Regular Paper

An Analysis of Rescue Requests on Twitter in a Disaster

Yuki Koizumi[†], Junji Takemasa[†], Toru Hasegawa[†], and Yoshinobu Kawabe[‡]

[†]Osaka University, Japan
[‡]Aichi Institute of Technology, Japan
{ykoizumi, j-takemasa, t-hasegawa}@ist.osaka-u.ac.jp, kawabe@aitech.ac.jp

Abstract - During catastrophic disasters like the Japan floods 2018, phone-based emergency call systems may not work as expected due to heavy congestion and network disruption. People needing help use social media, e.g., Twitter and Facebook, for delivering rescue requests, which complements phone-based emergency call systems, in disasters. Machine learning is a promising approach to automated rescue request extraction from a vast amount of social media posts. Since it is a critical task, such classifiers should produce few false negatives (rescue requests identified as non rescue requests). The objective of the study is to learn lessons to develop better classifiers for rescue request extraction. Hence, we analyze rescue-related tweets, which are tweets containing rescue-related keywords like "rescue," conduct a classification experiment of rescue requests using a classifier based on the bidirectional encoder representations from transformers (BERT) model, and investigate the classification results, particularly focusing on why the classifier produce false negatives. Furthermore, we construct an annotation mechanism based on a recurrent neural network to understand why the classifiers produce false negatives.

Keywords: Social Media, Twitter, Disaster, Analysis, Machine Learning

1 INTRODUCTION

Delivering rescue requests from citizens in need of help to the right persons, such as rescue authorities and first responders, is a key to effective disaster management. Phone-based emergency call services, however, may not work as expected during and after catastrophic disasters because of network disruption and congestion [1].

Circumstances of rescue requests during disasters have been changing. Some citizens needing help use social media, e.g., Twitter and Facebook, for delivering rescue requests, especially when phone-based emergency call services are inadequate, and social media complements existing phone-based emergency call services [1], [2]. For instance, in the U.S., several studies reported the use of social media during Hurricane Harvey. Stelter [2] reported that hundreds of stranded Texas residents sought help by posting on Facebook and Twitter during Hurricane Harvey, and Facebook and Twitter were clearly used as a supplement for traditional emergency services. Jahanian et al. [1] also reported that people tweeted their addresses for seeking rescues during hurricane Harvey, especially, when they felt that traditional aid-seeking methods was not adequate. Japan, for instance, had several catastrophic floods in 2018 and 2019. One is the *heavy rain of July 2018*, also referred to as *Japan floods 2018* [3] and another is the *19th typhoon of 2019* [4]. We observed that many rescue requests were posted on Twitter during both the two disasters.

Few rescue requests on social media, nevertheless, contributed to actual rescue activities. Though our analysis, discussed in Section 3, reveals that 312 rescue requests were posted during the Japan floods 2018, we did not find any news that reported any of the rescue requests directly contributed to rescue activities. As another example, a local government, Nagano Prefecture, deployed several workers to capture rescue requests from Twitter [5] during the 19th typhoon. On the one hand this activity finally contributed to saving about 50 victims, but on the other hand it consumed many workers of the local government, who might have contributed to other tasks. These tales imply that it is indispensable to extract rescue requests on social media automatically to utilize them for rescue activities. Machine learning is a promising technique for filtering rescue requests on social media [6]–[8].

Rescue-related social media posts, which are defined as tweets with rescue-related keywords, such as rescue, are a mixture of good and bad; that is, rescue-related tweets contain non rescue requests as well as rescue requests, as described in Section 3. Understanding actual rescue-related tweets is a key to realizing good rescue request classifiers. Hereafter, we refer to social media posts as tweets since this study focuses on Twitter as social media. Several studies analyzed rescue requests on Twitter, especially focusing on tweets during flood disasters in Japan [9], [10]. Sato and Imamura [9] analyzed rescue requests found on Twitter during two flood disasters in Japan and claimed that tweets tagged with #rescue include many non-rescue requests. Song and Fujishiro [10] interviewed a news article writer and a member of the social media listening team of a Japanese television company about rescue requests observed during flood disasters in Japan. In addition to the effort, we in this study analyze rescue requests on Twitter from the perspective of classification with machine learning.

The main contribution of this paper is to draw lessons for developing machine learning based classifiers identifying rescue requests through analyses on actual tweets during flooding disasters. One of the severe requirements for automated rescue request classification is reducing (ideally minimizing) false negatives, which are rescue requests identified as non rescue requests, because missing rescue requests is a matter of life and death for citizens needing help. Although much effort has been devoted to machine learning based tweet classification

ISSN1883-4566 © 2023 - Informatics Society and the authors. All rights reserved.

(not rescue requests) [6]–[8], few studies focus on reasons for misclassification, including false negatives and even false positives. Kshirsagar et al. [8], for example, tried to detect states of crisis, such as suicide, self-harm, abuse, or eating disorders, using machine learning. However, they focused on identifying what phrases contribute to classification results using an attention mechanism. Similarly, Ma et al. [6] and Ruchansky et al. [7] tried to classify fake news on Twitter. However, they focused on selecting appropriate features fed to machine learning to improve classification accuracy. In contrast to the existing studies, we analyze classification results, aiming at understanding reasons for misclassification, particularly false negatives.

