Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)

CrisisMMD: Multimodal Twitter Datasets from Natural Disasters

Firoj Alam, Ferda Ofli, Muhammad Imran
Qatar Computing Research Institute, HBKU, Doha, Qatar
{fialam, fofli, mimran}@hbku.edu.qa

Abstract

During natural and man-made disasters, people use social media platforms such as Twitter to post textual and multimedia content to report updates about injured or dead people, infrastructure damage, and missing or found people among other information types. Studies have revealed that this online information, if processed timely and effectively, is extremely useful for humanitarian organizations to gain situational awareness and plan relief operations. In addition to the analysis of textual content, recent studies have shown that imagery content on social media can boost disaster response significantly. Despite extensive research that mainly focuses on textual content to extract useful information, limited work has focused on the use of imagery content or the combination of both content types. One of the reasons is the lack of labeled imagery data in this domain. Therefore, in this paper, we aim to tackle this limitation by releasing a large multimodal dataset collected from Twitter during different natural disasters. We provide three types of annotations, which are useful to address a number of crisis response and management tasks for different humanitarian organizations.

Introduction

At times of natural and man-made disasters, social media platforms such as Twitter and Facebook are considered vital information sources that contain a variety of useful information such as reports of injured or dead people, infrastructure and utility damage, urgent needs of affected people, and missing or found people among others (Houston et al. 2015; Alam, Ofli, and Imran 2018). Information shared on social media has a wide variety of applications (Imran et al. 2014; Ashktorab et al. 2014; Reuter et al. 2015; Poblet, García-Cuesta, and Casanovas 2014; Kishi et al. 2017; Laudy 2017; Meissen et al. 2017). One application that also motivates our work is "humanitarian aid" where the primary purpose of humanitarian organizations such as The United Nations Office for the Coordination of Humanitarian Affairs (OCHA) is to gain situational awareness and actionable information to save lives, reduce the suffering of affected people, and rebuild communities (Castillo et al. 2016).

Processing social media data to extract life-saving information which is also helpful for humanitarian organizations in preparedness, response, and recovery of an emergency involves solving multiple challenges including handling information overload, information classification and determining its credibility, prioritizing certain types of information, etc. (Imran et al. 2015). These challenges require building computational systems and methods useful for a number of information processing tasks such as information classification, clustering, and summarization among others.

Information on social media is mainly shared in two forms: textual messages and images. Most of the past studies and systems mainly focused on using textual content to aid disaster response. However, in addition to the usefulness of textual messages, recent studies have revealed that images shared on social media during a disaster event can help humanitarian organizations in a number of ways. For example, Nguyen et al. used images shared on Twitter to assess the severity of infrastructure damage (Nguyen et al. 2017). Peters and Joao reported that the existence of images within on-topic messages were more relevant to the disaster event based on their analysis of tweets and messages from Flickr and Instagram for the flood event in Saxony in 2013 (Peters and Joao 2015). Similarly, Jing et al. investigated the usefulness of image and text and found that they were both informative. For their study, they collected data from two sources related to flood and flood aid (Jing et al. 2016). A similar study has been conducted by (Kelly, Zhang, and Ahmad 2017) to extract useful information from flood events occurred in Ireland during December 2015 to January 2016.

Despite extensive research that mainly focuses on social media text messages, limited work has focused on the use of images to boost humanitarian aid. One reason that hinders the growth of this research line is the lack of ground-truth data. There exist a few repositories such as CrisisLex (Olteanu et al. 2014) and CrisisNLP (Imran, Mitra, and Castillo 2016) which offer several Twitter datasets from natural and man-made disasters, but all of them share only textual content annotations. To overcome this limitation, we present human-labeled multimodal datasets collected from Twitter during seven recent natural disasters including earthquakes, hurricanes, wildfires, and floods. To the best of our knowledge, these are the first multimodal Twitter datasets ever shared publicly with ground-truth annotations.[1]

[1] The dataset is available at https://dataverse.mpi-sws.org/dataverse/icwsm18

To acquire ground-truth labels, we employed paid workers from a well-known crowdsourcing platform (i.e., Figure Eight[2]) and asked them to annotate data based on three humanitarian tasks. The first task aims to determine the informativeness of a given tweet text or an image for humanitarian aid purposes. Given the fact that millions of tweets are shared during disasters, focusing only on the *informative* messages or images help reduce information overload for humanitarian organizations. The second task aims to further analyze the set of messages and images that have been identified as informative in the first task to determine what kind of humanitarian information they convey (see Section *Humanitarian Tasks and Manual Annotations* for detailed categories). Finally, the third task aims to assess the severity of damage to infrastructure and utilities observed in an image.

The rest of the paper is organized as follows. In the next section, we provide a summary of the related work. Then, we provide details about the disaster events and the data collection procedure in our study. Next, we elaborate on the humanitarian tasks as well as their annotation details and results. Furthermore, we present possible applications and discussion in the later section. Finally, we conclude the paper in the last section.

Related Work

The use of social media such as Twitter, Facebook, and Youtube, has been explored in numerous studies (Imran et al. 2014; Vieweg et al. 2010; Imran et al. 2015; Terpstra et al. 2012; Tsou et al. 2017) for curating, analyzing and summarizing crisis-related information in order to make some decisions and responses. Current literature does not only highlight its importance but also provides directions for possible research avenues. Among them, one of the important research avenues is the exploitation of textual and visual content to extract useful information for humanitarian aid, which has been remained unexplored to a large extent. One of the important limitations of this line of research is the lack of ground-truth data. Below, we describe works that provide crisis-related datasets.

