WARWICK

LECTURE 19

Event Detection in Social Media

Arkaitz Zubiaga, 5th December, 2018

LECTURE 19: CONTENTS

- Social Media and Event Detection. Definition and Motivation.
- Event Detection for Scheduled/Predictable Events.
- New Event Detection.
  - Twitcident.
  - TopicSketch.

THE VALUE OF SOCIAL MEDIA DATA

- Thanks to **social media**, people report online what's **happening in the offline world**.
- Ubiquitous digital devices + social media → **citizen journalism**.

THE VALUE OF SOCIAL MEDIA DATA

- We can indeed use social media as a data source to identify **what's going on in different parts of the world**.
  - Social media is **not just about people eating a burger**.

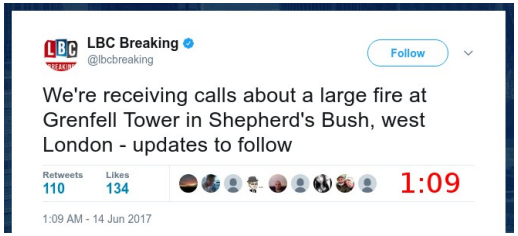
- **Event detection**:
  - Detect breaking news.
  - Identify events of potential impact to stock markets.
  - Early detection of flu outbreaks.
  - etc.

CITIZEN JOURNALISM: EXAMPLE

- 2017: Grenfell Tower fire in London.

CITIZEN JOURNALISM: EXAMPLE

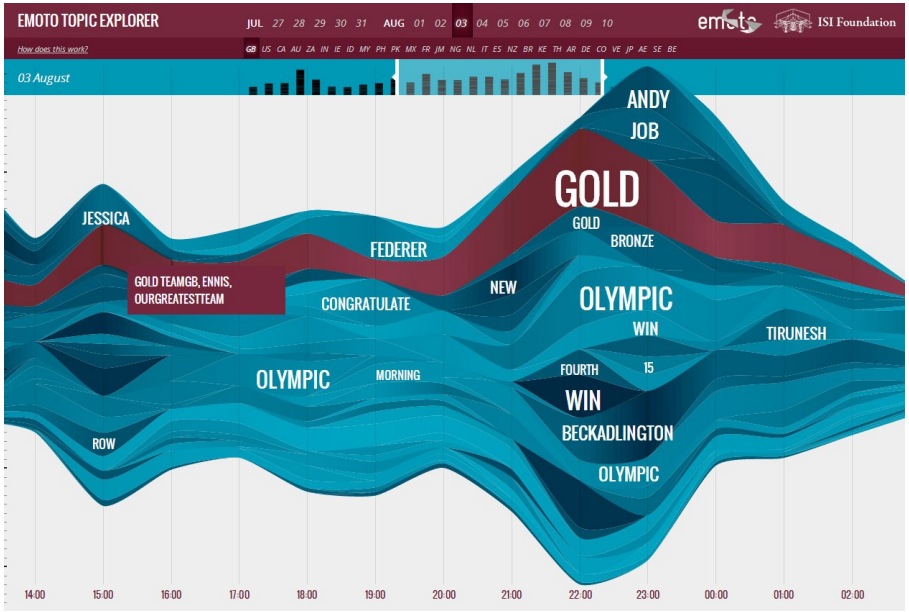
0:54 Firefighters were called at 00:54 BST



1:43 BBC Breaking News @BBCBreaking

TRENDING TOPICS: EXAMPLE

- Likewise, we may want to track trending/hot topics over time.



EVENT DETECTION

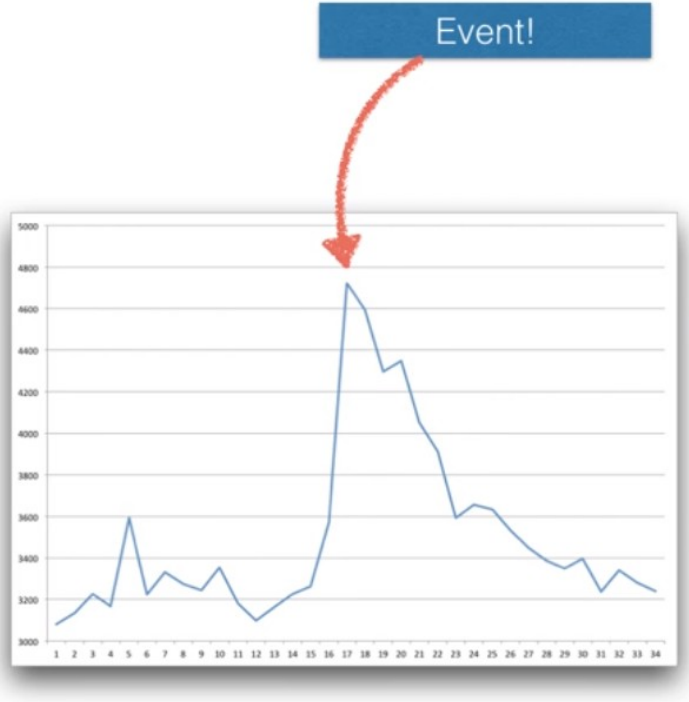
- The **event detection** task consists in **identifying emerging events of interest**.