We in this study conduct analyses on actual tweets in the following steps: First, we captured tweets that include disaster-related keywords (e.g., flooding and rescue, as well as the names of disaster-affected areas) on Twitter during several recent flooding disasters in Japan. Second, we analyze the captured tweets similarly to the work in [9] to understand how rescue-related tweets evolve during a disaster. Third, we analyze rescue-related tweets to understand what kinds of tweets are included in rescue-related tweets. In this step, we categorize rescue-related tweets into several categories, such as rescue requests, disaster information, and sympathy. Finally, we conduct experiments classifying rescue requests from numerous tweets using machine learning. One of the essential lessons learned from the experiment is that the variety of textual contexts of rescue-related tweets is a reason for producing false negatives. The analysis in the third step helps investigate this lesson.

This paper is extended from its conference version [11] from the following two aspects: First, we have updated the analysis based on machine learning by using the state-of-the-art machine learning model, i.e., the bidirectional encoder representations from transformers (BERT) model [12]. Second, we provide more detailed analysis results and observations with introducing actual tweets in disasters.

The paper is organized as follows: We first analyze disaster-related tweets in Section 2 and rescue-related tweets in Section 3. Based on the observations found in the analyses, we build a classifier to extract rescue requests from Twitter in Section 4. Section 5 briefly summarizes related work and Section 6 finally concludes this paper.

2 ANALYSIS ON DISASTER-RELATED TWEETS

2.1 Objective and Overview of Analysis

This section analyzes disaster-related tweets captured during the recent floods in Japan. The objective of the analysis in this section is to understand tweets mentioning a catastrophic disaster, especially during and in the aftermath of the disaster. More precisely, we investigate how many tweets mentioning the disaster exist, how the number of such tweets evolves, and when citizens post tweets having rescue-related keywords. We should note that though our analysis method is similar to that in [9], our dataset is different from that in [9], i.e., our dataset contains both rescue-related tweets and disaster information (non rescue requests) unlike the dataset in [9] that only includes tweets tagged with #rescue. Thus, this section's results will add value to the existing studies. More precisely, the results in this section help us understand statistics and time-series information about disaster-related tweets and rescue-related tweets.

2.2 Dataset

We collected tweets from three flood disasters in Japan: the Japan floods 2018 [3], the 15th [13], and the 19th typhoons [4] of 2019 in Japan. We use the tweets from the Japan floods 2018 for the analyses in this section and those from the other two disasters for the classification in the next section.

The tweets were captured via the Twitter search API [14] by specifying disaster-related keywords, which were selected so that we were able to capture as many disaster-related tweets as possible. Tweets retweeted by means of the twitter official API were eliminated from the dataset. Let us note that all the keywords are Japanese, and therefore all the tweets are also written in Japanese. The keywords are categorized into five classes: rain disaster, rescue request, first responder, volunteer, and infrastructure, as summarized in Table 1. We refer to tweets containing at least one keyword in any of the classes as *disaster-related tweets*. In the same way, tweets containing keywords in the classes of rescue request, first responder, infrastructure, and volunteer are referred to as *rescue-*, *first responder-*, *infrastructure-*, and *volunteer-related tweets*, respectively. Since most disaster-related tweets contain keywords in the rain disaster class, we do not focus on tweets in this class. Each class may have tweets unrelated to the class because they were captured with keyword search. For instance, tweets that are not rescue requests, although having rescue-related keywords, are categorized into rescue-related tweets.

The number of collected tweets, i.e., disaster-related tweets, during and in the aftermath of the Japan floods 2018 is 6,978,389, and the number of tweets in each class is summarized in Table 2. As we discuss later in Section 4, we captured 246,807 rescue-related tweets.

2.3 Temporal Analysis

We investigate how the number of tweets grows in accordance with situations in a flood disaster. Before explaining the result, we briefly explain the Japan floods 2018 to understand this analysis. From the end of June to the middle of July 2018, it had been raining heavily and steadily in western Japan. As a result, it caused widespread and catastrophic floods throughout western Japan, especially in Okayama and Hiroshima. From July 5th to 7th, the most severe floods occurred in Okayama and Hiroshima.

The time series of the number of disaster-related tweets is plotted in Fig. 1a. The horizontal and vertical axes represent the date and the number of tweets in each day. The number of disaster-related tweets steeply increases around July 5th as floods became severe, and its peak is on July 6th and 7th, the most intense days of floods. The numbers of tweets on July 6th and 7th are 591,514 and 584,692, respectively. Note that

Table 1: Search keywords used for collecting disaster-related tweets

Category	Keywords
Rain disaster	Heavy rain, rain disaster, disaster, flood, flood disaster, disaster-hit area, flood-hit area, river burst, evacuation, and (confirmation of someone's) safety
Rescue request	Rescue, rescue request, SOS, and help (me)
First responder	Rescue team, fire fighting team, police, Japan self-defense forces, hospital, and local government
Infrastructure	Infrastructure, lifeline, water supply, electricity supply, gas supply, (network) disconnection, (network) congestion, (network) failure, and recovery
Volunteer	Volunteer, support, and relief supply

Table 2: The number of tweets captured during and in the aftermath of the Japan floods 2018 (July 1–29, 2018)