In crisis informatics, one of the earliest and publicly-available datasets is CrisisLex (Olteanu et al. 2014). It consists of tweets collected during six disaster events occurred in USA, Australia, and Canada between October 2012 and July 2013. The dataset was collected using keywords and geo-graphical information from Twitter. The annotations of the dataset consist of i) directly related, ii) indirectly related, iii) not related, and iv) not in English or not understandable. In another work (Olteanu, Vieweg, and Castillo 2015), the authors provide a dataset that consists of tweets from 26 crisis events that took place between 2012 and 2013. In this work, they first characterize the datasets along different crisis dimensions: 1) hazard type (i.e., natural vs. human-induced) and their subtypes, 2) temporal development (i.e., instantaneous vs. progressive), 3) geographic (i.e., focalized vs. diffused). Then, they characterized the datasets by 1) informativeness, 2) information type, and 3) source. Similar to the previous study they employed crowd-source workers to annotate the dataset. The dataset is publicly available at CrisisLex site[3].

Another initiative to provide crisis-related data is CrisisNLP[4]. Currently, this site has published three major data resources. For instance, Imran et al. provide tweets collected during the Joplin tornado, which hit Joplin, Missouri (USA) on May 22, 2011, and the tweets collected during the Hurricane Sandy, which hit Northeastern US on October 29, 2012 (Imran et al. 2013). The annotated dataset consists of 2,000 tweets for Hurricane Sandy and about 4,400 for Joplin Tornado. Recently published dataset by (Imran, Mitra, and Castillo 2016) consists of tweets from 19 different crisis events that took place between 2013 to 2015. A particular focus of this dataset is humanitarian information categories annotated by Stand-By-Task-Force (STBF) volunteers and crowd-workers from CrowdFlower. In another study, Ashktorab et al. report a dataset that has been collected from 12 different crises occurred in the United States (Ashktorab et al. 2014). The annotation of this dataset consists of infrastructure damage and human casualty. Wang, Hovy, and Dredze present a corpus of 6.5 million geotagged tweets collected during 2012 Hurricane Sandy (Wang, Hovy, and Dredze 2015). However, this corpus does not provide any human labeled annotations. Lagerstrom et al. present the utility of image classification to support emergency situation by utilizing tweets collected from the event of 2013 New South Wales bushfires (Lagerstrom et al. 2016). They have ∼5,000 images with labels "fire" or "not-fire" and present an image classification accuracy of 86%. The limitation is that their data is not publicly available for research.

Despite all the initiatives for providing crisis-related datasets that are mainly useful for natural language processing tasks, no multimodal dataset consisting of combined textual and visual annotations has been published yet. In this paper, we try to bridge this gap by releasing multimodal datasets that are collected from Twitter during seven natural disasters in 2017 and annotated for several tasks. We hope the research community will take advantage of this multimodal dataset to advance the research on both image and text processing.

Natural Disaster Events and Data Collection

We used Twitter to collect data during seven natural disasters. The data collection was performed using event-specific keywords and hashtags. In Table 1, we list the keywords used and the data collection period for each event. Next, we provide details of data collection for each event.

Hurricane Irma 2017

Hurricane Irma[5] caused catastrophic damage in Barbuda, Saint Barthelemy, Saint Martin, Anguilla, and the Virgin Islands. On Friday, September 8, a hurricane warning was issued for the Florida Keys and the Florida governor ordered

[2] https://www.figure-eight.com/, previously known as CrowdFlower, http://crowdflower.com/

[3] http://crisislex.org

[4] http://crisisnlp.qcri.org

[5] https://en.wikipedia.org/wiki/Hurricane_Irma

Table 1: CrisisMMD dataset details including event names, keywords used for data collection, and data collection period.

Crisis event	**Keywords**	**Start date**	**End date**
Hurricane Irma	*Hurricane Irma, HurricaneIram, Irma storm,...*	Sep 6, 2017	Sep 21, 2017
Hurricane Harvey	*Hurricane Harvey, Harvey, HurricaneHarvey,...*	Aug 26, 2017	Sep 20, 2017
Hurricane Maria	*Hurricane Maria, Maria Storm, Maria Cyclone,...*	Sep 20, 2017	Nov 13, 2017
Mexico earthquake	*Mexico earthquake, mexicoearthquake,...*	Sep 20, 2017	Oct 6, 2017
California wildfires	*California fire, California wildfires,...*	Oct 10, 2017	Oct 27, 2017
Iraq-Iran earthquake	*Iran earthquake, Iraq earthquake, halabja earthquake,...*	Nov 13, 2017	Nov 19, 2017
Sri Lanka floods	*SriLanka floods, FloodSL, SriLanka flooding,...*	May 31, 2017	Jul 3, 2017

all public schools and colleges to be closed. The Irma storm was a Category 5 hurricane, which caused \$66.77 billion in damage. We collected Hurricane Irma-related data from Twitter starting from September 6, 2017, to September 19, 2017, and the resulted collection consists of $\sim$3.5 million tweets and $\sim$176,000 images.

Hurricane Harvey 2017

Hurricane Harvey was a Category 4 storm when it hit Texas, USA on August 25, 2017[6]. It caused nearly \$200 billion in damage, which is record-breaking compared with any natural disaster in the US history. As can be seen in Table 1, we started the data collection on August 25, 2017, and ended on September 5, 2017. In total, $\sim$7 million tweets with $\sim$300,000 images were collected during this period.

