

Discussion of “The Effect of Big Data on Recommendation Quality. The Example of Internet Search”

Daniel Ershov

Toulouse School of Economics

March 29, 2018

Motivation

- Do some tech companies (i.e., Google) have an insurmountable competitive advantage due to having more data?
- Answer depends on how much your product improves with more data
- Some claim that data is a significant barrier to entry and competition
 - Google has an advantage (huge amounts of data) that keeps accumulating
- Google claims that you don't need that much data to compete
 - But you do need to invest in order to use the data more effectively
 - Contrasts itself with Bing - also lots of data, but arguably worse search results

Question and Findings

- Research Question: What is the effect of more data on the quality of internet search results?
- Use a month of anonymized Yahoo search engine queries
- Construct a quality measure - share of consumers who click on top link
- Regress this measure on the number of query searches
- More query searches are associated with increased quality
- In preferred specification effect is driven by searches with cookies (user data)

Big Picture Thoughts

- I like the paper
- Obviously high policy relevance
- Part of a nascent literature of the effects of data on markets
 - Useful reference: Bajari, Chernozhukov, Hortacsu and Suzuki (2018)

Comment (1)

- Some concerns about the quality measure
- Potentially large heterogeneity across queries
- Some queries have obvious answers that are “discoverable”
- Other queries are inherently ambiguous
- For example...

Comment (1)

YAHOO!
FRANCE

joe pesci


Rechercher


À tout moment
Hier
Semaine dernière
Mois dernier


Sur tout le Web
En français
En France


Joe Pesci — Wikipédia
fr.wikipedia.org/wiki/Joe_Pesci En cache
Biographie Jeunesse. D'origine italienne, **Joe Pesci** commence sa carrière d'acteur dès son plus jeune âge, en apparaissant au milieu des années 1950 dans la série ...

Joe Pesci - Résultats vidéo


Guy Movie Hall of Fame: Casino - youtube.com 3:48


Joe Pesci and Robert De Niro youtube.com 2:03


Wiseguy dailymotion.com 4:04


HOW I SECURED THE PART in youtube.com 3:57

Plus de vidéos pour Joe Pesci

Joe Pesci : Filmographie - AlloCiné
www.allocine.fr/personne/fichepersonne-2875/filmographie En cache
Quelle est filmographie de **Joe Pesci**? Découvrez tous les films et séries de la filmographie de **Joe Pesci**. De ses débuts jusqu'à ses projets à venir.

Joe Pesci - IMDb
www.imdb.com/name/nm0000582/?nmdp=1 En cache
Joe Pesci, Actor: Goodfellas. Compact Italian-American actor **Joe Pesci** was born February 9, 1943 in Newark, New Jersey, to Mary (Mesce), a part-time barber, and ...

- What about a more flexible version of the quality measure?
Say, clicks on top 2 or 10 sites?

Comment (2)

- Some identification concerns
- Things can happen relatively quickly on the internet
- Even within the bandwidth (and in a month), coefficients could be downward/upward biased by changing popularity of some websites and increasing/decreasing searches for those websites
- Searches with cookies are not randomly assigned - effect could be due to selection of consumers of a certain type who search certain queries with cookies
- Controlling for “initial quality” likely not enough to account for heterogeneity
 - Query/time fixed effects?
 - Still some problems with fundamental heterogeneity of queries

Comment (3)

- Everyone (including Google) agrees that data matters
- Question is how much does the complementary algorithmic technology matters
- The Microsoft/Bing event seems to be important
- Variation in search technology
- Not clear what happens exactly - why does quality measure fall?
- Should be looked at in more detail

Other Thoughts

- Some evidence of diminishing returns in the more flexible specifications
 - Negative interaction between cookies and “initial quality”/total searches
 - This is what Bajari et al (2018) find
 - Should be quite relevant for policy
- Is there learning across queries? Some sort of aggregate effects?
 - i.e. “Joe Pesci” and “Joe Pesci movies”
 - i.e. “Joe Pesci” and “Home Alone”
- More non-linearity? To allow for compounded learning, or stronger diminishing returns
 - If only care about fit, maybe ML more appropriate than linear regression model