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

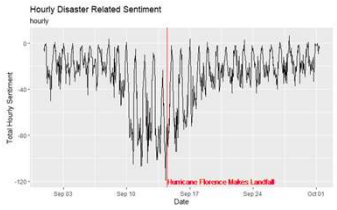
Tweet-Pooling Method

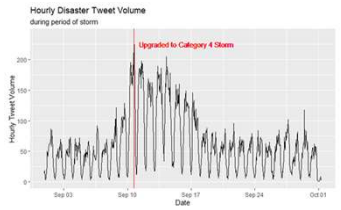
topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	topic_10	topic_11	topic_12	topic_13	topic_14	topic_15
hurricane	carolina	florence	hurricane	hurricaneflorence	leave	people	hurricaneflorence	watch	wilmington	watching	flooding	storm	hurricane	storm
winds	watching	hurricane	watching	flooded	police	family	crazy	power	damage	evacuated	night	realdonaldtrump	evacuate	north
watched	friends	north	coming	talking	roads	family	supposed	water	beach	start	prepare	doesnt	prepared	hours
emergency	youre	coast	gonna	weather	flood	flooding	makes	beach	weather	season	praying	female	supplies	hurricane
tropical	video	storm	didnt	destroyed	school	flood	coverage	ready	morning	gotta	house	county	evacuation	trump
river	local	closed	tonight	community	coast	tomorrow	scared	florence	storm	drive	friday	wilmingtonnc	activist	wrightsville
hopeimhurricane	females	weatherchannel	evacuation	monday	fired	beach	house	northcarolina	hurricanes	tornado	preparing	update	myrtle	south
beautiful	happy	storms	hoping	change	cancelled	starting	middle	fayetteville	hurricaneflorence	beach	flood	power	shelter	category
ready	classes	weather	staying	windy	favorite	carolina	whats	party	myrtle	items	started	watch	wilmington	raleigh
thursday	tomorrow	watch	matthew	shelter	playing	inland	bitch	students	myrtlebeach	stuff	games	check	times	tornado
tryna	living	shc	landfall	prepared	stupid	thinking	families	worse	wanna	prayers	trees	south	finally	helping
expected	county	house	warning	amazing	glory	pretty	greensboro	flash	south	florencian	dont	clemson	world	ocean
coast	event	literally	[redacted]	media	wednesday	lives	south	house	worried	guess	front	heart	warning	emergency
national	couldnt	phone	movie	mandatory	information	black	continues	county	continue	wrong	evacuating	bless	support	raining
residents	bring	worst	follow	honestly	arent	president	edpiotrowski	ahead	folks	foxnews	straight	waiting	window	taking

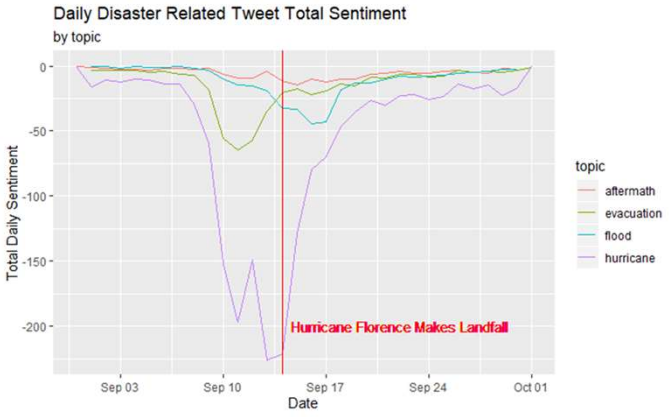
State-Pooling Method

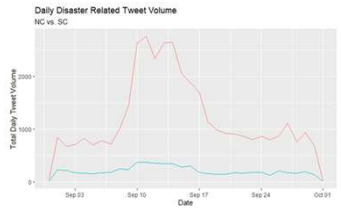
topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	topic_10	topic_11	topic_12	topic_13	topic_14	topic_15
carolina	north	nc	carolina	carolina	cb	jc	xo	mph	wx	news	fox	life	jr	mb
north	carolina	mph	north	north	football	michael	tr	ke	cc	av	bn	mike	jones	trump
south	south	vn	south	south	uf	hw	florence	wi	sp	ym	md	ay	pn	lk
edjunkiecom	edjunkiecom	sc	prima	edjunkiecom	chris	dx	trump	tz	qg	nq	py	jm	jk	pg
florenceyousuck	florenceyousuck	im	fascism	florenceyousuck	ncat	de	tg	lo	oe	wo	eq	bp	ft	pk
miraculously	miraculously	qk	edjunkiecom	miraculously	sn	ia	kj	weather	xm	gj	kr	gm	rl	hn
pdys	pdys	em	florenceyousuck	pdys	xh	kn	tv	buckeye	boo	qp	truth	jb	ca	si
jamieisperfect	jamieisperfect	qe	miraculously	jamieisperfect	ns	lu	lady	yq	jh	qh	qb	ms	vd	ncdot
itsdavahn	itsdavahn	fa	pdys	itsdavahn	gc	zo	zi	ac	xx	pm	vl	id	lq	ds
conflict	conflict	ny	jamieisperfect	conflict	aq	um	dm	qy	gv	ya	nj	kk	wb	ka
releases	releases	ej	itsdavahn	releases	cheyenne	qf	qw	ce	ir	mt	st	sf	ss	dc
traded	traded	di	conflict	traded	yl	rc	ul	racing	bq	ki	family	dg	jd	vj
antmasiello	antmasiello	news	releases	antmasiello	qo	ky	cq	morning	forever	american	zf	gj	dj	golf
ropomjcym	ropomjcym	sq	traded	ropomjcym	uq	cs	kp	jg	gw	brian	cm	iw	ii	stormtrack
uufcs	uufcs	dy	antmasiello	uufcs	lc	unc	abc	hurricaneflorence	ab	rt	robert	sweet	od	ray

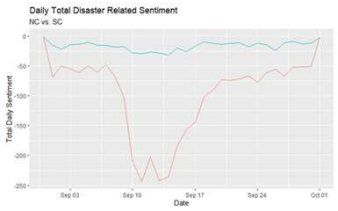
Sentiment/Volume Analysis











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