Twitter Emergency Response System

...

Benjamin Brown, Ryan Dorris, Kate Dowdy, Ryan Fuller, Sam Singh

January 18, 2019

Problem Statement

Design and implement a tool using social media activity to develop a rapid alert response system from a disastrous event.

Project Overview

1. Framework
2. Data Collection
3. Cleaning/Modeling
4. Prototype Demonstration
5. Conclusion

Framework

1. Which social media, and why?
2. What situations or disasters would this alert system make sense for?
3. How do you identify signal from noise in social media?
   a. Guessing disaster-related words
   b. Classification modeling based on real data

Data Collection

- Pulled FEMA database for disasters
    - Narrowed dataset to more recent disasters from 2007-present
    - Filtered FEMA results by only counties that were affected by major disasters.

- Scraped historic tweets locations during FEMA disasters
    - Scraped Twitter for 497 county disasters resulting in 44,472 total tweets
    - Pulled tweets from 2 days prior to disasters and two days following (including day of disaster)
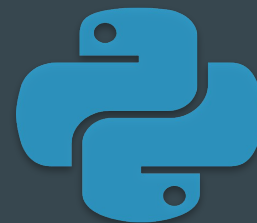
Tableau Visualization

FEMA Database Visualization

Data Cleaning

1. Removed hyper-links
2. Removed punctuation
3. Removed all non-letters and extra spaces
4. Set all text to lowercase

Natural Language Processing

1. Tokenized each word
2. Lemmatized each token
3. Train-Test Split
4. Vectorized data
   a. Count Vectorizer
   b. TF-IDF Vectorizer
5. Fit and transform Vectorizer training data
6. Transform testing data

Congrats to CB4 Trucking Inc. on the purchase of this 2012 Utility flatbed! If you're looking for a flatbed, give us a call today and let us help you! New and used trailers in...
https://www.instagram.com/p/BnbMHefF8RT/?utm_source=ig_twitter_share&igshid=8jhheo2oxng9\xa0...

congrats to cb trucking inc on the purchase of this utility flatbed if youre looking for a flatbed give u a call today and let u help you new and used trailer in

Top Phrases

Pre-Disaster

just came uniform sitting week

left right outt

house campaigned platform sticking

ha charleston state emergency evacuation

video past weekend

just classless disgusting

wonderful community better thr people

ha criminal justice reform

retweeted senator harry

central houston houston texas

house campaigned platform sticking little

wayne dyer quote

temperature humidity wind mph mph

temperature rain today

temperature rain today humidity

Post-Disaster

just posted photo

happy labor day

happy valentine day

sun oct hfd

virtual tour listing

retweeted rachel maddow

retweeted rachel maddow msnbc

rachel maddow msnbc

just posted video

humidity wind mph

motor vehicle incident

coral sky amphitheatre

panama city beach

mph pressure mb

humidity wind mph pressure

Model Types

- Logistic Regression: Logistic function with values between 0 and 1
- kNN: Classifies each point by the class of its neighbors
- **Random Forest: Group of multiple trees from different segments of data**
- AdaBoost: Combines several simple models
- Support Vector Classification (SVC): Finds a geometric shape that separates data

Optimal Model

TF-IDF Vectorizer

- 30,000 features
- 1 to 4-word phrase length

Random Forest

- 100 component trees
- Unlimited tree depth

	Predicted 0	Predicted 1
Actual 0	5175	1926
Actual 1	3515	2726

Prototype Demonstration

Demo

Sample Output

Searching 'fire' terms across the U.S., output from a 3-minute window:

@ATLairport Every TSA agent that doesn't show for work should be fired! (North Carolina, USA)

Why is @lifetimetv app not working on @amazonfiretv? I've tried everything... (Houston, TX)

All of em, i'm setting the Monday's on fire, the I'm bout to hit the club lashed going up in flames... (Victorville, CA)

Hunterdon - High Bridge - 3 Washington Ave - 2nd Alarm - machinery fire inside of a building (East Brunswick, NJ)

BRUSH FIRE A/0 Broadway Bl toward the end. North Cocoa. Approx ½ acre brush fire, slow moving (Rockledge, FL)

Challenges and Limitations

- Limited access to tweets (historic and real-time)
- Inability to accurately track locations
- Modeling with limited data, imperfect classification

Future Steps

- Gather more relevant data from counties affected the most by past disasters for deeper analysis and training our model
- Invest in the Twitter Firehose API to have access to all streaming tweets (which we could use to further train our model)
- Pull in official alerts for comparison
- Introduce image recognition for photos of disasters
- Get access to other social media like Facebook and Instagram
- Developing our alert prototype into an interactive online dashboard

Questions?