Actionability Detection on Twitter During Disastrous Events

Alessandro Rennola
arenno2@uic.edu

Edoardo Savini
esavin3@uic.edu

INTRODUCTION

Disasters are unforeseen, calamitous event that have caused great damage and losses in term of goods and lives since the dawn of our existence. Societies have always attempted to limit the impact of such calamities by developing measures that address the initial impact and handle the consequences.

Our project focuses on this last purpose. We believe that collecting and analyzing information about a certain disaster event would increase the understanding about the needs of the affected communities and, consequently, help responders plan the next actions to be taken to increase safety and limit additional damages.

Thanks to the advent of information technology, social networks can be a relevant resource to find this kind of information, as they have become the first source of knowledge when dealing with anomalous events.

However, the amount of data social media provide is huge and contains a lot of irrelevant messages. Thus, retrieving useful data from them may result in finding a needle in haystack.

Our approach proposes to overcome these shortcomings through Information Retrieval techniques, generating a set of words recurrent in Actionable tweets, expanding them and scoring them according to some relevance criteria that account also for word frequency (tf-idf) in relevant and irrelevant tweets.

RELATED WORK

To better understand the concept of "actionable tweets" we refer to the work done by (Derczynki et al. 2017 "Helping Crisis Responders Find the Informative Needle in the Tweet Haystack") defining it as an "Informative message that would be helpful to the authorities or the general public with saving lives or assisting the other response teams in dealing better with the incident".

Our research to detect actionable information from disaster data was also influenced by some of the previous work by (Munro 2011, "Subword and spatiotemporal models for identifying actionable information in Haitian Kreyol") and (Spasojevic et Rao 2015, "Identifying Actionable Messages on Social Media").

DATASET

Our analysis is focused on Twitter. The dataset consists of a set of tweets regarding hurricanes and earthquakes that affected US and Mexico in 2017. They have been selected by using a relevance classifier (general classifier trained on CrisisLex26 data) to identify those tweets that are relevant to a disaster and only English tweets have been kept.

It includes over 2M tweets about the Irma hurricane, 200K about Harvey, 500K about Maria, and 200K about Mexico earthquake. There are also other files with some miscellaneous data from 2 or three of these catastrophes.

Notice that the files are not labelled for Actionability, hence a manual research of about 600 tweets has been made to detect some recurrences or other relevant information in actionable tweets.

METHOD

Our experiment can be divided in 4 main phases: preprocessing, synonyms expansion, scoring and similarity evaluation.

1. Preprocessing

In this first phase, our purpose is to create a vocabulary data structure suitable for the computation of tf-idf and able to match the scored words that will be found in step 2.

To do that we decided to avoid stemming and lemmatizing, as it could cause some loss of information and decrease the precision of our research. In fact, in our case, some words are more relevant when conjugated than in their "root form" (e.g. The word "stuck" will retrieve more actionable tweets than its verb infinitive "stick").

For the same reason, we also decided not to remove stopwords, as for example some possessive forms like "my" could result particularly informative.

So, after removing punctuation marks (included hashtags '#'), we collected all the words in the corpus with their frequency in a vocabulary that comprises all the input datasets (therefore including all the words in one single dictionary, regardless of the calamity) and then scored each term using tf-idf. There is a flag in the tf-idf library, that allows the user to decide whether to perform tf-idf normalization or not; the results are quite similar in both cases, but normalization provides, generally, better results.

2. Synonyms expansion

Since the dataset was not labelled for actionable tweets, we have manually analyzed about 600 random tweets detecting the actionable ones. Doing that, we have discovered some words that are frequent in actionable tweets and collected them into a file.

Our approach aims to filter the dataset using these words to retrieve a ranking with significant actionable tweets.

We have also decided to expand the words in the list with some of their synonyms (also scored) to evaluate whether adding some synonyms would increase accuracy.
For this purpose, we used the WordNet Library, a package within the NTLK library.

Our expansion phase consists in:
- Assigning to each term in the manually created list a score equal to 1;
- Expanding each term in the former list with its synonyms for each semantic field
- Weighting the synonyms with a score that corresponds to the maximum value (if the synonym appears multiple times with different semantic meanings) of "wup_similarity" between our term and the synonym itself
- E.g.:
    a)   first word selected from the list: 'lost';
    b)   the expansion of 'lost' ("wordnet.synsets(lost)") brings up 22 terms among which there are 7 definition of 'lose' ('lose.v.01', 'lose.v.02', 'lose.v.03', 'lose.v.05', 'lose.v.06', 'lose.v.07', 'lose.v.08') and 5 definitions of 'lost' itself ('lost.a.01', lost.a.03', 'lost.a.04', 'lost.s.05', 'lost.s.06');
    c)   for the first definition of 'lost' we compute its 'wup_similarity' with the other 17 (=22-5) expansions;
    d)   we do the same procedure for the other four definitions of 'lost' updating the values found in point c) only if their similarity is higher than the one found before;
        (e.g. if 'lost.a.01'.wup_similarity('doomed.n.01')= 0 and
        lost.s.05'.wup_similarity('doomed.n.01') = 0.5 then update score['doomed'] = 0.5);
    e)   at this point there will be 7 different scores for each definition of 'lose', we take the maximum value;
    f)   we add the new term to the final map to be retrieved and restart a) with the next word from the list
- At the end, the procedure returns a map containing all the terms in the list, together with their associated synonyms, scored using the method described above.

It is possible to evaluate the performance of tf-idf without expanding the terms with the synonyms by setting the flag 'enabled' to False, we are going to present the results for both methods and compare them in step 4.

