

Toh, l'imprevisto!

Cristiana Pulcinelli

L'umanità cerca di prevedere il futuro da sempre. Negli ultimi anni la predizione è una scienza in crescita. Si esercita su questioni più diverse: da dove è meglio allocare le risorse a cosa accadrà all'ambiente fino al risultato delle prossime elezioni. L'apprendimento automatico dei computer è uno degli strumenti a disposizione, ma basterà? Potremo fare a meno dell'elemento umano? A che punto siamo?

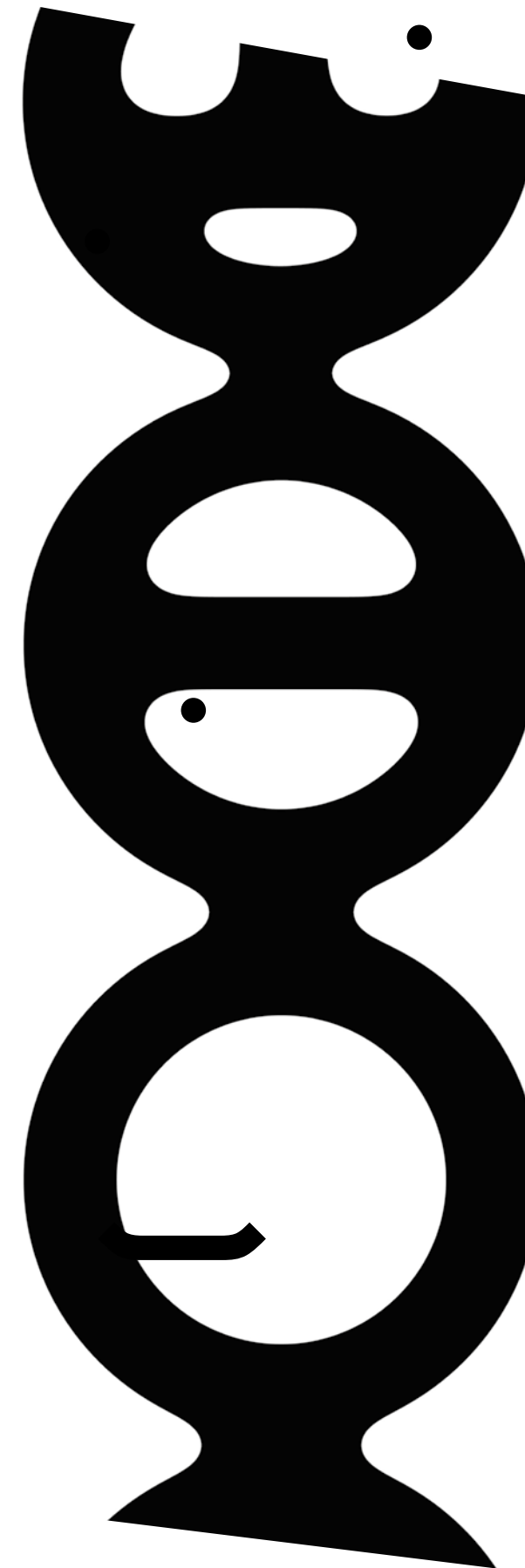


«*With four parameters I can fit an elephant, and with five I can make him wiggle his trunk*» (Con quattro parametri posso descrivere un elefante e con cinque gli faccio muovere la proboscide). Così era solito dire il matematico e informatico John von Neumann, almeno secondo quanto affermò Enrico Fermi durante un incontro con Freeman Dyson il quale riportò poi la storiella in un articolo pubblicato su *Nature* molti anni dopo. In sostanza, possiamo leggere la frase di von Neumann come il fatto che possiamo descrivere qualsiasi cosa, se solo siamo in possesso di dati sufficienti per creare un modello che funzioni. Oggi di dati ne abbiamo molti di più di quanti Neumann, che è morto nel 1957, potesse mai immaginare, tanto che si parla di "big data" ovvero un dataset talmente grande da richiedere strumenti non convenzionali per estrapolare, gestire e processare informazioni. I "big data" inoltre intrecciano fonti eterogenee, quindi non solo dati strutturati come i database, ma anche non strutturati come immagini, email, dati GPS, informazioni prese dai social network. La capacità di elaborare questi dati è cresciuta in modo esponenziale, basti pensare al *machine learning*, l'apprendimento automatico, ovvero l'insieme di metodi sviluppati negli ultimi anni che forniscono ai computer l'abilità di apprendere senza essere stati esplicitamente programmati e che vengono utilizzati per la costruzione di algoritmi che possano apprendere da un insieme di dati e fare delle predizioni. Pensiamo ad esempio ai motori di ricerca su internet.

