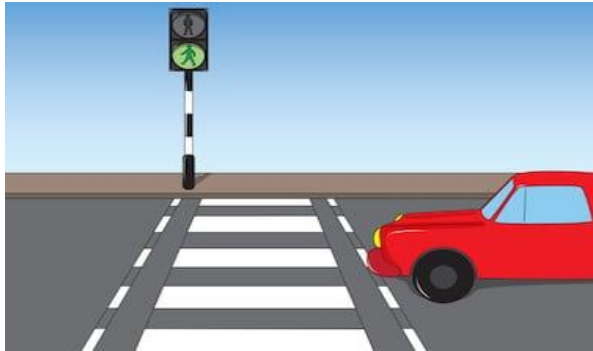# Event Recommendation using Social Media Signals
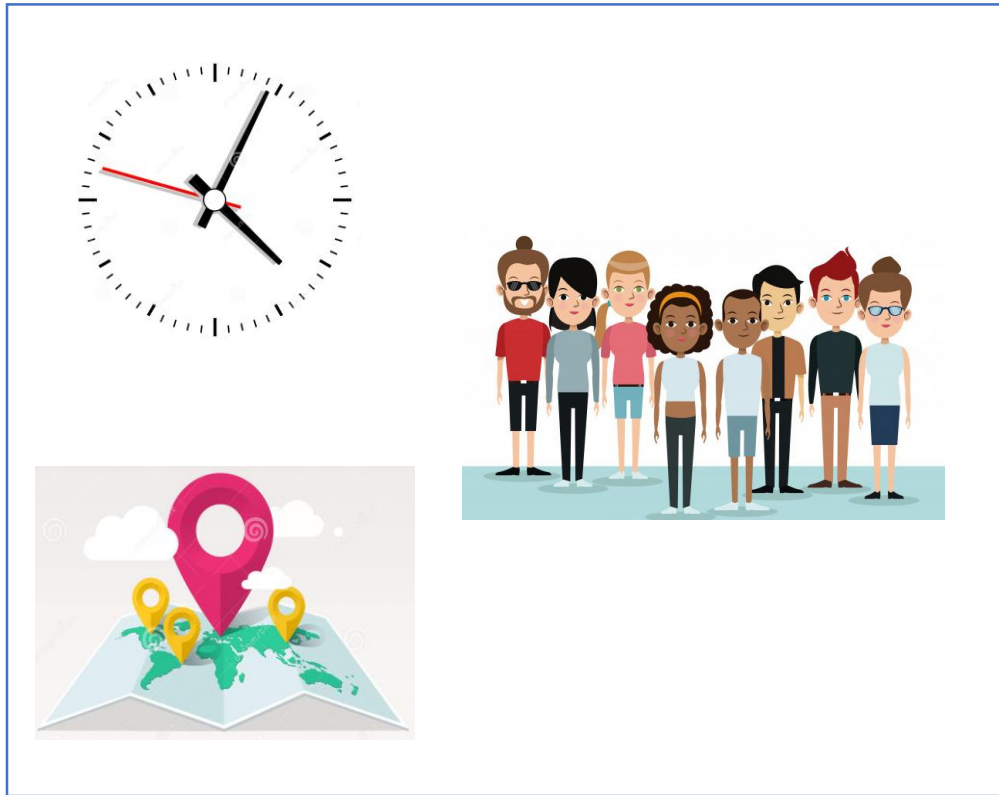
Dr. Maunendra Sankar Desarkar

