Raccoon: A Query System for Social Media Signals

Dolan Antenucci, Michael R. Anderson, Michael Cafarella
University of Michigan

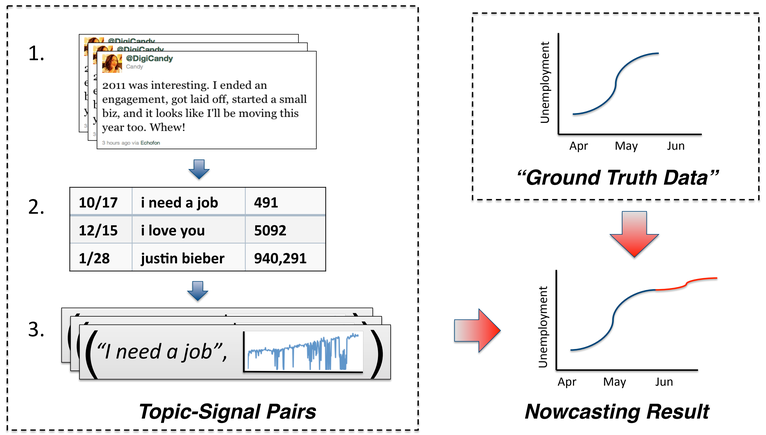
Motivation: Forecasting the Present via Social Media

Real world phenomena like unemployment and disease behavior have been estimated using social media – a process known as ***nowcasting*** [1].

These estimates are **cheaper** and **faster** than traditional data collection methods like phone surveys.

Faster and cheaper means **saved money** and **better policy**.

Problem: Nowcasting Requires Ground Truth Data

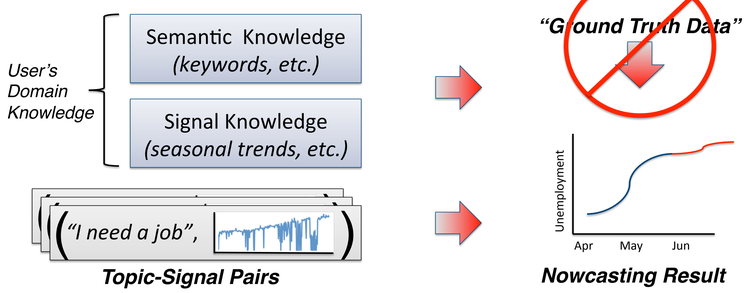


Yet, economists [3] still want estimates for targets lacking ground truth data:

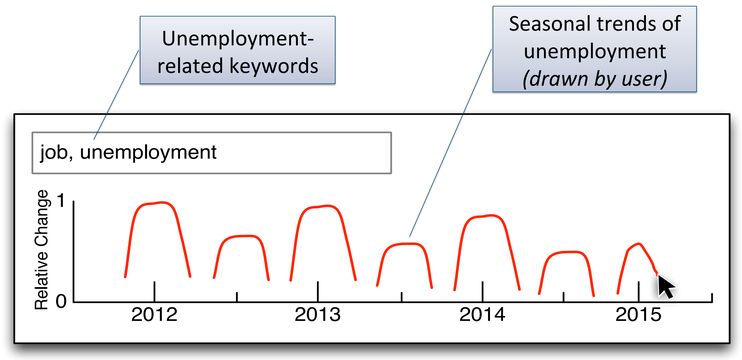
- 10-day auto sales (once used, but no longer available [2])
- Physical movement (e.g., out of parents house, for job)
- *etc.*

Solution: Substitute Domain Knowledge for Ground Truth

Our target users have some ***semantic*** and ***signal*** domain knowledge about their target phenomena:

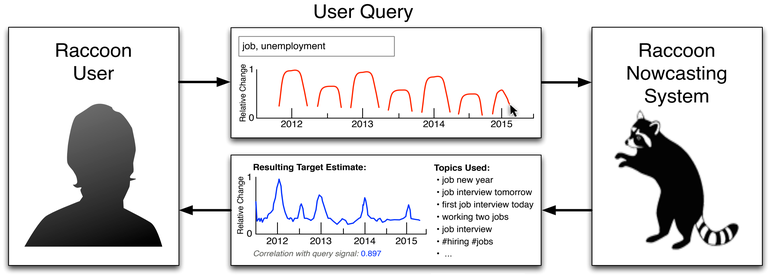


Example User Query (Target: *Unemployment Behavior*)



