

Dimmi cosa googli e ti dirò chi siamo

scritto da Alessio Guandalini, Andrea Fasulo, Marco Terribili | 27 Aprile 2018



I flussi delle ricerche su Google possono aiutare a predire, pressoché in tempo reale, fenomeni sociali ed economici rilevanti, come i consumi delle famiglie o l'andamento della disoccupazione in Italia.

Guandalini, Fasulo e Terribili suggeriscono che anche la statistica ufficiale potrebbe giovare dall'utilizzo di queste informazioni.

Google è ad oggi il più diffuso motore di ricerca al mondo e viene interrogato circa 3 miliardi di volte al giorno dagli utenti del world wide web. I dati relativi a queste interrogazioni vengono messi a disposizione del grande pubblico di Internet attraverso uno strumento chiamato *Google Trends*, che permette di conoscere la frequenza di ricerca di una, o più, parole chiave. I dati relativi a queste ricerche, la tendenza che esse seguono nel corso del tempo e la loro distribuzione geografica consentono di valutare come questi trends siano correlati con fenomeni socio-economici e, in alcuni casi, di predirli. L'Istituto Nazionale di Statistica stima in maniera cadenzata l'entità dei principali fenomeni socioeconomici italiani, tra i quali il PIL, l'inflazione, la disoccupazione e la spesa delle famiglie relativa all'acquisto di beni e servizi. Per realizzare tali obiettivi, l'Istat sta studiando modelli previsivi che, utilizzando le informazioni provenienti da Google Trends, consentano di migliorare le previsioni sul breve periodo (*nowcasting*) dei fenomeni d'interesse per la statistica ufficiale nazionale.

Come funziona Google Trends

Lo strumento messo a disposizione dal motore di ricerca permette agli utenti di studiare, attraverso una semplice interfaccia grafica, la frequenza di ricerca di una *keyword*. I dati forniti sono aggiornati in tempo reale e possono essere circoscritti ad un preciso periodo d'osservazione o ad una certa area geografica d'interesse; questo permette agli utenti più esigenti di indirizzare le loro analisi già in fase di acquisizione dei dati.

Google Trends non fornisce numerosità assolute sulla quantità di interrogazioni svolte ma le variazioni percentuali del numero di ricerche, rispetto alla settimana in cui si è registrato il massimo numero di ricerche per di una certa parola chiave. L'utente può studiare l'andamento di una o più parole chiave

contemporaneamente (al massimo 10).

Nella Figura 1, è riportata la serie storica di Google Trends per due parole chiave: “cerco lavoro” e “mutuo”. Alcuni studi hanno dimostrato che la keyword “cerco lavoro” è particolarmente legata al livello di disoccupazione, mentre “mutuo” – che è legata prevalentemente all’acquisto di una casa – fornisce una misura indiretta della fiducia delle condizioni economiche individuali (D’Amuri e Marcucci, 2012). L’andamento delle ricerche delle due keyword su Google è molto simile. Infatti, ai picchi negativi nelle ricerche di “cerco lavoro” sono associati i minimi nella ricerca di “mutuo”, sintomo della scarsa fiducia degli individui nelle proprie condizioni economiche.

Figura 1. Il grafico dei Google Trends relativi alle ricerche “cerco lavoro” e “mutuo” svolte in Italia negli ultimi 5 anni.

