Microblog Track

Jaime Arguello
jarguell@email.unc.edu

February 8, 2016

Track Goals

- Streaming task: evaluate systems that monitor a stream of social media updates (public tweets) and "push" content in response to an interest profile (query).

    ‣ relevant, novel (not redundant), and timely

Track Goals

- **Scenario A:** at most 10 tweets per day ASAP

- **Scenario B:** at most 100 tweets per day at midnight (batch mode).

Topics

<top>
<num> Number: MB297

<title>
Wimbledon Tennis Tournament

<desc> Description:
Find reactions to the latest Wimbledon Tennis Tournament.

<narr> Narrative:
The user wants to follow commentary about the Wimbledon Tennis Tournament. Relevant tweets include opinions/assessments of the competitors, refereeing controversies, and viewer and attendee experiences.
</top>

4

Relevance Assessment

- Retweets: relevance label was propagated to all retweets and the original tweet.

- Clustering: tweets were assigned to topical clusters in order to punish redundant tweets

  ‣ only the first relevant tweet from each cluster received credit.

Evaluation Metrics

- 10-day evaluation period

- For each topic profile, average performance across days in the evaluation period

- For each run, average performance across profiles

Evaluation Metrics
scenario A

- Daily gain, discounted by the difference (in minutes) between the time the tweet was published and the time it was predicted to be relevant (i.e., "pushed")

- Gain = 0, 1, 2

- Redundant tweets = 0

$$\sum G(t) \times \text{MAX}(0, \frac{100 - d}{100})$$

Evaluation Metrics
scenario B

- NDCG@10:

$$\frac{1}{\mathcal{Z}} \left(G(t_1) + \sum_{i=2}^{10} \frac{G(t_i)}{\log_2(i)} \right)$$

Evaluation Metrics
edge cases for scenarios A and B

- If an interest profile had relevant tweets for the day

  ▸ if the system returned tweets, give day score as usual

  ▸ otherwise, give day score of 0

- If an interest profile had **<u>no</u>** relevant tweets for the day

  ▸ if the system returned tweets, give day score of 0

  ▸ otherwise, give day score of 1

9

Challenges
scenarios A and B

- **Filtering:** removing non-english tweets, tweets with lots of hashtags and URLs, tweets with lots of stopwords

- **Topic profile representation:** selecting the important terms from the title, description, and narrative fields based on TF.IDF-like measures

- **Tweets representation:** tokenizing the tweet text (http://www.cs.cmu.edu/~ark/TweetNLP/)

- **Query expansion:** adding additional terms to the query in order to increase recall

- **Novelty:** filtering tweets that are similar to a previously "pushed" tweet

Challenges
scenarios A and B

- **Topic drift:** modeling changes in the "relevant" vocabulary over time

- **Push notifications:** predicting whether to "push" a highly scoring tweet without knowledge of future tweets

University of Waterloo
scenarios A and B

- Topic profile representation:

  ‣ select top-10 terms from title, description, and narrative

  ‣ select 10 terms and 5 hashtags from pseudo-relevant tweets

  ‣ term weighting (point-wise KL divergence):

$$P(w|\mathcal{Q}) \log_2 \left(\frac{P(w|\mathcal{Q})}{P(w|\mathcal{C})} \right)$$

University of Waterloo
scenarios A and B

- **Ranking:** different weights for terms originating from the title, description/narrative, and expansion

$$\frac{N_t}{|T|} \sum_{i=\{t,n,e\}} w_i \times N_i$$

scenarios A - push notifications

- Fix time window:

  ▸ profile-specific threshold = retrieval score of 50th ranked tweet from previous day

  ▸ every 100 minutes, push top-scoring tweet if: (1) above threshold and (2) sufficiently different from previously pushed tweets

scenarios A - push notifications

- Dynamic emission:

  ‣ *k1* = score of 5th ranked tweet from previous day

  ‣ *k2* = score of 10th ranked tweet from previous day

  ‣ dynamic time window

  ‣ if *score* > *k2*: push tweet, re-start window time, clear heap

  ‣ if *k2* > *score* > *k1*: add tweet to heap

  ‣ if end of window: push stop-scoring tweet, re-start window time, clear heap

15

University of Delaware
scenarios A and B

- Topic profile representation:

  ‣ topic, description, narrative

  ‣ text from pseudo-relevant web results

  ‣ text from pseudo-relevant tweets

  ‣ hashtags from pseudo-relevant tweets

  ‣ authors of pseudo-relevant tweets

University of Delaware
scenarios A and B

- Ranking:

    ‣ combine TF.IDF scores between tweet and different parts of the interest profile

    ‣ add text from predicted relevant tweets to the interest profile

- Push notifications:

    ‣ single threshold across all interest profiles

- Novelty:

    ‣ on-line clustering of top-scoring tweets based on cluster-similarity threshold

Peking University
scenarios A and B

- Topic profile representation:

  ‣ TREC topic text and text from top-5 google results (titles and snippets)

  ‣ combined using linear interpolation

- Ranking:

  ‣ Query-likelihood model

  ‣ Linear combination of two smoothing techniques

- Push notifications:

  ‣ profile-specific threshold = retrieval score of 10th ranked tweet from previous day

18