Discovering Topical Experts in Twitter Social Network

Naveen Kr. Sharma
IIT Kharagpur

Advised By

Niloy Ganguly
IIT Kharagpur

Krishna P. Gummadi
MPI-SWS
Outline

- Introduction - Motivation
- Background - State-of-the-art
- Our Methodology
  - Twitter Lists
  - Quality of Inference
- Cognos Search System
  - Ranking Experts
  - Evaluating our results
- Summary - Publications
- References
Introduction

Twitter Online Social Network

- Primarily an “interest” based social network
- Allows users to send and read text-based posts of up to 140 characters, informally known as “tweets”
- A nice analogy: A **very very large** hall full of people talking about **all sorts of things**

How to find someone who's talking about a topic of my interest?
Challenges

- Very very Large System
  - 200 million active users.
  - 150 million tweets everyday.
  - ~2000 tweets per second !
  - Cannot check each and every one out there!
- Real-time nature of Twitter
  - Things propagate very fast
- First step to retrieving “relevant” content
- Finding “trustworthy” content
State-of-the-art

- Twitter's Official "Whom To Follow"
  - Searches the name and bio for the given query term
  - Gives heavy preference to "Verified Users"

- Pros
  - Works extremely well if query is a popular term (politics, music..)
  - If there exists a "verified" expert, he/she will certainly show up.

- Cons
  - Only 11,186 Verified Users, and a very Cryptic process.
  - If query outside this domain, almost zero meaningful results.
State-of-the-art Research

- 17 Features extracted from the Twitter graph.
- Content of the tweets posted by the users.
  - Number of original tweets.
  - Number of mentions by others.
  - Number of topically active friends/followers
- Extremely computation intensive.
- Results for only 3 topics known publically.
Twitter “Lists”

- Provide a way to organize the people you’re following on Twitter.
  - Similar to any other “groups” feature.
- Anybody can create lists (name, description).
- Add any other Twitter user to the list.

Why create a list?

- Tweets from all people within the list, available as a separate stream.
How Lists work?

mashable's lists

@mashable/news
A curated list of news organization's Twitter accounts.

@mashable/tech
Experts and sources to keep up with the latest in tech.

@mashable/design
Tweets and tips from designers.

@mashable/food
Love food? Here are chef's, cooks and others in food to follow.

@mashable/celebrity
Celebrities on Twitter.

@mashable/journalism
Journalists interested in the future of news media.

@mashable/music
Musicians on Twitter.
Annotating people using Lists?

- If someone is an “expert” in specific area,
  - Decent number of followers (well known)
  - He/She is added to “lists” pointing to “area” of expertise
Extracting meaningful “Tags”

- Crawl “Lists” information (name, description)
- Apply language processing techniques:
  - Case-folding
  - Stemming
  - Removal of stop words
- Split CamelCase words. (e.g. TennisPlayers)
- Identify nouns/adjectives using a POS-tagger.
- We consider only unigrams and bigrams as topics.

A set of topics for each user, along with frequency.
Sample “Tags”

- **Music, Celebrity, Friend, Love, Artist**
  - Britney Spears
    - @britneyspears
    - Los Angeles, CA

- **Sport, Cycle, Celeb, Athlete, Bike**
  - Lance Armstrong
    - @lancearmstrong
    - Austin, TX

- **News, Media, World, Update, Break**
  - CNN Breaking News
    - @cnnbrk
    - Everywhere

- **Cricket, Sport, Celeb, Indian, Journalist**
  - Harsha Bhogle
    - @bhogleharsha
    - Mumbai, India
Comparison with Twitter WTF

- We obtained the top 20 experts returned by the Twitter WTF service for a few hundred queries.
- Out of the 3495 users, the topics inferred include the query for 83.4% (2916) of the users.

- Query – Dining
  - User – dineLA (official Twitter account of dineLA)
  - Tags - restaurant, food, los angeles, chefs

- Query – Astrophysics
  - User – JimmyFallon
  - Tags - celebs, comedy, funny, actors, humor
Comparison with Pal et. al.

- The paper lists the top 10 experts for three specific topics – **iPhone**, **oil spill** and **world cup**.
- Our methodology could recall -
  - 8 out of 10 for “iphone”,
  - 7 out of 10 for “world cup”, and
  - 6 out of 10 for “oil spill”.