Category	Number
Total (disaster-related tweets)	6,978,389
Rescue-related tweets	246,807
First responder-related tweets	932,605
infrastructure-related tweets	889,889
Volunteer-related tweets	324,935

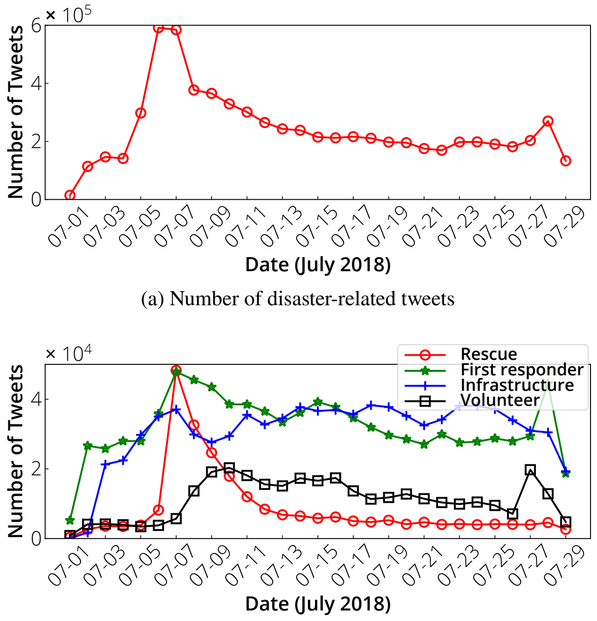
the number of disaster-related tweets again increases at the end of July because another typhoon came to Japan.

Figure 1b indicates the number of rescue-, first responder-, infrastructure-, and volunteer-related tweets. Though rescue-related tweets also increased as floods became severe, the trend is shifted slightly behind compared to that of disaster-related tweets. That is, the number of rescue-related tweets suddenly increases on July 7th, and the peak is also on the same day. We captured 48,272 rescue-related tweets on July 7th. First responder and infrastructure-related tweets were captured almost invariably throughout the disaster. In contrast, the number of volunteer-related tweets increases in the aftermath of the disaster.

3 ANALYSIS ON RESCUE-RELATED TWEETS

3.1 Objective and Overview of Analysis

This section explains observations obtained through reading tweets containing rescue-related keywords. The primary objective of the analysis is to understand what kinds of tweets are included in rescue-related tweets. We categorize rescue-related tweets into eight categories. In the classification experiment in Section 4, we will reveal that a mixture of texts in different contexts is one of the causes of misclassification in rescue request extraction. The categorization is indispensable for drawing such observations. Furthermore, we observe that the locations where real rescue requests in our dataset were posted were concentrated in a narrow area. The observation is also essential for drawing a lesson discussed in the next section.



(a) Number of disaster-related tweets

(b) Number of rescue, first responder, infrastructure, and volunteer-related tweets

Figure 1: Time series analysis of the number of tweets

3.2 Dataset

Since it is nearly impossible to read an enormous number of tweets containing rescue-related keywords (246,807, as shown in Table 2), we further sampled 5,304 tweets from the rescue-related tweets using hashtags. Specifically, we selected tweets that contains hashtags related with rescue requests (i.e., *#rescue*, *#rescue request*, *#SOS*, *#help*, and *#help_me*) from the rescue-related tweets. The reasons for filtering tweets with the hashtags are twofold. One is that Twitter Japan (@TwitterLifeline) recommended using the #rescue hashtag [15]. The other is that many of the actual rescue requests eventually had the hashtags since other Twitter users re-posted the tweets, especially if they did not contain the hashtags, by adding the hashtags.

3.3 Taxonomy

We found that a keyword search with rescue-related keywords is not a good approach for extracting rescue requests because it results in a mixture of rescue and non-rescue requests.

Therefore, as the first step of extracting rescue requests from Twitter, we need to understand what kinds of tweets are included in rescue-related tweets. We read all the sampled tweets to achieve the objective. As a byproduct of reading the tweets, we develop a taxonomy of disaster-related tweets. This section explains the taxonomy categories, introduces actual disaster-related tweets categorized according to the taxonomy, and discusses observations.

3.3.1 Categories

This section overviews the taxonomy categories, which are summarized in Table 3. Alam et al. [16] have similarly developed a taxonomy about tweets captured during disasters. While their taxonomy categorizes disaster-related tweets, our taxonomy focuses only on rescue-related tweets and analyzes them in detail. As shown in Table 3, rescue-related tweets are categorized into eight categories. We explain the criteria for categorizing tweets in the rest of this section, and we introduce several examples and discuss observations in the subsequent sections.

The rescue request category is based on a recommendation about rescue requests on Twitter, which Twitter Japan (@TwitterLifeline) posted [15]. The recommendation suggests that victims post a rescue request indicating the detailed current situation and their postal address (or their geographical location) as well as the hashtag #rescue. Hence, we define *rescue requests* as tweets containing the following information:

1. Help message: A message representing that the persons posting the tweet themselves are victims and are in need of rescue.
2. Disaster situation: A message explaining the situation of the victims.
3. Location: The location of the victims.

