Compass: Comprehensive Analytics on Sentiment for Spatiotemporal Data

Use Case: Spatio-temporal Sentiment Analysis of US Election

Debjyoti Paul, Feifei Li, Murali Krishna Teja, Xin Yu, Richie Frost
[deb, lifeifei, teja614, xiny, rfrost]@cs.utah.edu
School of Computing, University of Utah

KDD2017

Abstract

With the widespread growth of various social network tools and platforms, analyzing and understanding societal response and crowd reaction to important and emerging social issues and events through social media data is increasingly an important problem. However, there are numerous challenges towards realizing this goal effectively and efficiently, due to the unstructured and noisy nature of social media data. The large volume of the underlying data also presents a fundamental challenge. Furthermore, in many application scenarios, it is often interesting, and in some cases critical, to discover patterns and trends based on geographical and/or temporal partitions, and keep track of how they will change overtime. This brings up the interesting problem of spatio-temporal sentiment analysis from large-scale social media data. This paper addresses this problem with a framework called Compass.

Motivation

- *People are open to share views on public events*
  - Twitter, Facebook, news blogs
  - Emergency and calamity services
  - Apps with location information
- *Spatio-temporal factors:*
  - Crowd reaction/response reveals pattern with spatial information.
  - Crowd response on events changes with time.
- *Collective crowd opinion matters*:
  - Gauge community influence of events
  - Local/global socio-economic concerns
  - Local Health-Food indicators
  - Local Ad-targeting framework

Challenges

- Large volume of data
- Realtime
- Clean up/filter non-relevant data
- Generic Topic Classification techniques
- Sentiment Analysis
- Adhoc query support
- Information representation

Demo: http://estorm.org/
Code: Will be released soon in Github

Compass

A framework with **minimal** human hand curation

- Scalable
- Streaming and batch
- Easy configurable
- Pluggable
  - Data Sources
  - Machine learning models
- Adhoc pluggable modules
  - Bursty Event Detection module
- Analytics Support
  - SIMBA- Spatial In-memory Big Analytics (SIGMOD 2016)
- APIs for User Interface (UI)