- Different to most NLP tasks we've seen in the module, time is crucial here.
  - Using a dataset as a bag of documents/posts is not enough.
  - Posts associated with a particular event are expected to occur within a period of time.
  - Ideally we treat the data as an incoming stream.

EVENT DETECTION: BASELINE

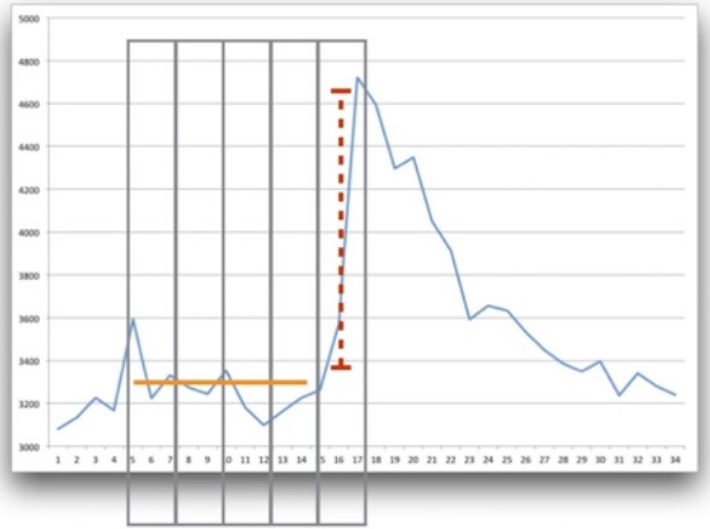
- The simplest approach to event detection is by identifying **sudden bursts of specific keyword(s)**.
  - e.g. we track mentions of the keyword ‘earthquake’ on Twitter → if we observe an unusually high number of mentions at a specific point in time, potential new event occurred.

SIMPLE EVENT DETECTION: EXAMPLE



SIMPLE EVENT DETECTION

- Keep counts of keyword occurrences over time.
- We can track one or several keywords.
- If we observe an increase that exceeds a threshold, that's a new event.



SIMPLE EVENT DETECTION

- **+:** very **easy to implement**, just keyword tracking.
- **-:** we may **miss many related keywords**, synonyms, or in other languages.
- **+:** can be good to track **known events**
  e.g. track tweets about the Olympics, identify spikes in mentions of “gold”
  e.g. track tweets about a particular company, identifying spikes that indicate I need to buy/sell shares
- **-:** not always suitable to **track unexpected events**
  e.g. terrorist attack or earthquake

APPROACHES TO NEW EVENT DETECTION

- **Clustering: Cluster similar social media posts** regularly (or even online).
  - Then score those clusters to determine the **likelihood that they represent new events**.

APPROACHES TO NEW EVENT DETECTION

- Build **probabilistic language models**:
  - Pre-train a probabilistic **language model for regular, eventless streams**.
    e.g. what is the language model for tweets in a normal day/hour in the UK?
  - If we observe emerging **vocabulary that differs from that language model** (e.g. high perplexity), that's an event.

APPROACHES TO EVENT DETECTION

- We will see two approaches to event detection:
  - **Twitcident** (Abel et al., 2012).
  - **TopicSketch** (Xie et al., 2016).

TWITCIDENT

TWITCIDENT

- **Intuition:** Twitter is an interesting data source to get information from eyewitnesses; however, difficult to clean and identify events.
  - There is a lot of data, and just a tiny bit is about events.

- **Idea:** use emergency broadcasting services to identify events.

