Vocabulary cross-contamination between entertainment content and disaster-related social media posts

1st Yutaka Morino
Graduate school of Informatics
Kansai University
Takatsuki, Japan
k790414@kansai-u.ac.jp

2nd Mitsunori Matsushita
Graduate school of Informatics
Kansai University
Takatsuki, Japan
m_mat@kansai-u.ac.jp

3rd Hiroyuki Fujishiro
Faculty of Social Sciences
Hosei University
Machida, Japan
fujisiro@hosei.ac.jp

***Abstract*—During disasters, local governments use Twitter (X) to obtain information on damages and rescue requests. However, due to the large volume of entertainment information posted on Twitter, the collection of disaster information can be hindered even during disasters. Specifically, while entertainment posts such as those related to video games contribute to content and community activation, in case of a disaster, they can obstruct the collection of disaster information by local governments. We refer to this case as contamination by entertainment content and aim to analyze it separately. Specifically, this study investigates the contamination of information collected via Twitter during disasters by entertainment content and identifies methods to prevent the impact of this contamination. As a preliminary step, it analyzes the Twitter posts related to entertainment content during disasters and examines 1) which types of content are easily contaminated and 2) whether these types can be separated from others.**

***Index Terms*—X(Twitter), information contamination, information reliability, noise sweeping, information triage**

I. Introduction

In Japan, following the Great East Japan Earthquake that occurred in 2011, social media garnered attention for the diffusion and sharing of information as an alternative to congested telephone communication lines, and its use in disaster rescue operations and support has been anticipated [1]. This has led to increased expectations for using social media in rescue operations and disaster support during emergencies [2]. A social network service X[1] (formerly known as Twitter), in particular, is recognized as a convenient means of communication in disaster situations because (1) the posts (tweets) are limited to 140 Japanese characters, so lengthy texts are not necessary, and (2) unlike services such as Messenger that are designed for communication with specific individuals, Twitter allows for spreading information to an unspecified large number of people.

According to Japan's Ministry of Internal Affairs and Communications, the usage rate of Twitter in Japan was 15.7% in 2012 and is expected to increase to 45.3% in 2024, making it a prominent social media service [3]. The platform's simplicity and broad reach thus make it a valuable tool for sharing information and mobilizing resources during disasters, with users increasingly posting about their situations and seeking help. Indeed, the number of posts related to the 2016 Kumamoto earthquake increased 23-fold compared with those related to the Great East Japan Earthquake in 2011.

In response to these realities, local governments have started using social networking services (SNS) to disseminate information to a large number of unspecified individuals, as they have a high diffusion property for information gathering and notifying evacuation information during actual disasters. There were actual cases in which local governments have succeeded to rescue people based on Twitter rescue requests. For example, during Typhoon Hagibis in 2019, the Nagano Prefecture government facilitated approximately 50 rescues based on Twitter posts.

The use of SNS for information dissemination and collection has become prevalent, making it possible to understand the extent of the damage and victims' rescue requests. However, collecting information from SNS during disasters is challenging. The problems include: (1) the difficulty for local governments to organize and select only the necessary information from a vast and diverse amount of duplicate information and (2) the inclusion of many unverified and unreliable posts. The phenomenon of information sharing and diffusion during disasters, along with the noise caused by general users retransmitting mass media reports, has led to the burying of information, as well as the spread of false information [4]. Therefore, when local governments collect information via SNS, human verification of this information is required despite the overwhelming number of irrelevant posts. This necessitates the deployment of dedicated staff and imposes a significant human burden, especially in time-sensitive and understaffed disaster situations [5] [6]. Additionally, the collaboration between humans and computers is crucial because of the limited time [7].

[1] In this paper, to improve identification, the social network service X will be referred to as its former name, Twitter.