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

Event

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.
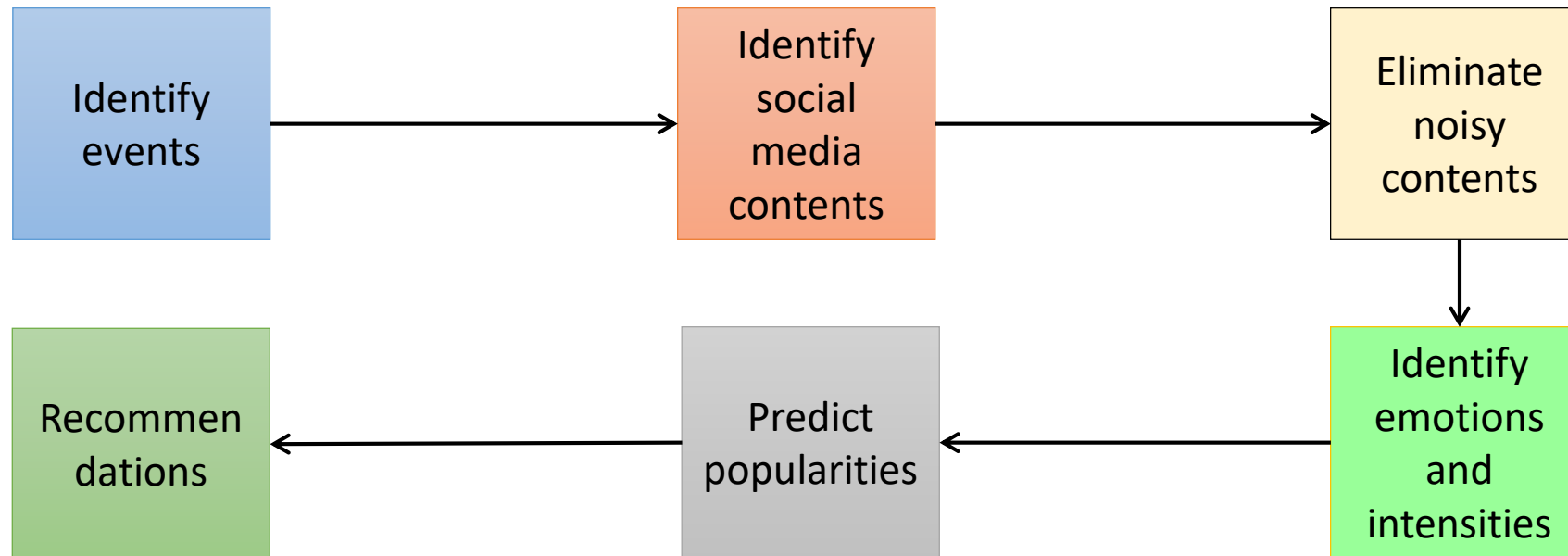
Usefulness of Event Popularity Estimation

- **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

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

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

Identify events → Identify social media contents → Eliminate noisy contents

Recommendations ← Predict popularities ← Identify emotions and intensities

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

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 from events

Example tweets related to planned events
Ready for the show @organicbananas1 live #Trans2015 **@TransMusicales** #festival #vielleelectro https://t.co/p7ZaJqQiQg
Looking for a qualifier run for Airtel **Hyderabad Marathon** (AHM)? Come, participate in #Whitathon2019, now a qualifier run for Airtel Hyderabad Marathon.
I'm already blown away by the **@TransMusicales festival** and it hasn't even really started. The venue alone is mind blowing.
This is in reference to the **SPIC MACAY VIRASAT** series being held at IIT Bombay from 10th-12th March 2019. I am sharing the the program schedule for the series as below for your ready reference.
Smart India Hackathon 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 www. http://aicte-india.org @HRDMinistry

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.

Sreekanth Madisetty, Maunendra Sankar Desarkar: **Identification of Relevant Hashtags for Planned Events Using Learning to Rank.** Revised Selected Papers from KDIR 2017, Springer Nature Switzerland AG, 2019.