Hurricane Maria 2017

Hurricane Maria[7], was a Category 5 hurricane that slammed Dominica and Puerto Rico and caused more than 78 deaths including 30 in Dominica and 34 in Puerto Rico, while many more left without homes, electricity, food, and drinking water. The data collection for Hurricane Maria was started on September 20, 2017, and ended on October 3, 2017. In total, we collected $\sim$3 million tweets and $\sim$52,000 images.

California Wildfires 2017

A series of wildfire took place in California in October 2017[8] causing more than \$9.4 billion losses of property. We started our tweet collection on October 10, 2017 and continued until October 27, 2017. As can be seen in Table 2, the collected dataset contains $\sim$400,000 tweets and $\sim$10,000 images.

Mexico Earthquake 2017

The Mexico earthquake[9] on September 19, 2017 was another major earthquake with a magnitude of 7.1. The earthquake caused death of around 370 people. For this event, our data collection started on September 20, 2017 till October 6, 2017. In total, we collected $\sim$400,000 tweets and $\sim$7,000 images.

Iraq-Iran Border Earthquake 2017

On November 12, 2017, a strong earthquake with a magnitude of 7.3 struck the border of Iran and Iraq[10]. The earthquake caused around 630 casualties, seventy thousand became homeless and eight thousand were injured. For our study, we collected tweets from November 12, 2017 to November 19, 2017, which resulted in $\sim$200,000 tweets and $\sim$6,000 images.

Sri Lanka Floods 2017

Due to heavy monsoon on southwest, Sri Lanka faced severe flooding in May 2017[11]. Furthermore, the flooding situation got worsened due to the Cyclone Mora[12], which caused more floods and landslides throughout Sri Lanka during the last week of May 2017. Our tweet data collection started on May 31, 2017 until July 3, 2017, which resulted in $\sim$41,000 tweets and $\sim$2,000 images.

Data Filtering and Sampling

To prepare data for manual annotation, we perform the following filtering steps:

1. As we build a multimodal dataset, we are interested only in tweets *with* images. Thus, our first filtering step is to discard all the tweets that do not contain at least one image URL. Since a tweet can contain more than one image, we extract all image URLs from the *"extended_entities"* element of the retrieved JSON record of a tweet.
2. We discard all non-English tweets using Twitter-provided language meta-data for a given tweet.
3. We retain tweets that contain at least two or more words or hashtags. In other words, we remove tweets containing a single word or hashtag since single-word tweets are less likely to convey any useful or meaningful information. We do not consider URLs or numbers as proper English words.
4. We remove duplicate tweets using tweets' textual content. For this purpose, we use the cosine similarity measure to compute tweet similarity scores. Two tweets with a similarity score greater than 0.7 are considered duplicate.

After performing the above mentioned filtering steps, we take a random sample of N tweets containing one or more

[6] https://en.wikipedia.org/wiki/Hurricane_Harvey

[7] https://en.wikipedia.org/wiki/Hurricane_Maria

[8] https://en.wikipedia.org/wiki/2017_California_wildfires

[9] https://en.wikipedia.org/wiki/2017_Central_Mexico_earthquake

[10] https://en.wikipedia.org/wiki/2017_IranIraq_earthquake

[11] https://en.wikipedia.org/wiki/2017_Sri_Lanka_floods

[12] https://en.wikipedia.org/wiki/Cyclone_Mora

Table 2: Event-wise data distribution. The numbers inside the parentheses in the last column represent the total number of images associated with tweets. Total number of images can be larger than total number of tweets as some tweets contain more than one image.

Crisis name	# tweets	# images	# filtered tweets	# sampled tweets (images)
Hurricane Irma	3,517,280	176,972	5,739	4,041 (4,525)
Hurricane Harvey	6,664,349	321,435	19,967	4,000 (4,443)
Hurricane Maria	2,953,322	52,231	6,597	4,000 (4,562)
California wildfires	455,311	10,130	1,488	1,486 (1,589)
Mexico earthquake	383,341	7,111	1,241	1,239 (1,382)
Iraq-Iran earthquake	207,729	6,307	501	499 (600)
Sri Lanka floods	41,809	2,108	870	832 (1,025)
Total	**14,223,141**	**576,294**	**36,403**	**16,097 (18,126)**

images from each dataset. Due to budget limitations, we sample around 4,000 for Hurricanes Irma, Harvey, and Maria. For the rest, we take all of the filtered tweets, as they are already low numbers. Table 2 describes all the datasets with details including total number of tweets initially collected, total number of images associated with the initial set of tweets, and the total number of tweets retained after the filtering and sampling steps for each dataset. In particular, the last column of the table shows the number of tweets and corresponding images (in parentheses) for each disaster event in our dataset. A tweet can contain more than one image, and hence, the number of images (shown in parentheses) are slightly larger than the actual number of sampled tweets.

Humanitarian Tasks and Manual Annotations

We perform the manual annotations of the sampled data along three humanitarian tasks. The first task aims to categorize the data into two high-level categories called "Informative" or "Not informative". During disasters and emergencies, as thousands of tweets arrive per minute, determining whether or not a tweet contains crucial information useful for humanitarian aid is an important task to reduce information overload for humanitarian organizations.

The second task, on the other hand, aims to identify critical and potentially actionable information such as reports of injured or dead people, infrastructure damage, etc. from the tweets. For this purpose, we use seven humanitarian categories. The third task is specific to damage severity assessment from images. Determining severely-damaged critical infrastructure after a major disaster is a core task of many humanitarian organizations to direct their response efforts.

Next, we present the exact instructions provided to the human annotators for all three tasks.

Task 1: Informative vs. Not informative

The purpose of this task is to determine whether a given tweet or image, which was collected during *"event name"*, is useful for humanitarian aid purposes as defined below. If the given tweet/image is useful for humanitarian aid, it is considered as an "Informative" tweet/image, otherwise as a "Not informative" tweet/image.

