Identifying Fact-Checkable Tweets

Siddharth Ghelani[1], Divyank Barnwal[1]

and Rohit Krishna[1]

[1] University of Engineering and Management, Kolkata, India
sghelani29@gmail.com

Abstract. Microblogging sites like twitter are increasingly playing an important role in real time disaster management. There are many miscreants who would want to derail the rescue and relief operation by spreading rumors and thereby creating panic. Therefore, it becomes imperative to correctly identify such rumors and nip them in the bud. This paper describes our approach on "Identifying factual or fact checkable tweets" as an attempt in the shared task of the Microblog Track at Forum for Information Retrieval Evaluation (FIRE) 2018 [6]. Our approach uses a version of Stanford's POS Tagger [1] trained especially on tweets to extract features from the tweets for training the classifier. The system was evaluated on the Twitter dataset consisting of 50000 odd tweets provided by the FIRE 2018 shared task. Our system showed encouraging performance. We had submitted two separate runs, each using a different approach. The performance in each case is separately mentioned and explained.

Keywords: Microblog, Disaster, Classification.

1 Introduction

Social media has become increasingly important in disseminating real-time information in case of disaster outbreaks. Efficient processing of information from social media websites such as Twitter can be challenging due to the noisy nature of the tweets, but if pursued properly can be very helpful in disaster management. A lot of research regarding extraction of situational information from microblogs during disasters already exists [3, 4, 5] .Many such Natural Language Processing techniques have been used in the past to solve this problem. We have modelled this problem as a classification task and use SVM (Support Vector Machine) to achieve the said classification. SVM has been shown to classify text data very efficiently [2] in the past .To give the reader some sense of the of the problem at hand, we present some examples of fact checkable and non-fact checkable tweets.

Example of fact-checkable tweets

1. ibnlive:Nepal earthquake: Tribhuvan International Airport bans landing of big aircraft
[url: https://twitter.com/Michael_Vasanth/status/594840493244194816]

2. #Nepal #Earthquake day four. Slowly in the capital valley Internet and electricity beeing restored . A relief for at least some ones
[url: https://twitter.com/navyonepal/status/592901901479505920]

3. @mashable some pictures from Norvic Hospital *A Class Hospital of nepal* Patients have been put on parking lot.
[url: https://twitter.com/masterashim/status/592089990512807936]

4. @Refugees: UNHCR rushes plastic sheeting and solar-powered lamps to Nepal earthquake survivors
[url: https://twitter.com/AbdulHai23/status/643051227991904256]

5. @siromanid: Many temples in UNESCO world heritage site Bhaktapur Durbar Square have been reduced 2 debris after recent earthquake
[url: https://twitter.com/siromanid/status/594876694592299009]

6. @SamitLive: Nepal has requested for Drinking water. @RailMinIndia has decided to send 1 Lak liter of Rail Neer over night.
[url: https://twitter.com/SamitLive/status/591999777237180416]

Examples of non-fact-checkable tweets

1. Students of Himalayan Komang Hostel are praying for all beings who lost their life after earthquake!!! Please do...
[url: https://twitter.com/komang28645362/status/596961034772029441]

2. We humans need to come up with a strong solution to create earthquake proof zone's.
[url: https://twitter.com/_GraceBaldwin/status/1042075740982915074]

3.really sad to hear about d earthquake. praying for all the ppl who suffered & lost their loved ones. hope they get all the h…
[url: https://twitter.com/vrinda_90/status/591954205696331776]

4.@Gurmeetramrahim Msg helps earthquake victims
[url: https://twitter.com/drtinamehta/status/739792214599966720]

5.Nepal earthquake Students light candles offer prayers for victims: Students in Amritsar led a candle light vig...
[url: https://twitter.com/nepalnewsnet/status/592658008066359297]

6.I am so deeaking scared omg i dont even know what should i tweet.. This could possibly be my last tweet if the earthquake doesnt stop
[url: https://twitter.com/blackmoondior/status/592662979864350720]

2   Task Definition

A set of fifty thousand tweets were given and the task is to classify each tweet as either fact checkable or non-fact checkable. The tweets given in the task were posted during the Nepal earthquake in April 2015.