Based on the definition, we next define *incomplete rescue requests* as tweets containing a help message but lacking either a location or a disaster situation (or both). In contrast, *disaster situations* are defined as tweets containing disaster situations but not containing a help message. The fundamental idea behind these definitions is that people posting tweets in the incomplete rescue request category are likely to be facing a dangerous situation, whereas people posting tweets in the disaster situation category themselves do not face a dangerous situation. Nonetheless, tweets in the disaster situation category are helpful in knowing the disaster.

The categories of *sympathy*, *advice*, and *volunteer* are straightforwardly defined as the definitions mentioned above. In contrast, tweets in the categories of *exploitation* and *unrelated* are similar because tweets in both categories are unrelated to the disaster. A difference lies in the fact that rescue-related keywords used in tweets in the exploitation categories are intentionally used. Tweets in the exploitation category, for instance, include a political opinion and a service advertisement and intentionally put #rescue tag so that the tweets appear on the trend line. In contrast, tweets in the unrelated category use rescue-related keywords in non-disaster situations, such as smartphone games, books, movies, and TV shows.

3.3.2 Actual Tweets

This section introduces examples of actual tweets classified into each category. Note that all the tweet examples in this paper were originally written in Japanese, but we translated them into English. In addition, privacy information, such as complete postal addresses, building names, and victim names, is concealed, which is indicated by the dash sign (—).

Rescue Request: A typical example of rescue requests is as follows:

> #rescue_request URGENT! Please rescue us, **— (full postal address), — (name of this victim)**. My elderly father, mother, and I are isolated in our house. The first floor of the house is fully flooded, and we have taken refuge in a closet on the second floor. Please help us!

The tweet includes all the three kinds of information (a help message, a disaster situation, and a location).

Incomplete Rescue Request: The following tweets are categorized into incomplete rescue requests.

> 1. Please help me! **— (full postal address)**
> 2. #rescue The mountain is likely to collapse, and we are stuck. #Hiroshima

Although tweet #1 contains the exact location of this victim, the victim's situation is still unknown. Similarly, rescue teams need to know the exact location of the victim in tweet #2. In this way, tweets in this category lack any of the three kinds of information, and thus rescue teams (or local government officers) need to communicate with victims to collect the missing information.

Disaster Situation: Tweets in the disaster situation category do not contain a help message though the tweets contain somewhat helpful information for disaster management. Since tweets do not contain a help message, posters of the tweets themselves or acquaintances of the posters, such as their family, relatives, colleagues, and friends, are not likely to be in need of rescue. Another situation is that tweets describe general people rather than specific people even though they may face a dangerous situation. The following tweet is an example of tweets in this category.

> None of the news media have been reporting anything about Ehime. Rescue in mountainous areas has been delayed. People there are almost isolated. They are sending out lots of messages and requesting rescue. Please help them. #rescue #Ehime #— (town name)

This is a borderline tweet between the incomplete rescue request and disaster situation categories. The reason why we categorize it into the disaster situation category is that it mentions that some people are requesting rescue but they seem to be general people in disaster-stricken areas.

Sympathy: Posters of tweets in the sympathy category express in their tweets that they are praying for the safe rescue of the victims. Though such tweets have rescue-related keywords, they seldom contain information helpful for disaster management. A typical example of tweets in this category is as follows:

Table 3: A taxonomy of rescue-related tweets

Category	Tweets	Ratio
Rescue request	Tweets indicating all the following information a rescue request, situations of victims, and locations of victims.	24.5%
Incomplete rescue request	Tweets indicating a rescue request but lacking either the situation and the location of victims (or both).	8.7%
Disaster situation	Tweets reporting a situation of the disaster.	7%
Sympathy	Tweets praying for the safe rescue of victims.	25.9%
Advice	Tweets giving advice to victims or adding supplementary information to existing tweets.	20.0%
Volunteer	Tweets offering or requesting voluntary contributions, e.g., voluntary work and donations of supplies like foods, water, and clothes.	1.2%
Exploitation	Tweets mentioning something unrelated to rescue requests by intentionally exploiting rescue-related keywords so that the tweets are widespread.	3.5%
Unrelated	Tweets that is not related to the disaster while it contains rescue-related keywords, which are used in a different context from rescue requests, e.g., tweets regarding a smartphone game application.	9.5%

> A town I have visited several times faces a catastrophic situation. Please rescue people there. #— (town name)

Advice: Tweets in the advice category offer supplementary information, particularly for victims requesting rescue. Since such tweets are for victims needing rescue, they contain rescue-related keywords. However, they seldom offer helpful information for disaster management. An example of tweets in this category is as follows:

> People who are waiting for rescue should indicate something conspicuous so that rescue teams easily find you.

Exploitation: An example of tweets in the exploitation category is as follows:

> @— (politician account name) Please search Twitter for the hashtags “#rescue” instead of having a dinner party, and you will know the current situation that would be more critical than your party.

Tweets in this category tend to exploit rescue-related hashtags, particularly for satisfying their own purposes. For example, a typical type of tweet expresses criticisms of politicians like the example above. Another type of tweet exploits rescue-related hashtags to satisfy the posters’ desire, such as advertising their services, increasing the number of followers, and disseminating their tweets.

Unrelated: Other tweets are categorized into the unrelated category. An example in this category is as follows:

> Please help me with assignments in my university lectures. #SOS

Tweets in this category use rescue-related keywords in a context different from disaster management, such as daily life, smartphone games, TV shows, and movies.

