Event Recommendation using Social Media Signals

Dr. Maunendra Sankar Desarkar
Department of CSE
IIT Hyderabad

IIT-H and RIKEN-AIP Joint Workshop on Machine Learning and Applications
Events: Different interpretation

• What is an event?
Events: What do we mean
Important factors for recommendation

• **Local information**
  • User interest in different (known/unknown) categories

• **Global information**
  • Popularity
  • Can be obtained from feedbacks
  • Easy to get for movies, music, books, news, ...
  • *Can the same thing be done for events?*
Challenges in planned event recommendation

• How to get popularity information?
  • Can not wait for feedback after the event
  • Alternative ways?
  • Assumption: Social media
Using social media: the motivation

• Lots of user generated contents in social media
• If an event generates a lot of discussion in social media, then it might be popular
  • Always?
• Given an Event E, predict the future popularity of the event. Also, develop an event recommendation algorithm that uses this predicted popularity as a feature to recommend future events to the users.
Usefulness of Event Popularity Estimation

• Event popularity estimation and recommendation

• How does it help?
  • Assisting event organizers with outreach
  • Assisting civic authorities for traffic planning
  • Helping users to know about the upcoming events
Event Recommendation using social media

Identify events

Identify social media contents

Eliminate noisy contents

Predict popularities

Identify emotions and intensities

Event aggregation sites such as Eventbrite, Eventful, last.fm

... shortcuts, code-mixing data, lot of typos, aggressive contents, spam

Using context features of the event, features of the extracted data etc.

Recommen...
Identifying Social Media Contents

• Given an event E retrieve all relevant tweets related to the event

  <topics>
  ...  
  <topic>
    <id> 1 </id>
    <title/>
    <artist> Anna calvi </artist>
    <festival> charrues </festival>
    <startdate> 16/07/15 - 18:45 </startdate>
    <enddate> 16/07/15 - 19:45 </enddate>
    <venue> Kerouac </venue>
  </topic>
  ...
  </topics>
## Example tweets related to planned events

<table>
<thead>
<tr>
<th>Tweet</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ready for the show @organicbananas1 live #Trans2015 @TransMusicales festival vielleelectro <a href="https://t.co/p7ZaJqQiQg">link</a></td>
<td></td>
</tr>
<tr>
<td>Looking for a qualifier run for Airtel Hyderabad Marathon (AHM)? Come, participate in #Whitathon2019, now a qualifier run for Airtel Hyderabad Marathon.</td>
<td></td>
</tr>
<tr>
<td>I'm already blown away by the @TransMusicales festival and it hasn't even really started. The venue alone is mind blowing.</td>
<td></td>
</tr>
<tr>
<td>This is in reference to the SPIC MACAY VIRASAT series being held at IIT Bombay from 10th-12th March 2019. I am sharing the program schedule for the series as below for your ready reference.</td>
<td></td>
</tr>
<tr>
<td><strong>Smart India Hackathon</strong> is a non-stop product development competition, where problem statements are posed to technology students for innovative solutions. Register yourself to participate in Smart India Hackathon 2019. For more details, visit <a href="http://aicte-india.org">www.http://aicte-india.org</a> @HRDMinistry</td>
<td></td>
</tr>
</tbody>
</table>
Identifying Relevant Hashtags for Planned Events

• If we know hashtags, then those can be used to pull relevant tweets
• By using hashtags we can identify the topic of the discussion.
  • E.g. #iPhoneXLaunch to iPhone X Launch event, #rio2016, #rio to the Rio Olympics 2016, #InternationalYogaDay2018 for International Yoga Day 2018 etc.
• However, manual selection of these hashtags is **not a scalable** approach.

• **Hashtag Identification:** Given metadata of an event E, find a list of hashtags relevant for the event E.

Sreekanth Madisetty, Maunendra Sankar Desarkar: *Exploiting Meta Attributes for Identifying Event Related Hashtags*. 9th International Conference on Knowledge Discovery and Information Retrieval (*KDIR 2017*), Madeira, Portugal.