"Humanitarian aid" definition: In response to humanitarian crises including natural and man-made disasters, humanitarian aid involves providing assistance to people who need help. The primary purpose of humanitarian aid is to save lives, reduce suffering, and rebuild affected communities. Among the people in need belong homeless, refugees, and victims of natural disasters, wars, and conflicts who need basic necessities like food, water, shelter, medical assistance, and damage-free critical infrastructure and utilities such as roads, bridges, power-lines, and communication poles.

Moreover, the tweet/image is considered "Informative" if it reports/shows one or more of the following: cautions, advice, and warnings, injured, dead, or affected people, rescue, volunteering, or donation request or effort, damaged houses, damaged roads, damaged buildings; flooded houses, flooded streets; blocked roads, blocked bridges, blocked pathways; any built structure affected by earthquake, fire, heavy rain, strong winds, gust, etc., disaster area maps.

Images showing banners, logos, and cartoons are *not* considered as "Informative".

- **Informative:** if the tweet/image is useful for humanitarian aid.
- **Not informative:** if the tweet/image is not useful for humanitarian aid.
- **Don't know or can't judge:** due to non-English tweet or low-quality image content.

Task 2: Humanitarian Categories

The purpose of this task is to understand the type of information shared in an image/tweet, which was collected from Twitter during *"event name"*. Given an image/tweet, categorize it into one of the following categories.

- **Infrastructure and utility damage:** if the tweet/image reports/shows any built structure affected or damaged by earthquake, fire, heavy rain, floods, strong winds, gusts, etc. such as damaged houses, roads, buildings; flooded houses, streets, highways; blocked roads, bridges, pathways; collapsed bridges, power lines, communication poles, etc.

- **Vehicle damage:** if the tweet/image reports/shows any type of damaged vehicle such as cars, trucks, buses, motorcycles, boats, ships, trams, trains, etc.
- **Rescue, volunteering, or donation effort:** if the tweet/image reports/shows any type of rescue, volunteering, or donation effort such as people being transported to safe places, people being evacuated from the hazardous area, people receiving medical aid or food, people in shelter facilities, donation of money, blood, or services etc.
- **Injured or dead people:** if the tweet/image reports/shows injured or dead people.
- **Affected individuals:** if the tweet/image reports/shows people affected by the disaster event such as people sitting outside; people standing in queues to receive aid; people in need of shelter facilities, etc.
- **Missing or found people:** if the tweet/image reports/shows instances/pictures of missing or found people due to the disaster event.
- **Other relevant information:** if the tweet/image does not belong to any of the above categories, but it still contains important information useful for humanitarian aid, then select this category.
- **Not relevant or can't judge:** if the image is irrelevant or you can't judge, for example, due to its low-quality.

Task 3: Damage Severity Assessment

The purpose of this task is to assess the severity of damage reported/shown in an image. The severity of damage is the extent of physical destruction to a build-structure. We are only interested in physical damages like broken bridges, collapsed or shattered buildings, destroyed or cracked roads, etc. An example of a non-physical damage is the sign of smoke. Damage severity categories are discussed below:

- **Severe damage:** Substantial destruction of an infrastructure belongs to the severe damage category. For example, a non-livable or non-usable building, a non-crossable bridge, or a non-driveable road are all examples of severely damaged infrastructures.

  Specifically,

  - **Building:** If one or more buildings in the focus of the image show substantial loss of amenity/roof. If the image shows a building that is unsafe to use, it should be marked as severe damage.
  - **Bridge:** If a bridge is visibly not safe to use because parts of it are collapsing and should not be driven or walked upon, it should be listed as severe damage.
  - **Road:** If a road should not be used because there has been substantial damage, it should be marked as severe damage. Examples: due to an avalanche, there may be huge rocks piled up and you cannot drive or only a narrow part of the road is open. Due to an earthquake, you see a sinkhole, a substantial part of the road has sunk and the road cannot be navigated safely, that is severe damage.
- **Mild damage:** Partially destroyed buildings, bridges, houses, roads belong to mild damage category.
  - **Building:** Damage generally exceeding minor [damage] with up to 50% of buildings in the focus of the image sustaining a partial loss of amenity/roof. Maybe only part of the building has to be closed down, but other parts can still be used.
  - **Bridge:** If the bridge can still be used, but, part of it is unusable and/or needs some amount of repairs.
  - **Road:** If the road is still usable, but part of it has to be blocked off because of damage. This damage should be substantially more than what we see due to regular wear or tear.
- **Little or no damage:** Images that show damage-free infrastructure (except for wear and tear due to age or disrepair) belong to the little-or-no-damage category.
- **Don't know or can't judge:** Due to low-quality image.