3.3.3 Observations

The taxonomy implies that extracting rescue requests via keyword searches is challenging since many tweets are unrelated to rescue requests while containing rescue-related keywords. As shown in Table 3, more than 65% of tweets in our dataset are not rescue requests though they contains rescue-related keywords.

Another observation is that many rescue requests and incomplete rescue requests were describing the same victims and the same incidents. We discovered 312 original rescue requests in the tweet dataset, whereas more tweets are categorized into rescue requests. The reason is that voluntary citizens re-posted such an incomplete rescue request by adding missing information, e.g., the hashtag #rescue.

3.4 Spatial Analysis of Rescue Requests

We further analyze the 312 original rescue requests, focusing on the locations where they were posted. Because few tweets have geographical metadata like geotag information, we extract a postal address in the text field of the rescue requests. We put pins pointing the extracted postal addresses on the map in Fig. 2.

A crucial observation is that rescue requests were concentrated in severely flood-hit areas, which were very narrow, in the case of the Japan floods 2018. Although we present a coarse-grained map in Fig. 2 to protect the privacy of victims, the 312 rescue requests contain the exact location of the victims. For instance, there were 235 rescue requests indicating a postal address within a 7 × 5 km rectangle area in a certain town in Okayama.

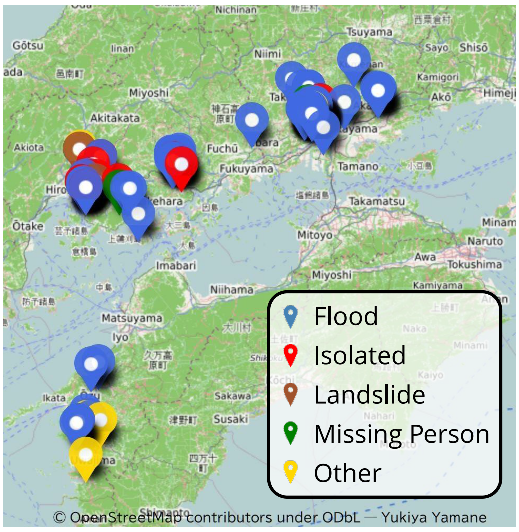


Figure 2: A map pointing rescue requests during the Japan floods 2018

4 CLASSIFICATION BASED ON MACHINE LEARNING

4.1 Objective and Overview of Classification Experiment

Rescue request classification is a critical task that directly impacts human lives, and hence reducing (ideally minimizing) false negatives, which are rescue requests classified into non rescue requests, is one of the severe challenges compared to the classification of other types of tweets. Thus, we focus on understanding why classifiers produce false negatives, which contributes to developing classifiers that produce fewer false negatives.

Note that the objective of the analysis is neither to develop a new machine learning technique nor to review the feasibility of state-of-the-art machine learning techniques for rescue request classification. Instead, we focus on learning lessons for extracting rescue requests from Twitter using machine learning. Thus, we build our classifiers using an existing machine learning model for natural language processing and analyze why the classifier identifies rescue requests as non rescue requests in two ways. First, we manually read misclassified tweets. Second, we use the attention mechanism of neural network models in the same way as Kshirsagar et al. [8], which contributes to understanding the reasons for the misclassification. The lessons drawn through the experiments are summarized in Section 4.6.

4.2 Dataset

We use the same dataset as used in Section 2. We use tweets from the Japan floods 2018 and the 19th typhoon for training and testing because the two disasters caused more severe damage than recent other typhoons in Japan. For training and testing, we further extract tweets containing rescue-related hashtags, i.e., #rescue, #rescue request, #SOS, #help, and #help_me. We should note that all the hashtags are specified in Japanese and they are translated into English in this paper. We finally selected 3424 and 1501 rescue-related tweets from the Japan floods 2018 and the 19th typhoon dataset, respectively.

Each tweet t is labeled by z_t, where 1 indicates t is a rescue request and 0 otherwise. One of the authors gives labels to tweets, and the labels are verified by two of the other authors. There are 328 and 74 rescue requests in the the Japan floods 2018 and the 19th typhoon dataset, respectively. The datasets are summarized in Table 4.

4.3 Model of Classifier

We use two classifiers in this study, and the two classifiers are based on a neural network: One is the bidirectional encoder representations from transformers (BERT) model [12] and the other is a recurrent neural network consisting of gated recurrent unit (GRU) cells. Hereafter, we refer to the recurrent neural based on GRU cells as the GRU model. We use the BERT model because it is a state-of-the-art neural network model for natural language processing. We use the GRU model because we adopt the same analysis method as used by Kshirsagar et al. [8].

We use the BERT Japanese pre-trained model developed by National Institute of Information and Communications Technology, Japan [17]. The BERT model is pre-trained using Japanese Wikipedia pages, and the number of its vocabulary is 100 thousand. The BERT model is fine-tuned using the Japan floods 2018 dataset.

We build another classifier using the GRU model to analyze why machine learning misclassifies rescue requests. Specifically, we adopt the same analysis method as Kshirsagar et al. [8], where they use an attention mechanism to analyze what parts of a tweet contribute to the classification result. They analyzed hidden states produced by GRU cells to investigate what phrases contribute to the classification results. Hidden states are referred to as attention values, hereafter. A high attention value represents that the corresponding phrase (or word) is related to a rescue request. In the same way, we use the attention values of the BERT classifier as well to analyze why machine learning misclassifies rescue requests.