Identifying relevant hashtags – two step approach

**Phase 1:** Retrieve a set of candidate hashtags for an event from Twitter.
- Precision Query to Twitter
- Pool results to build candidate hashtag set

**Phase 2:** Rank the hashtags from this candidate set according to their relevances with the event.
- Supervised approach
- Identify features for <event, hashtag> pair
- Take weighted combination of feature scores to predict relevance
- Use RankSVM to learn the weights
The pipeline

- Event Metadata (EM)
- Post collection (e.g. Twitter)
- Machine Learning Model
- Identify relevant Hashtags
- Top-k Hashtags

Build Precision Query

Posts

Candidate Hashtags
### List of features for (event, hashtag) pair

<table>
<thead>
<tr>
<th>Feature</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>Frequency of Hashtag</td>
<td>Computes the frequency of the hashtag in tweet corpus of event $E$</td>
</tr>
<tr>
<td>$f_2$</td>
<td>Bigram Feature</td>
<td>Number of common character-level bigrams present in the hashtag $HT$ and event metadata $EM$</td>
</tr>
<tr>
<td>$f_3$</td>
<td>Trigram Feature</td>
<td>Number of common character level trigrams present in the hashtag $HT$ and event metadata $EM$</td>
</tr>
<tr>
<td>$f_4$</td>
<td>Bigrams of Abbreviated Title</td>
<td>Number of common character level bigrams in the hashtag $HT$ and abbreviated title $T$</td>
</tr>
<tr>
<td>$f_5$</td>
<td>Trigrams of Abbreviated Title</td>
<td>Number of common character level trigrams in the hashtag $HT$ and abbreviated title $T$</td>
</tr>
<tr>
<td>$f_6$</td>
<td>Bigrams of Top-K trigrams</td>
<td>Number of bigrams that are common in both hashtag $HT$ and Top-K word level trigrams of tweet corpus of an event $E$</td>
</tr>
<tr>
<td>$f_7$</td>
<td>Subsequence Feature</td>
<td>Checks whether hashtag $HT$ is a subsequence of event metadata $EM$ or not</td>
</tr>
<tr>
<td>$f_8$</td>
<td>Substring</td>
<td>Tests whether hashtag $HT$ is a substring of event metadata $EM$ or not</td>
</tr>
</tbody>
</table>
Learning the weights

\[
\text{minimize } \frac{1}{2} w^T w + C \sum_{i,j,k} \varepsilon_{i,j,k}
\]

\(\forall k\) and \(i \neq j \in \{1, \ldots, n_k\}\) with \(h_{ki} >_{E_k} h_{kj}\)

\[
w^T \Phi(E_k, h_{ki}) \geq w^T \Phi(E_k, h_{kj}) + 1 - \varepsilon_{i,j,k}
\]

\(\varepsilon_{i,j,k} \geq 0\)
### Dataset

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of events</th>
<th>Tweet volume (in Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euro Cup 2016</td>
<td>51</td>
<td>14.4</td>
</tr>
<tr>
<td>Celebrity Birthdays</td>
<td>10</td>
<td>0.97</td>
</tr>
<tr>
<td>Festivals</td>
<td>5</td>
<td>1.62</td>
</tr>
<tr>
<td>Movie Launches</td>
<td>13</td>
<td>1.70</td>
</tr>
<tr>
<td>International Days</td>
<td>11</td>
<td>2.58</td>
</tr>
<tr>
<td>Politics and Governance</td>
<td>4</td>
<td>0.62</td>
</tr>
</tbody>
</table>

- **94 Events, 21.37 Million tweets**
- For each event, *most frequent hashtags* were identified and assigned a *relevance label* in \( \{0, 1, 2\} \)
Subjective Results: Example hashtags Retrieved