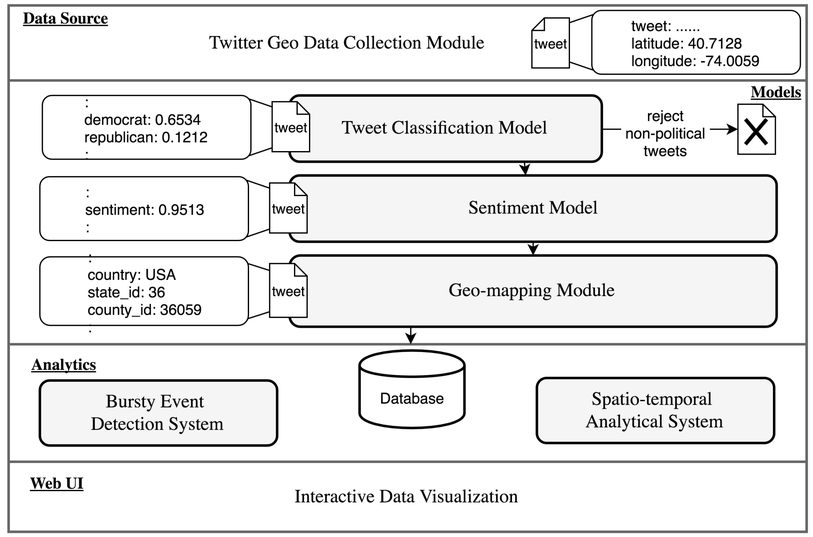
Bursty events results

Date	Keywords	b_{r+}	b_{r-}	b_{d+}	b_{d-}
		Event Description			
19th Oct	debatenight, debate Trump, answer the question, power transfer	11.8	8.3	5.6	7.4
		Flash Poll Trump Wins Final Presidential Debate			
12th Oct	sexual assault, sexual predator, grabbing, dressing rooms	2.2	2	1.8	1.5
		Women Accuse Trump of inappropriately touching them			
9th Oct	debate tonight, jumping ship, tape	15.8	10.0	5.9	7.7
		Trump and Clinton face off second presidential debate			
7th Oct	Women, sexual assault, Respect, abuse, hated	3.4	4.8	1.6	2.02
		Tape leaks showing Trump talking about groping women			
4th Oct	Pence, Kaine, crazy system, Poor Black, tax returns	4.0	2.6	2.2	2.7
		Kaine and Pence face off in vice presidential debate			
28th Sep	Clinton, answer questions, FBI, won the debate	2.3	1.9	1.9	1.72
		Clinton slams Trump in blistering presidential debate			
10th Sep	BasketOfDeplorables, Trump supporters, voters	1.72	1.3	1.3	1.7
		Clinton called out for Trump supporter deplorable			
29th Aug	Huma Abedin, Separate, Joe, Breitbart	1.85	1.4	0.7	0.8
		Clinton aide Abedin separates from Weiner			
25th Jul	DNC convention, Democratic, Hillary, Michelle, emails	1.3	1.1	2.3	2.0
		Democratic National Convention Philadelphia			
21th Jul	Make America Great, trumps speech, supporters, jobs	2.2	1.1	2.0	2.3
		Trump accepts Republican presidential nomination			
18th Jul	GOP convention, Trump, RNC, RNC in CLE	2.5	1.4	1.1	0.97
		Republican National Convention in Cleveland			

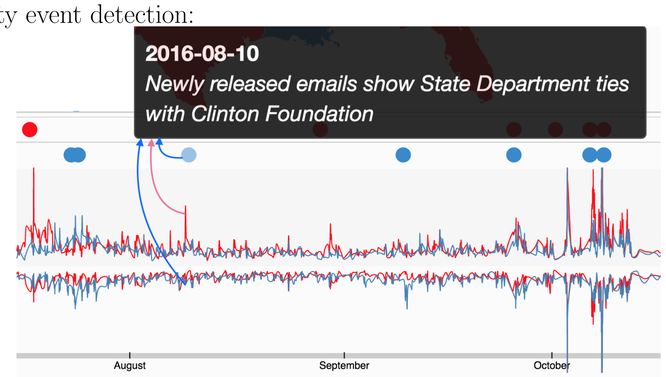
Sample of bursty event scores from Bursty Event Detection module

Compass Architecture – US Election

US Election Compass architecture:



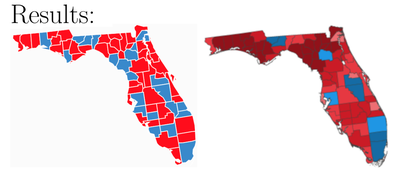
Bursty event detection:



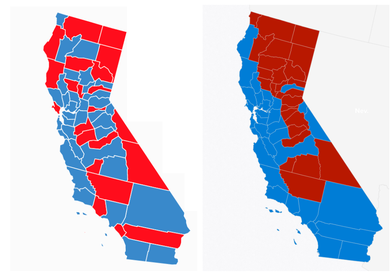
Compass modules

- Geotagged tweet collection
  - 1% Streaming API
  - Location based search API
  - Prioritize queries based on density
  - Partition of query location
- Tweet Classification Model
  - Political classification
  - Party classification
  - Minimal human curation
- Sentiment Analysis Model
  - State-of-the-art sentiment models
  - Generic classification excluding emoticons
- Bursty Event Detection Model
  - A novel approach, to define events likely to be popular in near future.
- Spatio Temporal Analysis Framework
  - Databases with support from SIMBA and Apache Spark.
- Visualization

Results:



Florida State Left: Compass; Right: Actual .



California State Left: Compass; Right: Actual .