4.4 Classification Results

We use accuracy, precision, recall, and F-measure as performance metrics. In our case, a true positive (negative) corresponds to a (non-)rescue request that is (not) identified as a rescue request. A false negative corresponds to a rescue request that is not identified as a rescue request and a false positive corresponds to a non-rescue request that is identified as a rescue request. The most critical metric for classifiers of rescue requests is recall, which represents the ratio of tweets identified as rescue requests among tweets that are actual rescue requests. A low recall value means that many rescue requests will be missed, i.e., many false negatives. Since rescue requests must not be missed, the recall must be high for classifiers of rescue requests.

We first investigate the performance of the classifier by using the same dataset. Specifically, we use 80%, 10%, and 10% of the tweets in the Japan floods 2018 dataset for fine-tuning, validating, and testing purposes, respectively. Table 5 summarizes the classification results and Table 6 shows the

Table 4: Summary of datasets used for classification analysis

Dataset	Number of tweets	Number of rescue requests
Japan floods 2018	3424	328
19th typhoon	1501	74

Table 5: Classification result

Predict\Label	1	0
1	34	6
0	17	321

Table 6: Performance metrics

Accuracy	Precision	Recall	F-measure
0.94	0.85	0.67	0.75

Table 7: Classification result

Predict\Label	1	0
1	45	26
0	28	1402

Table 8: Performance metrics

Accuracy	Precision	Recall	F-measure
0.96	0.63	0.62	0.63

four performance metrics. Although the accuracy is very high, the recall is low, as shown in Table 6. This result implies that the classifier misses several rescue requests.

Next, we investigate the applicability of the trained classifier to other disaster situations. We use the 19th typhoon dataset for testing, while we use the same classifier as the previous evaluation, which is fine-tuned using the Japan floods 2018 dataset. The classification results and the performance metrics are summarized in Table 7 and Table 8, respectively. The recall value is further decreased compared to the classification results in Table 5.

4.5 Analysis based on Attention Values

To understand the reasons why the recall is low, we investigate attention values of these classification results. Attention values indicate where the classifier focuses to filter (non-)rescue requests. Analyzing attention values, we obtain four observations regarding the misclassification: The first observation is that words of high attention values for non-rescue requests is one of the causes of false negatives. The second observation involves that similar expressions are used in both rescue requests and tweets of in the three categories, disaster situation, sympathy, advice and supplement. The third one involves that many citizens retweeted rescue requests by adding text, which is likely to belong to the three categories. Finally, the last observation involves that location names cause misclassification. The rest of this section explains the observations. We should again note that all tweets analyzed in this paper are written in Japanese and they are translated in English for this manuscript.

The results regarding the first observation are shown in Table 9, which summarizes the words with high attention values for the four classification results. The words are selected in the following steps. First, we selected words of the top five highest attention values from each tweet. Second, we counted the number of appearances of the selected words. Finally, we sorted the words in descending order of the number of appearances and selected frequently used words. Note that we eliminated postpositional particles, and privacy information, such as postal addresses, commas, and periods from the table. Both tweets classified as rescue requests and those not classified as rescue requests have words indicating postal addresses, such as -town and -city. However, the hashtag mark # is more frequently used in tweets classified as non-rescue requests regardless of true and false negatives than tweets classified as rescue requests. We observed that many Twitter users quote actual rescue requests and add some information to the quoted tweets, which often contains the hashtag mark and the mention mark. Hence, rescue requests are misclassified as non-rescue requests if the poster unintentionally use the hashtag mark and the mention mark.

Regarding the second observation, for example, some rescue requests use guessing expressions, which are often used in tweets reporting a disaster situation. The following tweet is a typical example of such rescue requests:

> Help me. My house is flooded above the 2nd floor level. Nobody comes to my rescue. *Roads appear to be blocked due to landslide*.

The last sentence is similar to tweets reporting disaster situations. Such rescue requests tend to be misclassified.

The third observation comes from the fact that there were incomplete rescue requests during the Japan floods 2018, as discussed in the previous section. Many voluntary citizens added several pieces of information, such as hashtags, to incomplete rescue requests to make them complete and retweeted them. On retweeting rescue requests, the citizens often add comments praying for the safety of the victims. The following tweet is a typical example:

> *I hope the victim will be immediately rescued*. #rescue RT: Help me. My house is flooded above the 2nd floor level.

Another example is as follows:

> *Add #rescue in case you need rescue*. RT: Help me. My house is flooded above the 2nd floor level.

Table 9: Examples of words of a high attention value

Classification Result	Words of a high attention value
True positive	town, -, floor, #, people, left (leave), flooded, water, city, isolated, -Chome (Japanese postal address)
False positive	town, #, -, floor, city, San (a Japanese title for general people, like Mr./Mrs.), -Chome
True negative	#, @, town, http, please, city, people, dissemination
False negative	#, evacuation, name, help, city

> This rescue request may not be delivered to rescue authorities unless you put hashtags like #rescue #SOS. RT: Help me. My house is flooded above the 2nd floor level.

Such tweets are misclassified into non-rescue requests since the added texts are similar to tweets in the category of either sympathy or advice and supplement.