Tutto questo potrebbe far pensare che il sogno di prevedere cosa accadrà – un sogno che ci accompagna da millenni, da

quando scrutavamo il futuro nelle viscere degli animali o nel volo degli uccelli – potrebbe essere vicino ad avverarsi. Peccato che poi eventi come le recenti elezioni del presidente degli Stati Uniti ci facciano precipitare dal soffice mondo dei sogni a quello molto più duro della realtà. Già, perché l'elezione di Trump non era stata prevista da nessuno, o quasi. Come, del resto, la crisi economica del 2008. O uno dei numerosi conflitti che sono scoppiati nel mondo dal 2000 ad oggi. E come non potremo prevedere lo scoppio della prossima epidemia o come sarà il tempo a Roma fra tre settimane.

La rivista *Science* ha recentemente dedicato una sezione speciale al tema della previsione vista come una scienza in evoluzione e il primo articolo è dedicato proprio all'elezione di Trump. In particolare, gli autori raccontano come il gruppo di lavoro di Hernan Makse, fisico statistico dell'Università di New York, avesse lavorato nei mesi precedenti alla votazione a un metodo per predire chi sarebbe stato il nuovo presidente. Niente di nuovo: i siti dei giornali e delle televisioni lo fanno continuamente. Ma in questo caso, invece di affidarsi agli *exit poll*, i ricercatori si sono basati sui dati raccolti da Twitter nei quattro mesi precedenti l'elezione. Se Makse avesse azzeccato le previsioni, avrebbe avuto motivo di vantarsi. Il sistema degli *exit poll* è laborioso e molto costoso, si tratta di sentire moltissime persone telefonicamente o andando porta a porta. Si calcola che questo lavoro solo negli Stati Uniti foraggi un'industria da 18 miliardi di dollari. Inoltre, presenta degli indubbi problemi: intanto i tassi di risposta sono spesso bassi, il che vuol dire che gli analisti si ritrovano con un campione esiguo su cui fare le previ-



sioni, e poi il metodo si espone anche alla falsificazione dei dati, come ha dimostrato un articolo pubblicato sempre su *Science* a marzo del 2016. Il gruppo di Makse, invece, ha tracciato le opinioni politiche di milioni di individui direttamente, secondo dopo secondo, per mesi. E gratuitamente. Un bel vantaggio. Se si pensa che Twitter non è il solo a fornire dati interessanti e gratuiti da inserire in modelli predittivi ma che anche Facebook e Google generano grandi masse di dati che possono essere utilizzati gratuitamente, ci si trova di fronte a un'enorme quantità di informazioni da cui attingere.

Purtroppo le previsioni di Makse erano sbagliate. Come, del resto, quelle degli *exit poll* del *Times* con i quali lo scienziato confrontava via via i suoi risultati. Entrambi fino alle 8 di sera davano Clinton in vantaggio con il 55% dei voti. Il che vuol dire che forse è troppo presto perché i dati raccolti su Internet ci dicano qualcosa del futuro. Il problema principale, probabilmente, è che ancora non si sa come tradurre quei dati in intenzioni umane.

Prevedere cosa la gente farà e perché costituisce un po' l'essenza delle scienze sociali. Un compito quasi impossibile, se si considera quanto sia difficile predire il comportamento anche di una sola persona. Tuttavia, dicono alcuni scienziati, si può pensare a un'analogia con la fisica: anche se non possiamo sapere dove andrà una singola particella, il comportamento di un gas composto di milioni di particelle è prevedibile. Dunque, l'idea per molti anni è stata quella di trattare le società come un problema di fisica. Del resto, lo scrittore di fantascienza Isaac Asimov negli anni

Il matematico David Orrell sostiene che i modelli predittivi di oggi sono governati dagli ideali classici degli antichi greci

Cinquanta del secolo scorso immaginava che, grazie a potenti computer e enormi banche dati, si potesse prevedere addirittura la nascita e la caduta di imperi. Ma le cose si sono rivelate più complicate.

Nel caso del modello basato su Twitter di cui parlavamo prima, lo stesso autore ha individuato alcuni problemi, dopo lo scivolone dei risultati elettorali. Ad esempio, dei 73 milioni di tweet sui due candidati che sono stati presi in

esame, quanti erano scritti da esseri umani? Twitter infatti permette di prendere parte alle discussioni online anche ai bot, programmi di computer che imitano gli esseri umani. I bot non sono etichettati come tali e appaiono a chi li segue come dei sostenitori entusiasti o dei detrattori accaniti (ma sempre in carne ed ossa) di una qualche tesi o posizione. Possono essere smascherati analizzando il loro account, ma eliminare i bot non basta: rimane il problema dei prezzolati. Persone che vengono pagate per prendere posizione e orientare i votanti e che per fare questo utilizzano finti account. E poi ci sono problemi più ampi: ad esempio, siamo sicuri che se una persona parla bene di un candidato poi lo voterà? Le motivazioni profonde rimangono sconosciute. E poi: quanti votanti sono sottorappresentati nella rete? La popolazione rurale che ha votato Trump, ad esempio, usa Twitter? Ma, del resto, scrive lo statistico Andrew Gelman, se potessimo predire i risultati delle elezioni in modo perfetto, votare diventerebbe un atto senza senso.

