Tweedr: Twitter for Disaster Response

Zahra Ashktorab Chris Brown Jit Nandi Aron Culotta

August 15, 2013

1 Introduction

I. Context

II. Problem

III. Solution Overview

2 Data

I. Unlabeled data from different disasters

II. Labeling for classification (and uniform vs keyword sampling)

III. Labeling for extraction

IV. Summary statistics (number labeled/unlabeled; number of each class; number by disaster)

3 Methods

I. Classification

II. Clustering

III. Extraction

4 Experiments

I. Classification results

A. overall precision, recall, f1

B. compared with predicting on unseen disasters

C. comparison of sLDA and vanilla classifiers

D. visualize important features (e.g., sLDA graph)

E. list some exemplary good/bad classifications

Correctly identified as damage or casualty
xxxx
Incorrectly identified as damage or casualty
xxxx

II. Clustering results (maybe don't need accuracy, but at least what percent is duplicate)

III. Extraction

A. overall precision, recall, f1, confusion matrix

B. compared with predicting on unseen disasters

C. visualize important features

D. list some exemplary good/bad classifications

5 Related Work

- Extracting Information Nuggets from Disaster-Related Messages in Social Media
- Practical Extraction of Disaster-Relevant Information from Social Media
- Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods
- TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief
- Natural Language Processing to the Rescue?: Extracting Situational Awareness Tweets During Mass Emergency

	All			New Disaster		
Method	**F1**	**Pr**	**Re**	**F1**	**Pre**	**Re**
sLDA	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10
SVM	F1	Pr	Re	F1	Pre	Re
LogReg	F1	Pr	Re	F1	Pre	Re

Table 1: Classification results

	All			New Disaster		
Features	**F1**	**Pr**	**Re**	**F1**	**Pre**	**Re**
All	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10	0.01 ± 0.10
feature1	F1	Pr	Re	F1	Pre	Re
feature2	F1	Pr	Re	F1	Pre	Re

Table 2: Extraction results

6 Conclusions and Future Work

I. Summarize what we did

II. Mention limitations

III. Summarize next steps