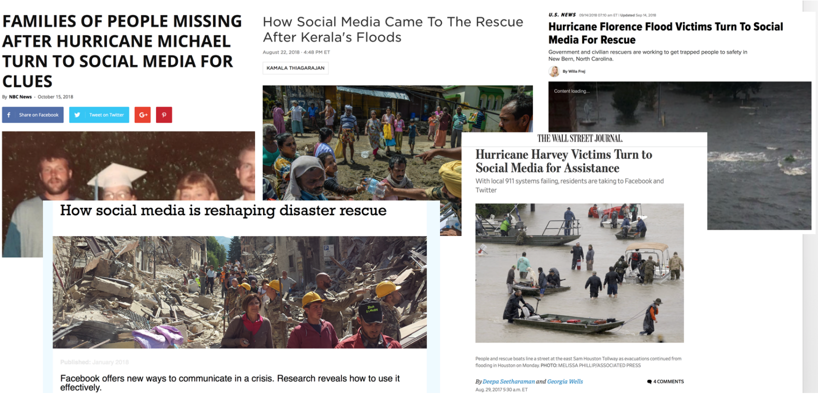
Crisis Eyewitness Tweets Classification

Hongmin Li

Department of Computer Science, Kansas State University

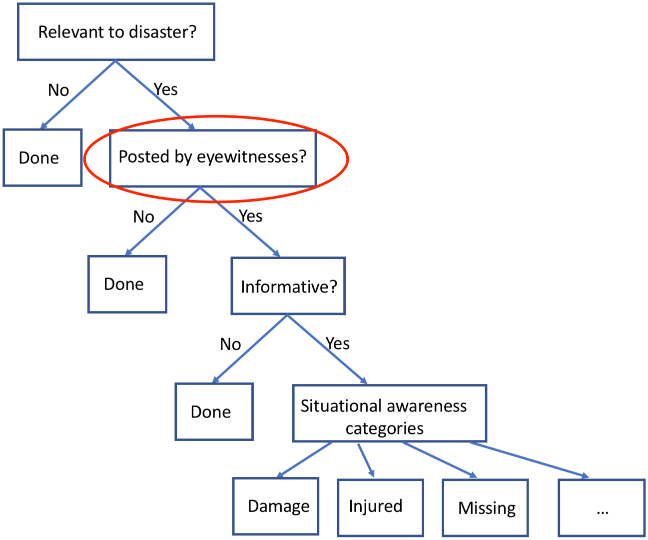
Social Media - Important Tool during Emergencies

First hand information, rescue requests etc. - useful for first responders gaining situation awareness and planning rescue operations



Problem: Information Overload

Auto-filtering to Reduce Information Overload



Labeled Datasets

- CrisisLexT26
- Our own labeled data
- CrisisNLP Eyewitness dataset
- In total about 5,000 English eyewitness tweets for mainly disasters of Earthquakes, Floods, Hurricanes and Wildfires

Labeled Datasets

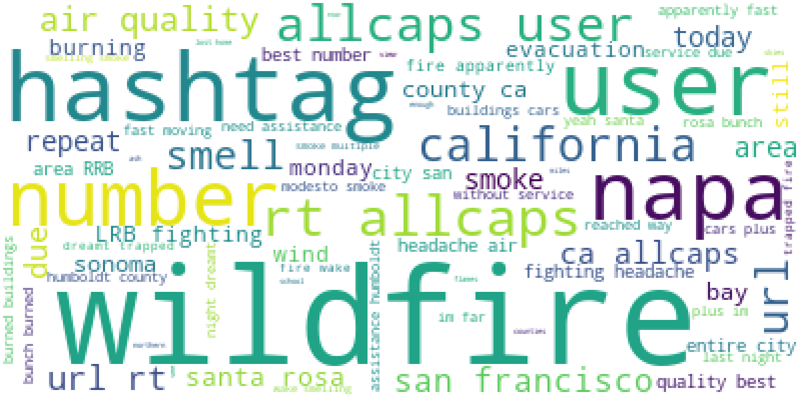
CrisisLexT26	Eyewitness	Non-eyewitness	Total
2013_Alberta_floods	235	736	971
2013_Queensland_floods	144	781	925
2013_Australia_bushfire	95	853	948
2012_Colorado_wildfires	66	886	952
2013_Colorado_floods	65	854	919
2013_LA_airport_shootings	42	878	920
2013_West_Texas_explosion	26	873	899
2013_Boston_bombings	13	901	914
Total	686	6762	7448

