CS272 Project Pitch: Generalizing Humanitarian-Centric Topics from Twitter Crisis Data for Multilabel Classification and Tagging
Team: NLTweetRelief

Sam Showalter
UC, Irvine
showalte@uci.edu

Edgar Robles
UC, Irvine
roblesee@uci.edu

Preethi Seshadri
UC, Irvine
preethis@uci.edu

1 Project Overview

1.1 Problem Setup

Swift and comprehensive response to disaster situations is crucial for maintaining safety in society. However, it can be difficult to quickly understand the full extent of a disaster, as severity may not immediately be known. With the proliferation of social media giving everyone unfettered access to the internet, disasters are often first reported on Twitter and other social, mobile-based platforms. Improvements in deep learning have enabled scientists to track, tag, and categorize Tweets such that disaster response may be swift. In particular, connecting social media information about disasters to humanitarian causes (loss of life, injury, caution, etc.) is an ongoing body of research. However, with existing datasets few researchers have framed the problem as a multilabel classification, assuming one semantic context per tweet or message. In social media platforms without character limits (Facebook, Instagram, etc.), it is both possible and likely that messages about a crisis may span multiple topics. Models that can tag text with several labels, as well as localize the information relevant to that tag, are indispensable for organizing information about a crisis for later dissemination.

Moreover, virtually no datasets exist that model crisis message classification as a multilabel objective. This is fundamentally flawed; nearly all humanitarian topics are tightly correlated in crisis messages, making organization of information difficult and topics entangled (e.g. loss of life statistics may include aid information, injury, etc.). In society, the separation of crisis events by humanitarian topic is essential as response agencies tends to naturally stratify along humanitarian needs (injury, donation, infrastructure, etc.).

Multilabel classification has been done before (Schulz et al., 2014), however, it was done using TF-IDF vectors with a neural network. Unsupervised information extraction from crisis twitter data has also been used to rank the importance of tweets (Interdonato et al., 2018). In addition, methods to decide whether tweets during crisis are relevant or not exist (Kruspe et al., 2020), as well as methods to map crises using twitter's geolocation data (Middleton et al., 2014) and create reports from extracted information (Di Corso et al., 2017).

However, none of these approaches directly tackle the task of explicitly separating crisis information by humanitarian topic in a multilabel objective. This is important because report generation, relevance, and information extraction performance will suffer if information is not semantically clustered. Below, we propose a method that would allow us to leverage data augmentation to generalize a standard classification objective for crisis tweets into a multilabel information extraction tool applicable to text beyond social media.

1.2 Proposed Approach

On April 8, 2021, HumAID (Imran and Ofli, 2021) - the largest collection of disaster related tweets - compiled and annotated for a variety of humanitarian sub-incidents (loss of life, injury, property damage, etc.), was released. We feel this presents an unique opportunity to develop a multilabel NLP system that can also extract text relevant to each detected tag. Since each tweet in our dataset is aligned with a single humanitarian label, we intend to define a data augmentation approach protocol that samples sets of tweets (1 to k). By randomly sampling a varying number of tweets, we then will create *passages* that will serve as our augmented training samples. After training this data on all disasters up to 2019, we will evaluate the system on *passages* from disasters occurring from 2019 onward. There are several benefits of this data augmentation objective, including a dataset of virtually

unlimited size. We ensure in our augmentation that tweets remain clustered by crisis.

We will likely make use of a transformer-based model trained on general text and may also embed the original tweets using a crisis-specific embedding model (Nguyen et al., 2017). As part of our experimentation, we will qualitatively apply our model to non-tweet text to qualitatively examine its generalization to crisis information in news articles and other sources.

1.3 Evaluation Plan

For the traditional classification performance on the multilabel objective, a log-loss or cross entropy can be utilized to extract how well the model generalizes in its multilabel objective. In addition, a word error rate across the extracted messages will determine how precise and complete the textual extractions were for each label. Baseline performance on the multilabel objective can be conducted with a model that tags each word with its humanitarian topic with the ultimate goal of recovering the original independent tweets of the HumAID dataset. Once a baseline performance has been attained, we intend to manually weight the importance of different classification objectives based on their humanitarian severity (e.g. loss of life would be considered more severe than infrastructure damage). Our original model will be evaluated under this regime, and then a new system will be tuned for the weighted objective and compared to the naive baseline.

Lastly, qualitative analysis will be conducted to see how well this multilabel model adapts to non-twitter text when applied to longhand articles documenting the events of a crisis. Ideally, this multilabel crisis classifier would be able to extract and organize information from all text data sources, not just Twitter and social media. To that end, we will also tune our model pretraining to attempt to bridge the lexical gap between Tweets and more proper news correspondences.

1.4 Potential Challenges

No members of our group have directly conducted research in Natural Language Processing, and only a few members have worked with state of the art language models (i.e. Transformers). It will take some time for us to understand the strengths and weaknesses of these SOTA models as well as how they fit to our objective. At the same time, our experiments are fairly reliant on our data augmentation protocol, which, though promising, may become the source of performance issues. Lastly, generalizing crisis data extraction to non-Twitter corpuses may results in performance dips as the lexicon's between these two domains can vary significantly. With that said, we maintain this approach is both promising as a starting point for future work and as an implementation in its own right.

1.5 Plan of Work

Currently, we have a basic data loader that reads and organizes our crisis Tweet information. Moreover, we have an environment that will support GPU accelerated training and, if needed, additional parallelization across machines. We hope that by the end of our proposal we can have a fairly basic, pre-trained model finetuned for our objective and producing reasonable performance on the unilabel crisis tagging objective. From there, we think that it is reasonable to expect another implementation, this time including the data augmentation protocol, to establish baseline performance on the multilabel objective. It is unclear how much additional qualitative and qualitative finetuning we will have time for after this, but ideally we would like to establish at least one qualitative result and a few model / data ablations, ultimately honing in on a tuned model.

1.6 Computational Requirements

Our datasets will likely fit in 10GB of RAM during experimentation and we will probably leverage GPU support via one of the ICS Computing servers available to us. Storage space for our solutions will represent less than 2 GB.

References

E. Di Corso, F. Ventura, and T. Cerquitelli. 2017. All in a twitter: Self-tuning strategies for a deeper understanding of a crisis tweet collection. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 3722–3726.

Firoj Alam Umair Qazi Muhammad Imran and Ferda Ofli. 2021. Humaid: Human-annotated disaster incidents data from twitter. In *15th International Conference on Web and Social Media (ICWSM)*.

R. Interdonato, A. Doucet, and J. Guillaume. 2018. Unsupervised crisis information extraction from twitter data. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 579–580.

A. Kruspe, J. Kersten, and F. Klan. 2020. Review article: Detection of informative tweets in crisis events.

Natural Hazards and Earth System Sciences Discussions, 2020:1–18.

S. E. Middleton, L. Middleton, and S. Modafferi. 2014. Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, 29(2):9–17.

Dat Nguyen, Kamela Ali Al Mannai, Shafiq Joty, Hassan Sajjad, Muhammad Imran, and Prasenjit Mitra. 2017. Robust classification of crisis-related data on social networks using convolutional neural networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11.

A. Schulz, E.L. Mencía, T.T. Dang, and B. Schmidt. 2014. Evaluating multi-label classification of incident-related tweets. 1141:26–33.