

Predicting consumer behavior with web search

Sharad Goel, Jake M. Hofman, Sébastien Lahaie, David M. Pennock, Duncan J. Watts



Motivation

- Hypothesis: users browse online about stuff they might do in near future, example- watch a movie, buy a stock, or a vacation to LA.
- Study goal: find the correlation between user online search and future behavior.
- Past work: Ettredge et al. predicted unemployment rates in 2001-03, Cooper et al – Cancer, Eysenbach- Canadian flu, Polgreen et al-Influenza, Google flu trends
- Importance: prediction performance is relative.



Method

Data collection:

data for online search regarding movies, video games and song ranking is collected

Movie (IMDb)

Production budget :range - \$1.5M- \$250M
Opening screens: 1-4325
Movies: 119
HSX estimates

Video Games (VGChartz)

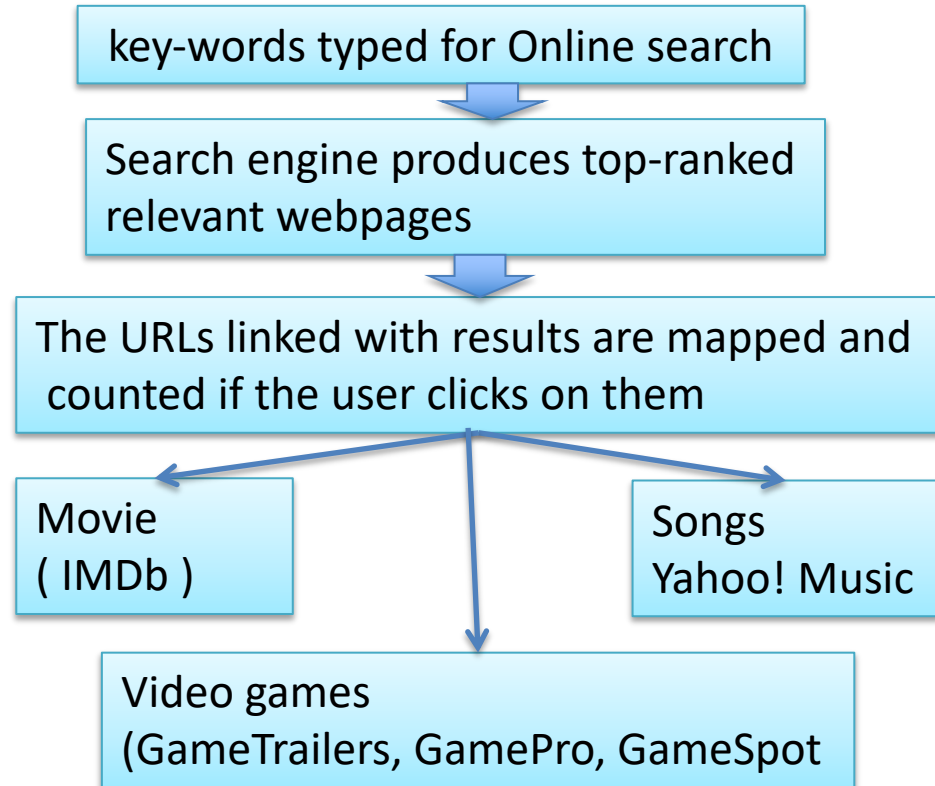
Sales: \$4K-\$2M
Critic ratings: 3.1-9.5

Music : Hot 100 Charts by Billboard:June - Aug 2009

Artist, Song title, Rank

Data analysis:

Linear regression is performed using three models: Baseline model, online search data, and augmented model (data from internal sources + online search)



Take-away

Learning:

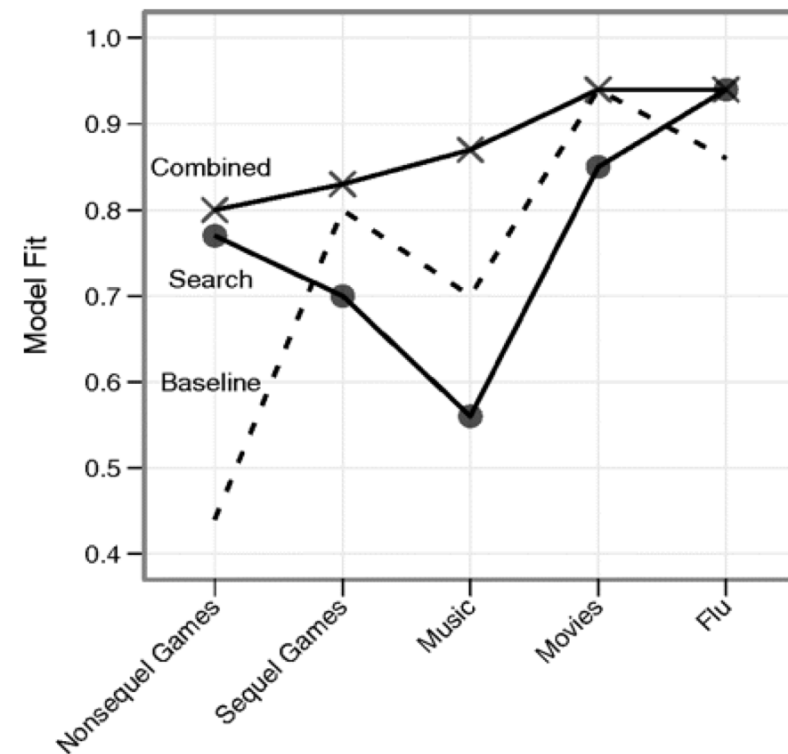
- The paper showed how search based predictions perform
- Alternative information sources perform equally well or even better than search data
- Search data is useful when key indicators are unavailable

Loop-holes:

- Amount of data available for analysis varies
- Online search can be misleading: user looking for lyrics might be counted as a future buyer for album
- User might put wrong key-words

Future applications:

- Financial analysis: small gains are also helpful
- Search data can be collected across many domains, merged, and analyzed in real time.



(% accuracy)	Baseline model fit	Combined model fit
Movie	Same	
Video Game	0.44	0.8
Music	0.7	0.87