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
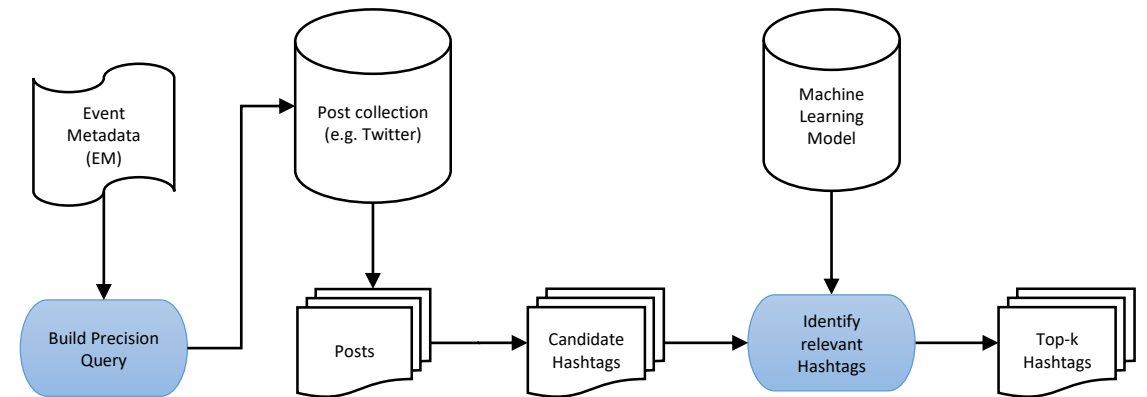
The pipeline
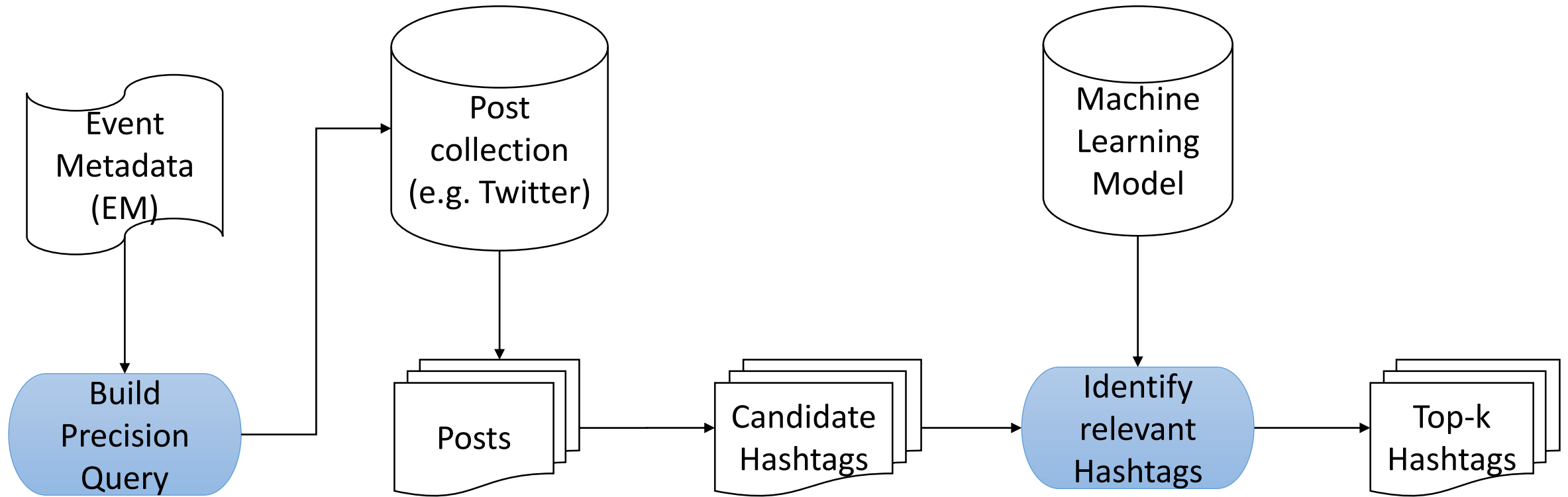
List of features for (event, hashtag) pair

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

Learning the weights

$$minimize_{w} \frac{1}{2} w^T w + C \sum_{i,j,k} \varepsilon_{i,j,k}$$

$$\forall k \text{ and } i \neq j \in \{1,...,n_k\} \text{ with } h_{ki} >_{E_k} h_{kj}$$