The fourth observation is that tweets containing the names of disaster-hit locations tend to be misclassified into rescue requests.

> The situation of — *(town name)* — *(city name)* looks terrible.

During the Japan floods 2018, many rescue requests are generated in a town, which is one of the severely damaged areas, and therefore many rescue requests contain the name of the town and the city of the area.

Finally, we pick up typical phrases contributing to prediction of rescue requests. The following list summarizes phrases of a high attention value of rescue requests:

- Characteristics of victims like age and sex (e.g., elderly persons, babies, children, woman, man, octogenarian (80s), septuagenarian (70s), grandfather, and grandmother).
- Situations of flood-hit buildings (e.g., the water level has been rising gradually and my house is flooded above the 2nd floor level).
- Location names of flood-hit areas during the Japan floods 2018.
- Numbers in postal addresses (e.g., *x*-*y* (*x* and *y* represent a block and a house number and they are used like Mabi-town *x*-*y* in Japanese style postal addresses)).

In contrast, the attention values of phrases often used in rescue requests (e.g., rescue request, SOS, help, share it if you can, and retweet it if you can) are very low because they are also used in non-rescue requests. Furthermore, the attention values of location names of typhoon-hit areas during the 19th typhoon are also very low. If we replace a location name in a false negative rescue request of the 19th typhoon to that of a flood-hit area during the Japan floods 2018, they are identified as a rescue request.

4.6 Lessons Learned

This section summarizes four lessons learned through the classification experiments.

1. First, the entire text of rescue requests should not be used for training classifiers because a part of a rescue request is often unrelated to the rescue request but it is similar to tweets reporting disaster situations.

2. Second, retweeted rescue requests should be carefully handled in the same way as the first lesson because persons retweeting rescue requests often add several texts expressing sympathy for the victims. If texts added on retweeting are unrelated to rescue requests, they should be eliminated for training a classifier.

3. Third, the entire text of tweets should not be used for predicting whether they are rescue requests or not for the same reason behind the first and the second lesson. One way to realize this lesson is to classify tweets using attention values as well as a predicted result of classifiers. Using attention values allows us to identify texts related to rescue requests in a tweet and omit texts unrelated to the rescue requests, which cause misclassification, as demonstrated in the previous section.

4. Finally, location names should be handled independently for each disaster. One way to follow this lesson is to convert location names in tweets to a particular reserved word.

5 RELATED WORK

This section compares the present study with related studies.

Several studies analyzed disaster-related social media posts, especially focusing on Twitter. Alam et al. [16] analyzed disaster-related tweets in three hurricanes in the U.S. in 2017, i.e., Harvey, Irma, and Maria. They conducted both textual content analysis and multimedia content analysis on tweets from the three hurricanes. They define a taxonomy for disaster-related tweets and this study inspired us to categorize rescue-related tweets. Yang et al. [18] also analyzed tweets from the hurricane Harvey and proposed a framework to estimate the credibility of events reported by tweets. They also captured disaster-related tweets by searching Twitter for predefined keywords. While those studies focuse on disaster-related tweets, our study focuses on rescue-related tweets.

Next, we introduce studies that proposed classifiers for social media posts with machine learning. Studies in [6] and [7] propose classifiers based on neural networks for detecting fake information or rumors. Though both studies adopt recurrent neural networks, they develop slightly different models. Ma et al. [6] use intervals between social media posts reporting the same event as input values for their classifiers because of the fact that tweets are too short to identify their context with machine learning. Ruchansky et al. [6] propose to use relation between users who post tweets regarding the same event as well as texts of tweets to identify fake news. Kshirsagar et al. [8] propose a classifier for detecting crises like suicide, self-harm, abuse, or eating disorders by using a recurrent neural network. They also develop an annotation mechanism for explaining how social media posts are related to the crises. Their model uses texts of tweets as input values of their classifier. Our model is based on their model.

Finally, we introduce our previous study [19], which develops a communication framework for disaster management. The key idea behind the proposed framework is to utilize social media for collecting information regarding disaster situations. The framework delivers social media posts to right persons by using a machine learning technique. While the motivation of the previous study is to develop a communication framework that utilizes social media, the present study focuses on rescue requests in social media.

6 CONCLUSION

Circumstances of rescue requests during disasters have been changing. Citizens in need of help use social media, like Twitter, for expressing their rescue requests. To utilize such rescue requests on social media, it is a key to understand real rescue requests on social media. This study captured real disaster-related tweets from several flood disasters in 2018 and 2019 in Japan and analyzed the tweets. We observed that tweets having rescue-related keywords are classified into the eight categories, which include not only rescue requests but also non-rescue requests. Furthermore, most of the rescue-related tweets are unrelated to rescue requests. Next, we conducted preliminary experiments of classifying rescue requests from tweets using a BERT-based classifier. Moreover, we built a classifier based on GRU and LSTM-based recurrent neural networks, and analyzes the reason why our classifier based on machine learning misses rescue requests. The experiments revealed several lessons for building classifiers of rescue requests.

ACKNOWLEDGMENTS

The work has been supported by NICT under grant agreement no. 19303.

REFERENCES

[1] M. Jahanian, Y. Xing, J. Chen, K. K. Ramakrishnan, H. Seferoglu, and M. Yuksel, "The evolving nature of disaster management in the internet and social media era," in *Proceedings of IEEE International Symposium on Local and Metropolitan Area Networks* (2018).

