January 14, 2013

Twitter

Speaker: Jake Mannix: Tech Lead in Interest Modeling

Wendy Lee, send resumes to:

Recruiter: [wendylee@twitter.com](mailto:wendylee@twitter.com)

Apply for jobs at twitter.com/jobs/university

**Personalization, and recommender systems**

Recommend:

* Who to follow (40-50% of followings come throught the widget)
  + Try to 7-10 for new users
  + Importing contacts can really help with topics
  + Other classified topics can’t be talked about
* Webdocs (#discover): A snippet of a web page
* Tweets (Ads: Promoted Tweets)

Personalization:

* You can create the top results of the day in a relatively chronological and then map reduce that for each user based on their interests
* Search for people who you don’t know and who you already follow but don’t remember the name

Inputs:

* Follow graph (>20B edges, 40% bidirectional)
  + Direct follows
    - Giving biased search results similar to who you follow generally gives better results
  + Extended Network
* Interest graph (topical classification + clicks/RT/favorites): Interests tend to be long term for users
  + Text classifiers
    - Define topic hierarchy
      * Sports -> football
      * Politics -> US Elections
    - How many do you need?
      * Depends on engagement metrics, not necessarily more = better
    - How do you do topical classification
      * Collect labeled training data
      * Mechanical Turk
  + p(engagement | interest) of user + Topic of (target item) + Historical engagement data = Can create a topic-based personalized ranker
* Raw engagement (traditional collaborative filtering)

Personalized PageRank: Take random walks along the list and create clusters. Find the most popular groups and reset the vertex at the user index. Pick 20 topics and find the top tweets and engagements

90% of employees are doing machine learning

Does web crawling ~3 layers deep