Review article: Detection of informative tweets in crisis events

Anna Kruspe[1], Jens Kersten[1], and Friederike Klan[1]

[1]German Aerospace Center (DLR), Jena, Germany

Correspondence: Anna Kruspe (anna.kruspe@dlr.de)

Abstract. Messages on social media can be an important source of information during crisis situations, be they short-term disasters or longer-term events like COVID-19. They can frequently provide details about developments much faster than traditional sources (e.g. official news) and can offer personal perspectives on events, such as opinions or specific needs. In the future, these messages can also serve to assess disaster risks.

5   One challenge for utilizing social media in crisis situations is the reliable detection of informative messages in a flood of data. Researchers have started to look into this problem in recent years, beginning with crowd-sourced methods. Lately, approaches have shifted towards an automatic analysis of messages. In this review article, we present methods for the automatic detection of crisis-related messages (tweets) on Twitter. We start by showing the varying definitions of importance and relevance relating to disasters, as they can serve very different purposes. This is followed by an overview of existing, crisis-related social media

10   data sets for evaluation and training purposes. We then compare approaches for solving the detection problem based (1) on filtering by characteristics like keywords and location, (2) on crowdsourcing, and (3) on machine learning techniques with regard to their focus, their data requirements, their technical prerequisites, their efficiency and accuracy, and their time scales. These factors determine the suitability of the approaches for different expectations, but also their limitations. We identify which aspects each of them can contribute to the detection of informative tweets, and which areas can be improved upon in the future.

15   We point out particular challenges, such as the linguistic issues concerning this kind of data. Finally, we suggest future avenues of research, and show connections to related tasks, such as the subsequent semantic classification of tweets.

1 Introduction

During a crisis situation, quickly gaining as much information as possible about the tide of events is of crucial importance.

20   Having access to information is necessary for developing situational awareness, and can mean the difference between life and death. This has become obvious once again in the ongoing COVID-19 pandemic. One source of such information that has started gaining interest in the last couple of years is social media. Twitter users, as an example, write about disaster preparations, developments, recovery, and a host of other topics (Niles et al., 2019). Retrieving this information could lead to significant improvements in disaster management strategies. In contrast to most other information sources, social media posts show up

25 nearly immediately whenever there is a new occurrence (as long as telecommunication infrastructure is still intact), and as such can deliver information very quickly. Such messages can also provide new perspectives that would not be available any other way at this speed, e.g. ground photos. In addition to factual information, social media can offer personal insights into the occurrences, as well as a back-channel to users for relief providers, government agencies, and other official institutions as well as the media. From a user perspective, 69% of Americans think that emergency response agencies should respond to

30 calls for help sent through social media channels according to a 2010 Red Cross study (American Red Cross, 2010). A very comprehensive overview of social media usage in crisis situations is given in (Reuter and Kaufhold, 2018).

The crux of this matter lies in the retrieval and classification of relevant messages. Twitter users worldwide generate 5,800 tweets per second on average[1]. In any given event, the majority of these posts will not be relevant to the event, or useful to service providers. The question is thus: What messages should be detected during a crisis event, and how can such a detection be

35 implemented? This review article will provide an overview over existing approaches to this problem. We will focus on Twitter data as most other social media sources do not offer a possibility to obtain large amounts of their data to outside researchers, or are not commonly used in a way that facilitates gaining information quickly during a disaster.

In the next section, we will examine the problem definition more closely. Section 3 introduces some already existing social media data sets useful for analyzing the task of retrieving informative tweets and for training as well as testing modeling

40 approaches. In section 4, we will then show how such approaches have been implemented so far, grouped into filtering, crowd-sourcing, and machine learning methods. Section 5 then goes into detail about the challenges these approaches frequently face, while section 6 briefly describes some related problems. We finish with suggestions for new developments in section 7, and a conclusion in section 8.

2 Problem definition

45 The task of finding social media posts in a crisis may appear clearly defined at first, but quickly becomes more convoluted when attempting an exact definition. Existing publications have gone about defining their problem statement in a variety of ways. An overview is provided in table 1.

What emerges from this table is a trichotomy between the concepts "related", "relevant", and "informative". As a tentative definition, we subsume that "related" encompasses all messages that make implicit or explicit mention of the event in question.

50 The "relevant" category is a subset of these, comprised of messages that contain actual information pertaining to the event. "Informative" messages, finally, offer information useful to the user of the system. Not all publications necessarily follow this pattern, and lines between these categories are blurry. In reality, many border cases arise, such as jokes, sarcasm, and speculation. In addition, the question of what makes a tweet informative, or even relevant, is highly dependent on who is asking this question, i.e. who the user of this system is. Such users are often assumed to be relief providers, but could also be members

55 of the government, the media, affected citizens, their family members, and many others. Building on top of this, each of these users may be interested in a different use case of the system. For instance, humanitarian and governmental emergency man-

---

[1] https://www.omnicoreagency.com/twitter-statistics/

Table 1. Overview of class definitions for filtering crisis-related tweets

Article	Class	Definition
(Imran et al., 2013)	*Personal*	A message only of interest to its author and her immediate circle of family/friends - does not convey any useful information to people who do not know its author
	Informative	Messages of interest to other people beyond the author's immediate circle
	Other	Not related to the disaster
(Parilla-Ferrer et al., 2014)	*Informative*	A tweet provides useful information to the public and is relevant to the event
	Uninformative	Tweets that are not relevant to the disaster and these do not convey enough information or are personal in nature and may only be beneficial to the family or friends of the sender
(Caragea et al., 2016)	*Informative*	Useful information
	Not informative	Not relevant to the event and no useful information
(Win and Aung, 2017)	*Informative*	Useful information
	Not informative	Not relevant to the event and no useful information
	Other information	Messages related to the event but without useful information
(Nguyen et al., 2017a)	*Useful/Relevant*	Information that is useful to others
	Not related or irrelevant	Not related to the event or does not contain useful information for others
(Burel and Alani, 2018)	*Crisis related*	Message related to a crisis situation in general without taking into account informativeness or usefulness
	Non-crisis related	Message that is not related to a crisis situation
(Stowe et al., 2018)	*Relevant*	Any information that is relevant to disaster events, including useful information but also jokes, retweets, and speculation
	Irrelevant	Not related to a disaster event

agement organizations are interested in understanding "the big picture", whereas local police forces and firefighters desire to find "implicit and explicit requests related to emergency needs that should be fulfilled or serviced as soon as possible" (Imran et al., 2018). Moreover, some of these use cases may require a high precision of the detected tweets while possibly missing

60   some important information; others may be more accepting of false alarms while focusing on a high recall.