Our labeled	Eyewitness	Non-eyewitness	Total
2017_California_Fire	42	2552	2594
2017_Harvey_Hurricane	313	4568	4881
2017_Irma_Hurricane	260	8789	9049
Total	615	15909	16524

CrisisNLP-Sample 1	Eyewitness	Non-eyewitness	Total
Floods	148	113	261
Earthquakes	367	321	688
Hurricanes	296	100	396
Total	811	534	1345

CrisisNLP-Sample 2	Eyewitness	Non-eyewitness	Total
Floods	627	551	1178
Earthquakes	1600	200	1800
Hurricanes	465	1199	1664
WildFire	189	1379	1568
Total	2881	3329	6210

2017 California WildFire



2017 Harvey Hurricane

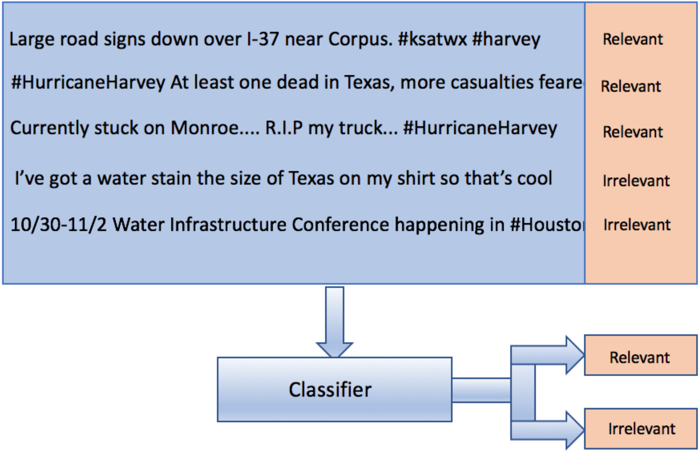
Supervised Learning

Labeled dataset

Large road signs down over I-37 near Corpus. #ksatwx #harve	Relevant
#HurricaneHarvey At least one dead in Texas, more casualtie:	Relevant
Currently stuck on Monroe.... R.I.P my truck... #HurricaneHarv	Relevant
I've got a water stain the size of Texas on my shirt so that's co	Irrelevant
10/30-11/2 Water Infrastructure Conference happening in #Housto	Irrelevant

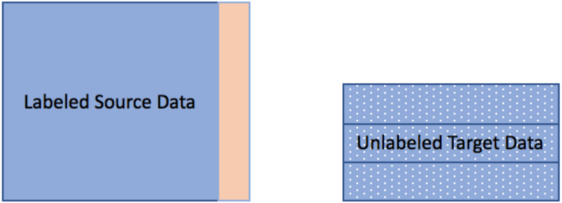
Supervised Learning

Learn a classifier



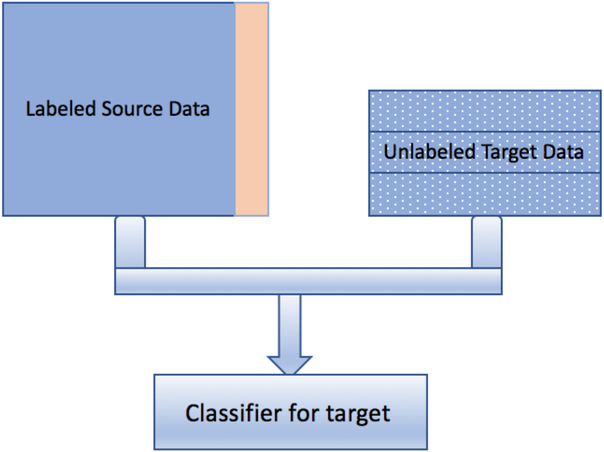
Domain Adaptation

Labeled source data and unlabeled target data



Domain Adaptation

Learn a classifier for the target



Final Goal

- We can build one model for each type of disaster in the training set separately or combine all available eyewitness tweets to just build one model which hopefully can achieve comparable results
- Build a web app with the best classification model
- Integrate Twitter API to allow users to search real-time filtered eyewitness tweets when a disaster occurs

Thanks for your time

Questions?