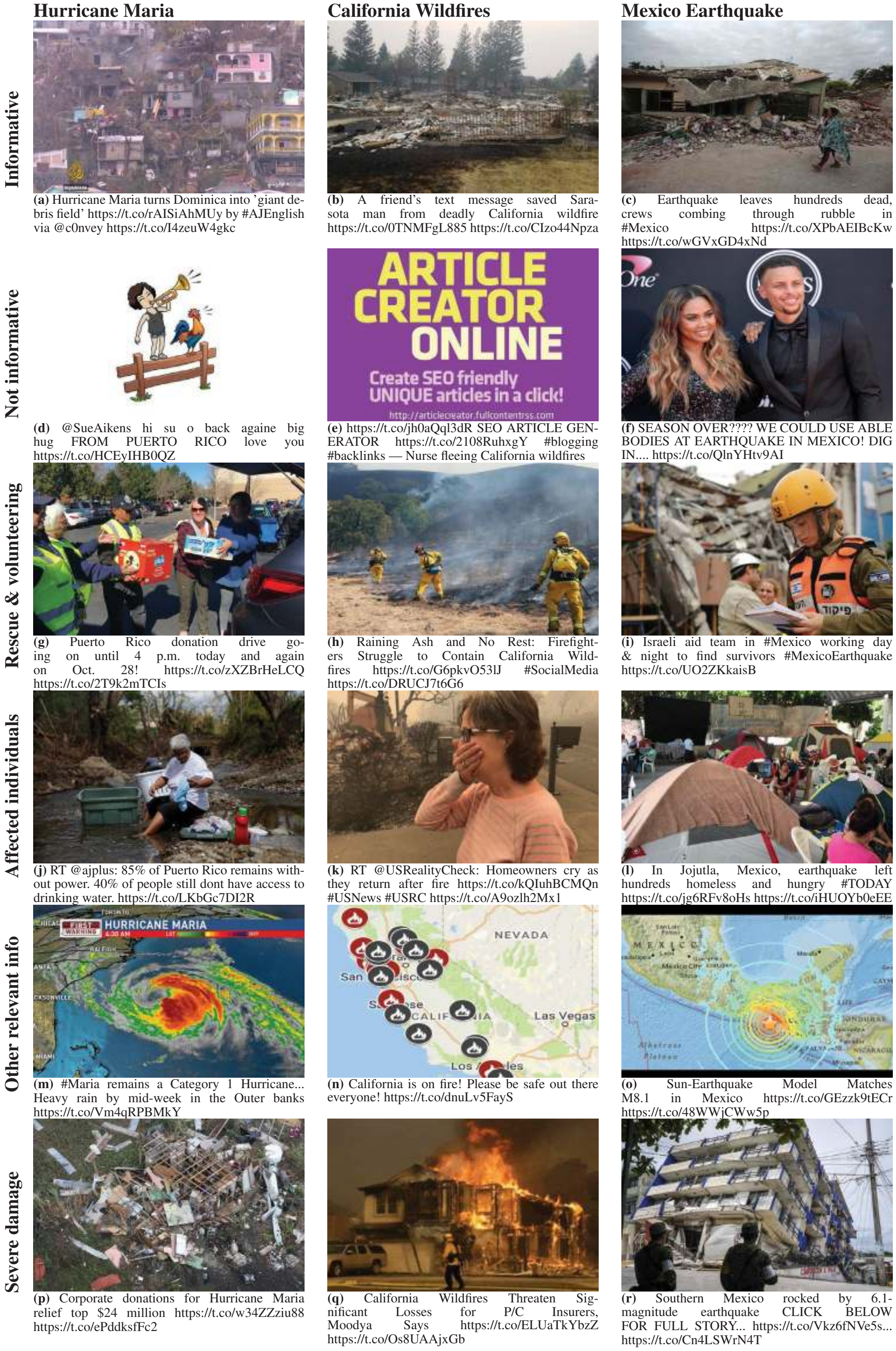
Manual Annotations using Crowdsourcing

Given the above specified tasks and instructions, we used Figure Eight, which is a well-known paid crowdsourcing platform previously known as CrowdFlower, to acquire manual annotations of the sampled data. Manual annotations for tweets (textual content) and images were acquired separately. In this case, a task consisted of an image or a tweet along with task instructions and a list of categories (e.g., informative and not informative). We first ran Task 1 (i.e., informative or not informative) for all the events. For Task 2 (i.e., humanitarian categories), we only used the data which was labeled as "informative" (i.e., either text or image was informative) in Task 1. We dropped tweets where neither text nor image was informative. For Task 3 (i.e., damage assessment), we only used images from Task 2 which were labeled as "Infrastructure and utility damage". For each task, we created at least 40 test questions to keep good quality annotators while excluding annotators that do not perform well on the test questions. We sought an agreement of three different human annotators to decide a final label/category for a tweet or an image. Human annotators with English language expertise were allowed to perform the tasks.

Crowdsourcing Results and Discussion

Figure 1 illustrates example tweet text and image pairs with different annotations from different disaster events. Figures 2 and 3 show the distribution of tweet text and image results into the informative categories task, respectively. Similarly, Figures 4 and 5 show the results of tweet text and image annotations for the humanitarian categories task, respectively. Lastly, Figure 6 shows the manual annotation results for the damage severity assessment task.

From around 25% to 35% of the tweet text data of all the events is considered as "not informative" with an exception of the Sri Lanka floods event in which case the "not informative" category is around 60% (see Figure 2). This finding is in line with previous studies that analyze Twitter data during disasters. The prevalence of the "not informative" category in the image informative task is higher than the text informative task (see Figure 3). All informative tweets and images were selected for the second task (humanitarian categories) which we examine next.