These questions are commonly not explicitly taken into account in publications focusing on technical solutions. In practical application scenarios, however, they add complexity to the system. Moreover, these varying definitions make existing approaches and data sets difficult to compare.

3 Data sets

65 Collections of social media data created during crises are necessary to study what users write about, how this develops over time, and to create models for automatic detection and other tasks. For these reasons, several such data sets have already been created. As mentioned above, Twitter is the most fruitful source of data for this use case; therefore, available data sets have mainly focused on Twitter data.

Table 2 lists an overview of available Twitter data sets collected during disaster events. These mainly focus on the text content

70 of tweets, except for *CrisisMMD* which provides tweets with both text and images. Some of these data sets only contain data for one event, while others aggregate multiple ones. Based on various existing data sets, Wiegmann et al. (2020a) recently proposed a balanced compilation of labeled Tweets from 48 different events covering the ten most common disaster types. A distinction can also be made for corpora focusing on natural disasters and those also including man-made disasters. *Events2012* goes even further, containing around 500 events of all types, including disasters.

75 Annotations vary between these data sets. Some of them do not contain any labels beyond the type of event itself, while others are labeled according to content type (e.g. "Search and rescue" or "Donations"), information source (first-party observers, media, etc.), and priority or importance of each tweet (*CrisisLexT26* and *TREC-IS 2019A*).

A general issue with these data sets lies in the fact that researchers cannot release the full tweet content due to Twitter's redistribution policy[2]. Instead, these data sets are usually provided as lists of tweet ID's, which must then be expanded to the

80 full information ("hydrated"). This frequently leads to data sets becoming smaller over time as users may choose to delete their tweets or make them private. Additionally, the teams creating these corpora have mainly focused on English- and occasionally Spanish-language tweets to facilitate their wider usage for study. More insights would be possible if tweets in the language(s) of the affected area were available. However, Twitter usage also varies across countries. Another factor here is that less than 1% of all tweets contain geolocations (Sloan et al., 2013), which are often necessary for analysis.

85 In recent months, specific data sets for the COVID-19 pandemic have been collected; we also list these in the table. This is a unique case, as the crisis has been progressing for much longer than all other covered events, and much higher numbers of tweets have consequently been produced. It has also affected most of the world, rather than just particular regions.

The following sections provide descriptions of the data sets in more detail:

Events2012 This data set contains 120 million tweets, of which around 150,000 were labeled to belong to one of 506 events

90 (which are not necessarily disaster events) (McMinn et al., 2013).

CrisisLexT26 *CrisisLex* was first published by Olteanu et al. (2014) and expanded later to *CrisisLexT26* (Olteanu et al., 2015). It contains tweets collected during 26 crises, mainly natural disasters like earthquakes, wildfires and floods, but also human-induced disasters like shootings and a train crash. Amounts of these tweets per disaster range between 1,100 and 157,500. In total, around 285,000 tweets were collected. They were then annotated by paid workers on the *CrowdFlower*

95 crowdsourcing platform[3] according to three concepts: Informativeness, information type, and tweet source.

---

[2]https://developer.twitter.com/en/developer-terms/agreement-and-policy

[3]Later named *Figure Eight*, https://www.figure-eight.com/; acquired in 2019 by *Appen*, https://appen.com

Table 2. Overview of crisis-related Twitter data sets

Name	# Tweets	Covered events
Events2012 (McMinn et al., 2013)	~150,000	Disasters and accidents, 500 other topics like sports, arts, culture and entertainment
CrisisLexT26 (Olteanu et al., 2014, 2015)	26,000	Earthquake, flood, wildfire, meteor, typhoon, flood, train crash, explosions, building collapse, bombings
CrisisNLP (Imran et al., 2016a)	23,000	Earthquake, hurricane, flood, typhoon, cyclone, ebola, MERS
CrisisMMD (Alam et al., 2018b)	16,000 (with images)	Hurricane, earthquake, wildfire, flood
Epic (Stowe et al., 2018)	~25,000	Hurricane
Florence (Kersten et al., 2019)	~600,000	Hurricane
Disaster Tweet Corpus 2020 (Wiegmann et al., 2020b, a)	~150,000	Earthquake, flood, biological, hurricane, wildfires, industrial, societal, transportation, tornadoes, other
TREC-IS 2019A (McCreadie et al., 2019)	~30,000	Bombing, earthquake, flood, typhoon/hurricane, wildfire, shooting
Appen Disaster Response Messages (Appen Ltd., 2020)	~30,000	Earthquake, flood, hurricane
Kaggle covid19 (Smith, 2020)	~1,500,000 (as of May 1st, 2020)	COVID-19 pandemic
covid19_twitter (Banda et al., 2020)	~400,000,000 (as of mid-June, 2020)	COVID-19 pandemic
GeoCoV19 (Qazi et al., 2020)	~524,000,000 (as of May 1st, 2020)	COVID-19 pandemic

CrisisNLP Similar to *CrisisLexT26*, the team behind *CrisisNLP* collected tweets during 19 natural and health-related disasters and published them for research (Imran et al., 2016a). Collected tweets range between 17,000 and 28 million per event, making up around 53 million in total. Out of these, around 50,000 were annotated both by volunteers and by paid workers on *CrowdFlower* with regard to information type.

100 **CrisisMMD** *CrisisMMD* is an interesting special case because it only contains tweets with both text and image content. 16,000 tweets were collected for seven events that took place in 2017 in five countries. Annotation was performed by *Figure Eight* for text and images separately. The three annotated concepts are: Informative/Non-informative, eight semantic categories (like "Rescue and volunteering" or "Affected individuals"), and damage severity (only applied to images) (Alam et al., 2018b).

