EPIC

Empowering the Public with Information in Crisis

Reason for EPIC

In disasters, citizens are the first responders

- Emergency crews come only later
- volunteers tend to self organize and assist others

"There is no THEY to tell people what to do, people need to rely on themselves to do it."

People are constantly connected to technology

- There is constant information being produced and made instantly public by these first responders

The Goal

Take all of the publicly-available, grassroots, peer-generated information, decide if it's trustworthy, secure, and accurate and combine it with official sources for optimal, local decision making by members of the public

Who This Requires

To make this possible, experts in the following fields are required:

Software engineering
Information science
Human-computer interaction
Information extraction
Natural language processing
Network security
Telecom policy
Emergency management

Where is the Information

Massive amounts of information is uploaded realtime to the internet

Blogs
Image Sites
Video Sites
Social Networking Sites
Everywhere

What Kind of Information is being generated?

Where and when crisis-based info arises?

Who participates in creating the information?

Is the information accurate?
- If not, will it be corrected?

Is the information timely?

What the information is focusing on?
- is it helpful to the situation?

How Can All The Data Be Combined

How much data is created for a single event?

How many different mediums are used?

- Text
- Pictures
- Video

Where is all of this data stored?

How can Information Extraction and NLP be Used

The diversity on volume of usefull data provides a challenge for current NLP and information Extraction Techniques.

It is important to develop a robust system that can accurately extract relevant information to be used.

How Security Plays a Role

The data people put on the internet is not always publicly available

In certain events, people and locations could suffer from broadcasting information

Anonymity in a system like this can result in individuals providing false information

Public Policy

The federal government, states, and other countries all have different laws and policies regarding information dissemination

As a service, these laws, policies, and rules need to be maintained

EPIC Specifically

EPIC is focusing on Twitter data

Simply put:

1) Comb the massive amounts of tweets
2) Identify relevant tweets regarding an event
3) Process and return useful information to the public

Initial Work

Picking out relevant information: warnings, road closures, and evacuations

The first step was Named Entity Tagging

-Person

-Location

-Organization

-Facility

Initial annotation task was to identify the syntactic span and entity class

Dataset was 200 Tweets from the 2009 Oklahoma Grassfires

Used Knowtator, a tool built within the Protege Framework

The tool is data driven, so certain annotations never emerged

Velma area residents: say to take Old Hwy to Speedy G to Safely evacuate. Stephens Co Fairgrounds in Duncan for Shelter

[PERSON Velma area residents]: [PERSON Officials] say to take [FACILTY Old Hwy 7] to [FACILTY Speedy G] to safely evacuate. [LOCATION Stephens Co Fairgrounds]
in [LOCATION Duncan] for shelter

Results

If both span and class required to be the same

F-Score of 56.27

Same class and overlapping spans

F-Score of 72.85

Span matching is difficult for all entity classes

Discussions

These are lower than other published results BUT, crisis communication is different from printed text or broadcasts.

Tweets are only 140 characters which liit the syntactic and semantic context

Can lead to ambiguity

This is a new application of annotation practices in a uniquely challenging domain!

The Next Step

Use four different events
- Red River Flood 2009
- Oklahoma Grassfires 2009
- Haiti Earthquake 2010
- Red River Flood 2010

500 tweets were gathered from each even for training data

Preliminary Analysis

Correlation between tweets that were expressed objectively, and those that contributed to situational awareness (Location of hazzard or state of recovery)

Subjective tweets seemed to correlate with tweets that didn't contribute to situational awareness (Sympathy and support/prayer)

Annotation

Two annotatores independently coded tweets with four qualities.

- Communication of situational awareness
- Objective vs. Subjective
- Formal vs. Informal in style
- Personal vs. Impersonal

Expert annotator adjudicated the results into a gold standard.

Situational Awareness Tweets

Provides SA

- Niece and her dad are being evac'd in MWC, fire is headed towards SIL and her boyfriend in Harrah

Does not provide SA

- Tweeps please pray for the families and firefighters battling these crazy fires in Oklahoma

Objective vs. Subjective Tweets

Objective

OHP is responding to Cater County to assist with RD closures and evacuations due to fires in and near Tutums. Approx 20 homes in danger

Subjective

so proud to be from Oklahoma. The outpouring of support for those devastated by the fires is amazing

Formal vs. Informal Tweets

Formal

Staging for fires in Elk City Area is currently at Elk City FD

Informal

landed in fargo today...locals say red river will crest at 43 feet...worse then 97 flood

Personal vs. Impersonal Tweets

Personal

Our best hopes and wishes go out to the folks in Manitoba. As the Red River is about to crest.

Impersonal

Canadian and Oklahoma Counties under RED FLAG FIRE WARNING until 10 p.m. This warning means conditions are ripe for wildfire outbreaks.

Annotation Results

Agreement on objectivity and personal/impersonal codes are consistent

Agreement on formal/informal is lower

- This is more of an abstract idea

Depending on the event, some ideas are more straightforward than with other events.