(a) Hurricane Maria turns Dominica into 'giant debris field' https://t.co/rAISiAhMUy by #AJEnglish via @c0nvey https://t.co/I4zeuW4gkc

(b) A friend's text message saved Sarasota man from deadly California wildfire https://t.co/0TNMFgL885 https://t.co/CIzo44Npza

(c) Earthquake leaves hundreds dead, crews combing through rubble in #Mexico https://t.co/XPbAEIBcKw https://t.co/wGVxGD4xNd

(d) @SueAikens hi su o back againe big hug FROM PUERTO RICO love you https://t.co/HCEyIHB0QZ

(e) https://t.co/jh0aQql3dR SEO ARTICLE GENERATOR https://t.co/2108RuhxgY #blogging #backlinks — Nurse fleeing California wildfires

(f) SEASON OVER???? WE COULD USE ABLE BODIES AT EARTHQUAKE IN MEXICO! DIG IN.... https://t.co/QlnYHtv9AI

(g) Puerto Rico donation drive going on until 4 p.m. today and again on Oct. 28! https://t.co/zXZBrHeLCQ https://t.co/2T9k2mTCIs

(h) Raining Ash and No Rest: Firefighters Struggle to Contain California Wildfires https://t.co/G6pkvO53lJ #SocialMedia https://t.co/DRUCJ7t6G6

(i) Israeli aid team in #Mexico working day & night to find survivors #MexicoEarthquake https://t.co/UO2ZKkaisB

(j) RT @ajplus: 85% of Puerto Rico remains without power. 40% of people still dont have access to drinking water. https://t.co/LKbGc7DI2R

(k) RT @USRealityCheck: Homeowners cry as they return after fire https://t.co/kQIuhBCMQn #USNews #USRC https://t.co/A9ozlh2Mx1

(l) In Jojutla, Mexico, earthquake left hundreds homeless and hungry #TODAY https://t.co/jg6RFv8oHs https://t.co/iHUOYb0eEE

(m) #Maria remains a Category 1 Hurricane... Heavy rain by mid-week in the Outer banks https://t.co/Vm4qRPBMkY

(n) California is on fire! Please be safe out there everyone! https://t.co/dnuLv5FayS

(o) Sun-Earthquake Model Matches M8.1 in Mexico https://t.co/GEzzk9tECr https://t.co/48WWjCWw5p

(p) Corporate donations for Hurricane Maria relief top $24 million https://t.co/w34ZZziu88 https://t.co/ePddksfFc2

(q) California Wildfires Threaten Significant Losses for P/C Insurers, Moodya Says https://t.co/ELUaTkYbzZ https://t.co/Os8UAAjxGb

(r) Southern Mexico rocked by 6.1-magnitude earthquake CLICK BELOW FOR FULL STORY... https://t.co/Vkz6fNVe5s... https://t.co/Cn4LSWrN4T

Figure 1: Example tweet text and image pairs with different annotation labels from different disaster events.

In the humanitarian task for tweet text annotations, the "Other relevant information" and "Rescue, volunteering, or donation effort" categories appear as the most prevalent ones among others. The "Missing or found people" category is one of the rarest, as can be seen in Figure 4. It appears that earthquakes cause more injuries and deaths compared to hurricanes and floods (see Figure 4). In particular, the reports of "Injured or dead people" both in the tweets and images in the Iraq-Iran earthquake event are significantly more than any other event. A small proportion of the "Not relevant or can't

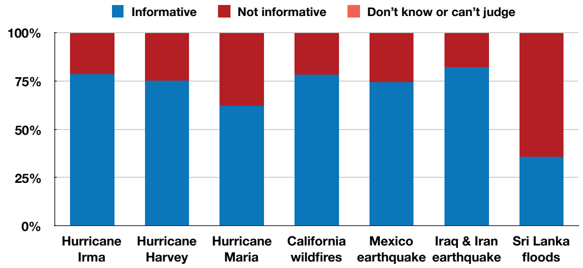


Figure 2: Manual annotation results of tweets (text) for the informative task.

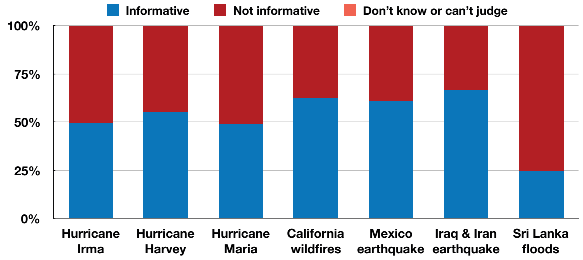


Figure 3: Manual annotation results of images for the informative task.

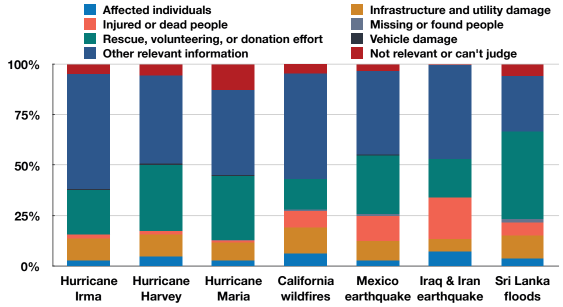


Figure 4: Manual annotation results of tweets (text) for the humanitarian task.

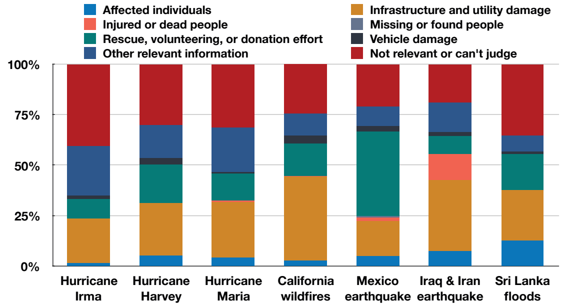


Figure 5: Manual annotation results of images for the humanitarian task.

judge" tweets can still be seen in Figure 4. This is mainly due to our sampling strategy for this annotation task (i.e., if either text or image of a tweet is informative, it is selected to be annotated for the humanitarian task).