A. This study

Information overload can be partly attributed to the varied everyday uses of social media. Notably, there is a significant overlap between entertainment and social media; for example, it has been reported that 60% of typical Twitter posts in Japan are related to entertainment. Indeed, even during disasters, posts related to hobbies and entertainment persist, complicating local governments' efforts to respond swiftly and effectively. The keywords and hashtags related to disasters, such as "rescue," "damage," "support," and "help," are utilized to gather information from social networks. However, these terms are also commonly used in posts related to video games, leading to challenges in distinguishing genuine disaster-related communication. The act of seeking help during disasters mirrors the dynamics of video games, in which users also request assistance, complicating the task of filtering relevant information based on keyword matching, a method currently employed by local governments. We term this phenomenon "cross-contamination"; it refers to the prevalence of entertainment content on social media during a disaster, regardless of its relevance to the actual damage. In this study, we thus hypothesize that the prevalence of entertainment content during disasters contributes to information overload. If we could effectively distinguish entertainment content from disaster-related information on social media without restricting the posts with such content, it could significantly reduce the workload of the local governments tasked with collecting disaster information. This study investigates the feasibility of mitigating cross-contamination by separating entertainment content from disaster-related content. Our primary concern is to determine whether "cross-contamination" occurs on social media during the disaster and, if so, how it can be separated from disaster-related information.

The rest of this paper is organized as follows. Section II describes about our target disaster and datasets which includes about classification results and analysis method. Section III describes analysis results and discussion points. Section IV describes related researches. Section V concludes this paper.

II. Target Disaster and Datasets

Specifically, this study analyzes the "heavy rainfall in July 2020," which occurred in Japan, especially in the Kyushu area, from July 3 to July 31, 2020. According to the Japan Meteorological Agency, this was historically the maximum rainfall recorded in the region, which has led to flooding, sediment disasters, land subsidence, and house collapses, resulting in the loss of 84 lives. From a dataset of 476,827 posts collected via the Twitter API using the keywords "rescue" and "evacuation," we extracted 47,434 posts that included images. These posts were gathered from July 1 to July 15, which corresponded to the period of the heaviest rainfall and the subsequent week.

A. Classification

Twitter Japan established guidelines for posting during disasters. It is recommended that posts requesting rescue should include detailed and precise information, such as the specific nature of the request, accompanying photographs, and precise address or location details, using the hashtag "#rescue." Therefore, it can be argued that reliable information should include photographs.

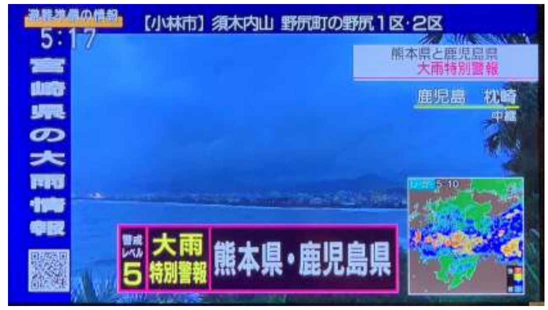

Fig. 1. Example of "Inverse L"

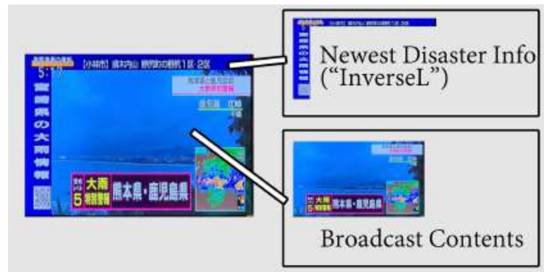

Fig. 2. Description of "Inverse L"

In this study, the information collected from social media during disasters was posted by individuals present in the affected areas describing the situation and damage. In other words, we excluded posts by news agencies such as newspapers and TV posts, as well as expressions of concern or empathy from individuals not located in disaster zones. The images that we aimed to collect as primary disaster information were photographs taken by posters in disaster areas. The availability of such photographs implies that the individual was indeed present in the disaster zone and had directly observed the damage, which serves as reliable evidence that they were confronted with the issue. Based on the above, we focused on posted photographs, and classification was thus conducted based on images rather than text alone. The procedure for classifying the posted images is described below. Four categories of images were presented as examples to four university students (hereafter, "classifiers") tasked with performing the classification. Images related to news and games were selected using a pre-visual sampling process that identified them as frequently observed.