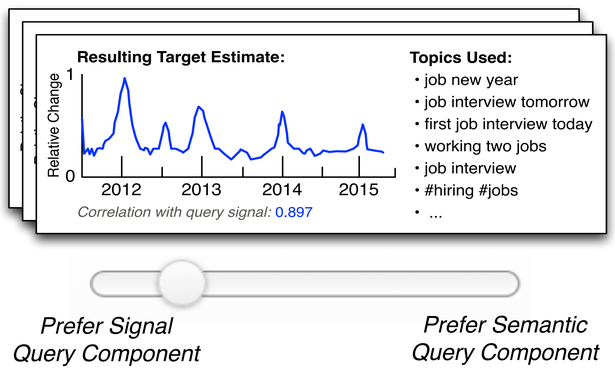
User Interaction Loop

Like a search engine, **users iteratively submit queries:**

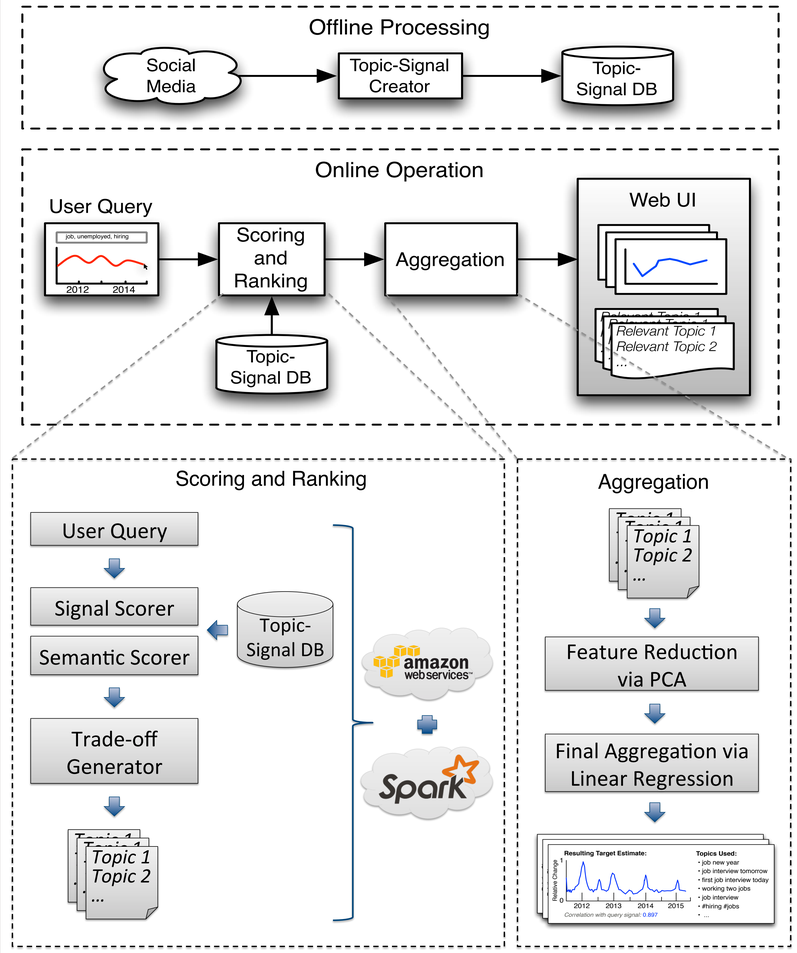


Users Given Trade-Off Between Query Components

Users can **explore a set of query results in real-time**, produced by varying the relative influence of the semantic and signal query components:



Architecture



Scalability with the Cloud

Current Prototype System:

- 40 billion tweets *(collected 2011-2015)*
- 150 million topic-signal pairs *(after threshold filtering)*
- Query processing runtime: ~20s *(on 10 EC2 c3.8xlarge servers)*
- Topic-signal pairs support daily updating

Related Work

[1] S. L. Scott and H. Varian. Bayesian variable selection for nowcasting economic time series. In *Economics of Digitization*. University of Chicago Press, 2014.
[2] A. Greenspan. *The Age of Turbulence*. Penguin Press, 2007.
[3] D. Antenucci, M. Cafarella, M. C. Levenstein, C. Ré, and M. D. Shapiro. Using social media to measure labor market flows. Working Paper 20010, National Bureau of Economic Research, March 2014.

dol@umich.edu ◆ http://web.eecs.umich.edu/~dol