105 **Epic** This data set with a focus on Hurricane Sandy was collected in a somewhat different manner than most others. The team first assembled tweets containing hashtags associated with the hurricane, and then aggregated them by user. Out of these users, they selected those who had geotagged tweets in the area of impact, suggesting that these users would have been affected by the hurricane. Then, 105 of these users were selected randomly, and their tweets from a week before landfall to a week after were assembled. This leads to a data set that in all probability contains both related and unrelated

110 tweets by the same users. Tweets were annotated according to their relevance as well as 17 semantic categories (such as "Seeking info" or "Planning") and sentiment (Stowe et al., 2018).

Florence The Florence data set contains 600,000 tweets collected in the area affected by Hurricane Florence in the week of September 10, 2018, to September 17, 2018. These were not originally pre-filtered in any way; therefore, only a subset of them is related to Hurricane Florence. Such possible subsets were determined in a number of ways, including a

115 filtering with various approaches. The overlap between these results was interpreted to contain related tweets with high confidence; this leaves around 20,000 tweets (Kersten et al., 2019).

Disaster Tweet Corpus 2020 This dataset contains tweets collected, annotated, and published in several other works (Imran et al., 2014; Olteanu et al., 2014, 2015; Imran et al., 2016b; Alam et al., 2018b; Stowe et al., 2018; McMinn et al., 2013), and covers 48 disasters over 10 disaster types. This balanced collection is intended as a benchmarking dataset for

120 filtering algorithms in general (Wiegmann et al., 2020b, a).

TREC-IS 2019A A crisis classification task named "Incident Streams" has been a part of the Text REtrieval Conference (TREC) organized by NIST since 2018 (McCreadie et al., 2019). In the first iteration, tweets for six events were first collected automatically using a pre-defined list of keywords, and then annotated with one of 25 information type categories. Further iterations were conducted twice in 2019, for which the data set was expanded each time. The format was

125 also changed to allow multiple labels per tweet. The most recent version of the corpus contains around 30,000 tweets from 15 events. For the Hurricane Florence subset, 2,000 tweets from the previously described data set were used.

Appen Disaster Response Messages This dataset was published in an open-source format originally by *Figure Eight*, now part of private company *Appen* (Appen Ltd., 2020). It contains 30,000 messages split into training, test, and validation

130   sets collected during various disaster events between 2010 and 2012. These tweets are annotated according to 36 content
       categories, such as "Search and rescue", "Medical help", or "Military", as well as with a "Related" flag. These messages
       contain multiple languages plus English translations. The data set also includes news articles related to disasters. The
       data set is used in a Udacity course[4] as well as a Kaggle challenge[5].

Kaggle covid-19   This data set is currently still in the process of being published (Smith, 2020). It contains tweets of users
       who have used one of nine COVID-19-related hashtags collected since March 2020. So far, there has been an aggregated

135   amount of around 1 million tweets per month across more than 200 countries. Several analysis tasks are being worked
       on via Kaggle at the moment, the original one being "Are tweets indicative of infection rates?".

covid19_twitter   This is a larger data set of COVID-19-related tweets. Collection started on January 1st with very few found
       tweets, and was expanded on March 11. Tweets are collected worldwide from the API using 13 keywords. 238 countries
       are covered. There is also a version without retweets, which contains around 100,000 messages. The top 1000 terms,

140   bigrams, and trigrams are also contained in the data set (Banda et al., 2020).

GeoCoV19   This is an even larger data set of coronavirus-related tweets. Messages were collected on a basis of 800 multilin-
       gual keywords, starting on February 1, 2020. For nearly all of these tweets, geolocation information was retrieved, either
       from the tweet's own geolocation metadata, from user location or place information metadata, or from place mentions in
       the tweet's text. Non-coordinate information has been geocoded to coordinates (Qazi et al., 2020).

145 4   Approaches

As described above, users generate huge amounts of data on Twitter every second, and finding tweets related to an ongoing
event is not trivial (Landwehr and Carley, 2014). Several detection approaches have been presented in literature so far. We will
group them into three categories: Filtering by characteristics, crowdsourcing, and machine learning approaches.

4.1   Filtering by characteristics

150 The most obvious strategy is the filtering of tweets by various surface characteristics as shown in (Kumar et al., 2011), for
example. Keywords and hashtags are used most frequently for this and often serve as a useful pre-filter for data collection.
The Twitter API allows searching directly for keywords and hashtags or recording the live stream of tweets containing those,
meaning that this approach is often a good starting point for researchers. This is especially relevant because only 1% of the
live stream can be collected for free (also see section 5) - when a keyword filter is employed, this 1% is more likely to contain

155 relevant tweets.

Olteanu et al. (2014) developed a lexicon called *CrisisLex* for this purpose. However, the keyword-filtering approach easily
misses tweets that do not mention the keywords specified in advance, particularly when changes occur or the attention focus

---

[4]https://www.udacity.com/course/data-scientist-nanodegree--nd025

[5]https://www.kaggle.com/jannesklaas/disasters-on-social-media

shifts during the event. To tackle this problem, the authors propose a method to update the keyword list based on new messages.
Another problem with keyword lists is that unrelated data that contains the same keywords may be retrieved (Imran et al., 2015).

160 Nevertheless, such approaches have been used in insightful studies, e.g. in (de Albuquerque et al., 2015), where keyword-filtered tweets during a flood event were correlated with flooding levels.

Geolocation is another frequently employed feature that can be useful for retrieving tweets from an area affected by a disaster. However, this approach misses important information that could be coming from a source outside the area, such as help providers or news sources. Additionally, only a small fraction of tweets is geo-tagged at all, leading to a large amount of

165 missed tweets from the area (Sloan et al., 2013).

4.2 Crowdsourcing approaches

To resolve these problems, other strategies were developed. One solution lies in crowdsourcing the analysis of tweets, i.e. asking human volunteers to manually label the data. Naturally, due to the large amount of incoming tweets, many helpers are necessary. Established communities of such volunteers can be activated quickly in a disaster event. One example of such a

170 community is the *Standby Task Force*[6].

