

Dominique Cardon

CHE COSA
SOGNANO
GLI ALGORITMI

Le nostre vite al tempo dei *big data*

Traduzione di Chetro De Carolis

Calcolare da più vicino possibile

Se i calcoli sono sempre più attrattivi, è anche perché la società non si lascia più prendere le misure facilmente. La logica di potenziamento della personalizzazione, nella quale sono entrate le tecniche attuali, è una conseguenza dell'individualizzazione espressiva che accompagna lo sviluppo degli usi del digitale.

Nelle società gerarchiche, in cui l'accesso allo spazio pubblico era ristretto, era facile parlare in nome di individui tramite categorie rappresentative. Governatori, portaparola o statistici potevano «dire» la società attraverso gli insiemi che avevano elaborato per dipingerla. Oggi, questa parola astratta e disincarnata sembra sempre più forzata e arbitraria. Essa è sempre meno in grado di rappresentare la varietà delle esperienze individuali. Mentre il web ha aperto a tutti il diritto di prendere la parola in pubblico, il monopolio esercitato dai rappresentanti sulla descrizione della società è stato minato e, con esso, le categorie che le servivano a far parlare gli altri. Lo spazio pubblico digitale ha liberato le soggettività, permettendo agli individui di autorappresentarsi¹². In ogni spazio d'incontro della vita sociale, questi ultimi reclamano di non essere ridotti alla categoria che li rappresenta. Rifiutano di lasciarsi rinchiudere nelle categorie socio-professionali che hanno fatto comodo a una società basata sullo status. I pazienti non vogliono essere ridotti alla loro malattia, i clienti ai loro acquisti, i turisti ai loro percorsi, i militanti alla loro organizzazione, gli spettatori al silenzio ecc.

L'individualizzazione dei modi di vita e l'aumento delle opportunità sociali hanno contribuito ad aumentare la volatilità delle opinioni, la diversità delle traiettorie personali, il moltiplicarsi dei centri di interesse e la varietà dei consumi. Un numero crescente di comportamenti e atteggiamenti si lascia sempre meno correlare alle grandi variabili esplicative a cui erano abituati sociologi e professionisti del marketing. Benché la realtà delle determinazio-

¹² Cardon, Dominique, *La Démocratie Internet. Promesses et limites*, Seuil, «La République des Idées», Paris 2010.

ni sociologiche dei comportamenti e delle opinioni sia tutt'altro che scomparsa, queste non si lasciano più cartografare nel modo ovvio in cui lo aveva fatto Pierre Bourdieu ne *La Distinction* (1979)¹³, organizzando un insieme disparato di preferenze e pratiche politiche, culturali, culinarie o turistiche sui due assi di distribuzione del capitale economico e del capitale culturale.

Il moltiplicarsi delle pratiche, lo sviluppo del consumo di oggetti di svago, la diversificazione delle scale di giudizio, le variabili d'intensità e di accumulo delle pratiche hanno reso molto più complessa – e meno leggibile – la distribuzione delle scelte culturali¹⁴. Questa non è scomparsa, ma, per osservarla, ormai bisogna mettere in moto analisi molto più sofisticate ed entrare profondamente nelle traiettorie delle attività degli individui¹⁵. Certi sistemi classificatori che erano diventati categorie comuni di percezione del mondo sociale non funzionano più come strumenti di lettura condivisi da tutti¹⁶.

I recenti sviluppi delle tecniche statistiche cercano di mitigare queste difficoltà rinnovando sia la natura dei dati che i metodi di calcolo. Oggi è in atto una modificazione sistematica nella scelta dei dati da sottoporre ai calcolatori. Alle variabili stabili, perenni e strutturanti, che fissavano gli oggetti statistici in determinate categorie, gli algoritmi digitali preferiscono *catturare gli eventi* (un clic, un acquisto, un'interazione ecc.) che registrano al volo, per confrontarli con altri eventi, senza dover procedere a una categorizzazione. Piuttosto che variabili «lorde», cercano di misurare segnali, comportamenti, azioni, performance.

I nuovi sistemi pubblicitari sul web sono automi che funzionano sulla base di un sistema di *real-time bidding*. Mentre l'inter-

¹³ Éditions de Minuit, Paris 1979; trad. it. *La Distinzione*, Il Mulino, Bologna 2007.

¹⁴ Donnat, Olivier, *Les Français face à la culture. De l'exclusion à l'éclectisme*, La Découverte, Paris 1994.

¹⁵ Lahire, Bernard, *La Culture des individus. Dissonances culturelles et distinction de soi*, La Découverte, Paris 2004.

