

Evaluation of Text Mining Techniques Using Twitter Data for Hurricane

Disaster Resilience

Joshua Eason¹, Sathish Kumar²

¹Dept. of Physics, Creighton University

²Dept. 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

Sentiment/Volume Analysis

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

Results

LDA Model Performance Evaluation

Tweet-Pooling Method

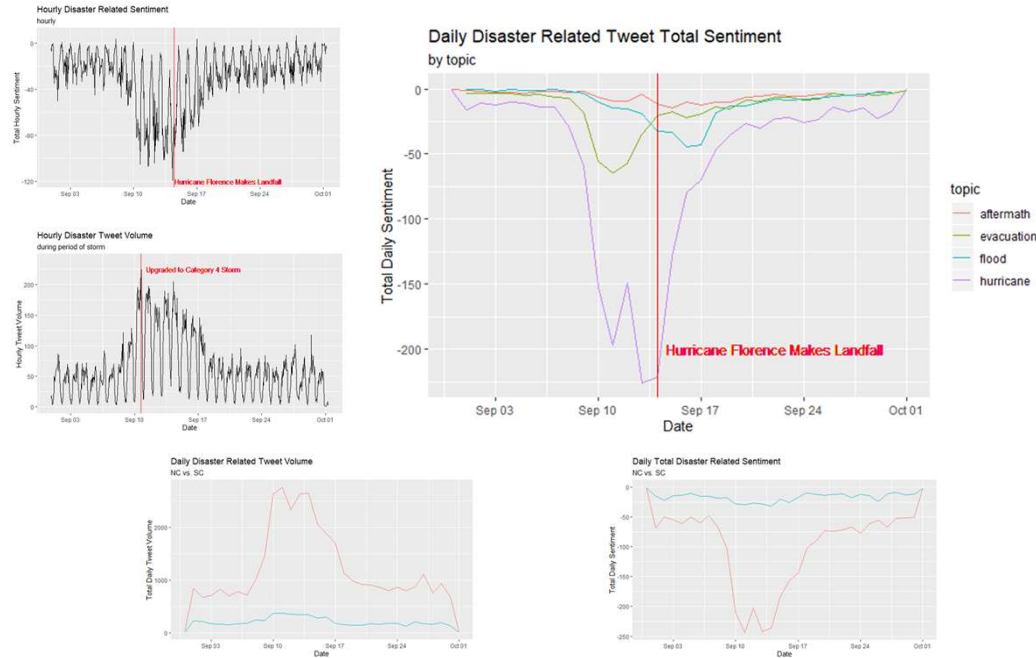
Pooling Method	Mean Coherence
Tweetid	0.667
Author	0.667
City	0.667
City/Period	0.667
State	0.557

topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	topic_10	topic_11	topic_12	topic_13	topic_14	topic_15
hurricane	carolina	hurricane	hurricane	hurricane	news	people	news	news	news	wilmington	evacuation	evacuation	evacuation	evacuation
evacuation	carolina	evacuation	evacuation	evacuation	news	people	news	news	news	wilmington	evacuation	evacuation	evacuation	evacuation
evacuation	carolina	evacuation	evacuation	evacuation	news	people	news	news	news	wilmington	evacuation	evacuation	evacuation	evacuation

State-Pooling Method

topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	topic_10	topic_11	topic_12	topic_13	topic_14	topic_15
carolina	north	nc	carolina	carolina	carolina	cb	nc	nc	nc	nc	nc	nc	nc	nc
carolina	north	nc	carolina	carolina	carolina	cb	nc	nc	nc	nc	nc	nc	nc	nc
carolina	north	nc	carolina	carolina	carolina	cb	nc	nc	nc	nc	nc	nc	nc	nc

Sentiment/Volume Analysis



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

- Funding for C-SURF was provided by NSF REU Award AGS 1560210
- Dr. Zhenlong Li, Department of Geography, University of South Carolina, Columbia
- Michael Bunker, Department of Computing Sciences, Coastal Carolina University