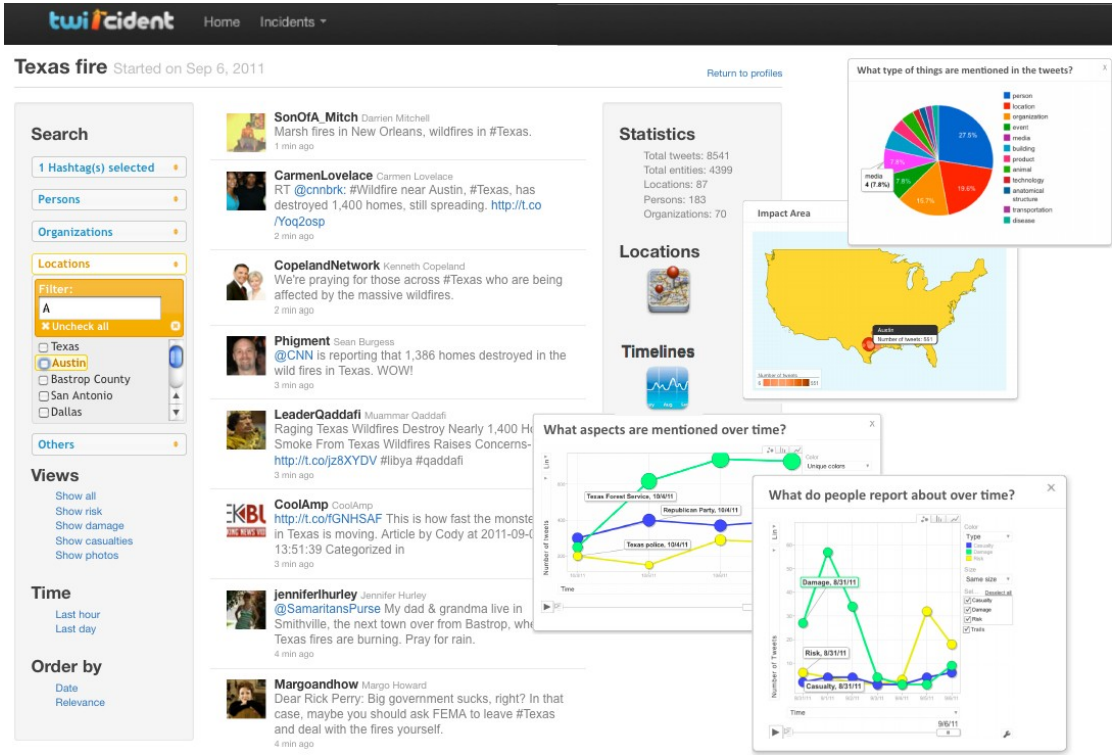
TWITCIDENT

- It was tested in the Netherlands, using the P2000 communication network (an emergency broadcasting system).

TWITCIDENT

- When Twitcident receives **new event from P2000**.
  1. Transforms the P2000 message into a **query**.
  2. This query is the input into Twitter to **track new incoming tweets**.
  3. This incoming tweets enable **tracking the event**.
  4. This tweets are processed to extract **linguistic features** and enable **visualisation in a dashboard**.

TWITCIDENT


twiTcident
Home
Incidents
Texas fire Started on Sep 6, 2011
Return to profiles
Search
1 Hashtag(s) selected
Persons
Organizations
Locations
Filter:
Uncheck all
Texas
Austin
Bastrop County
San Antonio
Dallas
Others
Views
Show all
Show risk
Show damage
Show casualties
Show photos
Time
Last hour
Last day
Order by
Date
Relevance
SonOfA_Mitch Darrien Mitchell
Marsh fires in New Orleans, wildfires in #Texas.
1 min ago
CarmenLovelace Carmen Lovelace
RT @cnnbrk: #Wildfire near Austin, #Texas, has destroyed 1,400 homes, still spreading. http://t.co/Yoq2osp
2 min ago
CopelandNetwork Kenneth Copeland
We're praying for those across #Texas who are being affected by the massive wildfires.
2 min ago
Phigment Sean Burgess
@CNN is reporting that 1,386 homes destroyed in the wild fires in Texas. WOW!
3 min ago
LeaderQaddafi Muammar Qaddafi
Raging Texas Wildfires Destroy Nearly 1,400 H
Smoke From Texas Wildfires Raises Concerns-
http://t.co/jz8XYDV #libya #qaddafi
3 min ago
CoolAmp CoolAmp
http://t.co/fGNHSAF This is how fast the monste
in Texas is moving. Article by Cody at 2011-09-
13:51:39 Categorized in
3 min ago
jenniferlhurley Jennifer Hurley
@SamaritansPurse My dad & grandma live in Smithville, the next town over from Bastrop, wh
Texas fires are burning. Pray for rain.
4 min ago
Margoandhow Margo Howard
Dear Rick Perry: Big government sucks, right? In that case, maybe you should ask FEMA to leave #Texas and deal with the fires yourself.
4 min ago
Statistics
Total tweets: 8541
Total entities: 4399
Locations: 87
Persons: 183
Organizations: 70
Locations
Timelines
What type of things are mentioned in the tweets?
Impact Area
What aspects are mentioned over time?
What do people report about over time?