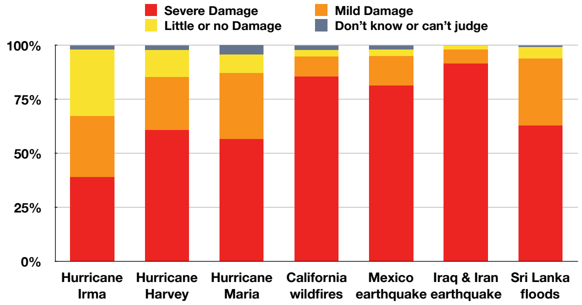


Figure 6: Manual annotation results of images for the damage severity assessment task.

We observe that images in tweets tend to contain more damage-related information compared to their corresponding text. For instance, according to Figure 5, the "Infrastructure and utility damage" category is generally prevalent in all the events, however, in the case of the California wildfires, it appears to be around 50% of the informative event data. Moreover, the "Vehicle damage" category, which does not appear at all in the text annotation results, appears in the image annotations of many events (see California wildfires and Hurricane Harvey bars in Figure 5). The other most prevalent information type present in images is "Rescue, volunteering, or donation effort". Mainly, these images show people that help or rescue others, or are involved in volunteering efforts during or after a disaster.

The results of the damage severity assessment task are shown in Figure 6. Since we used images that were already annotated as "Infrastructure and utility damage", the results do not show many "Don't know or can't judge" cases in this task. Most of the images were annotated as severely-damaged infrastructure in all the events. However, most of the severe damage seems to be actually caused by the earthquakes and wildfires as opposed to the hurricanes and floods. Figure 7 shows the inter-annotator-agreement of all the tasks. Generally, all the tasks have strong agreement scores.

Applications and Future Directions

The provided datasets have several potential applications in many different domains. First of all, CrisisMMD datasets can be used in any multimodal task involving computer vision and natural language processing. For instance, one can try to learn a joint embedding space of tweet text and images that can be used for text-to-image as well as image-to-text retrieval tasks. Another multimodal use case of CrisisMMD can be the image captioning task where the goal is to learn a mapping from the visual content to its textual description. Furthermore, more powerful event summarization models can be trained on these aligned and structured multimodal data to automatically generate a multimedia summary of a given event. Since we have developed CrisisMMD datasets mainly with the "humanitarian aid" use case in mind, we further discuss applications specific to the humanitarian domain in the rest of this section.

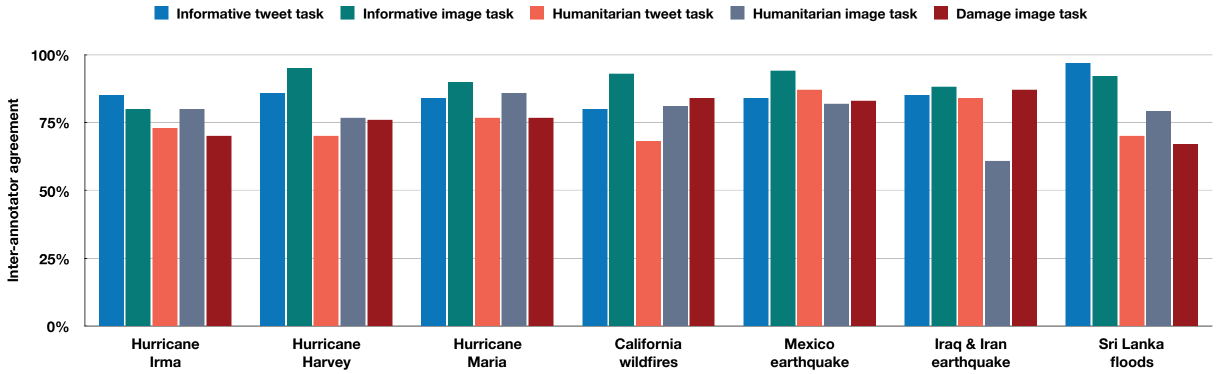


Figure 7: Inter-annotator-agreement scores for all the tasks and all the disaster events.

High-level Situational Awareness by Reducing Information Overload

Information posted on social media during natural and man-made disasters vary greatly. Studies have revealed that a big proportion of social media data consists of irrelevant information that is not useful for any kind of relief operations. Humanitarian organizations do not want a deluge of noisy messages that are of a personal nature or those that do not contain any useful information. Instead, they look for messages which contain some useful information. Among other uses, they use the informative messages and images to gain situational awareness. This dataset provides human annotations along informative and not informative messages from seven crisis events to help community to build more robust systems.

Critical and Potentially Actionable Information Extraction

Depending on their roles and mandate, humanitarian organizations differ in terms of their information needs. Several rapid response and relief agencies look for fine-grained information about specific incidents which is also actionable. Such information types include reports of injured or dead people, critical infrastructure damage (e.g., a collapsed bridge), and rescue demand among others. Our dataset provides human annotations along many such critical humanitarian information needs, which can prove to be life saving if more effective systems and computational methods are developed. Furthermore, the damage severity annotations are critical for many response organizations to direct their focus to, for example, severely damaged infrastructure to reduce suffering of affected people.

Furthermore, with several thousands of manually annotated pairs of tweets and images, we claim that CrisisMMD is the first and largest multimodal dataset to date published for research community to explore different approaches and build computational methods to help humanitarian cause.

Conclusions

Information available on social media at times of a disaster or an emergency is useful for several humanitarian tasks. Despite extensive research that uses social media textual content, little focus has been given to images shared on social media. One issue in this regard is the lack of labeled imagery data. To address this issue, in this paper, we introduced CrisisMMD, multimodal Twitter corpora consisting of several thousands of manually annotated tweets and images collected during seven major natural disasters including earthquakes, hurricanes, wildfires, and floods that happened in the year 2017 across different parts of the World. The provided datasets include three types of annotations: informative vs. not informative, humanitarian categories, and damage severity categories. We also presented a number of humanitarian use cases and tasks that can be fulfilled using these datasets if more robust and effective systems are developed.

