**Finding Influencers within Fuzzy Topics on Twitter**

Tal Stramer

---

**Training Data**

*limited to Twitter activity between 11/1/15 and 11/7/15*

1. Top 200 search queries by US searches e.g. “Adele” “Trump”
2. Counts of search queries contained in verified user’s tweets: e.g. vit_1 - “Adele”: 2, “Trump”: 1
3. Favorite, retweet, and reply counts for each user’s tweets.

- Total of ~190,000 tweets and ~5,000 users

---

**Detecting Fuzzy Topics**

- Detect topics through a variant of LDA
- Each user is a “bag-of-words” of search queries contained in their tweets
- Each topic is a list of search queries
- Model user as a multinomial distribution over topics
- Model topic as a multinomial distribution over search queries
- Remove search queries ranked in top 20 of more than 3 topics (likely too general) and re-run LDA e.g. “christmas”, “halloween”, “likes”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Greg Hardy</td>
<td>Trump</td>
<td>Royals</td>
<td>#websummit</td>
<td>Sharm</td>
</tr>
<tr>
<td>Ken Whisenhunt</td>
<td>Ben Carson</td>
<td>Mets</td>
<td>Star Trek</td>
<td>John Lewis</td>
</tr>
<tr>
<td>Vernon Davis</td>
<td>West Point</td>
<td>Harvey</td>
<td>Black Ops 3</td>
<td>#bonfirenight</td>
</tr>
<tr>
<td>Steve Smith</td>
<td>Bernie Sanders</td>
<td>Jonny Gomes</td>
<td>Warcraft</td>
<td>#millionmaskmarch</td>
</tr>
<tr>
<td></td>
<td>Fred Thompson</td>
<td></td>
<td></td>
<td>Guy Fawkes</td>
</tr>
</tbody>
</table>

---

**Detecting Topic Influencers**

- Use tweet engagements to model how influential a user is for each topic
- Tweet engagements tend to have a long tail (see right). However, we can model tweet engagements as being generated from a normal distribution given a ‘influencer’ level (1 to 4)
- Use EM to estimate parameters of gaussian mixture model + predict influencer level given engagements.
- Calculate a topic influencer score for each user and topic using the following heuristic:

  score = 0
  for each tweet by user:
  for each query in tweet:
    score += topicQueryProb(topic, query) * estimateInfluencerLevel(query, totalEngagementsOnTweet)

---

**Predicting Future Influencer Levels**

- In the real world, we don’t know engagements on tweets that have just been posted, However, we can use historical influencer levels for a user and topic to predict influencer levels on a new tweet by a user
- Train a multi-class SVM classifier that predicts influencer levels for a tweet
- Training set - for each query in each tweet, generate:
  - X = (userTopicProb, topicQueryProb, aveInfluencerLevelByUserForQuery)
  - Y = (influencerLevelByUserForTweet)
- Run 10-fold cross validation to estimate generalization error
  - Error rate = 7.6%. 

---