
00020 Canterano (RM)
(06) 45551463

ISBN 978-88-255-2325-6

*I diritti di traduzione, di memorizzazione elettronica,
di riproduzione e di adattamento anche parziale,
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

*Non sono assolutamente consentite le fotocopie
senza il permesso scritto dell'Editore.*

I edizione: aprile 2019

9 *Introduzione*

13 **Capitolo I**
La conoscenza automatizzata

1.1. Conoscenza automatizzata e nuovi paradigmi, 13 – 1.2. Il corpo dei sistemi automatici e il circolo virtuoso della conoscenza automatizzata, 23 – 1.3. La Cyber Physical Convergence e l'industria 4.0, 30 – 1.4. La necessaria evoluzione dei sistemi automatici, 36

55 **Capitolo II**
Il ruolo del corpo e del contesto nei processi conoscitivi

2.1. Introduzione, 55 – 2.2. Il ruolo del corpo e del contesto nei processi conoscitivi, 56 – 2.3. La coevoluzione della conoscenza nelle interazioni tra i soggetti e la realtà, 76 – 2.4. Le precomprensioni nelle scelte teoriche e metodologiche che condizionano la progettazione dei sistemi automatici, 85 – 2.5. Il ruolo attivo dello sguardo, 97 – 2.6. Il ruolo del corpo nei processi conoscitivi, 103

111 **Capitolo III**
Il passaggio dalla conoscenza del corpo all'emergere della coscienza

3.1. Introduzione, 111 – 3.2. Il passaggio dalla conoscenza del corpo alla coscienza primaria, 112 – 3.3. L'emergere della coscienza di ordine superiore, 121 – 3.4. L'emergere della coscienza e il ruolo delle emozioni, 128 – 3.5. L'interdipendenza dei sistemi della conoscenza, 145 – 3.6. La creazione di valore del processo di apprendimento circolare, 161

171 **Capitolo IV**
Processi mentali e dati oggettivi

4.1. La mente come sistema ecologico, 171 – 4.2. L'interazione continua tra il soggetto e l'oggetto della conoscenza, 186 – 4.3. L'emergere del soggetto e dell'oggetto come elementi inseparabili dal processo della conoscenza, 196

217 **Capitolo V**
La determinazione della mente in termini relazionali e le conseguenze sulla conoscenza automatizzata

5.1. Introduzione, 217 – 5.2. La determinazione relazionale della realtà, 219 – 5.3. La determinazione relazionale dell'io, 223 – 5.4. Abduzioni, analogie e correlazioni, 238 – 5.5. La determinazione della mente in termini relazionali e le conseguenze sulla conoscenza automatizzata, 253

263 *Conclusioni*

273 *Bibliografia*

Introduzione

Uno dei concetti principali in grado di caratterizzare i progressi tecnologici di questi ultimi anni è senza dubbio l'automazione.

L'innovazione introdotta dalla rivoluzione dei big data e la possibilità di raccogliere e immagazzinare grandi quantità di dati da cui estrarre conoscenza ha generato la fiducia nell'idea di poter analizzare la realtà partendo dai dati in modo da cogliere al loro interno relazioni significative in grado di comprendere la complessità della realtà.

L'aspetto che viene maggiormente evidenziato è il ricorso all'analisi di enormi quantità di dati in modo da giungere a conclusioni senza che queste siano condizionate da teorie o modelli concettuali a priori. Alla base vi è quindi una cieca fiducia nel dato considerato come fonte in grado di farci giungere ad una conoscenza oggettiva, sempre aggiornata e costantemente monitorata, rispetto alla conoscenza soggettiva data dalle teorie e dai processi simbolici.

Partendo da questi presupposti da una parte è diventato quindi importante sviluppare algoritmi in grado di analizzare grandi quantità di dati rintracciando al loro interno pattern emergenti in grado di evidenziare forme di correlazione. Dall'altra è diventato altrettanto importante che gli algoritmi siano in grado di effettuare quest'analisi in maniera totalmente meccanica in modo da rendere il processo sempre più efficiente e in grado di gestire dataset di grandi dimensioni.

Ciò ha dato origine a una serie di settori di ricerca aventi come obiettivo quello di sviluppare algoritmi che non solo possano analizzare una gran quantità di dati ma siano anche in grado di apprendere un dato comportamento a partire dai dati. Tali algoritmi vengono quindi usati per costruire sistemi automatici che, partendo da un campione di dati in

ingresso, possano prendere decisioni intelligenti e pianificare in modo efficiente le azioni da fare per raggiungere un preciso obiettivo.

