Social Data Research at a National Laboratory

Eric Bell
Eric Bell @chsbellboy · now

Happy Purrthday: bit.ly/25L6dnM (A first cat photo tweet for me) Find out why at #BigBoulder
Facebook Activity for the Finals teams during their Semi-Final matches | Total Tweets: 49.8 million

Lancet is not responsible for the accuracy of the data sources and is only displaying a subset of tweets sent during the Semi-Final games.
“Think, think, think.”
Twitter Activity for the Finals teams during their Semi-Final matches | Total Tweets: 49.8 million

BRA

GER

NED

ARG

Lancet is not responsible for the accuracy of the data sources and is only displaying a subset of Tweets sent during the Semi-Final games.
Twitter Analysis

• Example Twitter Analysis Problems (Vetted Use Cases)
  • Studying cross-linguistic transfer- the influence of non-English language on various levels of linguistic performance in English.
  • Using twitter data linked across hashtags, authors, geography, or time to learn synonyms for newly emerging words used in social media.
  • Concept drift/relatedness. Using word embeddings, we’re building representations of topics or concepts. However, these topics/concepts being discussed change over time. We’re exploring the representation necessary for following a fixed topic of conversation over time as the discussion and vocabulary evolves.
  • Studying the share of voice for mentions and references to national laboratories
  • We’re interested in understanding the degree to which language sophistication varies on a topic or over time.
Motivation

Positive
- Connect
- Communicate
- Spread information
- Share interests
- Disaster responses
- Crisis events
- Situational awareness

Negative
- Social bots
- Spammers
- Trolls
- Misinformation
- Deceptive content
- Propaganda
- Manipulative campaigns

Detection of suspicious accounts = more replicable dataset
Related Work

- Social bot prediction (Ferrara et al., 2014)
- Suspended account analysis (Thomas et al., 2011)
- Non-personal and spam user detection (Guo and Chen, 2014; Lin and Huang, 2013)

- Troll detection (Mihaylov et al., 2015):
  - Accused trolls, small data (< 1K trolls)

- Analysis of 20K pro-Kremlin Twitter accounts
  - Tweet similar statements during/around breaking news
Who are the trolls?

- Look like real users (avatars)
- Similar followers and friends
- Similar tweeting behavior

**TROLL**

Web slang meaning for the word "a message or a person, whose primary objective is to annoy and provoke people, cause controversy, write pointless messages, and troll." The term often refers to aggressive and anonymous web communication supporting the agenda of the Russian leadership.
Dataset Creation

Twitter Suspension Policy

- Spam: invitation spam, selling, phishing
- Account security at risk: compromised
- Abusive behavior: violent threats, harassment, hateful conduct, multiple account abuse, impersonation, self-harm

RU-UA Crisis Twitter Dataset:

- Crisis-relevant keywords in RU/UA
- Rounds of querying API: March, June, and Dec 2015
- Balanced set of 188K accounts, 20 tweets per account
- Active vs. Non-active: 85% suspended and 15% deleted

July 11, 2016
# Features

## Profile
- days since account creation, # followers, friends, favorites, tweets, friend-to-follow ratio, name, bio, screen name length in chars/words, number of tweets per hour

## Visual
- profile background, link, text, sidebar color, background tile, sidebar border color, default profile image

## Syntactic
- tweet length in words/chars, RT, uppercase, elongated, repeated mixed punctuation, mention, hashtag, link rate, prop. of tweets with links, RTs, mentions, hashtags, punctuation, emoticons

## Network
- mentions, hashtags, LSA on mentions/hashtags

## Text
- tweet ngrams (1–3grams, binary vs. frequency), LSA on tweets, LDA topics (50–1K), word2vec embeddings (30–2K)

## Affect
- number of emoticons, prop. of six emotions, mean scores, prop. of tweets with sentiments (Volkova et al., 2015)
## Classification Results

<table>
<thead>
<tr>
<th>Features</th>
<th>D-S-ND</th>
<th>DS-ND</th>
<th>D-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile</td>
<td>0.78</td>
<td>0.85</td>
<td>0.86</td>
</tr>
<tr>
<td>Style + Syntax</td>
<td>0.72</td>
<td>0.81</td>
<td>0.86</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tweets</td>
<td>0.82</td>
<td>0.87</td>
<td>0.83</td>
</tr>
<tr>
<td>Tweets + LSA</td>
<td>0.79</td>
<td>0.84</td>
<td>0.85</td>
</tr>
<tr>
<td>Topics</td>
<td>0.77</td>
<td>0.81</td>
<td>0.83</td>
</tr>
<tr>
<td>Embeddings</td>
<td>0.72</td>
<td>0.76</td>
<td>0.94</td>
</tr>
<tr>
<td><strong>Network</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hashtags</td>
<td>0.67</td>
<td>0.76</td>
<td>0.84</td>
</tr>
<tr>
<td>Mentions</td>
<td>0.69</td>
<td>0.78</td>
<td>0.85</td>
</tr>
<tr>
<td>Hashtags + LSA</td>
<td>0.63</td>
<td>0.73</td>
<td>0.84</td>
</tr>
<tr>
<td>Mentions + LSA</td>
<td>0.64</td>
<td>0.72</td>
<td>0.85</td>
</tr>
<tr>
<td><strong>Affect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentiment + Emotion</td>
<td>0.62</td>
<td>0.72</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Key Findings