	Objectivity		Register		Pers./Imp.		SA	
	ITA	κ	ITA	κ	ITA	κ	ITA	κ
RR09	.93	.86	.79	.57	.92	.84	.80	.60
RR10	.92	.83	.90	.80	.95	.89	.91	.81
Haiti	.89	.78	.86	.72	.86	.72	.98	.96
OK Fire	.97	.94	.90	.80	.93	.86	.91	.81

Table 1: ITA and Kappa statistics for inter-rater agreement across all datasets for each annotation category.

ITA is the agreement between annotators

Kappa is the ajusted value of ITA due to 0.5 chance or agreement

Feature Extraction

All non-tweet components were removed
- Hashtags
- URLs
- symbols(RT, @*)

The Stanford part of speech tagger was used to tag features

Machine Learning

All of this was used as training data

Both Naive Bays and Maximum Entropy were used though Maximum Entropy produced a better result

Final Results

Features used	RR09		RR10		Haiti		OKFire		Uniform	
	NB	ME	NB	ME	NB	ME	NB	ME	NB	ME
Unigrams (Baseline)	74.0	82.2	74.4	**89.0**	81.4	81.4	84.3	83.4	79.8	79.3
Unigrams, Bigrams	73.2	81.1	71.0	88.7	79.8	80.2	83.2	82.4	80.3	79.8
Unigrams, POS	75.3	79.1	80.0	87.8	83.3	83.9	83.2	82.6	80.8	83.3
Objective (annotated), POS, unigrams	77.7	84.0	82.2	88.7	82.8	86.5	83.8	84.1	81.0	83.0
Objective (predicted), POS, unigrams	76.5	**84.8**	81.0	88.4	82.2	85.3	79.4	84.7	79.2	82.3
Objective (OpinionFinder), POS, unigrams	71.5	82.4	76.4	85.8	82.0	84.4	80.2	82.6	82.3	81.9
Formal (annotated), POS, unigrams	76.4	82.7	82.2	88.8	84.2	86.2	82.3	**87.9**	80.0	81.8
Formal (predicted), POS, unigrams	75.2	82.1	80.8	87.4	83.3	84.7	79.8	84.4	79.7	81.3
Impersonal (annotated), POS, unigrams	78.2	84.0	82.2	89.0	84.5	86.3	83.0	85.5	83.8	83.5
Impersonal (predicted), POS, unigrams	75.9	83.9	81.4	87.1	83.1	83.7	80.2	85.1	79.8	81.5
All features (predicted), POS, unigrams	76.1	84.1	80.2	88.6	83.5	**88.8**	81.0	87.1	80.2	**84.5**

Table 3: Average 10-fold cross validation accuracies for SA classifier. Best performance for each dataset is in boldface.

Breakdown

High baseline accuracy using only word and raw frequency as a feature

Most emergencies us certain words when conveying situation awareness

- Floods - depth, water level, crested
- Fire - grass fire, wildfire, burn
- all events - safe, evacuation, rescue

Note: using bigrams didn't improve anything

Using POS Tags generally improved overall classification

Specific POS tagging was implemented, but didn't show any improvements

- Adjectives only
- Messages in a specific tone

An increase in performance was seen when the gold standard tag for Objectivity was added as a feature.

Formal/Informal had a similar result as objectivity did and helped improve performance

Overall, All features helped and together produced the best output

Linguistic features

Help varied from event to event, but including Linguistic features reduced the error.

With large events like Haiti, a lot of 'noise' exists with the data. Linguistic features help

Additions

Data Collection

Dataset	Search Terms
Gustav	*gustav, hurricane*
Haiti	*haiti, earthquake, quake, shaking, tsunami, ouest, port-au-prince, temblement, tremblement de terre*
Oklahoma Fires	*okfire, oklahoma, grass fire, grassfire*
Red River 2009	*red river, redriver*
Red River 2010	*fmflood, flood10, red river, redriver, ccflood, fargoflood*

Table 1: Dataset Search Terms

Catagorizing Tweets

1st Pass: remove off topic tweets and tweets that do not include situational awarness

2nd Pass - Coded as social environment, built environment or physical environment

3rd Pass - Adds empirical analysis of datasets from similar events from twitter research on what is communicated in disasters

Second Pass Code	Third Pass Code
Social Environment	Advice - Information Space
	Animal Management
	Caution
	Crime
	Death
	Evacuation
	General Population Info.
	Injury
	Missing
	Offer of Help
	Preparation
	Recovery
	Request for Help
	Request for Information
	Rescue
	Response - Community
	Response - Formal
	Response - Miscellaneous
	Respose - Personal
	Sheltering
	Status - Community/Population
	Status - Personal
Built Environment	Damage
	Status - Infrastructure
	Status - Personal Property
	Status - Personal
Physical Environment	General Area Information
	General Hazard Information
	Historical Information
	Predictions
	Status - Hazard
	Weather

Tagger Agreement is around 0.9 for all passes

Future Work

Explore features to reduce false positive and negative rates

Active learning techniques so less data can cover a wide range of disasters

Integration of propbank style semantic role labeling

Embed into a larger system that analyzes features outside of the tweet text

My Project

Event extraction

- Using sample tweets from the EPIC data
- Make use of annotated tweets