Questi sistemi devono sostanzialmente riuscire a replicare la capacità umana di generalizzare comportamenti o modelli a partire dall'osservazione. Lo scopo è quello di creare algoritmi in grado di acquisire le proprie conoscenze estraendo strutture ("pattern") da dati grezzi esterni in modo da replicare la capacità della mente di apprendere dalle interazioni con la realtà e con gli altri soggetti prendendo decisioni autonomamente.

Questi sistemi automatici intervengono ormai in molti ambiti della nostra vita. Li troviamo nei software dei nostri smartphone, nelle piattaforme digitali usate quotidianamente per dialogare e fare acquisti, nei software che usiamo negli uffici per accedere alle informazioni e gestire il lavoro, e — come vedremo — in diversi ambiti che riguardano la vita quotidiana di ogni individuo: l'istruzione, le banche, le assicurazioni, la selezione del personale, i sistemi adottati dalla polizia, ecc.

La loro forza pervasiva è legata al fatto che essi permettono di rendere i processi decisionali più veloci ed efficienti consentendoci così di ridurre drasticamente i tempi delle nostre scelte e delle nostre azioni.

Viviamo in un'epoca dove ormai è quasi tutto informatizzato e la diffusione dei big data tende a relegare sempre di più la gestione e l'analisi della realtà alle macchine. Ciò viene fatto sulla base di un presupposto di maggior efficienza e maggior aderenza ad un'analisi oggettiva del dato, e quindi del problema.

Questa diffusione dovrebbe poi essere quasi una premessa dei cambiamenti che la quarta rivoluzione industriale potrebbe portare in breve tempo. L'ideazione di macchine a guida autonoma e di robot in grado di cooperare con gli uomini in un contesto di fabbrica intelligente — ad esempio — fino a pochi anni fa sarebbero state considerate solo pura fantasia. Ora invece idee di questo tipo stanno diventando oggetto di studi sempre più avanzati.

L'applicazione in ambito industriale dell'IoT e dei Big Data ha dato vita a quella che è stata ribattezzata quarta rivoluzione industriale, l'era dell'Industria 4.0, caratterizzata da un contesto in cui macchine intelligenti, oggetti e persone, sono collegate fra loro. Questa interconnessione permette da una parte di creare nuove forme di business attraverso la creazione di oggetti e prodotti intelligenti (come ad es. le connected car, gli elettrodomestici che interagiscono con l'utente for-

nendo informazioni su di sé e risultando comandabili a distanza, ecc.). Dall'altra permette un notevole miglioramento dei processi produttivi dando vita a vere e proprie industrie intelligenti.

I sistemi automatici stanno cambiando notevolmente la nostra società. La loro progettazione e applicazione solleva tuttavia problematiche tecnologiche ed epistemologiche. L'evoluzione di questi sistemi, come vedremo, è legata infatti non solo a sviluppi tecnologici ma anche a problematiche che riguardano direttamente i processi della conoscenza.

In molti casi infatti, anche nell'analisi di problemi quotidiani, gli algoritmi tendono a giungere a risultati ancora poco affidabili. Inoltre, il loro sviluppo risulta necessario per raggiungere i livelli di efficienza cui mira l'industria 4.0.

Queste problematiche, come vedremo, sono legate al fatto che non esistono dati oggettivi. I dati che immettiamo come input nelle macchine sono sempre intrisi di significati e presupposti taciti che si legano al contesto pragmatico e sociale in cui essi hanno avuto origine e si sono poi sviluppati. Questi presupposti guidano, in molti casi, le scelte che facciamo a proposito di quali dati considerare e utilizzare. Queste scelte non sono dunque effettuate dall'algoritmo soltanto con criteri oggettivi e di efficienza.

Pensare di riuscire ad analizzare il dato prescindendo da tutte queste caratteristiche che lo contraddistinguono per sua natura, porta semplicemente ad errori fuorvianti.

I dati sono ben lungi dall'essere nudi e crudi; essi sono piuttosto oggetti situati in luoghi e collocati in un ambiente specifico, determinato sia spazialmente che temporalmente nel mondo.

