Determining Urgency in the Content of Disaster Tweets

Undergraduate Researchers:

Angela Timochina, Arpita Kumari, Elijah Rodriguez Beltran

Supervisor:

Cornelia Caragea

Early Research Scholars Mentors:

Joe Hummel, Renata Revelo Alonso, Taha Khan

1. Research Context and Problem Statement

Research into machine learning has been growing at an increasing rate within the computer science community. As machine learning has evolved, new methods have developed that allow for more efficient and accurate data retrieval. Furthermore, there is interest in the extraction of keywords or phrases in tweets that would be informative to the researcher and potentially a future organization in utilization for applications, etc. In particular, extraction of relevant disaster related data has become a goal in this field. Since tweets are short, loosely structured sentences that allow for frequent updates, this provides a robust data set for analysis.

During disasters communication usually becomes difficult and unreliable due to severe conditions. Many disasters result in cases of injuries or even death because emergency responders become unreachable. During Hurricane Florence one family was able to avoid such a fate when their tweet went viral. Their house was quickly filling with water when the hurricane hit and they were trapped in their attic. Since many people were being affected by this disaster, 911 calls were jammed, and the family was unable to ask emergency responders for help (1) . If this tweet had not gone viral, most likely the family wouldn't have been rescued and would be added to the statistics of those who die. Many people face similar situations, and social media is sometimes their last resort. This is highly motivational for creating systems/models that can extract informative tweets in order to be used to inform responders, emergency volunteer services, or any people able to help.

There are many aspects to classifying informative tweets, and defining what specifically makes it informative. Thus far, research into the insight of what makes an informative disaster tweet results from looking at the amount of retweets, the content of tweets, and the distance of the tweet from the disaster. In particular, content based analysis of disaster tweets is focused on keywords or phrases. This has been helpful in defining if a tweet is informative but has not been used to determine the urgency of an individual asking for assistance (2).

The problem in disasters, is not just the disaster but the emergency response rate. In any emergency, time of response can determine whether you survive or not. Therefore, it is important to find ways to classify the urgency of an emergency so that responders may be more efficient and effective in distributing resources or coming to aid.

Current research that touches on the topic of urgency focuses mainly on position data. GIS (Geographic Information System) is a system for analyzing and displaying positional data on a global scale. Social media users who voluntarily contribute to geospatial analysis are called "micro-mappers" (5). Along with spatial data provided by these micro mappers, there are other noticeable trends that are particularly insightful. A case study of a flood in South East Queensland, Australia examines the content of tweets related to flooding disasters, and how valuable the information gained from these tweets was in relation to disaster management. There is a direct correlation between the severity of the situation and the number of tweets concerning the disaster. In this study they focused on the recurrence rate of a tweet as a way of classifying its urgency.

Although this method is helpful, there are limitations. Looking at retweet rates as a way of categorizing urgency could potentially be faulty if a tweet isn't retweeted. Furthermore, not only is the rate of retweets informative on urgency, or the amount of tweets coming from a specifically affected area but the content of a tweet can be informative to the level of urgency.

2. Proposed Solution

The solution to improving models and modes of defining urgency is to look at the linguistics of disaster tweets. In particular, for our purposes, we analyzed keywords determined by hashtags. However, there is research that is conducted into various linguistic aspects to create a criteria for a tweet that warrants aid. Through looking at specific adjective ratios, and frequencies in comparison with word bank data bases. However, urgency ratings developed by these findings is still subjective (3). By looking at the keywords and hashtags, we hope to determine whether there is a vocabulary that can describe urgency as well as context, and be able to use this to categorize the level of urgency in the tweet.

We will be going through a series of 1000 tweets from various disasters and annotating them the following way:

- Urgent
    - Urgent within a few minutes
    - Urgent within a few hours
- Not Urgent

If we define a tweet as urgent, we will determine if it requires response within minutes, or hours. In addition, if it is classified as urgent, we will also annotate hashtags we think would be useful for identifying this tweet and summarizing the tweet within a phrase or a few words.