3. Scoring the results

Once we obtain our weighted list of relevant words, for each term in that list we performed a multiplication between its weight retrieved in the phase two and the tf-idf weight computed in step 1 (=tf-idf computation considering the list as "query"). The results of this calculation represent the final score on which we base the final ranking.

After that, for every tweet that contains one or more words from our expanded list, we performed the sum of all the weights and ranked them in decreasing order according to that value.

4. Evaluation

Since we are not in possession of a set of labeled data to compare our results with, we have performed our evaluation manually, considering the TOP-N ranked tweets retrieved from our list and comparing the results with the ones obtained after thesaurus expansion in phase 2.

We report the results obtained for the best 10 guesses, when building the model using only 'Chiapas_filtered' and 'HarveyMaria_filtered' dataset.

Legend: Non-Actionable, <u>Actionable</u>, *Border Line*
EXPANDED LIST:
- <u>beaumont woman unable to reach family in puerto rico as the island remains without power via 12newsnow</u>
- <u>my family on puerto rico has been without power since harvey hit and they have no way to get to the states</u>
- <u>please help out</u>
- <u>earthquake juchitan oaxaca residents beg please we need urgent help people hurried alive no aid no help</u>
- <u>not like houston pr is an island tough for neighbor to help neighbor limited resources total destruction send more shipspuertorico</u>
- *prayers to mexico enduring the earthquake i know a lot of my listeners out here have family there please call check on them*
- no power in puerto rico no drinking water in san juan why cant we give pr just a fraction of the concern we gave houston maria
- my heart is breaking for mexico may you all get support comfort shelter and help from everywhere in the world mexicoearthquake Mexico
- chiapis family earthquake help earthquake prayformexico
- <u>please send ur thoughts to my family in puerto rico please get ready to donate to them like you did for Houston</u>

NOT EXPANDED LIST:
- <u>beaumont woman unable to reach family in puerto rico as the island remains without power via 12newsnow</u>
- <u>my family on puerto rico has been without power since harvey hit and they have no way to get to the states</u>
- <u>please help out</u>
- <u>earthquake juchitan oaxaca residents beg please we need urgent help people hurried alive no aid no help</u>
- <u>not like houston pr is an island tough for neighbor to help neighbor limited resources total destruction send more shipspuertorico</u>
- *prayers to mexico enduring the earthquake i know a lot of my listeners out here have family there please call check on them*

- my heart is breaking for mexico may you all get support comfort shelter and help from everywhere in the world mexicoearthquake Mexico
- chiapis family earthquake help earthquake prayformexico
- <u>please send ur thoughts to my family in puerto rico please get ready to donate to them like you did for Houston</u>
- please rt

From the comparison in the previous step we can see that both the expanded and the non-expanded lists perform quite well. In both case, indeed, precision reaches 70% for the first 10 results. This was quite surprising if we consider the sparse nature of the label and the unsupervised nature of the tf-idf. In fact, we could say that the problem belongs to 'needle in a haystack' category, this implies that the distribution of 'positive' and 'negative' classes is not balanced.

Considering that we do not have labels it is quite difficult to evaluate recall, however, since the amount of actionable information in the datasets is extremely limited, we would expect recall to be reasonably high.

CONCLUSION

Finally, we think that the tf-idf approach turned out to be a discrete starting point, given the available means and the simplicity but effectiveness of the method. Nonetheless, there is room for improvement:

- We would like to work on associations of words. Often, there are terms that assume more significance when together with other terms. Examples could be: 'my daughter', 'no electricity', 'need help', etc.

- We would like to work more thoroughly on the actual effect that normalizing the tf-idf score would cause in the scores. From a superficial point of view, we noticed an improvement in performance, but we would like to document it to a larger extent.

- We would like to work on negation as part of speech analysis, including and differentiating our algorithm behavior in case of negation in a surrounding window of M-words. [Deep contextualized word representations, Matthew E. Peters et al.]

- We would like to work on the preprocessing of the datasets. When we went through the data we found out that many tweets are unrelated to the crisis dominion, some of them were not in English, and there were still some retweets.

- We would like to work on the correlation between information content and length of the messages, it appears that shorter tweets convey a smaller amount of information. For example, tweets like 'please help out' and 'please rt' in our list are more difficult to classify.

- We would like to work on scalability issues. Tf-idf is easily extensible, but it faces problems whenever the dimensionality of the data increases too much, we have tried to run different

instances of the algorithm with different input data, and we noticed that complexity explodes after a few hundred thousand tweets. This limits how much we can exploit the datasets that we could employ. Therefore, we have tested the algorithm on a small set of data: the datasets HarveyMaria and Chiapas, roughly 20'000 tweets.

A major development could be performing a preliminary clustering phase, in order to define the boundaries of the classes and obtain some preliminary data, subsequently proceed with our method or more sophisticated techniques, like Convolutional Neural Networks [Derczynki et al. 2017].

FINAL REMARKS

We think that this could be the starting point for very interesting developments. The purpose of helping people in need stands among the newest and most exciting topics that only contemporary Computer Science can pursue, with state-of-the-art techniques. Our method is both simple and effective and has large improvement margins. We are very grateful for the resources we were kindly provided.

REFERENCES

[1] R. Munro, "Subword and spatiotemporal models foridentifying actionable information in haitian kreyol," inProceedings of the Fifteenth Conference on Computa-tional Natural Language Learning, 2011, pp. 68–77

[2] N. Spasojevic and A. Rao. Identifying actionable messages on social media. In IEEE International Conference on Big Data, IEEE BigData '15, 2015

[3] Derczynski, Leon & Meesters, Kenny & Bontcheva, Kalina & Maynard, Diana. Helping Crisis Responders Find the Informative Needle in the Tweet Haystack. (2018)

[4] Peters et al., 2018 Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word prepresentations. In Proceedings of NAACL, 2018.