



TUTTO DI TE MI PARLA

Big Data e dati personali confluiscono in prodotti digitali più innovativi, in grado di prevedere tutte le nostre esigenze e tenere traccia di tutti i nostri spostamenti, ma solo se lo vorremo.

Antonio Regalado

Sareste disposti a cedere i vostri dati personali in cambio di un pezzo di futuro? Andreas Weigend l'ha fatto. Weigend, che è stato coordinatore scientifico di Amazon.com prima di diventare il direttore del Social Data Lab dell'Università di Stanford, racconta che una volta si è svegliato all'alba per prendere un volo da Shanghai. Ma un'app che aveva iniziato a usare da poco, chiamata Google Now, lo ha informato che il suo volo era in ritardo.

Il software setaccia l'account di posta elettronica Gmail e il calendario Google e incrocia i dati disponibili su spostamenti e voli aerei. Aveva riscontrato una discrepanza tra l'agenda e l'orario indicato per il volo e ha avvisato l'interessato che non c'era bisogno di affrettarsi. Al momento dell'imbarco, gli altri passeggeri avevano atteso molte ore in più.

Per Weigend, che lavora come consulente e tiene corsi sul comportamento dei consumatori, un simile episodio consente di immaginare «le potenzialità di una società basata su una quantità di dati dieci volte superiore a quella attuale». Se il secolo scorso è stato segnato dalla capacità di osservare le interazioni al livello della materia fisica, per esempio grazie a tecnologie come i raggi X e il radar, il nuovo secolo sarà definito dalla gestione dei dati che le persone condividono.

I cosiddetti *anticipatory systems*, come Google Now, rappresentano un esempio di quello che potremmo essere in grado di fare. Già nella pubblicità on-line si coglie il senso di quello che il mondo dei Big Data è in grado di realizzare, in tutte le situazioni nelle quali le azioni di milioni di persone possono venire osservate istante per istante. Gli studi sui dati stanno cercando di guadagnarsi un posto nella nostra vita quotidiana. Si può pensare a modelli statistici che aiutino a scegliere tra diverse offerte di lavoro, o avvisare, sulla base di dati raccolti elettronicamente, del probabile arrivo di un'influenza.

Ogni due anni raddoppia la quantità di dati creati a livello globale. A generarli sono in grande parte gli stessi consumatori, sotto forma di comportamenti come la scelta di scaricare un certo film, le e-mail, i tracciati telefonici eccetera. Tuttavia, appena lo 0,5 per cento di quei dati viene davvero analizzato. «La quantità di dati raccolti è tale per cui già adesso sarebbe possibile creare profili individuali», dice Patrick Wolfe, statistico e studioso di reti sociali alla University College di Londra. «Statisticamente, la forza dei dati è maggiore se sono riferiti a più persone, ma si può continuare ad analizzarli fino a giungere a risultati individuali».

Per le «raffinerie di dati» della Silicon Valley, come Google, Facebook e LinkedIn, l'incontro tra i Big Data e i dati personali è un obiettivo già noto e serve a creare strumenti utili agli inserzionisti per legare ogni utente ai suoi prodotti di riferimento. Facebook ci suggerisce chi possono essere i nostri amici. Google Now migliora quanto più gli forniamo i nostri dati. Con il boom nelle vendite di smartphone dotati di accelerometro, macchina fotografica e GPS, «ciascuno è equipaggiato per raccogliere e trasmettere dati personali», osserva Weigend. E questo è solo l'inizio.

Già oggi una piccola comunità di appassionati del cosiddetto *quantified-self movement* si è dotata di sensori, contapassi, misuratori di glucosio sottopelle. Uno degli appassionati è Stephen Wolfram, il creatore del motore di ricerca Wolfram Alpha, che per anni si è prestato a un progetto di massiccia raccolta di dati personali, catalogando le e-mail, le singole battute sulla tastiera del computer e i minimi movimenti del corpo. Wolfram è interessato alla creazione di app utili nella vita quotidiana, ma anche alle previsioni che le grandi raccolte di dati possono tentare sui comportamenti personali: un campo che chiama *personal analytics*. L'idea di Wolfram è che, come i motori di ricerca tentano di organizzare tutti i dati disponibili sul mondo reale, «nella *personal analytics* quello che devi fare è accumulare informazioni sulla vita di un individuo». L'intoppo, secondo Wolfram, è che alcuni dei dati più utili non vengono neanche raccolti, o resi almeno accessibili. Si tratta di un problema parzialmente tecnico, di mancanza d'integrazione. Ma molti dati sono immagazzinati da aziende private come Facebook, Apple, e Fitbit, il produttore di un contapassi molto diffuso. Ora che il valore dei dati personali diventa sempre più evidente, ci si possono attendere conflitti su questo terreno. Lo Stato della California quest'anno ha introdotto la norma sul «Diritto di sapere», che obbliga le società private a rivelare a ogni individuo le informazioni personali che lo riguardano.

La legge è l'esito della volontà di un movimento politico che richiede più privacy e responsabilità, ma anche un diverso assetto dei rapporti economici tra chi crea i dati e chi li utilizza. ■

Antonio Regalado è responsabile del settore marketing di MIT Technology Review.

Anonimato impossibile

La quantità di dati condivisi aumenta esponenzialmente, ma in massima parte i dati permettono di risalire a chi li ha creati.