To facilitate their work, platforms have been developed over the years. One of the most well-known systems is *Ushahidi*[7]. This platform allows people to share situational information in various media, e.g. by text message, by e-mail, and of course by Twitter. Messages can then be tagged with categories relevant to the event. *Ushahidi* was started by a team of Kenyan citizens during the 2007 Kenyan election crisis, and has since been used successfully in a number of natural disasters, humanitarian

175 crises, and election (for monitoring). Both the server and the platform software are available open-source[8]. Efforts were made to integrate automatic analysis tools into the platform (named "SwiftRiver"), but discontinued in 2015.

Such automatic analysis tools are the motivation for *AIDR* (Imran et al., 2015). *AIDR* was first developed as a quick response to the 2013 Pakistan earthquake. Its main purpose lies in facilitating machine learning methods to streamline the annotation process. In a novel situation, users first choose their own keywords and regions to start collecting a stream of tweets. Then,

180 volunteers annotate relevant categories. A supervised classifier is then trained on these given examples, and is automatically applied to new incoming messages. A front-end platform named *MicroMappers*[9] also exists. *AIDR* is available in an open-source format as well[10]. It has been used in the creation of various data sets and experiments.

Another contribution to crowdsourcing crisis tweets is *CrisisTracker* (Rogstadius et al., 2013). In *CrisisTracker*, tweets are also collected in real-time. Local Sensitive Hashing (LSH) is then applied to detect clusters of topics (so-called stories), so

185 that volunteers can consider these stories jointly instead of single tweets. The *AIDR* engine has also been integrated to provide

---

[6]https://www.standbytaskforce.org/

[7]https://www.ushahidi.com/

[8]https://github.com/ushahidi/Ushahidi_Web

[9]https://micromappers.wordpress.com/

[10]https://github.com/qcri-social/AIDR

topic filtering. As a field trial, the platform was used in the 2012 Syrian civil war. *CrisisTracker* is also available free and open-source[11], but maintenance stopped in 2016.

4.3 Machine learning approaches

As another avenue, various machine learning approaches for automatically detecting crisis-related tweets have been developed over the years. Employed algorithms include Latent Dirichlet Allocation (Resch et al., 2018; Kireyev et al., 2009), Naive Bayes models (Li et al., 2018), Support Vector Machines (Sakaki et al., 2010; Stowe et al., 2018), and Random Forests (Kaufhold et al., 2020). A 2015 overview over these more traditional machine learning approaches can be found in (Imran et al., 2015). In recent years, deep learning techniques, i.e. neural networks, have come to the forefront of research. We will focus on these here.

On a general level, the problem falls under the umbrella of event detection as shown, for example, in (Chen et al., 2015; Feng et al., 2016; Nguyen and Grishman, 2015). Caragea et al. (2016) first employed Convolutional Neural Networks (CNN) for the classification of tweets into those related to flood events and those unrelated. (Lin et al. (2016) also applied CNNs to social media messages, but for the *Weibo* platform instead of Twitter). In many of the following approaches, a type of CNN developed by Kim for text classification is used (Kim, 2014), such as in (Burel and Alani, 2018; Kersten et al., 2019). A schematic is shown in figure 1. These methods achieve accuracies of around 80% for the classification into related and unrelated tweets. In (Burel and Alani, 2018) as well as in (Burel et al., 2017a) and (Nguyen et al., 2016), this kind of model is also used for information type classification.
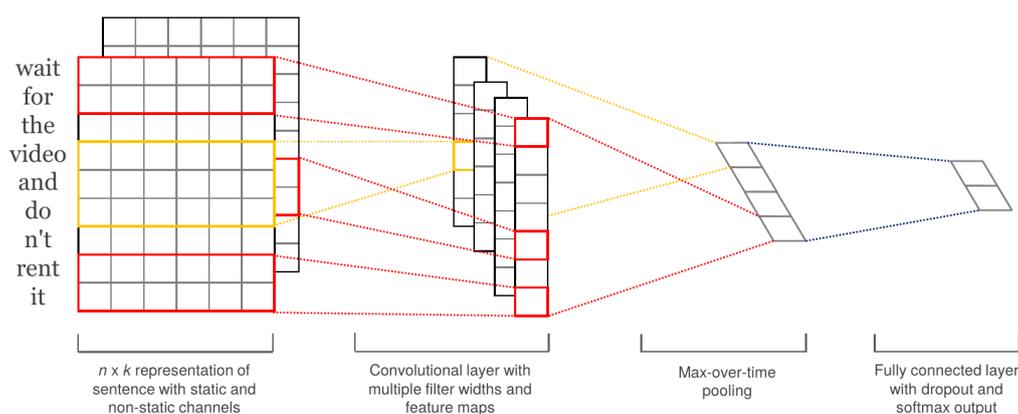
Figure 1. CNN for text classification as proposed by Kim (2014).

[11]https://github.com/JakobRogstadius/CrisisTracker/

Burel et al. (2017a) integrate semantic concepts and entities from *DBPedia*[12]. The resulting system is packaged as *CREES*
(Burel and Alani, 2018), a service that can be integrated into other platforms similar to *AIDR*. The model in (Nguyen et al.,
2016) has the ability to perform online learning, i.e. updating the model as new blocks of tweets arrive.

A crucial component of a social media classification model is the embedding of the text data at the input (i.e. how words or
sentences are mapped to numeric values that the model can process). Many approaches employ *word2vec*, a pre-trained word
embedding that was first presented in 2013 (Mikolov et al., 2013). A version specifically trained on crisis tweets is presented
in (Imran et al., 2016a). More sophisticated models have been developed in the meantime, such as BERT (Devlin et al., 2019),
for which a crisis-specific version is proposed in (Liu et al., 2020). A comparison of various word and sentence embeddings
for crisis tweet classification can be found in (ALRashdi and O'Keefe, 2019).