Se i "big data" possono essere usati (e lo sono) per trovare il giusto partner o la migliore stanza d'albergo o il giusto cliente per i prodotti di un'azienda, si è pensato potessero essere utilizzati anche per predire lo scoppio di conflitti. È così che alcuni ricercatori stanno applicando le tecniche di *machine learning* per cercare di capire dove e quando i fattori di rischio potranno portare alla violenza. Qualcosa in questo senso si sta muovendo. Ad esempio, lavorando ad un rapporto commissionato dalla *Political Instability Task Force* della CIA, un gruppo di analisti, tenendo sotto controllo proteste e ribellioni in 168 Paesi per 6 mesi, è riuscito a predire il colpo di stato dei militari in Thailandia nel 2014 un mese prima che avvenisse. E, con queste nuove tecniche, un altro team di ricercatori è riuscito a prevedere 9 guerre civili su 20 scoppiate tra il 2001 e il 2014.

I modelli precedenti non ne avevano previsto nessuno. E tuttavia, ancora non possiamo essere ottimisti. Il problema è che i processi che portano ai conflitti coinvolgono una serie di attori imprevedibili che interagiscono tra loro in modi che rompono gli sche-

mi. Insomma, uno scenario molto più complicato di quello di un insieme di consumatori da analizzare per le ricerche di mercato. I sociologi e la comunità degli scienziati che si occupano di *machine learning* stanno acquistando nuovi strumenti analitici per distinguere pattern significativi dal rumore di fondo. Però il rumore di fondo è ancora molto forte.

Lo stesso può dirsi per la "scienza delle scienze", ovvero la capacità di prevedere quali saranno le prossime scoperte scientifiche, o almeno i campi di ricerca più promettenti e soprattutto su cui si investirà di più. Anche qui si tratta di una disciplina in crescita, ma che ancora non ha prodotto una visione dei processi sociali che portano alla scoperta scientifica.

C'è chi poi lamenta che i modelli sono anche troppo potenti ma spesso non tengono conto della realtà. Come ad esempio il gruppo di ecologi autori di un appello comparso sulle pagine di *Science* nel 2016. Bisogna migliorare le previsioni sulle conseguenze del cambiamento del clima, dicono gli scienziati, ma non è un problema di modelli. A fronte di modelli sempre più sofisticati, a mancare sono i dettagli specifici sulla biologia delle specie. «Se vogliamo fare previsioni realistiche dobbiamo tirar su gli stivali, prendere il binocolo e tornare sul campo a raccogliere informazioni più precise», commenta uno dei firmatari. Molti aspetti cruciali, come la capacità di adattamento o le curve demografiche di una specie, sono spesso tagliati fuori dai modelli di previsione,

Secondo uno studio, tra l'85% e il 90% delle persone non vuole sapere in anticipo gli eventi negativi che stanno per accadere

che a lungo andare non corrispondono più al reale cambiamento in corso. Solo il 23% degli studi tengono conto dei meccanismi biologici, secondo gli scienziati che hanno aderito all'appello.

Forse però la difficoltà principale è nella nostra mente. Il matematico David Orrell sostiene che i modelli predittivi di oggi sono governati dagli ideali classici degli antichi greci: unità, stabilità, simmetria, eleganza.

za, ordine. Insomma, l'armonia matematica. Ma cose come il clima o la moderna economia non seguono queste regole. Invece di trovare nuovi modelli, però, si è cercato di adattare quelli vecchi con risultati discutibili. «Dobbiamo ammettere che alcune cose non si possono prevedere – dice Orrell – dobbiamo accettare l'incertezza dei sistemi viventi». Nuovi modelli stanno emergendo dalle scienze della vita che guardano al mondo come un organismo vivente piuttosto che una macchina. Modelli che danno vita a una nuova estetica che trova la bellezza nella complessità piuttosto che nell'eleganza della simmetria. Del resto, non sempre le persone vogliono sapere cosa accadrà. Uno studio pubblicato da poco su *Psychological Review* mostra proprio il fatto che spesso si sceglie di ignorare in modo deliberato in anticipo sia le cattive notizie come la morte o un divorzio sia quelle positive come il sesso del nascituro.

I ricercatori hanno stimato la prevalenza di quella che chiamano "ignoranza deliberata" per 10 eventi. Ne è emerso che tra l'85% e il 90% delle persone non vuole sapere in anticipo gli eventi negativi che stanno per accadere e tra il 40% e il 70% preferisce rimanere all'oscuro anche degli eventi positivi. Solo l'1% dei partecipanti allo studio voleva sempre e comunque conoscere il futuro. Il desiderio di preservarci da sentimenti forti spesso vince sulla nostra curiosità. Chi l'avrebbe previsto?