[2] B. Stelter, "How social media is helping Houston deal with Harvey floods." https://money.cnn.com/2017/08/28/media/harvey-rescues-social-media-facebook-twitter/index.html (2018).

[3] Wikipedia, "2018 Japan floods (heavy rain of July 2018)." https://en.wikipedia.org/wiki/2018_Japan_floods (2018).

[4] Wikipedia, "Typhoon hagibis (the 19th typhoon of 2019 in japan)." https://en.wikipedia.org/wiki/Typhoon_Hagibis_(2019) (2019).

[5] Japan Broadcasting Corporation (NHK), "Nagano-prefecture collected rescue requests from Twitter and it contributed to saving about 50 victims." https://bit.ly/3741noG (in Japanese) (2019).

[6] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," in *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 3818–3824 (2016).

[7] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in *Proceedings of International Conference on Information and Knowledge Management*, pp. 797–806 (2017).

[8] R. Kshirsagar, R. Morris, and S. R. Bowman, "Detecting and explaining crisis," in *Proceedings of ACM Workshop on Computer Linguistics and Clinical Psychology*, pp. 66–73 (2017).

[9] S. Sato and F. Imamura, "An analysis of tweet data tagged with "#rescue" in the 2018 west Japan heavy rain disaster: Comparative analysis with the case of 2017 north Kyushu heavy rain disaster," *Journal of Japan Society for Natural Disaster Science*, vol. 37 (2019). (in Japannese).

[10] C. Song and H. Fujishiro, "Toward the automatic detection of rescue-request tweets: analyzing the features of data verified by the press," in *Proceedings of International Conference on Information and Communication Technologies for Disaster Management* (2019).

[11] Y. Koizumi, J. Takemasa, T. Hasegawa, and Y. Kawabe, "An analysis of rescue requests on twitter in a disaster," in *Proceedings of International Workshop on Informatics* (2022).

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv1810.04805v2* (2019).

[13] Wikipedia, "Typhoon faxai (the 15th typhoon of 2019 in japan)." https://en.wikipedia.org/wiki/Typhoon_Faxai_(2019).

[14] Twitter, Inc., "Twitter api documentation—search tweets." https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.

[15] Twitter, Inc. (@TwitterLifeline), "Twitter post." https://twitter.com/TwitterLifeline/status/1016519147738419201 (2018).

[16] F. Alam, F. Ofli, M. Imran, and M. Aupetit, "A Twitter tale of three hurricanes: Harvey, Irma, and Maria," in *Proceedings of International Conference on Information Systems for Crisis Response and Management* (2018).

[17] National Institute of Information and Communications Technology, "NICT BERT japanese pre-trained model." https://alaginrc.nict.go.jp/nict-bert/index.html (2020).

[18] J. Yang, M. Yu, H. Qin, M. Lu, and C. Yang, "A Twitter data credibility framework—hurricane Harvey as a use case," *International Journal of Geo-Information*, vol. 8 (2019).

[19] M. Jahanian, T. Hasegawa, Y. Kawabe, Y. Koizumi,

A. Magdy, M. Nishigaki, T. Ohki, and K. K. Ramakrishnan, “Direct: Disaster response coordination with trusted volunteers,” in *Proceedings of International Conference on Information and Communication Technologies for Disaster Management* (2019).

(Received: October 31, 2022)
(Accepted: April 7, 2023)

Yuki Koizumi Yuki Koizumi is an associate professor at the Graduate School of Information Science and Technology, Osaka University, Japan. Having obtained his master’s degree and his doctoral degree from Osaka University in 2006 and 2009, respectively, he possesses a wealth of academic expertise. His primary research interests encompass the realms of programmable data plane, privacy, and security on the Internet.

Junji Takemasa Junji Takemasa is an assistant professor of Graduate school of Information and Science, Osaka University. He received the master’s and Ph.D. degrees in information science from Osaka University in 2016 and 2019, respectively. After receiving the Ph.D. degree, he worked as a research engineer at KDDI Research, Inc. for one and half years and moved to Osaka University. His research interests include programmable network, future Internet and anonymous communication. He is a member of IEICE, IPSJ and IEEE.

Toru Hasegawa Toru Hasegawa received the B.E., the M.E. and Dr. Informatics degrees in information engineering from Kyoto University, Japan, in 1982, 1984 and 2000, respectively. After he worked as a research engineer at KDDI R&D labs., he has become a professor of Graduate school of Information and Science, Osaka University. His current interests include future Internet. He has published over 100 papers in peer reviewed journals and international conference proceedings. He is a member of IEEE, ACM, IEICE and IPSJ.

Yoshinobu Kawabe Yoshinobu Kawabe received his B.E., M.E., and D.E. degrees in information engineering from Nagoya Institute of Technology in 1995, 1997, and 2003, respectively. He joined NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation in 1997. In 2002, he was a visiting research scientist at MIT Laboratory for Computer Science. Since 2008, he has been with Aichi Institute of Technology and is currently a Department of Information Science professor. His research interests include term rewriting systems, process algebras, network programming languages, formal methods, security/privacy verification, and computational trust. He is a member of ACM, JSSST, IPSJ, IEICE and SOFT.