All of the approaches mentioned above aim to generalize to any kind of event without any *a priori* information. The transfer-
ability of pre-trained models to new events and event types is thoroughly investigated in (Wiegmann et al., 2020b). However, a
real-world system may not need to be restricted in this way; in many cases, its users will already have some information about
the event, and may already have spotted tweets of the required type. This removes the need to anticipate any type of event. It
also directs the system towards a specific event rather than any event happening at that time. Li et al. (2018) and Mazloom et al.
(2019) showed that models adapted to the domain of the event perform better than generalized models. Alam et al. (2018a)
propose an interesting variant for neural networks: Their system includes an adversarial component which can be used to adapt
a model trained on a specific event to a new one (i.e. a new domain).

Kruspe et al. (2019) propose a system that does not assume an explicit notion of relatedness vs. unrelatedness (or relevance
vs. irrelevance) to a crisis event. As described above, these qualities are not easy to define, and might vary for different users
or different types of events. The presented system is able to determine whether a tweet belongs to a class (i.e. crisis event)
implicitly defined by a small selection of example tweets by employing few-shot models. The approach is expanded upon in
(Kruspe, 2019). Another perspective on this is shown in (Kruspe, 2020), where tweets are not classified into explicit "related"
or "unrelated" classes, but rather clustered by topic at time of publication; these clusters can shift over time.

5 Challenges

None of the approaches presented are able to solve the problem of detecting tweets in disaster events perfectly. In some respects,
this is due to technical limitations; however, there are several difficulties immanent to the task itself, which we will discuss in
this section.

Ambiguous problem definition As described in section 2, the task of tweet detection in disasters is ill-defined and heavily
dependent on the final user of the detection product. Annotation experiments also show that even if the goal is clearly
stated, inter-rater agreement is commonly low, with raters often interpreting both the problem statement as well as tweet
content very differently (Stowe et al., 2018). This problem becomes even more emphasized when annotating more fine-
grained labels, e.g. for content type classes or for priority.

---

[12]https://wiki.dbpedia.org/

Linguistic difficulties and language variety As mentioned above, most data sets and, accordingly, methods for automatic tweet detection focus on English-language data. This would often not be the best choice in a real-world scenario; multilingual methods are necessary.

Apart from the question of the language itself, Twitter users frequently utilize an highly idiosyncratic style of writing. Due to the character limitation, words are often abbreviated and grammar is compressed. In contrast to e.g. newspaper articles, user-generated content is relatively noisy, containing lots of erroneous or specialized spelling variations. Additionally, interpretation of tweet content frequently requires (cultural) context knowledge.

Data limitations, legal and privacy issues As mentioned above, Twitter is one of the few popular social media platforms providing an access API to its data to outside users. Despite this, however, limitations exist. For non-paying users, only 1% of the live data of each second can be collected automatically via Twitter's streaming API. For past events, the search API can be utilized, but this only returns tweets still in the search index, which is usually valid for around one week. Older tweets can be retrieved by their ID, but this does not allow for a flexible search. As a free user, the download rate is limited to 18,000 tweets per 15 minutes. Twitter also offers several paid options (called "firehoses") to access more live data, but these are somewhat intransparent. An in-depth analysis of the effect that these limitations can have on research is given in (Valkanas et al., 2014).

Twitter also forbids direct redistribution of tweet content, meaning that the described data sets are only available as lists of tweet IDs. This introduces two difficulties: One, retrieving the actual tweet content ("hydrating") can take a very long time for large data sets due to the rate limit. Two, tweets may become unavailable over time because their creator deleted them or their whole account, or because they were banned. In some cases of older data sets, this means that a significant portion of the corpus cannot be used anymore, impeding reproducibility and comparability of published research.

Apart from access limitations, Twitter and legal restrictions also regulate what researchers are allowed to do with this data. As an example, the Twitter user agreement states (Twitter, Inc., 2020):

"Unless explicitly approved otherwise by Twitter in writing, you may not use, or knowingly display, distribute, or otherwise make Twitter Content, or information derived from Twitter Content, available to any entity for the purpose of: (a) conducting or providing surveillance or gathering intelligence, including but not limited to investigating or tracking Twitter users or Twitter Content; (b) conducting or providing analysis or research for any unlawful or discriminatory purpose, or in a manner that would be inconsistent with Twitter users' reasonable expectations of privacy; (c) monitoring sensitive events (including but not limited to protests, rallies, or community organizing meetings); or (d) targeting, segmenting, or profiling individuals based on sensitive personal information, including their health (e.g., pregnancy), negative financial status or condition, political affiliation or beliefs, racial or ethnic origin, religious or philosophical affiliation or beliefs, sex life or sexual orientation, trade union membership, Twitter Content relating to any alleged or actual commission of a crime, or any other sensitive categories of personal information prohibited by law."

Many interesting research questions are not identical, but related to problematic usages described in this statement, e.g.

270   inference on a user basis or monitoring of protests. Researchers must therefore be careful not to step into prohibited

territory.

Lack of geolocation   In a disaster context, knowing exactly where a tweet was sent is often crucial to the usability of this

information. Twitter provides several ways of detecting geolocation. The most precise of them is the option for users to

275   send their coordinates along with the tweet. However, only about 1% of tweets contain this information (Sloan et al.,

2013). A tweet's location can also be estimated from the location stated in the user profile, or by analyzing the tweet's

content with regards to mention of geolocation. For operationalization, a geocoding to coordinates is then required, which

can be provided by services such as Google Maps or OpenStreetMap's Nominatim. Unfortunately, these geolocations

are prone to errors, e.g. because a user mentions a position other than their own, because they might be traveling, or

280   because the center coordinates of a city are to imprecise to be usable.

6   Related tasks

Once tweets related to a disaster event have been discovered, many further analysis steps are possible. We will only touch upon

those briefly here. As described in section 3, some of the available data sets have already been annotated with these additional

concepts.

285   A popular next step is the classification into semantic or information type classes. Such classes may include sentiments, affected

people seeking various types of assistance, media reports, warnings and advice etc. No common set of such classes exists; in

the *CrisisNLP* and *CrisisLexT26* corpora, 9 and 7 classes are used respectively with some overlap. For the TREC Incident

Streams challenge, potential end users were questioned about their classes of interests, resulting in a two-tier ontology with 25

classes on the lower tier. As an added difficulty, classes often overlap in tweets; for these reasons, TREC allows multiple labels

290   per tweet. Furthermore, annotators often disagree whether an information type is present in a tweet.