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

$$\varepsilon_{i,j,k} \geq 0$$

Dataset

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

- 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

| Event	FreqPearson	Proposed Method	Hashtags retrieved by our method but missed by other method
National Film Awards (Award Ceremonies)	#nationalfilmawards, #rustom, #nationalaward, #24themovie, #akshaykumar, #neerja, *#bestactor*, #nationalawards, #dangal, #zairawasim	#nationalfilmawards, #64thnationalfilmawards, #nationalfilmaward, #nationalfilmawards2017, ##nationalfilmawards, #64nationalfilmawards, #nationalaward, #nationalawards, #nationalfilmawardsindia, #64thnationalfilmaward	#64thnationalfilmawards, #64nationalfilmawards, #nationalfilmawardsindia
Flipkart Big Billion Days (E-commerce Events)	#bigbilliondays, #shoponbigbilliondays, #flipkart, #greatindianfestival, #mobilesonbigbilliondays, #bbd, #fashion, #unboxdiwalibestoffers, #unboxdiwalisale, *#amazon*	#bigbilliondays, #shoponbigbilliondays, #mobilesonbigbilliondays, #bigbilliondays2016, #flipkartbigbillionsale, #electronicsonbigbilliondays, #thebigbilliondays, #bigbilliondaystonight, #bigbilliondaysareback, #bigbilliondayssneakpeek	#bigbilliondays2016, #flipkartbigbillionsale, #electronicsonbigbilliondays, #bigbilliondaystonight, #bigbilliondaysareback, #bigbilliondayssneakpeek

Subjective Results: Example hashtags Retrieved

| Event	FreqPearson	Naive Bayes	Proposed Method	Hashtags retrieved by our method but missed by other methods
Wales vs Slovakia (Euro Cup)	#walsvk, #euro2016, #wal, #svk, *#ripchristina, #amjoy, #shapethefuturein5words, #thingsifindheartwarming, #togetherstronger,* #wales	#walsvk, #euro2016, *#ripchristina, #amjoy, #shapethefuturein5words, #thingsifindheartwarmingH,* #wal, #svk, *#engrus, #albsui*	#walesvslovakia, #walesvsslovakia, #walesslovakia, #walsvk, #stadedebordeaux, #euro2016, #bordeaux, #slovakia, #uefaeuro2016, #wal	#walesvslovakia, #walesvsslovakia, #walesslovakia
GST Bill (Politics and Governance)	#gst, #gstbill, #transformingindia, *#raghuramrajan, #rajanslastpolicy, #diljumlajumlahogaya, #fdi, #rbi, #aadhaar, #foodsecurity*	#gst, #gstbill, *#india,* #transformingindia, #tax, *#modi, #raghuramrajan, #rajanslastpolicy,* #gstcleared, #loksabha	#goodsandservicestax, #gstbill, #goodandservicesbill, #constitutionalamendmentbill, #goodsandservicetax, *#constitution,* #goodsandservice, #evilandservicesbill, #gst, #onenationonetax	#goodsandservicestax, #goodandservicesbill, #constitutionalamendmentbill, #goodsandservicetax, #onenationonetax

