



CHIARA DATTOLA

Google sbaglia le previsioni sull'influenza

Hal Hodson, *New Scientist*, Regno Unito

Google flu trends, un sistema che prevede i picchi influenzali, sovrastima i casi. Un errore che rivela le debolezze delle tecnologie usate per l'analisi di enormi quantità di dati

Lanciato nel 2008, Google flu trends rileva le ricerche compiute in rete negli Stati Uniti per individuare parole associate all'influenza come "tosse" o "febbre". Usa poi quelle ricerche per prevedere con nove settimane di anticipo il numero delle probabili visite mediche dovute all'influenza. Negli ultimi tre anni, però, le ha sempre sovrastimate ed è stato particolarmente inaccurato intorno al picco influenzale, quando i dati sarebbero più utili. Nella stagione 2012-2013 ha previsto il doppio delle visite mediche rispetto a quelle documentate dai Centri per la prevenzione e il controllo delle malattie (Cdc) degli Stati Uniti e nella stagione 2011-2012 ha sbagliato di oltre il 50 per cento.

Secondo David Lazer della Northeastern university, principale autore di uno

studio su Google flu trends (pubblicato da Science), risolvere i problemi sarebbe piuttosto semplice, un po' come correggere la taratura di una bilancia. "È un po' un mistero, perché non ci voleva molto per migliorare la prestazione di Google flu trends", sostiene. Proiettando gli attuali dati dei Cdc a tre settimane di distanza si ottengono risultati più precisi di quelli ricavati da Google. Combinando le due fonti si ottiene un modello accuratissimo. Per Lazer, Google flu trends potrebbe essere utile soprattutto per prevedere l'andamento dell'influenza in zone più circoscritte di quelle considerate dai Cdc, permettendo così a singole città o stati di prepararsi.

Neil Richards, un avvocato esperto di privacy della Washington university di St Louis, ritiene che lo studio evidenzi l'immenso potere che l'analisi dei *big data* (enormi e complesse quantità di dati) conferisce a colossi come Google e Facebook e del perché quel potere è pericoloso. "Oggi ci sono aziende informatiche con un potere per certi versi pari a quello di uno stato, che hanno un forte impatto sulla nostra vita quotidiana", dice Richards. Comprendere i difetti e l'attività di queste aziende diventa quindi sempre più importante.

Per Evan Selinger, esperto di etica tecnologica del Rochester institute of technology di New York, i difetti di Google flu trends rimandano a un problema più grande, ossia all'approccio algoritmico adottato dalle aziende informatiche per fornire i servizi. Il problema sta nell'idea che i dati e gli algoritmi usati per elaborarli siano neutri. Il principale interesse delle imprese informatiche come Google, o dei broker di dati come Acxiom, sta nell'usare quei dati in modo che portino soldi (per esempio, in che punto dello schermo piazzare una pubblicità perché più cittadini maschi di Boston di età compresa tra i 25 e i 30 anni la clicchino). Il modo in cui tutto questo è fatto condiziona le nostre vite, eppure non abbiamo idea di come funzionino.

Calcoli segreti

"Quello della trasparenza degli algoritmi è uno dei maggiori problemi del nostro tempo", sostiene Selinger. "Sempre più decisioni che ci riguardano sono basate su processi a cui non abbiamo accesso". Selinger fa notare che altre sfere della vita in cui si prendono decisioni determinanti per noi sono molto più trasparenti, per esempio abbiamo il diritto ad avere un giusto processo se accusati di un reato, o possiamo contestare il punteggio della nostra affidabilità creditizia se lo riteniamo sbagliato. Invece, un'azienda che si ritrova spostata nella seconda pagina dei risultati di una ricerca su Google o esclusa da Google maps non ne conoscerà mai il motivo, perché gli algoritmi che prendono quella decisione sono di proprietà di Google.

Quanto più le decisioni del mondo reale si basano su algoritmi, tanto più diventa importante la trasparenza di quei processi, dice Richards. Già ora spesso sono dei sistemi automatizzati a stabilire se si può ottenere un prestito o un lavoro e l'analisi algoritmica può essere usata per decidere, per esempio, chi finisce nella *no-fly-list* (la lista nera dei passeggeri aerei sospetti). Servendosi dei documenti diffusi l'anno scorso da Edward Snowden, il sito The Intercept ha riferito che l'analisi dei dati in possesso della National security agency è stata usata per individuare i bersagli degli attacchi con i droni. "Spero che chi elabora i dati riconosca che lo straordinario potere sociale di cui dispone comporta una grande responsabilità professionale, proprio come per i medici, gli avvocati o i giornalisti", conclude Richards. ♦ *sdf*