La conoscenza della realtà non può quindi limitarsi a una semplice analisi di dati trovando al loro interno correlazioni, proprio per il fatto che i dati non sono mai oggettivi e gli algoritmi delle macchine hanno al loro interno dei presupposti che sono intrinsecamente legati al contesto e alla realtà che analizzano.

Il contenuto della nostra esperienza non è neutro ma viene strutturato in riferimento ai nostri interessi e alla nostra capacità di cogliere quel che contiene. Un qualsiasi oggetto che sperimentiamo appare necessariamente all'interno di questo campo. Sono quindi i nostri interessi e il know how accumulato nel passato a determinare quello che verrà ignorato e quello che verrà considerato come essenziale nel corso della nostra esperienza della realtà.

Quando invece tentiamo di fornire bit alla macchina depurati da un contesto e da obiettivi, stiamo tentando di liberare i fatti della nostra esperienza da quella organizzazione pragmatica che li rende fruibili nel quotidiano.

Per accrescere le capacità cognitive dei sistemi automatici diventa quindi importante individuare ciò che caratterizza le dinamiche dei processi conoscitivi umani.

I modi in cui la mente coglie i significati della realtà e li rielabora prendendone coscienza sono infatti più sensibili al contesto riuscendo a cogliere maggiormente la complessità del dato.

I modelli epistemologici che delineano il modo di procedere dei processi conoscitivi mettono in evidenza come i significati della realtà con cui interagiamo quotidianamente siano il risultato di complesse costruzioni che partono dal corpo per giungere poi a una piena consapevolezza al livello della coscienza. Quest'ultima risulta fortemente determinata dalle interazioni con la realtà e con gli altri soggetti.

La mente dell'uomo esiste all'interno di un tutto integrato in cui corpo e cervello svolgono un'azione reciproca. Tale azione si sviluppa costantemente nel tempo ed è alla base dello sviluppo dell'individuo e dell'evoluzione dell'uomo.

Ci muoveremo dunque all'interno di questo contesto per cercare di delineare le caratteristiche su cui si basano i sistemi automatici mettendo così in evidenza gli aspetti che occorre ancora implementare.

Lo sviluppo di questi sistemi non riguarda infatti soltanto problematiche teoriche e cognitive ma può avere importanti ricadute sull'innovazione tecnologica e su quella industriale dei prossimi anni.

La conoscenza automatizzata

1.1. Conoscenza automatizzata e nuovi paradigmi

I progressi tecnologici hanno sempre influito sulla nostra maniera di pensare e sulle rappresentazioni che abbiamo di noi stessi e della realtà. Se si ripercorre la storia degli sviluppi compiuti dall'uomo è possibile constatare come essi abbiano avuto un riflesso sia sulla nostra vita quotidiana sia sul nostro modo di rapportarci agli altri soggetti e alla realtà.

In questi ultimi anni — grazie ai recenti progressi dell'information technology — è emerso un nuovo modo di considerare i processi conoscitivi. Questa maniera è stata addirittura indicata come un nuovo paradigma della conoscenza che andrebbe a sostituire quelli precedenti.

Esso sarebbe dato dalla sempre maggiore possibilità — legata all'avvento dei big data — di gestire le conoscenze che utilizziamo quotidianamente come dati calcolabili.

Secondo i sostenitori dei Big Data la possibilità di raccogliere e immagazzinare grandi quantità di dati da cui estrarre conoscenza ha generato la fiducia nell'idea di poter analizzare la realtà partendo dai dati in modo da cogliere al loro interno relazioni significative in grado di comprendere la complessità.

L'aspetto che viene maggiormente evidenziato è il ricorso all'analisi di enormi quantità di dati in modo da giungere a conclusioni senza che queste siano condizionate da teorie o modelli concettuali a priori. Alla base vi è quindi una cieca fiducia nel dato considerato come fonte in grado di farci giungere ad una conoscenza oggettiva, sempre aggiornata e costantemente monitorata, rispetto alla conoscenza soggettiva data dalle teorie e dai processi simbolici.

In un articolo apparso sulla rivista Wired nel 2008, il suo direttore Chris Anderson ha esplicitato questa posizione mettendo in evidenza come attraverso i big data sia possibile trovare correlazioni senza preoccuparsi di avere un modello che le spieghi. Non si cercano più le cause all'origine degli eventi o delle azioni dei soggetti, ma si cercano correlazioni senza preoccuparsi di avere un modello che le spieghi.