Objective Results
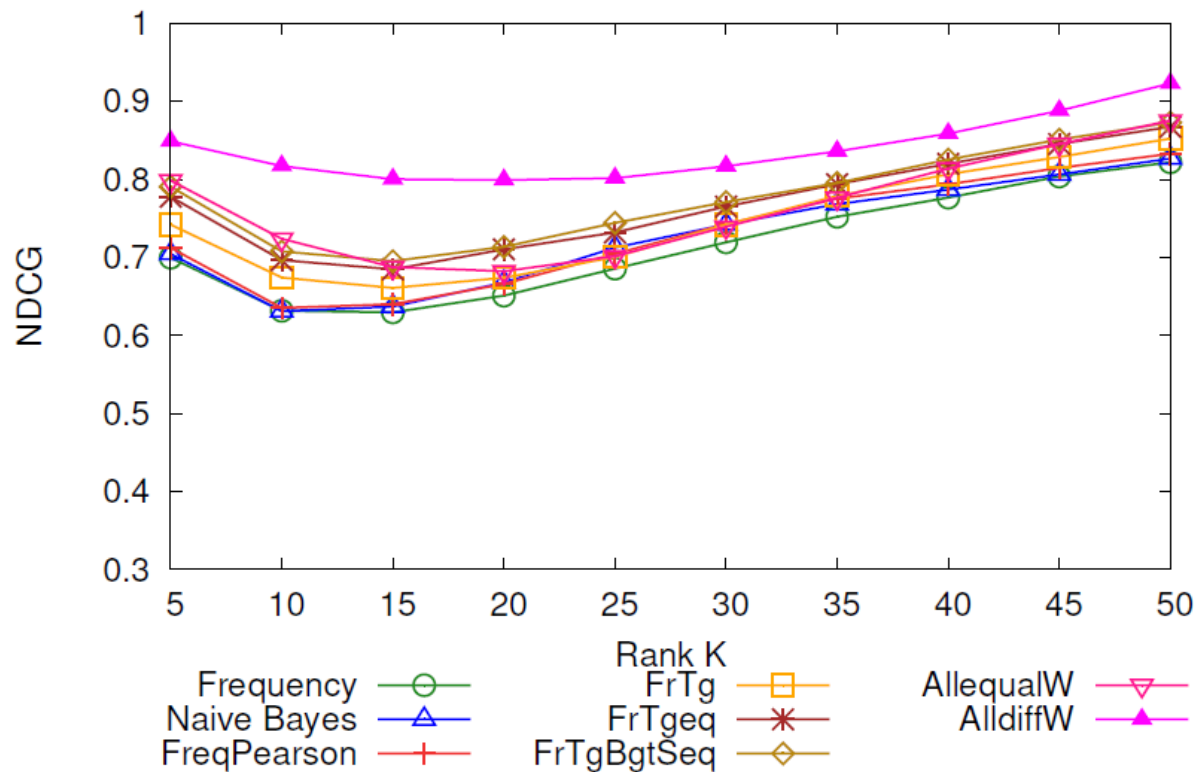
(a) NDCG
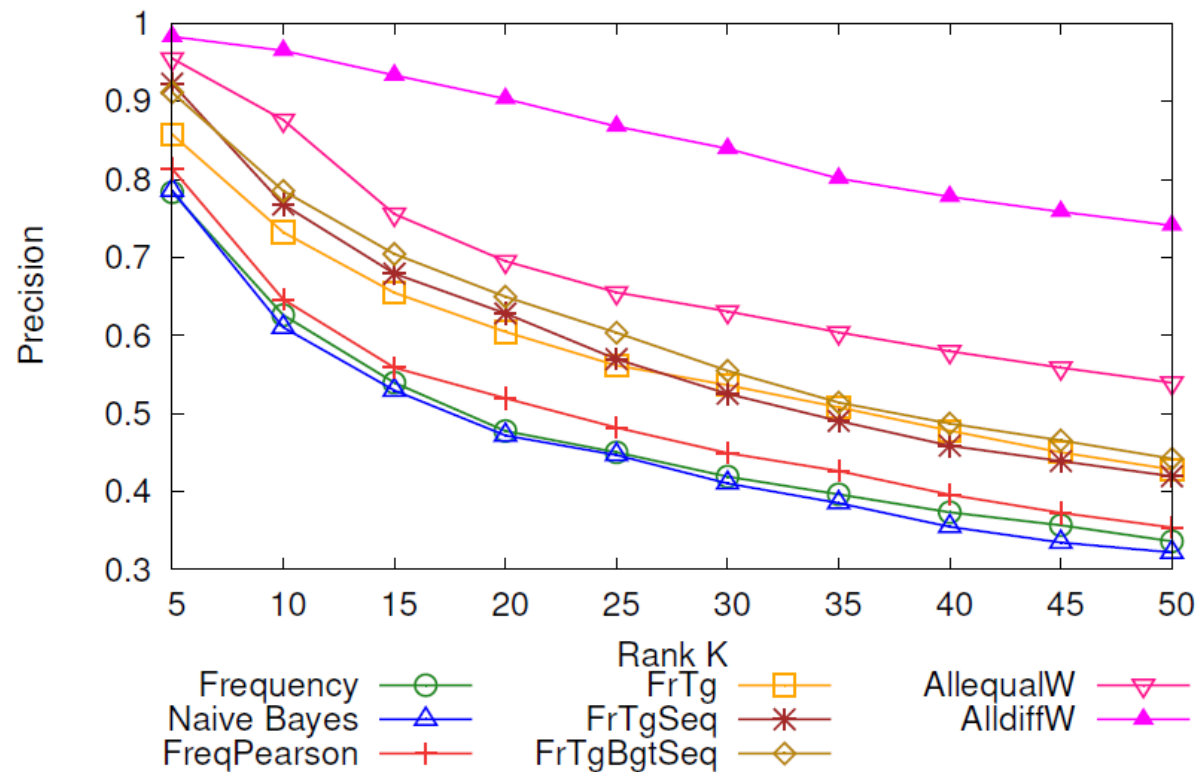
(b) Precision