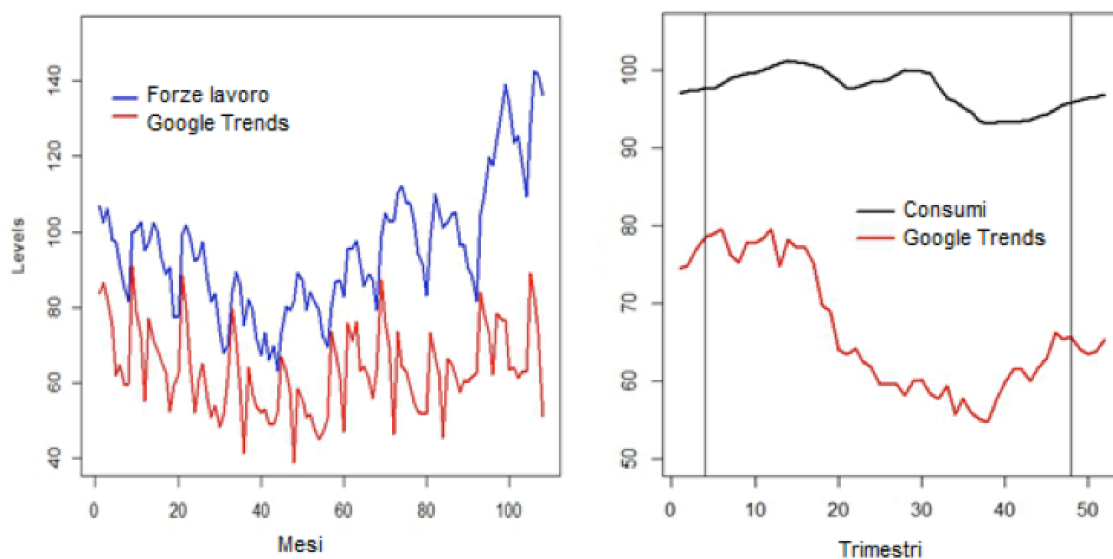


Le previsioni che sfruttano i Google Trends

Gli istituti nazionali di statistica, contestualmente alla pubblicazione di alcuni tra i principali indicatori statistici, rilasciano delle previsioni a breve periodo degli indicatori stessi. L’idea di fondo su cui l’Istat sta lavorando è quello di inserire nei modelli previsionali i dati provenienti da Google Trends, per ottenere previsioni affidabili e celeri. La velocità d’aggiornamento dei dati provenienti da Google rappresenta infatti un vantaggio non trascurabile quando si vuol fare nowcasting di un certo indicatore statistico. Questo è stato dimostrato da alcuni studi nazionali ed internazionali e, sulla base di questi lavori, sono stati studiati modelli previsivi, che sfruttano le informazioni derivanti dai flussi di Google Trends, per il tasso di disoccupazione (totale e giovanile) e per i consumi delle famiglie italiane (si veda D’Alò et al., 2015 e i riferimenti ivi contenuti). Nel primo caso la parola chiave di cui è stato osservato il trend è stata “cerco lavoro” mentre nel secondo caso è stato osservato l’andamento di un set di *keywords* che meglio possono descrivere la propensione di una famiglia al consumo di beni e servizi (“abbigliamento”, “bollette”, “finanziamento”, “hotel”, “IKEA”, “orologi”, “prestito”, “preventivo”, “ristrutturazione”, “voli”).

Scegliere le parole chiave richiede la massima attenzione: negli USA, per un problema di omonimia, il trend della keyword “jobs”, per lo studio della disoccupazione, salì vertiginosamente alla morte di Steve Jobs (Tkacz, 2013). La scelta di parole chiave molto specifiche, come quelle viste per la previsione del tasso di disoccupazione (totale e giovanile) e per i consumi delle famiglie italiane, preserva da erronei incrementi, dovuti da un aumento anomalo delle ricerche su Google, ma che non rispecchiano il reale andamento del fenomeno. (Figura 2)

Figura 2. Andamento delle stime delle Forze di Lavoro (mensili), delle Spese delle famiglie (trimestrali) e quello delle keywords che descrivono il rispettivo fenomeno d'interesse.



Il fascino delle stime tempestive

In entrambi i casi, si è osservato come l'utilizzo dei Google Trends abbia comportato una maggiore capacità previsiva dei modelli proposti e una loro migliore bontà di adattamento ai dati. I modelli, sfruttano dati sempre aggiornati e relativi ad un numero elevatissimo di utenti del web. Il nowcasting così ottenuto garantisce di colmare il gap informativo che intercorre tra il periodo di riferimento e la pubblicazione delle stime, offrendo all'utente finale stime affidabili che possono essere rilasciate in minor tempo.

L'uso di queste fonti, innovative e al tempo stesso attendibili, potrebbe riguardare lo studio di molti altri fenomeni economici, sociali e demografici che riguardano la popolazione e che la statistica ufficiale cerca di comprendere, da un punto di vista quantitativo, con tempestività e con la maggior esattezza possibile.

Riferimenti bibliografici

- D'ALÒ M., FALORSI S., FASULO A., 2015, Monthly unemployment rate prediction with Google Trends data: does Google search data improve the nowcast of Italian labour market?. *Big Data and the complexity of Labour Market Policies: New Approaches in Regional and Local Labour Market Monitoring for Reducing Skills Mismatches*, Larsen C., Rand S., Schmid A., Mezzanzanica M., Dusi S. (Eds.), Reiner Hamp Verlag., 99-114.
- FASULO A., GUANDALINI A., TERRIBILI M. D., 2017, Google Trends for nowcasting quarterly household consumption expenditure, *Rivista Italiana di Economia Demografia e Statistica*, 71 n. 1.
- D'AMURI F., MARCUCCI J., 2012, The Predictive Power of Google Searches in Forecasting Unemployment. *Banca D'Italia, Working Papers*, n. 891.
- TKACZ G. (2013). *Predicting recessions in real-time: Mining google trends and electronic payments data for clues.*