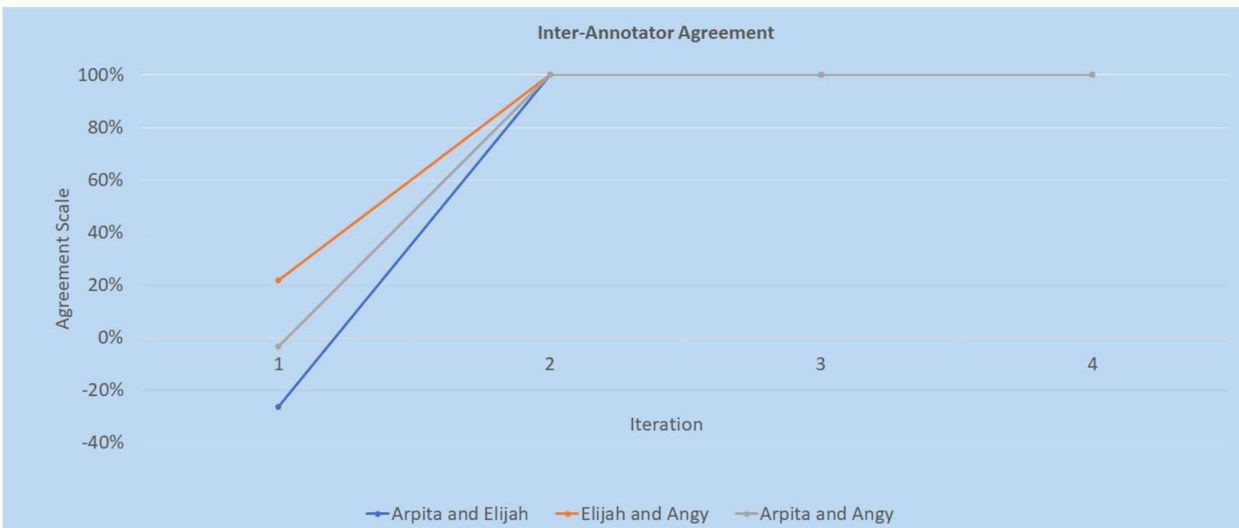
This is important because currently, we can see that when a disaster takes place, there are many tweets, but not all of them have accurate information. And because of this, the right resources don't reach affected people in time. We are trying to see if an urgency component is impactful in this case and if hashtags have a relationship with efficient responses to these tweets as well.

We also annotated other aspects of the Twitter data in search of patterns. In some cases, a Tweet was flagged and met disaster criteria but actually is an informative Tweet created by local government responses (3). During our annotation we will look for contextual relevance as well as urgency in search of informative patterns.

Through this process we hoped to find that there are ways to classify urgency based on the linguistics/content of the tweet. Based on the results, we developed a vocabulary/dictionary for classifying urgency to add into Dr.Caragea's existing model (4), or we will determine what prevented us from building this vocabulary.

3. Annotations

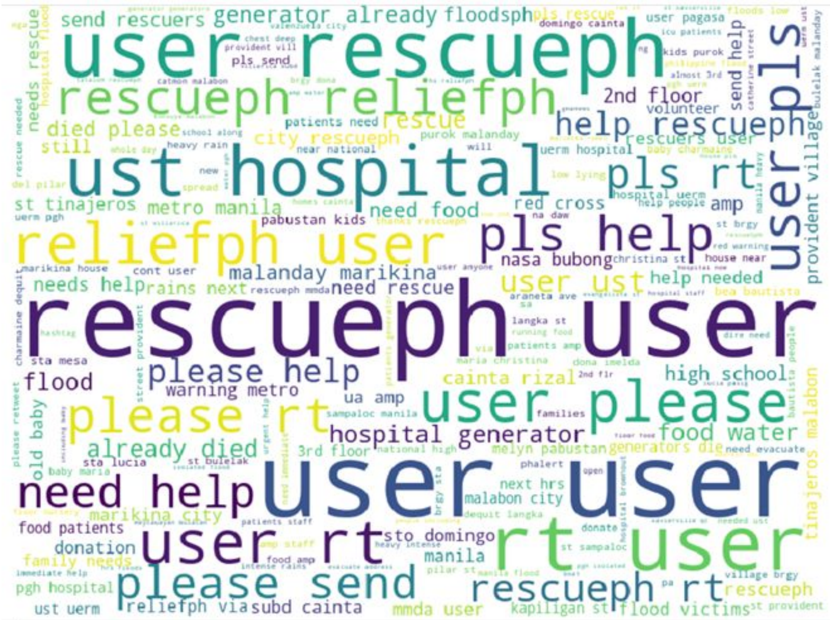
To ensure the development of objective criteria for urgent classification, we each annotated a data set of over 650 tweets concerning a hurricane in the Philippines. Upon the end of this original annotation we discovered discrepancies between each of annotations, however there existed enough similarities to warrant more research. Thus we decided to create specific criteria and narrow down our search for rating the urgency of a situation, to using a binary classification of urgent vs. not urgent. This next round of annotation saw greater similarity between annotations. To ensure our findings were objective, we utilized Cohen's Kappa to calculate the inner annotator agreement between annotations, in order classify the objectivity of our criteria. While this agreement was much greater than the first round of annotations, we decided to narrow down our criteria even further by eliminating certain themes from being urgent. Here is a graph of a compilation of our inter annotator agreement:



The final set of annotations had a criteria of being only actionable requests from humans (i.e no mentions of animals, no news related tweets, and volunteer requests/foreign aid). From this set of annotations we set out to find the common themes between urgent tweets. This heavily involved natural language processing, and phrase extraction.

4. Phrase Extraction

To begin our search for signifying features we looked at a broad overview of the commonalities between urgent tweets using a wordle graphic. The large items as seen below appear most frequently. Many of the largest items share commonalities, while the much smaller items are more specific and contextual.



Since many of these larger items are related, or differing by case or tense, we then began using natural language processing for the purpose of extraction. By lemmatizing every word, we removed any contextual discrepancies allowing for a more accurate view of common features between urgent tweets. The following phrases are extracted frequencies which occur with in more than 12% of urgently annotated tweets:

(' flood', 6)   12.76595744680851 %

(' generator', 6) 12.76595744680851 %

(' water', 7) 14.893617021276595 %

(' amp', 8) 17.02127659574468 %

(' rescue', 8) 17.02127659574468 %

(' ust hospital', 10) 42.5531914893617 %

(' please', 12) 25.53191489361702 %

(' food', 14) 29.78723404255319 %

(' retweet', 14) 29.78723404255319 %

(' help', 15) 31.914893617021278 %

(' user', 17) 36.17021276595745 %

(' rescueph', 27) 57.446808510638306 %

5. Machine Learning Model

Many of these common features were related, allowing for us to form 4 types of urgent features. This grouping is the type of information needed to run a machine learning model on the tweets. By creating a feature vector, where each value of the feature is indicative of whether or not that feature exists in the corresponding tweet. This type of vector is suitable for a Guassian Naive Bayes model, which bases a prediction on probability of a certain class given the existence of given features. Below are the 4 feature types and their corresponding bag of words:

People: [' family', ' patient', ' people', ' staff'],

Place : [' floor', ' house', ' st'],

Hashtag : [' reliefph',' rescueph'],

Need: [' food', ' generator', ' help', ' rescue', ' water']

We utilized an annotated sample of 150 tweets to train our model. And then tested the model on the larger 650 tweet sample. Initially with a binary feature vector (indicating only if a feature exists in a tweet) we achieved 72.8% prediction accuracy. By then using frequency (counting the number of each feature existent in the tweet) we achieved 75.8% accuracy.

6. Conclusion

Through annotation we concluded it is possible to develop criteria to determine the urgency of a disaster related tweet. Although the Naïve Bayes model wasn’t perfectly accurate, with more detailed phrased extraction this accuracy could very well improve. If we were to continue with this research in the future, we hope to be able to apply the same annotations and analysis procedure on a larger scale so that we can examine how the trends differ amongst other natural disasters - such as earthquakes, tornados, tsunamis etc. It would also be interesting to see how the trends differ based on locations in the world. Overall, further research into machine learning to determine urgency is promising.