Patrick Tucker

Nel 1995 l'Unione europea ha introdotto una normativa sulla privacy che definiva "dati personali" le informazioni in grado di identificare una persona, direttamente o indirettamente. Oggi, quella definizione include molte più informazioni di quelle che il legislatore europeo potesse immaginare. Molte più di quante ne fossero disponibili nel mondo digitale dell'epoca, appena 18 anni fa. Cosa è successo da allora? Prima di tutto, la quantità di dati creata annualmente è cresciuta esponenzialmente, arrivando a 2,8 zettabyte nel 2012, un numero gigantesco, e raddoppierà ancora entro il 2015, secondo uno studio firmato dal centro studi IDC. Circa tre quarti di questi dati sono generati dagli individui che in ogni istante creano e spostano nuovi file digitali.

Molti di questi dati sono invisibili alle persone e sembrano non appartenere a nessuno. Ma non è così. Lo studio dei dati rivela che ogni tipo di dati può venire sfruttato, proprio come le impronte digitali, per identificare chi l'ha generato: i film che si scaricano da Netflix, i dati di localizzazione che il telefono cellulare scambia con i ripetitori e anche l'itinerario di ogni spostamento dell'utente ricostruito dalle telecamere di sorveglianza. In effetti, più dati esistono, meno si possono considerare privati, dal momento che la ricchezza di quei dati rende "algoritmicamente possibile" l'identificazione delle persone, afferma Arvind Narayanan, informatico della Princeton University.

Il cammino su questa strada è già cominciato. Le informazioni che in passato venivano considerati dati personali — nome, indirizzo, carta di credito — sono già oggetto di compravendita da parte dei mercanti di dati. Sono dati che ciascuno di noi inserisce spon-

taneamente nei moduli dei sondaggi o dei contratti di acquisto, come la data e il modello dell'automobile, il reddito, gli investimenti personali, l'età, l'istruzione, il codice postale. Avete appena divorziato? Oppure i vostri figli sono andati a vivere da soli? Simili "eventi", che spostano le persone da una categoria di consumatori a un'altra, sono informazioni decisive per il settore pubblicitario. La Axiom, una società che ha raccolto informazioni su circa 500 milioni di consumatori, analizza i dati al fine di prevedere 3.000 propensioni al consumo distinte, rispetto per esempio a come una persona potrebbe comportarsi nella scelta tra diverse marche dello stesso prodotto. Ma questi mercanti di dati oggi sono considerati parte del passato, a confronto con società del mondo digitale come Facebook, che ha automatizzato la raccolta di informazioni personali per aggiornare i dati in tempo reale. Nel corso di alcuni procedimenti giudiziari in Europa, i ricorrenti hanno appreso che i registri delle loro interazioni con il sito, dai messaggi di testo ai contenuti che hanno contrassegnato con "Mi piace", agli indirizzi dei computer dai quali si collegavano, occupavano 800 pagine stampate, più alcuni megabyte ulteriori per ogni utente.

Con buona pace delle istanze dei sostenitori della privacy digitale, le raccolte di dati on-line e off-line aiutano gli esperti di marketing a indirizzare la pubblicità con maggiore precisione. A febbraio Facebook ha annunciato un accordo con la Axiom e altre società del settore per condividere i dati, collegando attività del mondo reale e attività sul Web.

Si sente dire spesso che queste raccolte di dati sono rese anonime in qualche modo, ma

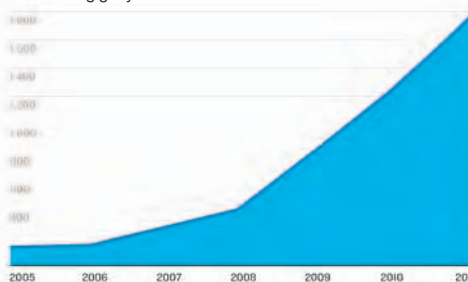
più i dati entrano in un file, meno questa affermazione appare vera. Gli operatori della telefonia mobile, per esempio, seguono i loro clienti passo per passo, creano un file che ricostruisce i loro spostamenti, poi cancellano il numero di telefono e vendono il file dei dati aggregati a società private interessate a studiare i percorsi delle persone. I ricercatori del MIT Yves-Alexandre de Montjoye e César A. Hidalgo hanno dimostrato che, persino quando i dati sulla posizione topografica sono anonimi, bastano quattro coordinate riconducibili a un telefono per identificarne il proprietario.

Più dati personali diventano disponibili, più informazioni se ne possono ricavare. Col giusto numero di dati a disposizione, si può anche tentare di vedere nel futuro di una persona. L'anno scorso Adam Sadilek, ricercatore alla University of Rochester, e John Krumm, ingegnere al Microsoft Research Lab, hanno predetto con un certo grado di approssimazione la posizione di un individuo con 80 settimane di anticipo, con un'accuratezza superiore all'80 per cento. Gli autori della ricerca hanno precisato che per ottenere questa previsione hanno fatto riferimento a "una raccolta di dati impressionante", mettendo insieme 32 mila giornate di rilevazioni GPS ottenute grazie a 307 persone e 396 veicoli. A quel punto hanno immaginato le applicazioni commerciali di questo sistema, che hanno significativamente chiamato *Far Out*, molto lontano. ■

Patrick Tucker è vicedirettore della rivista "The Futurist" ed esperto di tecnologia dell'informazione.

Informazioni digitali create ogni anno nel mondo

milliardi di gigabyte



Fonti: IDC, Radicati Group, Facebook, TR Research, Pew Internet

65 miliardi di dollari

Pagamenti localizzati, effettuati ogni anno in USA

154 miliardi

E-mail inviate quotidianamente

87%

Adulti americani la cui posizione è identificabile tramite telefono cellulare

2.000%

Aumento dei dati globali entro il 2020

111 megabyte

Video e foto conservate da Facebook per ogni utente

75%

Percentuale dei dati digitali creati dai consumatori