<table>
<thead>
<tr>
<th>Event</th>
<th>FreqPearson</th>
<th>Proposed Method</th>
<th>Hashtags retrieved by our method but missed by other method</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Film Awards (Award Ceremonies)</td>
<td>#nationalfilmawards, #rustom, #nationalaward, #24themovie, #akshaykumar, #neerja, #bestactor, #nationalawards, #dangal, #zairawasim</td>
<td>#nationalfilmawards, #64thnationalfilmawards, #nationalfilmaward, #nationalfilmawards2017, ##nationalfilmawards, #64nationalfilmawards, #nationalawards, #nationalawardsindia, #64thnationalfilmaward</td>
<td>#64thnationalfilmawards, #nationalfilmawardsindia</td>
</tr>
<tr>
<td>Flipkart Big Billion Days (E-commerce Events)</td>
<td>#bigbilliondays, #shoponbigbilliondays, #flipkart, #greatindianfestival, #mobilesonbigbilliondays, #bbd, #fashion, #unboxingdiwalibestoffers, #unboxdiwalsale, #amazon</td>
<td>#bigbilliondays, #shoponbigbilliondays, #mobilesonbigbilliondays, #bigbilliondays2016, #flipkartbigbillionsale, #electronicsonbigbilliondays, #thebigbilliondays, #bigbilliondaysareback, #bigbilliondaysneakpeek</td>
<td>#bigbilliondays2016, #flipkartbigbillionsale, #electronicsonbigbilliondays, #bigbilliondaysareback, #bigbilliondaysneakpeek</td>
</tr>
</tbody>
</table>
## Subjective Results: Example hashtags Retrieved

<table>
<thead>
<tr>
<th>Event</th>
<th>FreqPearson</th>
<th>Naive Bayes</th>
<th>Proposed Method</th>
<th>Hashtags retrieved by our method but missed by other methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wales vs Slovakia (Euro Cup)</td>
<td>#walsvk, #euro2016, #wal, #svk, #ripchristina, #amjoy, #shapethefuturein5words, #thingsifindheartwarming, #togetherstronger, #wales</td>
<td>#walsvk, #euro2016, #ripchristina, #amjoy, #shapethefuturein5words, #thingsifindheartwarmingH, #wal, #svk, #engrus, #albsui</td>
<td>#walesvslovakia, #walesvsslovakia, #walesvslovakia, #wals vk, #stadedebordeaux, #euro2016, #bordeaux, #slovakia, #uefaeuro2016, #wal</td>
<td>#walesvslovakia, #walesvsslovakia, #walesvslovakia</td>
</tr>
<tr>
<td>GST Bill (Politics and Governance)</td>
<td>#gst, #gstbill, #transformingindia, #raghuramraj an, #rajanslastpolicy, #diljumalajumlahogaya, #fdi, #rbi, #aadhaar, #foodsecurity</td>
<td>#gst, #gstbill, #india, #transformingindia, #tax, #modi, #raghuramraj an, #rajanslastpolicy, #gstcleared, #loksabha</td>
<td>#goodsandservicestax, #gstbill, #goodandservicesbill, #constitutionalamendmentbill, #goodsandservicetax, #constitution, #goodsandservice, #evilandservicesbill, #gst, #onenationonetax</td>
<td>#goodsandservicestax, #goodandservicesbill, #constitutionalamendmentbill, #goodsandservicetax, #onenationonetax</td>
</tr>
</tbody>
</table>
Objective Results

(a) NDCG

(b) Precision
Finally we want to retrieve the posts related to an event. Some of those relevant posts may contain hashtags, some may not. Proposed a method that involves:

- content based analysis
- hashtag and
- temporal information

CLEF 2017 Microblog Cultural Contextualization Dataset

- 70 million microblogs from 664 events. Each microblog has the following attributes.

  - **id**: unique id of the microblog
  - **from userid**: unique id of the author
  - **iso language code**: encoding of the tweet (en, es, fr, pt)
  - **wday**: week day
  - **created at**: tweet creation date
  - **content**: tweet content
  - ...
Scoring the tweets

- **Scoring method**
  - $S_{BD}(tweet)$: BM25+DFR
  - $S_{M}(tweet)$: Depends on whether the tweet contains
    - Festival name
    - Artist name
    - Top-K hashtags
  - $S_{T}(tweet) = \frac{\gamma^t + \lambda}{1 + \lambda}$
Retrieving Event Related Tweets: Results

Precision comparison

- BM25
- BM25+DFR
- BM25+DFR+Meta Attributes
- BM25+DFR+MetaAttributes+Time

K

Precision@K
Conclusions

• Focused on the problem of retrieving relevant tweets for given planned events
• Identified hashtags for the event
• Used this in conjunction with other signals (content, metadata) for the final retrieval
• Can work for other settings also, where the input context is not a planned event but
  • Virtual event
  • Discussion theme
  • State/situation after occurrences (e.g. natural calamities)
Thank You!!