TOPICSKETCH

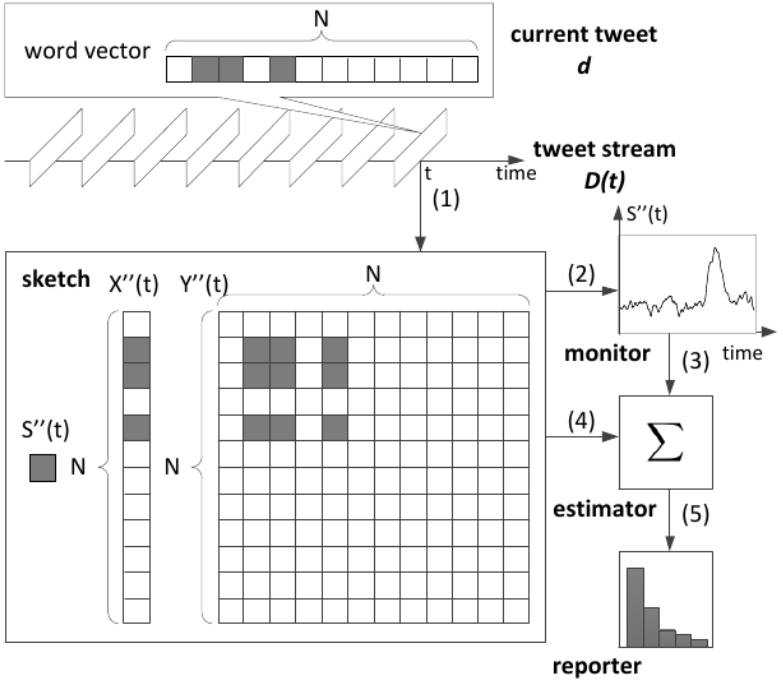
TOPICSKETCH

- TopicSketch deals with **three challenges** of event detection in social media:
  - How to **identify the bursty topics**, i.e., what are the keywords of the topics.
  - How to detect a bursty topic **as early as possible**.
  - How to perform the task efficiently in **a large-scale, real-time** setting.

TOPICSKETCH

- TopicSketch performs two steps:
  - **A sketch-based topic modelling step.**
    It looks for the acceleration of words and pairs of words over time.
  - **Hashing-based dimensionality reduction step.**
    The whole vocabulary of Twitter is too large. It proposes to only maintain tweets observed in the stream in the last 15 minutes. As this is still too large, they propose to hash all those words into buckets, and then identify frequent buckets.

TOPICSKETCH



EVENT DETECTION: SUMMARY

- Event detection in social media is still an **open research problem**.
- It has however shown **very promising results**.
  - Important events can be identified **quicker than from any other sources**.
  - This can in turn be leveraged for:
    - **social good**, e.g. to act quicker in emergencies.
    - **stock markets**, e.g. identify/predict impactful events.
    - etc.

REFERENCES

- Abel, F., Hauff, C., Houben, G. J., Stronkman, R., & Tao, K. (2012, April). Twitcident: fighting fire with information from social web streams. In Proceedings of the 21st International Conference on World Wide Web (pp. 305-308). ACM.

- Xie, W., Zhu, F., Jiang, J., Lim, E. P., & Wang, K. (2016). Topicsketch: Real-time bursty topic detection from twitter. IEEE Transactions on Knowledge and Data Engineering, 28(8), 2216-2229.

ASSOCIATED READING

- Goswami, A., & Kumar, A. (2016). A survey of event detection techniques in online social networks. Social Network Analysis and Mining, 6(1), 107. https://link.springer.com/article/10.1007/s13278-016-0414-1