Evaluation of Text Mining Techniques Using Twitter Data for Hurricane Disaster Resilience

Joshua Eason[1] and Sathish Kumar[2]

[1]Creighton University
[2]Coastal Carolina University

November 23, 2022

Abstract

Data obtained from social media microblogging websites such as Twitter provide the unique ability to collect and analyze conversations of the public in order to gain perspective on the thoughts and feelings of the general public. Sentiment and volume analysis techniques were applied to the dataset in order to gain an understanding of the amount and level of sentiment associated with certain disaster-related tweets, including a topical analysis of specific terms. This study showed that disaster-type events such as a hurricane can cause some strong negative sentiment in the period of time directly preceding the event, but ultimately returns quickly to normal levels. An analysis of the volume of tweets during the same time revealed that the public responds in near real-time to events with conversation on Twitter. This information can be an effective tool in which to arm emergency management personnel with vital human intelligence information to inform decision-making processes ahead of future storm, or disaster-related events. In addition, this study performed empirical performance evaluation experiments on Latent Dirichlet Allocation (LDA) topic models which were generated from Twitter data collected from Hurricane Florence. The performance evaluation experiments showed that LDA topic models struggle to accurately reflect the true latent conversation topics present within a medium-term, event-based dataset. Although the study successfully modeled LDA topic models, it could not produce models that were interpretable by human beings as distinct groups of topic words that were tightly coupled to one another.

Evaluation of Text Mining Techniques Using Twitter Data for Hurricane Disaster Resilience

Joshua Eason[1], Sathish Kumar[2]
[1]Dept. of Physics, Creighton University
[2]Dept. of Computing Sciences, Coastal Carolina University
josheason@Creighton.edu

Research Objectives

1. Analyze and evaluate Latent Dirichlet Allocation (LDA) topic model performance.
2. Analyze semantic interpretation, sentiment, and volume of tweets surrounding disasters.

Introduction

- Text mining has been used to model public opinion and sentiment about various topics
- Latent Dirichlet Allocation (LDA) topic models have limitations when used to model data collected from Twitter
- Terms in topic/word
- Sentiment and volume analysis can provide human intelligence to emergency managers

Data

Collection

- 2,728,730 tweets collected from Twitter Stream API
- Location: North Carolina & South Carolina
- Sept 1 – Oct 1, 2018 (Hurricane Florence)

Preprocessing

- Dataset was reverse geotagged
- Keyword searched and filtered using 31 hurricane-related terms
- Duplicate tweets removed
- Datasets created for experiments

Experimental Methodology

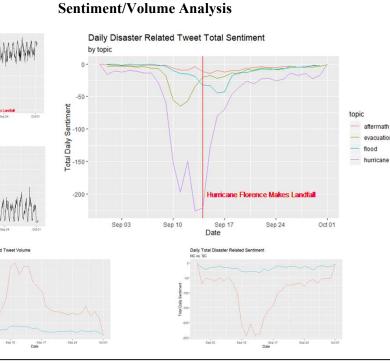### Latent Dirichlet Allocation Experimentation

- Coherence experimentation 120 models
- 5-60 topics, 5 pooling methods, 2 stemming methods
- Semantic interpretation (qualitative)

Sentiment/Volume Analysis

- Geospatial (state-level) analysis
- Temporal analysis
- Responses to specific events

Results

LDA Model Performance Evaluation

Pooling Method	Mean Coherence
Tweetid	0.667
Author	0.667
City	0.667
City/Period	0.667
State	0.557

Tweet-Pooling Method

State-Pooling Method

Sentiment/Volume Analysis

Hourly Disaster Related Sentiment

Hourly Disaster Tweet Volume

Daily Disaster Related Tweet Total Sentiment

Daily Disaster Related Tweet Volume — NC vs. SC

Daily Total Disaster Related Sentiment — NC vs. SC

Conclusions

LDA Topic Model Performance

- LDA topic models struggle to model narrow, event-based discourse with Twitter data
- Disaster-related tweets account for 0.7% of overall discussion in the states affected over the month of the hurricane event (Sept. 2018)
- Limitations exist in semantic interpretability across all pooling methods, but models become unintelligible when document sizes become too large

Sentiment/Volume Analysis

- Public sentiment is highly correlated to events
- Public sentiment and volume can be used to show to what extent the public cares about a topic or event
- Public sentiment and volume regarding the disaster closely follows the severity and temporal proximity of the event

Future Work

- Event-based methodology for improved topic model performance
- Use of audio-video-image data to provide more context
- Geospatial analysis/location accuracy
- Real-time analysis of sentiment per topic

Acknowledgments