Questo è un mondo dove quantità enormi di dati e la matematica applicata sostituiscono ogni altro strumento che si possa immaginare, insieme alle teorie sui comportamenti umani, dalla linguistica alla sociologia. Dimenticate tassonomie, ontologie e la psicologia. Chi sa perché la gente fa quello che fa? Il punto è che lo fa e che noi possiamo tracciarlo e misurarlo con una precisione senza precedenti. Con abbastanza dati, i numeri parlano da soli. [...]

Il metodo scientifico è basato su ipotesi verificabili. Questi modelli, per la maggior parte, sono sistemi visualizzati nelle menti degli scienziati. I modelli vengono poi verificati e gli esperimenti confermano o falsificano i modelli teorici di come funziona il mondo. Questo è il modo in cui la scienza ha lavorato per centinaia di anni.

Gli scienziati sono addestrati a riconoscere che la correlazione non è causale, che nessuna conclusione dovrebbe essere disegnata semplicemente sulla base della correlazione tra X e Y (potrebbe essere solo una coincidenza). Invece, è necessario comprendere i meccanismi sottostanti che collegano i due. Una volta che hai un modello, puoi connettere i set di dati con fiducia. I dati senza un modello sono solo rumori.

Ma di fronte a dati massicci questo approccio alla scienza — ipotesi modello, verifica — sta diventando obsoleto. [...]

Adesso c'è un modo migliore che ci permette di dire che la “correlazione è sufficiente”. Possiamo smettere di cercare modelli. Possiamo analizzare i dati senza ipotesi su ciò che potrebbero mostrare. Possiamo gettare i numeri nei più grandi cluster di calcolo che il mondo abbia mai visto e lasciare che gli algoritmi statistici trovino modelli dove la scienza non può.¹

Anderson mette quindi in evidenza come con i big data il mondo risulti finalmente misurabile e calcolabile. L'obiettivo è quello di misurare in modo sempre più esaustivo una realtà considerata prima inaccessibile. I dati sarebbero dunque in grado, secondo questo punto di vista, di rivelare verità che restavano prima ignote a causa della mancanza di sistemi di misura oggettivi.

¹ C. ANDERSON, *The end of the theory: the data deluge makes the scientific method obsolete*, in Wired, giugno 2008. <https://www.wired.com/2008/06/pb-theory/>.

Da questo punto di vista i big data rianimerebbero dunque il progetto di obiettività strumentale delle scienze naturali.

I *big data* rianimano il progetto di obiettività strumentale delle scienze naturali, ma questa volta senza laboratorio: è il mondo stesso che diventa direttamente misurabile e calcolabile. Quello cui ambiscono è misurare da più vicino possibile la “realtà”, in modo esaustivo, discreto e sottilissimo.²

Inoltre — proprio per l’accento posto sulla ricerca di semplici correlazioni — i Big Data individuerebbero un nuovo paradigma. Le correlazioni statistiche non procedono più dalla causa alla conseguenza, bensì, a partire dalle conseguenze, risalgono a una valutazione delle possibili cause.

Questo punto viene sottolineato anche da Cardon che mette in evidenza come, secondo i sostenitori dei big data, questa sia una diretta conseguenza del recente sviluppo delle risorse informatiche.

Visto che ormai le risorse informatiche lo permettono, non è più necessario affinare i modelli per filtrare la correlazione tra le variabili che servono loro come ipotesi. È sufficiente chiedere alla macchina di testare tutte le correlazioni possibili tra un numero sempre maggiore di variabili. Il modello non è più una voce di entrata nel calcolo, bensì una voce di uscita.³

Secondo i suoi sostenitori questo sarebbe un nuovo paradigma per la scienza dopo quelli precedenti: il metodo sperimentale, quello teorico matematico e quello computazionale, basato sulle simulazioni numeriche⁴.

Originariamente c’era solo la scienza sperimentale, poi c’è stata la scienza teorica con le leggi di Keplero, le leggi del Moto di Newton, le equazioni di Maxwell, e così via. Successivamente, per diversi problemi, i modelli teorici sono diventati troppo complicati per essere risolti in modo analitico, e gli scienziati hanno dovuto avviare la simulazione. Queste simulazioni ci hanno guidato per buona parte dell’ultima metà dello scorso millennio. Ora queste simulazioni stanno generando una gran quantità di dati, insieme a un grande aumento dei dati proveniente dalle scienze sperimentali. Gli scienziati ora non guardano attraverso i telescopi. Essi guardano attraverso complessi stru-

² D. CARDON, *À quoi rêvent les algorithmes*, Seuil, Paris 2015; trad. it. *Che cosa sognano gli algoritmi. Le nostre vite al tempo dei big data*, Mondadori, Milano 2015, p. 35.