- **Oil Spill**
  - Reuters: latest news from around the world
  - TIME: Breaking news and current events

- **World Cup**
  - Nikegoal: marketing, music, education, sport
  - Channel4news: exclusive stories & breaking news
Ranking Experts

- PostgreSQL Full Text Search
- `ts_vector` representation with **inverted index**. For example, "fat cats ate fat rats" stored as “ate:3 cat:2 fat:1,4 rat:5”
- First stage, similarity with query
  - We compute the **cover density ranking** for the given document vector and query.
- Second stage, weighting with number of lists
  - To give more weightage to users that have been “listed” more number of times.
- Final Score: \( \text{cover\_density}(v, q) \times \log(\text{lists}) \)
### Relevance Evaluation

<table>
<thead>
<tr>
<th>Website</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bbcscitech</td>
<td>BBC SciTech set up by @mario, supported by backstage.bbc.co.uk</td>
</tr>
<tr>
<td>Reuters_Science</td>
<td>Reuters Science News, From newly charted astronomical anomalies at the far reaches of the universe to the rise of nanotechnology, nobody covers science like Reuters.com.</td>
</tr>
<tr>
<td>newscientist</td>
<td>New Scientist is the world’s leading science and technology weekly</td>
</tr>
<tr>
<td>NatGeo</td>
<td>National Geographic, Since 1888, we’ve traveled the Earth, sharing its amazing stories with new generations. Official Twitter account of National Geographic.</td>
</tr>
<tr>
<td>science</td>
<td>Science news from ScienceNewsBlog.com</td>
</tr>
<tr>
<td>NASA</td>
<td>NASA</td>
</tr>
</tbody>
</table>
Relevance Results

- $1680/2186$ (78.7%) were judged relevant.
- 92.1% were judged relevant at least once.
Direct Comparison

Please compare the anonymized search results for "sports journalist"

- A is better
- B is better
- Both are equally good
- Both are equally bad

Search Engine A
- seaningle: Sean Ingle
  Sports editor of guardian.co.uk. Father of Football Weekly, editor of the Fiver, occasional writer.

- Pearcesport: James Pearce
  BBC Sports News Correspondent. This is an official @BBCSport account. Find our accounts at @BBCSport/Following

- owen_g: Owen Gibson
  guardian sports news correspondent

- JacquiOatley: Jacqui Oatley
  Football reporter/commentator for BBC Radio 5 Live/TV & the Premier League’s worldwide TV channel (IMG). Proud mum to Fraser (Mercury ode)

Search Engine B
- AdamSchefter: Adam Schefter
  ESPN NFL Insider.

- terrellowens: Terrell Owens
  Wide Receiver #81 Terrell Owens

- sportsguy33: Bill Simmons

- Pearcesport: James Pearce
  BBC Sports News Correspondent. This is an official @BBCSport account. Find our accounts at @BBCSport/Following

Direct side-by-side comparison with Twitter's Search Engine
Comparison Results

• Out of the 27 queries asked at least twice
  • 12 – Cognos was better!
  • 11 – Twitter WTF service better.
  • 4 – Tie

• Cognos - Linux, computer science, mac, India, Apple, Facebook, internet, ipad, markets, windows phone, photography, politic journalist

• Twitter WTF - politic news, music, Sachin Tendulkar, Twitter, Alka Yagnik, Anjelina Jolie, cloud computing, Delhi, Harry Potter, metallica

• Tie - Microsoft, Dell, Kolkata, Sanskrit as an official language
Summary

- We propose a methodology for finding topic experts using a recently introduced Twitter feature called Lists.
- We describe a novel method for inferring attributes that characterize individual Twitter users.
- Based upon this methodology, we build and deploy Cognos, a topical expert search system.
- We come up with an efficient solution to keep our Cognos system up-to-date.
- Cognos built using only the Lists feature, can compete with the who-to-follow search system deployed by Twitter.
Publications

- **Inferring Who-is-Who in the Twitter Social Network**
  In ACM SIGCOMM, Workshop on Online Social Networks 2012, Helsinki, Finland.

- **Cognos: Crowdsourcing Search for Topic Experts in Microblogs**
  In SIGIR 2012, Portland, USA. (under submission)
References

[1]. http://www.twitter.com, Twitter Online Social Network

[2]. https://dev.twitter.com/docs/api, Twitter API