2.1   Data and Resources

This section describes the dataset and resources provided to the shared task participants. The organizers provided a text file containing 50,068 tweet identifiers that were posted during the Nepal earthquake in April 2015. A Python script was provided to download the tweets using the Twitter API into a JSON encoded tweet file, which was processed during the task. A set of 80 fact checkable tweets was also given for testing the model.

3   System Description

3.1   Preprocessing

The raw tweets from jsonl file were taken into a separate file. All the tweets were pos tagged using a special version of Stanford's POS Tagger trained just on tweets. These tweets were broadly classified into four files one containing only the tweets with Retweets(RT) in them, one with tweets containing numerical values in them, one with tweets containing more than 2 proper nouns and other containing the rest of the tweets. The first three files were carefully and minutely examined to further filter out the redundancies and repetitions.5000 tweets were selected from this corpus with 1500 from the first file,2000 from the second file and 1500 from the third. By this stage, we had a corpus of 5000 fact-checkable labelled data for training. We separately examined the fourth file with the same objective. 5000 tweets were handpicked from this generating a corpus of non-fact checkable tweets.
The stopwords were filtered out from each of the tweets using NLTK (Natural Language Toolkit).

For the first submission

A Bag of Words model was created with the 26000 proper-nouns(obtained by POS tagging of the tweets) as features present among all the 50000 tweets. Top performing 10000 features were selected using the SelectKbest class in the sklearn library and a "linear" SVM model was trained with these 10000 features. The model made predictions on the corpus of 80 tweets provided with 80% accuracy. The following results were obtained when the model was tested on all the 50,000 odd tweets.

For the second submission

A Bag Of Words model was created with 6000 features obtained using the TFIDF vectorizer available in the sklearn library. A linear SVM model was trained using these features. The model made predictions on the corpus of 80 tweets provided with 93% accuracy. The following results were obtained when the model was tested on all the 50,000 odd tweets.

Run ID	UEM_DataMining_CSE_run1	UEM_DataMining_CSE_run2
Run Type	Automatic	Automatic
Precision@100	0.6400	0.6800
Recall@100	0.1069	0.1427
MAP@100	0.0340	0.0378
MAP Overall	0.0767	0.1178
NDCG@100	0.5237	0.5332
NDCG Overall	0.5276	0.6396

4   Conclusion

In this paper, we presented a brief overview of our system to address the issue of fact check ability of microblogging data. As a future work, we would like to explore more sophisticated techniques to classify the microblogs according to their fact check ability so that we can minimize the menace of fake news and false rumors in catastrophic situations.

References

1. Derczynski, L., Ritter, A., Clarke, S., Bontcheva, K.: "Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data" 2013. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing, ACL - Association for Computational Linguistics
2. Joachims, T.: Text Categorization with Support Vector Machines 1997: Learning with Many Relevant Features. In: Springer.
3. Imran, M., Castillo, C., Diaz, F., Vieweg, S.: Processing Social Media Messages in Mass Emergency: A Survey 2015. In: ACM Computing Surveys 47,4
4. Rudra, K., Ghosh, S., Goyal, P., Ganguly, N., Ghosh, S.: Extracting Situational Information from Microblogs during Disaster Events 2015: A Classification-Summarization Approach. In: Proc. ACM CIKM - Conference on Information and Knowledge Management
5. Basu, M., Ghosh, K., Das, S., Dey, R., Bandyopadhyay, S., Ghosh, S.: Identifying Post-Disaster Resource Needs and Availabilities from Microblogs. In: Proc. ASONAM - Advances in Social Network Analysis and Mining
6. Basu, M., Ghosh, S., Ghosh, K.: Overview of the FIRE 2018 track: Information Retrieval from Microblogs during Disasters (IRMiDis). In: Proceedings of FIRE 2018 - Forum for Information Retrieval Evaluation (December 2018)