³ Ivi, p. 37.

⁴ Cfr. T. HEY, S. TANSLEY, K. TOLLE, *The Fourth Paradigm. Data-Intensive Scientific Discovery*, Microsoft Corporation, Redmond, Washington 2009.

menti che trasmettono i dati ai datacenter, e solo allora guardano le informazioni sui loro computer. Il mondo della scienza è cambiato, e non c'è dubbio su questo. Il nuovo modello è per i dati che vengono catturati dagli strumenti o generati da simulazioni prima di essere elaborati da software e memorizzati nei computer. Gli scienziati devono guardare ai dati quasi alla fine di questa procedura. Le tecniche e le tecnologie per questa scienza dei dati sono così diverse che è utile distinguere la scienza dei dati dalla scienza computazionale, considerandola un nuovo *quarto paradigma* per la ricerca scientifica.⁵

Questo approccio sarebbe dunque l'evoluzione naturale dei precedenti paradigmi scientifici. Originariamente la scienza sarebbe stata empirica preoccupandosi di descrivere i fenomeni naturali. Successivamente, attraverso la formulazione di teorie, essa avrebbe utilizzato modelli e generalizzazioni. Lo sviluppo computazionale degli ultimi decenni avrebbe poi consentito di simulare fenomeni complessi. L'odierna analisi dei dati permetterebbe di unificare l'uso di teorie, esperimenti e simulazioni. I dati vengono acquisiti da strumenti ed elaborati poi da software. Le informazioni così elaborate vengono memorizzate in grandi data set che vengono analizzati dagli scienziati attraverso la gestione dei dati.

A una teoria e a un modello omogeneo, i big data sostituiscono dunque un mosaico costantemente rivedibile di micro-teorie contingenti che articolano spiegazioni locali dei comportamenti probabili. Tali calcoli sono destinati a guidare i nostri comportamenti verso gli oggetti più probabili: non hanno bisogno di essere compresi, e molto spesso, neppure possono esserlo⁶.

Anche Mayer-Schonberger e Cukier evidenziano questo punto:

Il mondo è subissato di informazioni come mai prima d'ora [...] Il cambiamento di dimensione ha prodotto un cambiamento di stato. Il cambiamento quantitativo ha prodotto un cambiamento qualitativo. [...] La società dovrà abbandonare almeno in parte la sua ossessione per la causalità in cambio di correlazioni semplici: non dovrà più chiedersi *perché*, ma solo *cosa*. Questo nuovo modo di affrontare i problemi ribalta secoli di prassi consolidate e mette in crisi il nostro approccio istintivo alle decisioni e alla comprensione della realtà.⁷

⁵ Ivi, pp. xviii–xix.

⁶ Cfr. D. CARDON, *À quoi rêvent les algorithmes*, cit.; trad. it. cit., pp. 43–44.

⁷ V. MAYER-SCHONBERGER e K. CUKIER, *Big Data: a revolution that will transform how we live, work, and think*, Houghton Mifflin Harcourt, New York 2013; trad. it., *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Garzanti, Milano 2013, pp. 15–16.

E poi, ancora più avanti, viene messo in evidenza il valore che emergerebbe dall'analisi dei dati.

C'è in corso un'autentica caccia al tesoro guidata dalle indicazioni che si possono estrarre dai dati e dal valore latente che si può liberare con il passaggio dalla causalità alla correlazione. Ma non è un tesoro unico. Ogni singolo gruppo di dati avrà un valore intrinseco, nascosto, non ancora portato alla luce, e l'obiettivo della competizione è scoprirlo e catturarlo per intero. I big data modificano la natura del business, dei mercati e della società.⁸

Anche Mayer-Schonberger e Cukier — come Cardon — sottolineano dunque come, secondo l'approccio dei big data, non sia più necessario affinare i modelli per filtrare la correlazione tra le variabili che servono loro come ipotesi. È sufficiente testare con un algoritmo enormi quantità di dati alla ricerca di correlazioni in modo da far emergere relazioni significative da sistemi estremamente complessi. Il modello non è più una voce di entrata nel calcolo, bensì una voce di uscita. Le correlazioni statistiche non procedono più dalla causa alla conseguenza, bensì, a partire dalle conseguenze, risalgono a una valutazione delle possibili cause.