References

Alam, F.; Ofli, F.; and Imran, M. 2018. Processing social media images by combining human and machine computing during crises. *International Journal of Human–Computer Interaction* 34(4):311–327.

Ashktorab, Z.; Brown, C.; Nandi, M.; and Culotta, A. 2014. Tweedr: Mining twitter to inform disaster response. In *ISCRAM*.

Castillo, C.; Imran, M.; Meier, P.; Lucas, J. K.; Srivastava, J.; Leson, H.; Ofli, F.; and Mitra, P. 2016. *Together We Stand—Supporting Decision in Crisis Response: Artificial Intelligence for Digital Response and MicroMappers*. Istanbul: Tudor Rose, World Humanitarian Summit. 93–95.

Houston, J. B.; Hawthorne, J.; Perreault, M. F.; Park, E. H.; Goldstein Hode, M.; Halliwell, M. R.; Turner McGowen, S. E.; Davis, R.; Vaid, S.; McElderry, J. A.; et al. 2015. Social media and disasters: a functional framework for social media use in disaster planning, response, and research. *Disasters* 39(1):1–22.

Imran, M.; Elbassuoni, S.; Castillo, C.; Diaz, F.; and Meier, P. 2013. Practical extraction of disaster-relevant information from social media. In *Proceedings of the 22nd international conference on World Wide Web companion*, 1021–1024. International World Wide Web Conferences Steering Committee.

Imran, M.; Castillo, C.; Lucas, J.; Meier, P.; and Vieweg, S. 2014. AIDR: Artificial intelligence for disaster response. In

ACM International Conference on World Wide Web, 159–162.

Imran, M.; Castillo, C.; Diaz, F.; and Vieweg, S. 2015. Processing social media messages in mass emergency: A survey. *ACM Computing Surveys* 47(4):67.

Imran, M.; Mitra, P.; and Castillo, C. 2016. Twitter as a lifeline: Human-annotated twitter corpora for nlp of crisis-related messages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris, France: European Language Resources Association (ELRA).

Jing, M.; Scotney, B. W.; Coleman, S. A.; McGinnity, M. T.; Zhang, X.; Kelly, S.; Ahmad, K.; Schlaf, A.; Gründer-Fahrer, S.; and Heyer, G. 2016. Integration of text and image analysis for flood event image recognition. In *Signals and Systems Conference (ISSC), 2016 27th Irish*, 1–6. IEEE.

Kelly, S.; Zhang, X.; and Ahmad, K. 2017. Mining multimodal information on social media for increased situational awareness.

Kishi, K.; Kosaka, N.; Kura, T.; Kokogawa, T.; and Maeda, Y. 2017. Study on integrated risk management support system application to emergency management for cyber incidents. In *ISCRAM*.

Lagerstrom, R.; Arzhaeva, Y.; Szul, P.; Obst, O.; Power, R.; Robinson, B.; and Bednarz, T. 2016. Image classification to support emergency situation awareness. *Frontiers in Robotics and AI* 3:54.

Laudy, C. 2017. Rumors detection on social media during crisis management. In *ISCRAM*.

Meissen, U.; Fuchs-Kittowski, F.; Voisard, A.; Jendreck, M.; Pfennigschmidt, S.; Hardt, M.; and Faust, D. 2017. Ensure: A general system for coordination of volunteers for agile disaster response. In *ISCRAM*.

Nguyen, D. T.; Ofli, F.; Imran, M.; and Mitra, P. 2017. Damage assessment from social media imagery data during disasters. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 1–8.

Olteanu, A.; Castillo, C.; Diaz, F.; and Vieweg, S. 2014. Crisislex: A lexicon for collecting and filtering microblogged communications in crises. In *ICWSM*.

Olteanu, A.; Vieweg, S.; and Castillo, C. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 994–1009. ACM.

Peters, R., and Joao, P. d. A. 2015. Investigating images as indicators for relevant social media messages in disaster management. In *International Conference on Information Systems for Crisis Response and Management*.

Poblet, M.; García-Cuesta, E.; and Casanovas, P. 2014. Crowdsourcing tools for disaster management: A review of platforms and methods. In *AI Approaches to the Complexity of Legal Systems*. Springer. 261–274.

Reuter, C.; Ludwig, T.; Kaufhold, M.-A.; and Pipek, V. 2015. Xhelp: design of a cross-platform social-media application to support volunteer moderators in disasters. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 4093–4102. ACM.

Terpstra, T.; De Vries, A.; Stronkman, R.; and Paradies, G. 2012. *Towards a realtime Twitter analysis during crises for operational crisis management*. Simon Fraser University Burnaby.

Tsou, M.-H.; Jung, C.-T.; Allen, C.; Yang, J.-A.; Han, S. Y.; Spitzberg, B. H.; and Dozier, J. 2017. Building a real-time geo-targeted event observation (geo) viewer for disaster management and situation awareness. In *International Cartographic Conference*, 85–98. Springer.

Vieweg, S.; Hughes, A. L.; Starbird, K.; and Palen, L. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *SIGCHI Conference on Human Factors in Computing Systems*, 1079–1088. ACM.

Wang, H.; Hovy, E. H.; and Dredze, M. 2015. The hurricane sandy twitter corpus. In *AAAI Workshop: WWW and Public Health Intelligence*.