Finding Relevant Tweets for Events

- 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

- Sreekanth Madisetty, Maunendra Sankar Desarkar: **IITH at CLEF 2017: Finding Relevant Tweets for Cultural Events**. Experimental IR Meets Multilinguality, Multimodality, and Interaction: 8th International Conference of the CLEF Association, **CLEF 2017**, Dublin, Ireland, September 11-14, 2017.

# 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
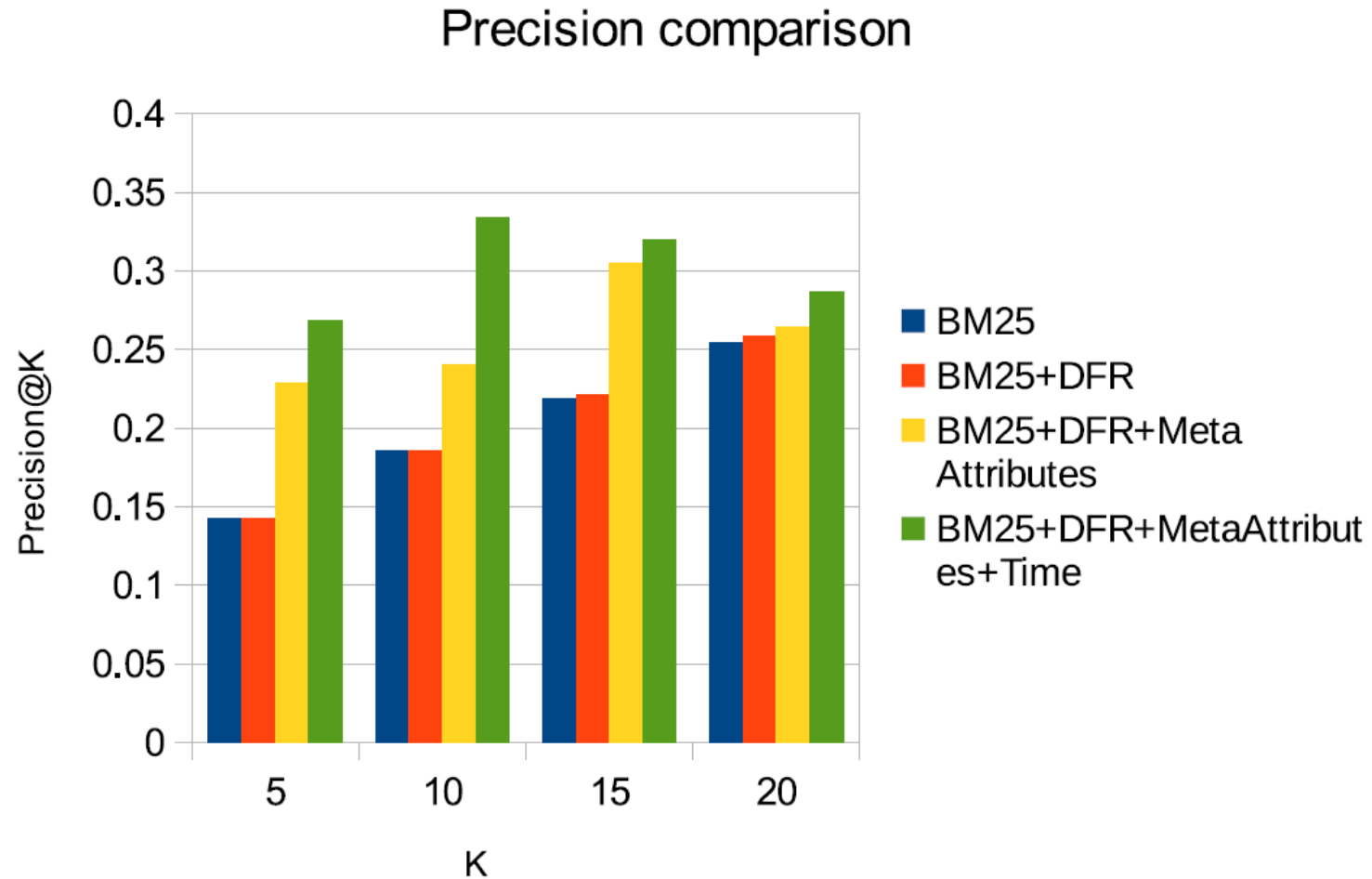 - $S_T(tweet) = \dfrac{\gamma^t + \lambda}{1 + \lambda}$
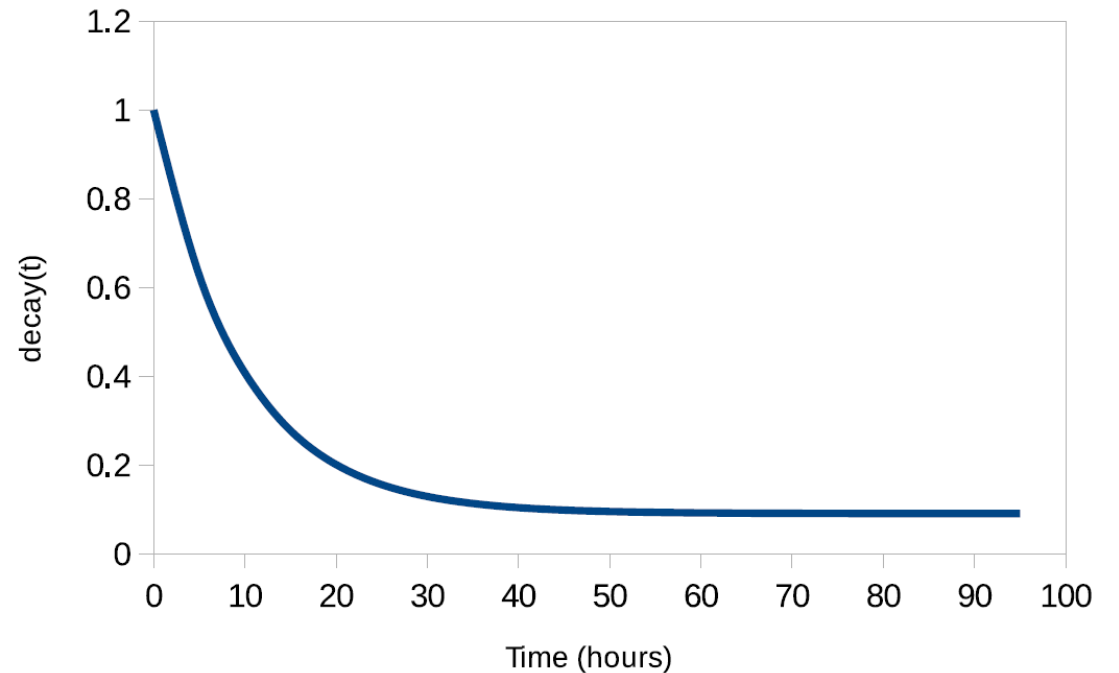
| Information Retrieval (Existing) | Meta Data (Contribution) | Temporal Impact (Contribution) |

Retrieving Event Related Tweets: Results

Precision comparison

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!!