La rivoluzione introdotta dai big data è dunque l'aver soppiantato la ricerca di una spiegazione dei fenomeni basata sulla relazione causale — ossia una condizione necessaria tra fatti, collegati tra loro da un rapporto di causa-effetto e secondo una precisa successione temporale — con una spiegazione dei fenomeni basata sulla correlazione.

La "logica dei Big Data" ci insegna che, in presenza di un campione di dati sufficientemente rilevante che attesti la presenza della condizione A, e osservato il verificarsi in concomitanza (o in stretta successione) di C, allora posso affermare che vi è una correlazione tra A e C senza necessariamente constatare il nesso B oppure dimostrare che C sia una conseguenza di A. Correlazione e causalità sono concetti ben diversi e non si deve incorrere nell'errore (grave) di scambiare i loro significati: la prima non spiega la relazione tra la serie di dati e il fenomeno osservato, ma semplicemente evidenzia un rapporto basato sull'osservazione empirica di "pattern" (schema o forma di regolarità) per così dire emergenti. Il grado (o indice) di correlazione, inoltre, può essere studiato in termini quantitativi come relazione tra due determinate variabili.⁹

⁸ Ivi, p. 28.

⁹ L. CAMICIOTTI, C. RACCA, *Creare valore con i big data*, LSWR, Milano 2015, pp. 55–56.

Nell'approccio Data Driven, il sistema oggetto di studio è dunque una scatola nera. Non interessa conoscere le cause dei fenomeni, ma soltanto modellare le loro correlazioni. Questi modelli generano decisioni basate su dati che in qualche modo sono altri dati. Il processo che porta a selezionare quali ritenere rilevanti per le decisioni non è a senso unico «bensì prevede un canale di ritorno altrettanto forte e importante. Le decisioni di oggi, infatti, influenzeranno i dati di domani, guidandoci nelle scelte relative a quali dati raccogliere, come acquisirli e come arricchirli. Il circolo virtuoso è, per certi aspetti, il motore perpetuo nell'approccio basato sui dati»¹⁰.

Secondo i sostenitori dei Big Data l'approccio alla scienza — ipotesi modello, verifica — sta dunque diventando obsoleto per via della possibilità di raccogliere e immagazzinare grandi quantità di dati da cui estrarre conoscenza. Ciò ha infatti generato la fiducia nell'idea di poter analizzare la realtà partendo dai dati in modo da cogliere al loro interno relazioni significative in grado di comprendere la complessità.

L'aspetto che viene maggiormente evidenziato è il ricorso all'analisi di enormi quantità di dati in modo da giungere a conclusioni senza che queste siano condizionate da teorie o modelli concettuali a priori. Alla base vi è quindi una cieca fiducia nel dato considerato come fonte in grado di farci giungere ad una conoscenza oggettiva, sempre aggiornata e costantemente monitorata, rispetto alla conoscenza soggettiva data dalle teorie e dai processi simbolici.

I dati sarebbero dunque in grado, secondo questo punto di vista, di rivelare verità che restavano prima ignote a causa della mancanza di sistemi di misura oggettivi. Inoltre — proprio per l'accento posto sulla ricerca di semplici correlazioni — i Big Data individuerebbero un nuovo paradigma. Le correlazioni statistiche non procedono più dalla causa alla conseguenza, bensì, a partire dalle conseguenze, risalgono a una valutazione delle possibili cause.

Da una parte diventa quindi importante sviluppare algoritmi in grado di analizzare grandi quantità di dati rintracciando al loro interno pattern emergenti in grado di evidenziare forme di correlazione. Dall'altra risulta altrettanto importante che gli algoritmi siano in grado di effettuare quest'analisi in maniera totalmente meccanica in modo da rendere il processo sempre più efficiente e in grado di gestire dataset di grandi dimensioni.

¹⁰ Ivi, p. 58.

Per far ciò occorre che l'algoritmo riesca non solo a condurre il procedimento in maniera completamente automatica mediante le istruzioni presenti al suo interno. È importante anche che esso sia in grado di apprendere a realizzare un compito pur senza essere programmato a farlo.

In sostanza si vuole che l'algoritmo, alla fine del processo di analisi dei dati, sia in grado di apprendere dall'esperienza avuta riuscendo così a migliorare la sua performance nelle successive analisi che dovrà compiere.