Another way of further discerning between tweets is a distinction between levels of informativeness or priority. This can be

implemented either with discrete classes (low/medium/high importance), on a continuous numerical scale, or as a ranking of

tweets. The *CrisisLexT26* and *TREC-IS 2019A* data sets contain such annotations.

Apart from approaches processing single tweets, the analysis of the spatio-temporal distribution and development of discussed

295   topics within affected areas at different scales may provide valuable insights (Kersten and Klan, 2020). Other research focuses

on the detection of specific events, or types of events (e.g. floods, wildfires, or man-made disasters) (e.g. Burel et al. (2017b)).

This can often be helpful when social media is used as an alerting system. Additionally, models specialized to event types

can be more precise and allow for different distinctions than general-purpose models (Kersten et al., 2019; Wiegmann et al.,

2020b); detection of the event type enables the automatic selection of such a more specialized method.

300   Apart from these text-based tasks, image analysis can also be a helpful source of information. As an example, images posted

on social media can be used to determine the degree of destruction in the aftermath of a disaster (Alam et al., 2017; Nguyen et al., 2017b).

7 Future work

Many very interesting new analysis tasks are thinkable based on the detection methods described so far, particularly when employing automatic methods. A good starting point to identify relevant practical issues related to acquisition tasks that could potentially be solved by analyzing social media data is provided in (Wiegmann et al., 2020c). Here, opportunities and risks of disaster data from social media are investigated by means of a systematic review of currently available incident information.

One aspect that has not been considered in research so far is how an event changes over time. New approaches could be used to analyze the spatiotemporal development of disasters, and how this could be utilized in disaster prevention. During the course of an event, clustering methods could be employed to rapidly detect novel developments such as sub-events or new topics. This is particularly relevant for relief providers, who require extremely fast situation monitoring.

As described in section 5, localizing information coming from Twitter is often a challenge. Approaches that are able to deal with this lack of information are necessary. This could be implemented either by deriving location by some other means, or by spatiotemporal and semantic analysis of large sets of tweets to cross-reference and check information.

As mentioned above, languages other than English have also usually not been included in research on this topic. Multilingual approaches would be a very helpful next step to facilitate usage of such methods in regions of the world where English is not the main language. Another aspect of the data that has not been used often so far are images posted by users. In particular, a multimodal joint analysis of text and images is very interesting from both the research as well as the usage perspective. The *CrisisMMD* data set is an interesting first step in this direction.

As described in section 4, some crowdsourcing approaches already integrate machine learning-based methods. In future work, expanding human-in-the-loop approaches would be very useful.

In general, social media is usually not the only source of information and cannot provide a full picture of the situation. Therefore, an integration with other information sources, such as earth observation data, media information, or governmental data, is highly relevant.

8 Conclusions

In this review paper, we gave an overview over current methods to detect tweets pertaining to disaster events. There are three main ways to approach this problem: Filtering tweets by characteristics such as location and contained keywords or hashtags, crowdsourcing, and machine-learning based methods. Each of these has its advantages and disadvantages, but machine learning appears to be the current main avenue of research with big improvements in the past few years. To train and test these models, various data sets spanning one or multiple events have been created and are available online. However, these usually only provide ID's of the tweets, which leads to changes in the data sets over time.

A central issue of this task is the problem statement. Existing publications focus on detecting "related", "relevant", or "informative" tweets, but struggle to define these concepts. Other difficulties include the subjectivity of classes and tweet interpretations, data limitations, linguistic difficulties, and legal issues.

335 Nevertheless, large strides have been made in the past years to tackle this problem, and research in this area remains highly active. Many related and novel analysis tasks are possible in the future.

Author contributions. Anna Kruspe wrote this paper with assistance and input from Jens Kersten and Friederike Klan.

Competing interests. The authors declare that they have no conflict of interest.

340

References

Alam, F., Imran, M., and Ofli, F.: Image4Act: Online Social Media Image Processing for Disaster Response, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17, p. 601–604, 2017.

Alam, F., Joty, S., and Imran, M.: Domain Adaptation with Adversarial Training and Graph Embeddings, in: 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, 2018a.

Alam, F., Ofli, F., and Imran, M.: CrisisMMD: Multimodal Twitter Datasets from Natural Disasters, in: Proceedings of the 12th International AAAI Conference on Web and Social Media (ICWSM), 2018b.

ALRashdi, R. and O'Keefe, S.: Deep Learning and Word Embeddings for Tweet Classification for Crisis Response, 2019.

American Red Cross: Social Media in Disasters and Emergencies, 2010.

Appen Ltd.: Multilingual Disaster Response Messages, https://appen.com/datasets/combined-disaster-response-data/, 2020.

Banda, J. M., Tekumalla, R., Wang, G., Yu, J., Liu, T., Ding, Y., Artemova, K., Tutubalina, E., and Chowell, G.: A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration, https://doi.org/10.5281/zenodo.3723939, https://doi.org/10.5281/zenodo.3723939, 2020.

Burel, G. and Alani, H.: Crisis Event Extraction Service (CREES) - Automatic Detection and Classification of Crisis-related Content on Social Media, in: Proceedings of the 15th International Conference on Information Systems for Crisis Response and Management (IS-CRAM), p. 12, 2018.

Burel, G., Saif, H., and Alani, H.: Semantic Wide and Deep Learning for Detecting Crisis-Information Categories on Social Media, in: International Semantic Web Conference (ISWC), Vienna, Austria, 2017a.

Burel, G., Saif, H., Fernandez, M., and Alani, H.: On semantics and deep learning for event detection in crisis situations, in: Workshop on Semantic Deep Learning (SemDeep), at ESWC 2017, 2017b.

Caragea, C., Silvescu, A., and Tapia, A.: Identifying informative messages in disaster events using Convolutional Neural Networks, in: ISCRAM 2016 Conference Proceedings - 13th International Conference on Information Systems for Crisis Response and Management, edited by Antunes, P., Banuls Silvera, V., Porto de Albuquerque, J., Moore, K., and Tapia, A., Information Systems for Crisis Response and Management, ISCRAM, 2016.

