

Big Data

Victor Mayer-Schonberger; Oxford University

Kenneth Cukier; The Economist

Garzanti 2013

Big Data

- sono enormi le quantità di dati, diversi tra loro per struttura (**ad esempio immagini, testi, video, numeri**), che vengono prodotti in qualsiasi campo attraverso l'utilizzo di dispositivi di vario genere come cellulari, sensori, carte di credito, interazioni sul web,
- l'autorevole gruppo di ricerca statunitense Gartner aveva inserito i Big Data tra i dieci problemi più importanti per il 2012, considerandolo la sfida del decennio 2010-2020, mentre il World Economic Forum, in un recente Rapporto ha classificato i Big Data in una nuova categoria di risorse economiche, alla stregua di una **materia prima**.

Big Data

- trasformare vasti oceani di dati disordinati in conoscenza è un compito estremamente difficile che necessita di un numero di professionisti qualificati in grado di affrontare la sfida, i **data scientist**,
- una figura professionale emergente e preziosa che unisce “le competenze di informatico, statistico e narratore per estrarre le pepite d’oro nascoste sotto montagne di dati” il mestiere che The Economist ha definito “**the sexiest job of the 21st century**”.

Big Data

Nella loro essenza i Big Data hanno a che fare con le previsioni.

Lo sfruttamento dei Big Data consiste nella applicazione della matematica a enormi quantità di dati per desumerne delle probabilità.

Spesso i Big Data sono caotici, variano nella qualità e sono distribuiti in una infinità di server sparsi in tutto il mondo. Ciò che perdiamo in accuratezza a livello micro lo recuperiamo in comprensione generale del fenomeno a livello macro.

Big Data

Tre cambiamenti rilevanti:

- Possibilità di analizzare enormi quantitativi di dati anziché essere costretti a lavorare su campioni.
- Disponibilità ad accettare la confusione intrinseca dei dati reali anziché privilegiarne l'esattezza.
- Disponibilità a spostare il fuoco sulle correlazioni anziché cercare di continuo le *cause* dei fenomeni sotto osservazione.

Usare l'identità $N=$ Tutti significa poter scavare in profondità all'interno dei dati. Con i campioni si perdono i dettagli, la specificità, la capacità di studiare da vicino alcuni sottogruppi.

Big Data

Stando ad alcune stime solo il 5% di tutti i dati digitali sarebbe «strutturato», ossia organizzato in maniera da inserirsi agevolmente in un data base tradizionale.

Se non si accetta la confusione il restante 95% dei dati destrutturati come le pagine web e i video sfugge alla rilevazione.

Le correlazioni ci aiutano a capire il presente e a prevedere il futuro: se A avviene spesso in concomitanza di B, l'utilizzo di B come indicatore ci aiuta a prevedere quello che sta accadendo presumibilmente ad A e ci aiuta a prevedere che cosa potrebbe accadere in futuro ad A.

Big Data

Molte aziende analizzano i Tweet per ottenere un feedback da parte dei clienti o valutare l'impatto delle campagne di marketing.

Al pari di Google tutta una serie di Social Media come Facebook, Twitter, LinkedIn e altri, siedono letteralmente su di una montagna di informazioni datizzate, che, una volta analizzate, faranno luce sulle dinamiche sociali a tutti i livelli.

Big Data

I dati si possono sfruttare molte volte per lo stesso scopo, ma anche e soprattutto per più scopi: il valore effettivo dei dati è molto superiore a quello ricavato dal suo primo utilizzo.

Oggi gli utilizzatori di Internet hanno familiarità con i «mashup» che combinano due o più fonti di dati: mappe, ad esempio, sui prezzi degli immobili incrociate con mappe dei distretti suburbani delle città americane.

Big Data

I big data possono essere un elemento di apprendimento per l'intelligenza artificiale:

Google ha costruito il correttore ortografico a costo zero, riutilizzando gli errori di digitazione (confrontati con le parole che gli utenti intendevano effettivamente digitare) inseriti nel suo motore di ricerca tra i tre miliardi di queries che gestisce ogni giorno. Ha usato dati in suo possesso per sviluppare l'intelligenza del suo motore di ricerca.

Big Data

Nei corsi educativi on line si possono raccogliere dati importanti.

Ad esempio se una grossa percentuale di studenti ha riguardato una parte di una lezione, è probabile che il docente in quella parte non sia stato chiaro.

Negli ultimi anni è venuta alla ribalta una nuova figura professionale, il «Data Scientist», che unisce le competenze dello statistico, del programmatore, del designer infografico e del narratore.

Gli statistici, i data base manager e gli esperti di apprendimento automatico saranno richiestissimi.

Big Data

Sta emergendo una nuova generazione di «angeli» ad esempio tra gli ex di Google, che insieme ad un gruppo ristretto di accademici sta finanziando le start up ad alta intensità di dati che oggi proliferano.

Stanno emergendo degli «intermediari di dati» che saranno in grado di raccoglierci da più fonti, aggregarli e impiegarli per scopi innovativi: i detentori di dati consentiranno agli intermediari di assolvere questo ruolo perché il valore di alcuni dati si può estrarre solo tramite la loro intermediazione.

Big Data

Qualche esempio di intermediari:

- Hitwise ha stretto accordi con vari provider per raccoglierne i dati sul Clickstream dei siti. I dati venivano concessi in licenza per una somma modesta mentre l'intermediario si accaparrava il grosso del valore.
- Quantcast misura il traffico in arrivo sui siti web per aiutarli a capire il profilo demografico dei loro visitatori.

Big Data

- Il produttore di motori per aerei Roll Royce, dal centro operativo situato in Gran Bretagna, monitora in continuazione la performance di oltre 3700 motori aerei in funzione in tutto il mondo per rilevare eventuali problemi prima che si verifichino dei guasti.
- I servizi così forniti generano attualmente un 70% dei ricavi annuali della divisione motori per l'aviazione civile.
- Oggi le automobili sono dotate di chip, sensori e software che inviano dati ai computer della casa produttrice tutte le volte che vanno in officina per una riparazione. Un meccanico si troverà a chiamare il cliente per avvisarlo che la sua macchina sta richiedendo un intervento direttamente a lui....

Big Data

- Sono sempre più numerose le città degli Stati Uniti che impiegano la sorveglianza preventiva: si usa l'analisi dei Big Data per selezionare le strade, i gruppi e gli individui da tenere particolarmente sotto controllo perché un algoritmo ha identificato in essi una più alta propensione alla criminalità. **Un importante utilizzo dei Big Data: la prevenzione del crimine.**
- Facebook è un generatore di Big Data che consentono di costruire profili di persone singole, utilizzando programmi di Intelligenza Artificiale, fortemente articolati e quindi utilizzabili per la selezione del personale.

Big Data

- I Big Data ci permettono di sperimentare più in fretta e di esplorare un maggior numero di opzioni. Questi vantaggi dovrebbero produrre più innovazione: ma la scintilla dell'invenzione sta in quello che i dati non dicono, almeno in modo diretto.....