1) "Inverse L" (Broadcast terms in Japan)
   Images showing an inverted L-shape, which is attributed to news bulletins about newest disaster information(see Fig. 1 and Fig. 2).[2]

[2]https://twitter.com/KANKAI79/status/1279148724766167041 (2021/11/21 confirmed)

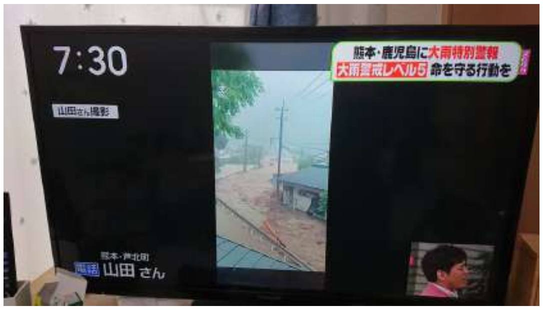

Fig. 3. Example of “TV news”

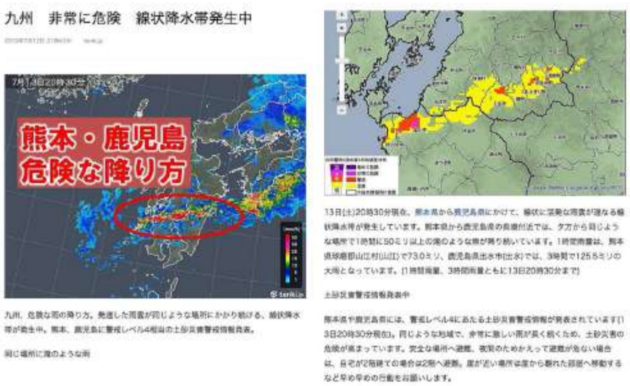

Fig. 4. Example of "Any kind of news"

2) “TV news”
   Images of TV news captured by a camera (see Fig. 3).[3]
3) “Any kind of news”
   Images related to news regardless of media (see Fig. 4).[4]
4) “2D images”
   Images of game screens, animations, cartoons, etc.

When an image was determined to fall into more than one class, it was assigned to all applicable classes; in such cases, the image was duplicated. Subsequently, considering the possibility of discrepancies in class names or the creation of multiple classes for the same type of image, interviews were conducted with the classifiers. Based on this discussion, classes deemed the same were integrated according to the author’s judgment.

B. Classification Result

The class with the largest number of images in this category was “games.” The images classified as “games” (6,855 posts) included mostly images and videos related to the mobile video game “Identity V: The Fifth Personality.” This may be due to the fact that the word “rescue,” which is commonly used in the game, was collected as a search result. The next most frequent images were classified into the “damage” class (4,135 posts). The class “damage taken by users,” which is assumed to be the most likely to reflect the circumstances of actual disaster victims, accounted for 2,689 posts. Table I lists the other classes as examples. Based on this classification, we selected specific categories of posts as training sets for the investigation, as discussed in the next section. These categories include “1: 2D images” (1,571 posts), “2: Damage” (4,135 posts), “3: Games” (6,855 posts), “4: Damage taken by users” (2,689 posts), “5: Merchandise” (746 posts), “6: TV news” (1,202 posts), “7: Inverse L” (325 posts), “8: Rescue” (641 posts), and “9: Live camera”(209 posts).

[3]https://twitter.com/TAKAYA05378399/status/1279181014711930880 (2024/4/26 confirmed)

[4]https://twitter.com/amasehimika147/status/1150073316863959041 (2024/4/26 confirmed)

TABLE I
EXCERPT FROM MANUAL IMAGE CLASSIFICATION CLASSES

Class	Number of posts
Any kind of News	1546
Twitter	1538
Game	6855
TV News	1202
2D Images	1671
People	2389
Maps	1205
Rescue	641
Damage	4135
Evacuation	3826
Landscapes	3168
Foods	1445
Weather	842
Disaster Supplies	618
Self-Defense Forces	439
Inverse L	325

C. Analysis Method

In the context of disaster response, constructing a model for rapid information collection using a vast corpus of posts is impractical because of time constraints. Consequently, this study used support vector machines (SVM) [15] for classification because they are known to yield effective results with relatively smaller datasets than other machine learning models. The normalization of each data count was set to 200 per category, in line with the live camera class, which had the fewest posts, totaling 1,800 posts used as training data. For classification purposes, the training data were divided into 70% for training and 30% for testing in each category. Morphological analysis was conducted on the posted texts using a Japanese morphological analyzer MeCab (Ver. 0.996) [8], from which the nouns were extracted. The extracted terms were then vectorized using the BoW method.