Chen, Y., Xu, L., Liu, K., Zeng, D., and Zhao, J.: Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks, in: Annual Meeting of the Association for Computational Linguistics (ACL), https://doi.org/10.3115/v1/P15-1017, http://aclweb.org/anthology/P15-1017, 2015.

de Albuquerque, J. P., Herfort, B., Brenning, A., and Zipf, A.: A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management, International Journal of Geographical Information Science, 29, 667–689, 2015.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019.

Feng, X., Qin, B., and Liu, T.: A language-independent neural network for event detection, in: 54th Annual Meeting of the Association for Computational Linguistics (ACL), Berlin, Germany, 2016.

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P.: Practical extraction of disaster-relevant information from social media, in: Proceedings of the 22nd International Conference on World Wide Web (WWW) Companion, pp. 1021–1024, ACM Press, Rio de Janeiro, Brazil, https://doi.org/10.1145/2487788.2488109, http://dl.acm.org/citation.cfm?doid=2487788.2488109, 2013.

Imran, M., Castillo, C., Lucas, J., Meier, P., and Vieweg, S.: AIDR: Artificial Intelligence for Disaster Response, in: Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion, p. 159–162, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/2567948.2577034, https://doi.org/10.1145/2567948.2577034, 2014.

Imran, M., Castillo, C., Diaz, F., and Vieweg, S.: Processing Social Media Messages in Mass Emergency: A Survey, ACM Computing Surveys, 47, 1–38, https://doi.org/10.1145/2771588, http://dl.acm.org/citation.cfm?doid=2775083.2771588, 2015.

Imran, M., Mitra, P., and Castillo, C.: Twitter as a lifeline: Human-annotated twitter corpora for NLP of crisis-related messages, in: Tenth International Conference on Language Resources and Evaluation (LREC), Portoroz, Slovenia, 2016a.

Imran, M., Mitra, P., and Srivastava, J.: Enabling Rapid Classification of Social Media Communications During Crises, Int. J. Inf. Syst. Crisis Response Manag., 8, 1–17, https://doi.org/10.4018/IJISCRAM.2016070101, https://doi.org/10.4018/IJISCRAM.2016070101, 2016b.

Imran, M., Castillo, C., Diaz, F., and Vieweg, S.: Processing Social Media Messages in Mass Emergency: Survey Summary, in: Companion Proceedings of the The Web Conference 2018, WWW '18, pp. 507–511, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2018.

Kaufhold, M.-A., Bayer, M., and Reuter, C.: Rapid relevance classification of social media posts in disasters and emergencies: A system and evaluation featuring active, incremental and online learning, Information Processing & Management, 57, http://www.sciencedirect.com/science/article/pii/S0306457319303152, 2020.

Kersten, J. and Klan, F.: What Happens Where During Disasters? A Workflow for the Multi-Faceted Characterisation of Crisis Events Based on Twitter Data, Journal of Contingencies and Crisis Management, special Issue: Knowledge, Semantics and AI for Risk and Crisis Management" of the Journal of Contingencies and Crisis Management, in press, 2020.

Kersten, J., Kruspe, A., Wiegmann, M., and Klan, F.: Robust Filtering of Crisis-related Tweets, in: International Conference on Information Systems for Crisis Response and Management (ISCRAM), Valencia, Spain, 2019.

Kim, Y.: Convolutional neural networks for sentence classification, in: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014.

Kireyev, K., Palen, L., and Anderson, K.: Applications of topics models to analysis of disaster-related twitter data, in: NIPS Workshop on Applications for Topic Models: Text and Beyond, 2009.

Kruspe, A.: Few-shot tweet detection in emerging disaster events, in: AI+HADR Workshop @ NeurIPS, 2019.

Kruspe, A.: Detecting novelty in social media messages during emerging crisis events, in: International Conference on Information Systems for Crisis Response and Management (ISCRAM), 2020.

Kruspe, A., Kersten, J., and Klan, F.: Detecting Event-Related Tweets by Example using Few-Shot Models, in: International Conference on Information Systems for Crisis Response and Management (ISCRAM), Valencia, Spain, 2019.

Kumar, S., Barbier, G., Abbasi, M. A., and Liu, H.: TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief, in: International AAAI Conference on Weblogs and Social Media (ICWSM), Barcelona, Spain, 2011.

Landwehr, P. M. and Carley, K. M.: Social Media in Disaster Relief - Usage Patterns, Data Mining Tools, and Current Research Directions, in: Data Mining and Knowledge Discovery for Big Data, edited by Chu, W. W., pp. 225–257, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-40837-3_7, 2014.

Li, H., Caragea, D., Caragea, C., and Herndon, N.: Disaster response aided by tweet classification with a domain adaptation approach, Journal of Contingencies and Crisis Management, 26, 16–27, https://doi.org/10.1111/1468-5973.12194, http://doi.wiley.com/10.1111/1468-5973.12194, 2018.

Lin, Z., Jin, H., Robinson, B., and Lin, X.: Towards an accurate social media disaster event detection system based on deep learning and semantic representation, in: Proceedings of the 14th Australasian Data Mining Conference, Canberra, Australia, pp. 6–8, 2016.

Liu, J., Singhal, T., Blessing, L. T. M., Wood, K. L., and Lim, K. H.: CrisisBERT: a Robust Transformer for Crisis Classification and Contextual Crisis Embedding, 2020.

Mazloom, R., Li, H., Caragea, D., Caragea, C., and Imran, M.: A Hybrid Domain Adaptation Approach for Identifying Crisis-Relevant Tweets, International Journal of Information Systems for Crisis Response and Management, 11, 2019.

McCreadie, R., Buntain, C., and Soboroff, I.: TREC Incident Streams: Finding Actionable Information on Social Media, in: International Conference on Information Systems for Crisis Response and Management (ISCRAM), Valencia, Spain, 2019.