Tasks: D – S > DS – ND > D – S – ND

Text: Tweet ngrams and embeddings are the most predictive

Network: Mentions are more predictive than hashtags

Frequency vs. Binary:

► **Tweet ngrams**: It is not only important what the users say but how much they say it

► **Mentions and hashtags**: It is not important how much the users use some hashtags or mentions, but whether they use them or not
Analysis: Verbal and Nonverbal Behavior Differences

**Verbal Behavior**
Deleted and suspended users generate:

- Shorter tweets
- Less elongated capitalized words and repeated punctuation
- Lower hashtag, mention and URL per word ratios
- Less RTs, tweets with hashtags, URL and mentions
- Less tweets with punctuations and emoticons

**Nonverbal Behavior**
Deleted and suspended users have:

- More friends
- Less followers and tweets
- Lower friend-to-follower ratio
- Shorter bios
- Longer user names
WHERE THIS BITLINK WAS SHARED

<table>
<thead>
<tr>
<th>Platform</th>
<th>Clicks</th>
<th>Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>2,622</td>
<td>5</td>
</tr>
<tr>
<td>Facebook</td>
<td>618</td>
<td></td>
</tr>
<tr>
<td>Tumblr</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Orkut</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Other Sites</td>
<td>4,942</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>10,431</td>
<td></td>
</tr>
</tbody>
</table>

GEOGRAPHIC DISTRIBUTION OF CLICKS

Top Countries (clicks / % of total)

<table>
<thead>
<tr>
<th>Country</th>
<th>Clicks</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>6,984</td>
<td>37%</td>
</tr>
<tr>
<td>Macedonia, The For...</td>
<td>1,306</td>
<td>7%</td>
</tr>
<tr>
<td>France</td>
<td>824</td>
<td>4%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>810</td>
<td>4%</td>
</tr>
<tr>
<td>Canada</td>
<td>713</td>
<td>4%</td>
</tr>
<tr>
<td>Germany</td>
<td>589</td>
<td>3%</td>
</tr>
<tr>
<td>Brazil</td>
<td>584</td>
<td>3%</td>
</tr>
<tr>
<td>Japan</td>
<td>497</td>
<td>3%</td>
</tr>
<tr>
<td>Argentina</td>
<td>404</td>
<td>2%</td>
</tr>
<tr>
<td>Mexico</td>
<td>373</td>
<td>2%</td>
</tr>
<tr>
<td>+90 more</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bitly Research Overview

• How does the design of the UI on various social media platforms manifest different styles of interaction propagation?

• Classic definitions of virality are based on large frequencies or potential reach, we’re instead looking to understand a model that lets us find events that are hyperlocally viral. This involves correlation of multiple data types

• Characterize URL types

• What do links look like as they move through time and spread geographically?
How do you think about Bitly clicks?
What does it mean to be viral?
Bitly Behavior By Country
Image Research Overview

• Image classification and multi-modal embeddings. Using convolutional neural networks, we’re building representations of objects and themes within images linked within social media data.

• Using language and visual embeddings, we’re exploring models for sense-making across data types for understanding how different data modalities are used to communicate ideas within a social context.
Second Quarter Highlights

- **Google+**: More than 24,000 new followers; content viewed 1.63 million times
- **LinkedIn**: An average of 64.6 engagements by unique users per day
- **Facebook**: Second among national labs in daily audience engagement
- **Twitter**: Highest audience engagement among national labs

Second Quarter Growth

- **G+**: 324,532 (+10.7%)
- **LinkedIn**: 17,638 (+5.5%)
- **Facebook**: 6,333 (+6.2%)
- **Twitter**: 5,518 (+10.7%)

Google+ FYTD Audience Share of Voice

Benchmark: as %, actual PNNL content views during period

- 82%

LinkedIn FYTD Audience Engagement

Benchmark: number of unique users engaged with PNNL content (clicked, liked, shared or commented)

Facebook FYTD Audience Engagement

Benchmark: daily audience engagement per 1000 page likes
Silver Bullet
not Just Ahead