A five-fold cross-validation was performed, and the best parameters were determined through a Grid Search. Candidates for the SVM cost parameter and gamma value were set to 0.01, 0.1, and 1.0, respectively, with linear and radial basis faction (RBF) as the kernel options. The preliminary validation of each parameter indicated that a cost parameter of 1.0, a gamma value of 0.1, and an RBF kernel were the optimal choices. The precision, recall, and f1-score for each class using the best parameters are shown in Table II. The overall accuracy was 0.820. The accuracy was calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

III. Results and Discussion

Table III shows the confusion matrix of the results.

During disasters, information related to games, illustrations, comics (2D images), and merchandise becomes cross-contaminated with entertainment content; thus, these were selected for analysis. Specifically, posts about games were the most numerous and showed the highest number of true positives (TP), with 129 cases and 11 false negatives (FN) being the highest among all categories in the SVM classification. For the 2D images, there were 104 true positives and 36 false negatives. Merchandise-related posts resulted in the highest TP count at 132 but also had a very high number of false positives (FP) at 96. Posts about damage captured by users had 102 TP, 34 FN (14 of which were related to merchandise), and five FP (all mixed with damage). Posts regarding damage not from actual disaster-affected users had 101 TP and 39 FN (11 of which were merchandise-related). The high TP values and precision of game-related information suggest that they are highly separable. However, while the TP for merchandise-related information was high, there was a tendency to classify posts from other categories as merchandise, indicating a high potential for cross-contamination. This could be attributed to the diverse nature of merchandise content, including anime-related items, rescue vehicle toys, and live-event towels, suggesting that a more detailed classification is necessary. From the perspective of saving lives during disasters, it is crucial to collect accurate information on damage and rescue requests. While high recall is essential for posts related to "damage taken by user" (class 4), precision and accuracy do not necessarily need to be 1.0. Increases in FP and TN can lead to a reduction in information. Therefore, reducing the counts of TN and FN in class 4 is necessary to prevent the omission of damage information.

As discussed in Section I, human verification is crucial in disaster situations from the perspective of saving lives, as missing a critical piece of information could directly result in loss of life. To enhance classification precision, several strategies can be employed. For instance, employing both image and text classifier models simultaneously, as detailed in Section IV, may prove highly effective. Alternatively, focusing on achieving a recall rate of 1.0 rather than emphasizing accuracy or precision might be advantageous. Another approach involves swqeeping out irrelevant information from the collected social media data during a disaster. This is essential because achieving 100% detection of key posts, such as "damage taken by user", solely through computer technologies is demanding problem. By concentrating on detecting and removing irrelevant information that contaminates disaster-related data before human verification, we are able to reduce the burden on human operators as a result.