McMinn, A. J., Moshfeghi, Y., and Jose, J. M.: Building a large-scale corpus for evaluating event detection on twitter, in: ACM International Conference on Information and Knowledge Management (CIKM), pp. 409–418, San Francisco, CA, USA, 2013.

Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient Estimation of Word Representations in Vector Space, CoRR, abs/1301.3781, http://arxiv.org/abs/1301.3781, 2013.

Nguyen, D. T., Joty, S. R., Imran, M., Sajjad, H., and Mitra, P.: Applications of Online Deep Learning for Crisis Response Using Social Media Information, CoRR, abs/1610.01030, 2016.

Nguyen, D. T., Al-Mannai, K. A., Joty, S., Sajjad, H., Imran, M., and Mitra, P.: Robust classification of crisis-related data on social networks using convolutional neural networks, in: Proceedings of the 11th International Conference on Web and Social Media (ICWSM), pp. 632–635, AAAI press, 2017a.

Nguyen, D. T., Ofli, F., Imran, M., and Mitra, P.: Damage Assessment from Social Media Imagery Data During Disasters, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17, p. 569–576, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3110025.3110109, https://doi.org/10.1145/3110025.3110109, 2017b.

Nguyen, T. H. and Grishman, R.: Event Detection and Domain Adaptation with Convolutional Neural Networks, in: 53rd Annual Meeting of the Association for Computational Linguistics (ACL), Beijing, China, https://doi.org/10.3115/v1/P15-2060, http://aclweb.org/anthology/P15-2060, 2015.

Niles, M. T., Emery, B. F., Reagan, A. J., Dodds, P. S., and Danforth, C. M.: Social media usage patterns during natural hazards, PLOS ONE, 14, 1–16, https://doi.org/10.1371/journal.pone.0210484, https://doi.org/10.1371/journal.pone.0210484, 2019.

Olteanu, A., Castillo, C., Diaz, F., and Vieweg, S.: CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises, in: AAAI Conference on Weblogs and Social Media (ICWSM), Ann Arbor, MI, USA, 2014.

Olteanu, A., Vieweg, S., and Castillo, C.: What to Expect When the Unexpected Happens: Social Media Communications Across Crises, in: Conference on Computer Supported Cooperative Work and Social Computing (ACM CSCW), Vancouver, BC, Canada, 2015.

Parilla-Ferrer, B. E., Fernandez, P., and T. Ballena IV, J.: Automatic Classification of Disaster-Related Tweets, in: International conference on Innovative Engineering Technologies (ICIET), 2014.

Qazi, U., Imran, M., and Ofli, F.: GeoCoV19: A Dataset of Hundreds of Millions of Multilingual COVID-19 Tweets with Location Information, ACM SIGSPATIAL Special, 12, 6–15, https://doi.org/10.1145/3404111.3404114, https://doi.org/10.1145/3404111.3404114, 2020.

Resch, B., Usländer, F., and Havas, C.: Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment, Cartography and Geographic Information Science, 45, 362–376, 2018.

Reuter, C. and Kaufhold, M.-A.: Fifteen Years of Social Media in Emergencies: A Retrospective Review and Future Directions for Crisis Informatics, Journal of Contingencies and Crisis Management (JCCM), 26, 41–57, http://tubiblio.ulb.tu-darmstadt.de/108144/, special Issue: Human-Computer-Interaction and Social Media in Safety-Critical Systems, 2018.

Rogstadius, J., Vukovic, M., Teixeira, C. A., Kostakos, V., Karapanos, E., and Laredo, J. A.: CrisisTracker: Crowdsourced social media curation for disaster awareness, IBM Journal of Research and Development, 57, 4:1–4:13, 2013.

Sakaki, T., Okazaki, M., and Matsuo, Y.: Earthquake shakes Twitter users: Real-time event detection by social sensors, in: Proceedings of the 19th International Conference on World Wide Web, pp. 851–860, ACM, 2010.

Sloan, L., Morgan, J., Housley, W., Williams, M., Edwards, A., Burnap, P., and Rana, O.: Knowing the Tweeters: Deriving Sociologically Relevant Demographics from Twitter, Sociological Research Online, 18, 7, 2013.

Smith, S.: Coronavirus (covid19) Tweets, https://www.kaggle.com/smid80/coronavirus-covid19-tweets, 2020.

Stowe, K., Palmer, M., Anderson, J., Kogan, M., Palen, L., Anderson, K. M., Morss, R., Demuth, J., and Lazrus, H.: Developing and Evaluating Annotation Procedures for Twitter Data during Hazard Events, in: Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018), pp. 133–143, Association for Computational Linguistics, http://aclweb.org/anthology/W18-4915, 2018.

Twitter, Inc.: Developer Agreement and Policy, https://developer.twitter.com/en/developer-terms/agreement-and-policy, 2020.

Valkanas, G., Katakis, I., Gunopulos, D., and Stefanidis, A.: Mining Twitter Data with Resource Constraints, in: 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, Poland, August 11-14, 2014 - Volume II, pp. 157–164, IEEE Computer Society, https://doi.org/10.1109/WI-IAT.2014.29, https://doi.org/10.1109/WI-IAT.2014.29, 2014.

Wiegmann, M., Kersten, J., Klan, F., Potthast, M., and Stein, B.: Disaster Tweet Corpus 2020, https://doi.org/10.5281/zenodo.3713920, https://doi.org/10.5281/zenodo.3713920, 2020a.

Wiegmann, M., Kersten, J., Klan, F., Potthast, M., and Stein, B.: Analysis of Detection Models for Disaster-Related Tweets, in: 17th ISCRAM Conference, edited by Hughes, A., McNeill, F., and Zobel, C., ISCRAM, 2020b.

Wiegmann, M., Kersten, J., Potthast, M., Klan, F., and Stein, B.: NHESS Special Issue: Groundbreaking technologies, big data, and innovation for disaster risk modelling and reduction, under review, 2020c.

Win, S. S. M. and Aung, T. N.: Target oriented tweets monitoring system during natural disasters, in: Proceedings of the IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), pp. 143–148, https://doi.org/10.1109/ICIS.2017.7959984, 2017.