Lo studio di queste tecniche prende il nome di apprendimento automatico o Machine Learning. Con questa denominazione si intende quel campo di ricerca che ha come obiettivo quello di implementare algoritmi che non solo possono analizzare una grande quantità di dati ma sono anche in grado di apprendere un dato comportamento a partire da questi ultimi.

Rispetto alla prima IA il Machine Learning usa un approccio differente: invece di dare al computer un insieme di istruzioni su come fare qualcosa, dà alla macchina istruzioni su come imparare a fare qualcosa.

Tali algoritmi vengono quindi usati per “costruire” modelli che, partendo da un campione di dati in ingresso, possano prendere decisioni intelligenti e pianificare in modo efficiente le azioni da fare per raggiungere un preciso obiettivo.

Lo sviluppo di questi algoritmi è centrale per i sostenitori del paradigma dei Big Data. Al cuore dell'analisi dei dati, come sottolineato anche da Cukier

c'è infatti il *machine learning* (o apprendimento automatico) che è una componente dell'Intelligenza Artificiale. Di AI si parla fin dagli anni '50, ma per decenni abbiamo pensato che non funzionasse, perché l'idea era fare una macchina che pensasse come un essere umano, il che è tuttora troppo difficile. Il machine learning invece cerca tendenze, legami causa-effetto, correlazioni nascoste, e oggi funziona perché abbiamo tanti dati. In fondo, se vogliamo, l'espressione “big data” è la popolarizzazione delle tecniche di machine learning: far lavorare i computer su enormi volumi di dati per scoprire legami e tendenze che gli uomini non possono vedere. E questo spiega tutte le grandi scoperte che stiamo facendo in questi anni.¹¹

¹¹ Cfr. K. CUKIER, *Artificial Intelligence? Grazie ai Big Data, dopo 70 anni funziona*, settembre 2018, <https://www.digital4.biz/executive/intelligenza-artificiale-big-data-analysis-cukier-the-economist/>.

Il machine learning non ha dunque come obiettivo quello di dare alla macchina un insieme di istruzioni in modo da farla pensare come un essere umano, ma ha come fine quello di realizzare una macchina che sia in grado di analizzare grandi quantità di dati in maniera automatica. Queste macchine hanno come scopo quello di analizzare grandi quantità di dati cercando al loro interno correlazioni nascoste, legami differenti e nuove tendenze.

Questi sistemi devono sostanzialmente riuscire a replicare la capacità umana di generalizzare comportamenti o modelli a partire dall'osservazione. L'obiettivo è quello di creare algoritmi in grado di acquisire le proprie conoscenze estraendo strutture ("pattern") da dati grezzi esterni in modo da replicare la capacità della mente di apprendere dalle interazioni con la realtà e con gli altri soggetti prendendo decisioni autonomamente.

Gli algoritmi del machine learning partono dall'analisi di esempi (i dati) per cercare di estrarre delle regole che siano in grado di descriverli. L'apprendimento diventa in questo modo un processo iterativo che, con l'aumentare dei dati e delle interazioni analizzate, migliora sempre di più le conoscenze del sistema.

L'obiettivo principale è quello di sviluppare sistemi che possano apprendere conoscenza attraverso l'analisi dei dati, estraendo da essi modelli ed essere così in grado di prendere decisioni¹².

Un esempio è dato dalla classificazione. Uno degli scopi principali del Machine Learning è la classificazione, cioè il problema di creare algoritmi che siano in grado di decidere a quale categoria appartiene un determinato dato. Sistemi di questo tipo sono detti classificatori. Essi estraggono dal dataset un modello che utilizzano successivamente per classificare i nuovi dati che incontrano, dando origine così a un albero di decisione. Le funzioni di classificazione vengono infatti descritte come coppie attributo–valore per le quali è nota la classe e si presentano in forma di albero. Esso viene costruito utilizzando tecniche di apprendimento a partire dall'insieme dei dati iniziali (training set) per i quali è nota la classe. Ogni nodo interno rappresenta una variabile e ogni foglia il valore per la classe a partire dai valori delle altre proprietà. L'albero è rappresentato dal cammino (path) che parte dal nodo radice (root) fino ai vari nodi–foglia.

¹² Cfr. D. FAGGELLA, *What is machine learning*, Techemergence 2018, <https://www.techemergence.com/what-is-machine-learning/>.