¹⁶ Boltanski, Luc – Thévenot, Laurent, *Comment s'orienter dans le monde social*, in «Sociologie», VI, 1, 2015.

nauta carica la pagina web che desidera consultare, il suo profilo viene messo all'asta da un automa affinché dei robot programmati dagli inserzionisti si contendano il miglior prezzo per inserire il loro banner pubblicitario. L'operazione dura meno di 100 millesimi di secondo. Il profilo messo all'asta non è uno di quei ritratti-tipo del marketing tradizionale. Le informazioni offerte ai robot degli inserzionisti sono le precedenti tracce di navigazione dell'internauta, navigazione che è stata registrata dai cookie. Come fulmini, i robot degli inserzionisti propongono un prezzo di acquisto stimando la probabilità che l'internauta clicchi sul banner pubblicitario a partire dai dati di attività di altri internauti. I promotori di questi sistemi automatizzati assicurano che il successo dell'affissione pubblicitaria è al 30% superiore quando è realizzata da un automa analista di tracce rispetto a quando è svolta da un *media planner* umano che si basa sulla sua conoscenza del mercato e della propria clientela.

Nel mondo senza frontiere di internet, lo sfumarsi del perimetro nazionale contribuisce a rendere più fragile la campionatura delle popolazioni per nazionalità. Gli internauti vengono identificati tramite altri criteri misurabili in rete, come il fatto di essere utenti o consumatori, o attraverso l'appartenenza a comunità etniche, religiose o culturali. Caduti i quadri nazionali, è venuta meno la maggior parte delle tecniche di campionatura che permettevano di classificare un fenomeno all'interno di una popolazione delimitata. Per rappresentare i fenomeni misurati, è così aperta la via a forme statistiche, come la classifica, l'indice e il barometro, alternative alla percentuale che paragona un'attività a se stessa senza reinserirla in un quadro più generale e spesso senza conoscere il perimetro preciso della popolazione in questione.

Il decadere delle tecniche di campionatura ha incoraggiato una rottura radicale nelle metodologie statistiche. Forti del potere del calcolo computerizzato, i promotori dei *big data* reclamano dei corpus completi di dati e si dilettono del catturarli allo stato «grezzo». Certi assicurano che è preferibile prelevare tutti i dati senza doverli scegliere con precisione. L'abbandono dell'esigenza metodica di selezione dei dati nei calcoli digitali, comporta diverse conseguenze.

Innanzitutto, le registrazioni riguardano solo gli attivi, coloro che hanno lasciato tracce; gli altri, quelli non connessi, quelli non agenti, quelli non sottoposti al *tracking*, risultano semplicemente esclusi da una struttura di dati on line. Poi, a contribuire alla personalizzazione del calcolo, è l'assenza di infrastrutture categoriali che rendano le registrazioni statistiche passibili di essere incluse in un insieme. La maggior parte dei servizi web che impiegano tecniche di trattamento dati massiccio, non fanno che restituire all'internauta le informazioni che gli corrispondono. Significativo è il fatto che l'unico strumento di rappresentazione che assicuri una generalizzazione dei dati resti la carta geografica. La geolocalizzazione che permette di zoomare sulla propria situazione e poi decentrarsene, è l'ultimo strumento di totalizzazione superstita, laddove è scomparsa ogni nomenclatura¹⁷. Il mio quartiere è protetto dai criminali? Il valore immobiliare della mia via sta crescendo? Il servizio di raccolta rifiuti della mia circoscrizione è efficace? Lo zoom permette all'individuo di vedersi nei dati, ma non gli permette più di liberarsi dal calcolo delle sue tracce per risalire la catena di interdipendenza tra locale e globale e scoprire le connessioni che inchiodano le azioni di ognuno al sistema. L'algoritmo incolla l'internauta alle proprie tracce e non gli consente di distanziarsene.

Correlazioni senza cause

Replica delle precedenti, una terza scossa fa vacillare i punti di riferimento della statistica standard: le correlazioni non necessitano di cause. In un articolo che ha fatto scalpore, Chris Anderson, uno dei guru della Silicon Valley, ha annunciato la «fine della teoria». Egli afferma che ormai i calcolatori dei *big data* possono cercare correlazioni senza preoccuparsi di avere un modello che dia loro una spiegazione. I dati massicci e quelli matematici consentirebbero di fare economia delle scienze umane.

¹⁷ Cardon, Dominique, *Zoomer ou dézoomer? Les enjeux politiques des données ouvertes*, in Bernard Stiegler (a cura di), *Digital Studies. Organologie des savoirs et technologies de la connaissance*, FYP Éditions, Paris 2014.

«Chi sa perché la gente fa quello che fa? Fatto sta che lo fa e che lo si può registrare con una fedeltà senza precedenti. Con un numero sufficiente di dati, le cifre parlano da sole»¹⁸. Prendendo atto della nostra ignoranza delle cause all'origine delle azioni degli individui, i calcolatori abbandonano la ricerca di un modello che permetta di spiegarle a priori. Così, in certi settori della statistica, si è instaurato un nuovo rapporto con la causalità, che conferisce ai modelli detti «di Bayes» una vittoria postuma sulla statistica «frequentistica» sviluppata sulla scia di Quetelet.

