Detection and Analysis of Disaster-Related Tweets

Daniel Solomon
EMAIL@mail.tau.ac.il

Gal Ron
galr1@mail.tau.ac.il

Omri Ben-Horin
EMAIL@mail.tau.ac.il

Abstract

TODO

1   Introduction

The popular microblogging service Twitter is a fruitful source of user-created content. With hundreds of millions of new tweets every day, Twitter has become a probe to human behavior and opinions from around the globe. The Twitter 'corpus' reflects political and social trends, popular culture, global and local happenings, and much more. In addition, tweets are easy to access and aggregate in real-time. Therefore, we experience an increased interest in natural language processing research of Twitter data.

As one of the world's most widely used social networks, Twitter is an effective channel of communication and plays an important role during a crisis or emergency. The live stream of tweets can be used to identify reports and calls for help in emergency situations, such as accidents, violent crimes, natural disasters and terror attacks (which we all refer to as 'disasters' in this paper).

In this work we present a model trained to identify disaster-related tweets from other messages, using a natural language processing pipeline adjusted to the special features of Twitter tweets. In addition, we present two experiments conducted on disaster-related tweets. First, we separate *subjective* tweets (example: TWEET) from *objective* reports on disasters (example: TWEET). We also recognize named-entities in disaster-related tweets to enrich our knowledge on the disaster (mostly location).

1.1   Tweets vs. Traditional Corpora

Tweets have some unique features that differ from traditional corpora (such as WSJ corpus). These features should be taken into consideration when implementing natural language processing techniques.

Here's a tweet:
RT This is an #awsome tweet lmao :O
TODO

2   Analysis Wokrflow

keywords   TODO
- A
- B
- C

3   Tweet Classification

keywords   TODO

4   Named-Entity Recognition in Tweets

keywords   TODO

5   Experimenting with Recent Tweets

keywords   Twitter's Search API

6   Conclusions

Future work   TODO

References

[1] KEVIN GIMPEL, NATHAN SCHNEIDER, B. O. D. D. D. M. J. E. M. H. D. Y. J. F., AND SMITH, N. A. Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proc. ACL* (2011).

[2] RITTER, A., CLARK, S., MAUSAM, AND ETZIONI, O. Named entity recognition in tweets: An experimental study. In *Proc. EMNLP* (2011).