IV. Related Work

Because of these characteristics, Twitter is suitable for collecting "current" information during a disaster. In particular, tweets on damage and images/videos of the disaster situation are expected to provide useful information confirming the disaster situation and the formulation of support measures. However, many posts are unrelated to the disasters or simply sympathize with disaster victims [9] [10]. Currently, attempts are being made to create new hashtags to make it easier to collect information and use natural language processing technology to automatically collect and classify information [11] [12] [13]. Chowdhury et.al focused on twitter hashtags(#) and proposes an LSTM-MTL model to automatically identify and extract useful hashtags using Twitter data. This research constructed a dataset containing 67,288 disaster-related tweets and the hashtag extracting model have achieved a maximum 92.22%of F1 score [11]. Shimauchi et.al proposed methods for efficiently extracting useful disaster information from tweets during typhoons in Japan using binary classification(Bi-Normal Separation) and clustering techniques(DBSCAN and k-means++) [12]. The experiments conducted with tweets collected during two different typhoons that occurred in 2016 achieved a 0.59 F1 score. Imran et al. proposed a method using Conditional Random Fields (CRF) to automatically extract relevant information from disaster-related tweets. This research conducted experiments using tweets from the Joplin tornado and Hurricane Sandy, achieving 40% to 78% precision and 90% recall. However, a multimodal approach is necessary to construct a classification model for posted images related to damage [14]. Kumar et.al developed a deep multi-modal neural network to classify informative content from Twitter during emergencies. This network combines text and image data from tweets using Long Short-Term Memory (LSTM) and VGG-16 neural network models, respectively. The results demonstrated that using both text and images from tweets significantly improves the model's ability to identify informative content. The research indicates that integrating multiple types of data (text and image) provides a more robust approach for classifying and filtering informative content on social media during emergencies.

As mentioned above, previous research has focused on how to extract useful information during disasters. While this is undoubtedly an important perspective, I point out that the challenge of disaster information extraction differs from general information retrieval and extraction tasks. In that the paramount issue is ensuring that no critical information is missed from the perspective of saving lives and assessing damage as I mentioned in section III. A common issue in aforementioned studies is the occurrence of missed disaster information. Additionally, there is a difficulty in that the accuracy of the constructed models decreases when applied to other datasets or different disasters. Considering these issues, this study focuses on extracting and excluding unnecessary information to retain the necessary information rather than directly extracting specific useful information.

V. Conclusion

Our findings show that entertainment content, particularly video game content, can easily cross-contaminate disaster-related information on social media platforms. This study explored the complex circumstances of social media use during

TABLE II
RESULTS OF SVM

	1	2	3	4	5	6	7	8	9
precision	0.743	0.721	0.921	0.729	0.943	0.707	0.87	0.907	0.843
recall	0.889	0.878	0.772	0.953	0.579	0.961	0.847	0.858	0.901
f1-score	0.809	0.792	0.840	0.826	0.717	0.815	0.859	0.882	0.871

TABLE III
CONFUSIN MATRIX OF RESULTS(VERTICAL:TRUE LABEL, HORIZONTAL:PREDICTED LABEL)

classNo.	1	2	3	4	5	6	7	8	9
1	104	0	11	0	21	0	3	1	0
2	1	101	7	5	11	1	5	3	6
3	0	0	129	0	10	0	0	1	0
4	4	8	4	102	14	0	1	5	2
5	0	0	3	0	132	0	3	2	0
6	5	1	5	0	11	99	9	8	2
7	0	0	0	0	14	1	122	1	2
8	1	0	7	0	4	0	0	127	1
9	2	5	1	0	11	2	1	0	118

disasters, focusing on the challenges posed by the contamination between entertainment and disaster-related content on Twitter. We demonstrated that the prevalence of entertainment content, even during crises, contributes to a significant information overload, complicating the efforts of local governments and rescue organizations to respond effectively. Our investigation showed that the difficulties in filtering relevant from irrelevant posts using conventional keyword searches are due to cross-contamination from gaming and entertainment-related communications. However, while our model can effectively separate most categories, the precision for critical categories such as "damage taken by user" still requires improvement. Overall, our findings advocate a more strategic approach to social media management during disasters, thus emphasizing the need for noise reduction solutions that can parse and prioritize data effectively and ensure that rescue efforts are not hampered by information overload or misinformation.

Furthermore, it is essential to emphasize trustworthy news and SNS information during disasters. The presence of reliable and verified information becomes crucial, as it forms the basis of 'trust' among local government and users seeking support or help. Reducing noise not only clarifies the transmission of news but also enhances the visibility and reliability of critical information. This is important for ensuring that timely and accurate updates reach those in need, supporting effective decision-making during disasters. It is also in need to find out if vocabulary cross-contamination could possibly happens in different disaster and in different culture.