Il caso Target ha fatto scalpore negli USA e offre un esempio semplificato del principio dei metodi di apprendimento¹⁹. Questa insegna di supermercati, tramite i data base della sua clientela, dispone delle informazioni di acquisto dei suoi clienti e conosce, grazie a un registro delle nascite, una parte delle sue clienti che hanno dichiarato di aver avuto un figlio. Il metodo delle tecniche di apprendimento consiste nel dividere i dati in due corpus diversi. A partire dai cambiamenti dei comportamenti di acquisto effettuati dal sotto-corpus delle donne di cui si sa che hanno avuto un figlio di recente, è possibile trovare correlazioni tra variabili di acquisto e farne un modello. Questo viene in seguito applicato all'altro sotto-corpus, allo scopo di *predire*, tra le clienti di cui non si sa se siano incinte, quelle che potrebbero esserlo.

L'algoritmo ha *appreso* il suo modello a partire dal primo sotto-corpus per *predire* un evento relativo al secondo sotto-corpus. Alcune delle correlazioni che consentono di fare tali predizioni sono estremamente banali (tipo comprare dei prodotti per neonati). Altre, in compenso, possono essere molto meno ovvie e spesso non sono altro che variabili nascoste (dai *proxies*) di altri fattori che non figurano tra i dati. Gli assicuratori, per esempio, avrebbero constatato, a partire dai dati di acquisto dei loro clien-

¹⁸ Anderson, Chris, *The End of Theory: the Data Deluge Makes the Scientific Method Obsolete*, in «Wired Magazine», 2008.

¹⁹ Siegel, Eric, *Predictive Analytics. The Power to Predict who Will Click, Buy, Lie or Die*, John Wiley & Son, Hoboken 2013; trad. it. *Analisi predittiva. Sapere in anticipo chi clicca, compra, mente o muore*, SLWR, Milano 2015.

ti, che coloro che compravano feltrini da mettere sotto le zampe di tavoli e sedie per non fare rigare il parquet, avevano un comportamento automobilistico molto prudente e che potevano, senza rischi, proporre loro riduzioni sul premio.

Questo non è un calcolo «individuale». Se è possibile è solo in quanto esiste un volume importante di comportamenti di acquisto. La predizione è solo una stima statistica e non presenta alcuna certezza. Tuttavia, è proprio quella cliente che un giorno riceverà un buono sconto per «donna incinta», pur non avendo confidato a nessuno di esserlo. La «predizione» sembra avere indovinato la sua intimità. In realtà, il calcolo ha soltanto presunto che potesse esserlo, osservando il comportamento delle altre. Attraverso una strumentazione ben più elaborata, è questo il principio che viene ormai applicato per individuare l'infedeltà dei clienti, gli accoppiamenti amorosi sui siti di ricerca dell'anima gemella, la recidiva giudiziale o la diagnosi preventiva di certe malattie nei data base medici.

I modelli statistici dei nuovi *data scientists* provengono dalle scienze esatte. Con un metodo induttivo, procedono alla ricerca di regolarità facendo meno ipotesi possibile. Le capacità di calcolo ormai sono talmente potenti da permettere di testare tutte le correlazioni possibili, senza risparmiarne nessuna col pretesto che l'ipotesi che vi conduce non la si farebbe mai. Tuttavia, sarebbe errato pensare che tali metodi siano alla caccia di correlazioni «che funzionano» senza preoccuparsi di spiegarle. In realtà, essi producono, sì, modelli di comportamenti, ma questi non emergono che *ex post* e si presentano come una serie aggrovigliata di spiegazioni le cui variabili funzionano a seconda dei profili. Per esempio, è possibile che, secondo i profili, la predizione che piaccia la cosa A dipenda, in un caso dal colore degli occhi, dall'origine sociale e dal numero dei traslochi, e, in un altro caso, dal fatto di aver viaggiato in Estonia e aver letto le opere complete di Balzac.

A una teoria comportamentale omogenea, i calcolatori sostituiscono un mosaico costantemente rivedibile di micro-teorie contingenti che articolano pseudo-spiegazioni locali dei comportamenti probabili. Tali calcoli sono destinati a guidare i no-

stri comportamenti verso gli oggetti più probabili: non hanno bisogno di essere compresi e, molto spesso, neppure possono esserlo. Questa maniera invertita di fabbricare il sociale testimonia il rovesciamento della causalità operato dal calcolo statistico per far fronte all'individualizzazione delle nostre società e alla sempre maggiore indeterminazione di ciò che determina le nostre azioni. In effetti, colpisce constatare che le logiche odierne dei calcolatori cercano di re-inquadrare la società, ma, in un certo senso, alla rovescia e dal basso, partendo dai comportamenti individuali per poi inferirne le peculiarità che li rendono statisticamente probabili.