This conclusion does not imply that disaster-related information dissemination should be separated and established independently from the daily use of social media networks. It is precisely because these platforms are utilized in daily life that individuals in crisis situations can communicate information in a timely manner. In addition, relationships formed under normal circumstances serve as the foundation for mediating and addressing information during crises. Therefore, it is important to thoroughly investigate adequate separation methods.

ACKNOWLEDGMENT

This work was supported by JST RISTEX, Grant Number JPMJRS23L2.

REFERENCES

[1] Simon, T., Goldberg, A., and Adiniab, B. "Socializing in emergencies —A review of the use of social media in emergency situations," *Int. J. Information Management*, Vol. 35, No. 5, pp. 609–619, 2015.

[2] Sato, S. "Analysis of Tweets Hashtagged "# Rescue" in the 2017 North Kyushu Heavy Rain Disaster in Japan," *2018 5th International Conference on Information and Communication Technologies for Disaster Management*, pp. 1-7, 2018, doi: 10.1109/ICT-DM.2018.8636370.

[3] Institute for Information and Communications Policy, "Survey Report on Information and Communication Media Usage Time and Information Behavior in FY2022 <Outline>," *The Fire and Disaster Management Agency of the Ministry of Internal Affairs and Communications*, 2022 (in Japanese).

[4] Morino, Y., Yasuo, M., Matsushita, M., and Fujishiro, H. "Investigating the influence of Web-Media in disaster situations by analyzing diffusion of tweets", The 34th Annual Conference of the Japanese Society for Artificial Intelligence, 3G5ES103, 2020.

[5] Song, C. and Fujishiro, H. "Toward the automatic detection of rescue-request tweets: analyzing the features of data verified by the press," *2019 International Conference on Information and Communication Technologies for Disaster Management*, pp. 1-4, 2019. doi: 10.1109/ICT-DM47966.2019.9032895

[6] Fujishiro, H., Matsushita, M., and Ogasawara, M. "Effective Use of Social Media in Large-Scale Disasters: The Applicability of Information Triage," *Socio-Informatics*, Vol. 6, No. 2, pp. 49-63, 2018. doi:10.14836/ssi.6.2 49 (in Japanese).

[7] Yasuo, M. and Matsushita, M. "Maintaining Homogeneity of Severity Rating among Triage Officers in Collaborative Information Triage," Proc. 6th Asian Conference on Information Systems, pp.68-73, 2017.

[8] Kudo, T., Yamamoto, K., and Matsumoto, Y. "Applying Conditional Random Fields to Japanese Morphological Analysis," *Proc. 2004 Conference on Empirical Methods in Natural Language Processing*, pp.230-237, 2004.

[9] Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J., Palmer, M., Schram, A., and Anderson, K. "Natural language processing to the rescue? extracting situational awareness tweets during mass emergency," *Proc. International AAAI Conference on Web and Social Media*, Vol. 5, No. 1, pp. 385–392, 2011.

[10] Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A., and Chakraborty, B.: "A review on application of data mining techniques to combat natural disasters," *Ain Shams Engineering Journal*, Vol. 9, No. 3, pp. 365–378, 2018.

[11] Chowdhury, J. R., Caragea, C., and Caragea, D. "On identifying hashtags in disaster Twitter data," In *Proc. AAAI Conference on Artificial Intelligence*, Vol. 34, No. 01, pp. 498–506, 2020.
[12] Shimauchi, T., Taguchi, N., Nambo, H., and Kimura, H."A study on extracting disaster information from tweets," *Journal of Global Tourism Research*, Vol. 2, No. 2, pp. 93–98, 2017.
[13] Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P. "Practical extraction of disaster-relevant information from social media," In *Proc. 22nd International Conference on World Wide Web*, pp. 1021-1024, 2013. doi:10.1145/2487788.2488109
[14] Kumar, A., Singh, J.P., Dwivedi, Y.K., and Rana, N.P. "A deep multimodal neural network for informative Twitter content classification during emergencies," *Annals of Operations Research*, Vol. 319, pp, 791–822, 2022. doi:10.1007/s10479-020-03514-x
[15] Vapnik, V. N. *Statistical learning theory*, Wiley, 1998.