DisasterMM: Multimedia Analysis of Disaster-Related Social Media Data Task at MediaEval 2022

Stelios Andreadis[1,*], Aristeidis Bozas[1], Ilias Gialampoukidis[1], Anastasia Moumtzidou[1], Roberto Fiorin[2], Francesca Lombardo[2], Thanassis Mavropoulos[1], Daniele Norbiato[2], Stefanos Vrochidis[1], Michele Ferri[2] and Ioannis Kompatsiaris[1]

[1] *Information Technologies Institute - Centre of Research and Technology Hellas, Greece*
[2] *Eastern Alps River Basin District, Italy*

Abstract
This paper describes the "DisasterMM: Multimedia Analysis of Disaster-Related Social Media Data" Task at MediaEval 2022. Social media data have been widely used in disaster management and have been proven valuable to all phases of a disaster: from early warning to nowcasting and from response to recovery. Nevertheless, the large streams of user-generated content can be very noisy and usually lack geoinformation, which is an essential attribute. The goal of DisasterMM is to tackle these challenges with two subtasks: RCTP (Relevance Classification of Twitter Posts), which asks participants to build a classifier that will predict whether a tweet is relevant or not to a disaster, in particular floods, and LETT (Location Extraction from Twitter Texts), which calls for a text analysis technique that detects which words inside a Twitter message refer to a location.

1. Introduction

Flooding is the most common disaster occurring worldwide [1] and their impact is expected to grow due to climate change [2]. Besides loss of lives and property damage, floodwaters pose immediate dangers to human health [3] and long-term effects resulting from displacement and worsened living conditions. With the prominent rise of the social media, the ability to get real-time crowdsourced data, including text, photos and videos, becomes integrated into daily activities. Social media data are already explored by first responders and civil protection authorities as an alternative source of information [4, 5, 6, 7], complementary to traditional means such as telephone, in order to raise the situational awareness and support their operations. In parallel, the scientific society has been proposing AI and Machine Learning solutions that improve the quality of the incoming social media data [8, 9].

However, the utilization of large and continuous streams from social media platforms comes with two significant limitations. First, the vast amount of user-generated posts carries lots of noise, with messages that may contain flood-related words, but are actually irrelevant to floods (e.g., words used in a metaphorical way). Second, the majority of posts are not geotagged (i.e. not associated with a geographic position) or their geoinformation is questionable [10].

The automatic prediction of a post's relevance could reduce the social media noise and thus assist the interested parties in receiving only useful information, without spending time on

MediaEval'22: Multimedia Evaluation Workshop, January 13–15, 2023, Bergen, Norway and Online
*Corresponding author.
andreadisst@iti.gr (S. Andreadis); arbozas@iti.gr (A. Bozas); heliasgj@iti.gr (I. Gialampoukidis); moumtzid@iti.gr (A. Moumtzidou); roberto.fiorin@distrettoalpiorientali.it (R. Fiorin); francesca.lombardo@distrettoalpiorientali.it (F. Lombardo); mavrathan@iti.gr (T. Mavropoulos); daniele.norbiato@distrettoalpiorientali.it (D. Norbiato); stefanos@iti.gr (S. Vrochidis); michele.ferri@distrettoalpiorientali.it (M. Ferri); ikom@iti.gr (I. Kompatsiaris)

Figure 1: A tweet that is considered relevant to flooding

filtering out unrelated messages. In addition, recognizing the locations that are mentioned inside the post's text could enhance the post with geographic information, which would allow the automatic positioning of a potential incident. By receiving solely high-quality and geotagged social data, disaster management practitioners will be able to manage their resources more efficiently, which could even lead to saving more human lives.

The above has motivated the organisation of the "DisasterMM: Multimedia Analysis of Disaster-Related Social Media Data" Task[1] at MediaEval 2022, which swaps the focus back to floods from water quality (2021's WaterMM [11]), following the Multimedia Satellite Task (2017-2019) [12, 13, 14] and the Flood-related Multimedia Task (2020) [15]. The goal of DisasterMM is to tackle two individual challenges: to identify posts that are related to floods using textual, visual and metadata information and to detect mentioned locations inside a post text. For both subtasks, the datasets are in Italian language, so as to encourage the research community to move beyond the English language for text analysis.

2. Task Description

The DisasterMM task concerns the multimedia analysis of social media data, specifically posts from the popular platform of Twitter, that relate to the disaster of floods. The participants of this task are provided with a set of Tweet IDs in order to download textual as well as visual information and other metadata of tweets that have been selected with keyword-based search that involved words/phrases about flood. DisasterMM includes two subtasks:

- *Relevance Classification of Twitter Posts (RCTP)*: The objective of this subtask is to build a binary classification system that will be able to distinguish whether a tweet is relevant or not to flooding incidents. An example of a relevant tweet is shown in Fig. 1, while an irrelevant tweet in Fig. 2
- *Location Extraction from Twitter Texts (LETT)*: In this subtask, participants are asked to develop a named-entity recognition model in order to identify which words (or sequence of words) inside a tweet's text refer to locations. An example is shown in Table 1, with the tags being explained in the next section.

[1]https://multimediaeval.github.io/editions/2022/tasks/disastermm/

Figure 2: A tweet that is considered irrelevant to flooding

3. Dataset Description

DisasterMM involves two datasets that have been retrieved from Twitter by searching for flood-related keywords (e.g., "alluvione", "allagamento", "esondazione" – all translated as flood). The complete list of keywords can be found along with the datasets for RCTP[2] and LETT[3] in the task's repository, which will be made public after MediaEval 2022.

The dataset for RCTP is a set of 6,672 tweets collected between May 25, 2020 and June 12, 2020. The ground truth of the RCTP dataset refers to the relevance of a tweet, i.e., relevant (1) or not relevant (0), and is provided to participants in the form of key-value pairs of Tweet ID and relevance label. The dataset is separated randomly[4] into two sets: the *development-set* that contains 5,337 posts and the *test-set* with 1,335 posts.

The dataset for LETT consists of 4,992 tweets collected between March 25, 2017 and August 1, 2018. The ground truth of the LETT dataset is a string per tweet that contains one of the following labels for each word of its text: "B-LOC" for the first word of a sequence that refers to a location or a single-word location, "I-LOC" for the subsequent word of a sequence that refers to a location, and "O" for any non-location word (as mentioned before, an example can be seen in Table 1). Furthermore, in order to have a fairer evaluation, the word tokens deriving from the processing of the Twitter text (e.g., removal of multiple spaces, new lines, etc.) are also shared. The *development-set* of this dataset includes 3,993 posts and the *test-set* 999 posts (again split randomly).

Both datasets (RCTP/LETT) have been manually annotated by native speakers that are employed by the Eastern Alps River Basin District, which is responsible for the hydrogeological defense and flood risk management in the Eastern Alps partition of North-East Italy. Furthermore, only the ground truth for the development-sets is released, since the ground truth for the test-sets is used in the evaluation stage and will be available only after the completion of the challenges. Finally, it should be noted that only the IDs of the tweets are distributed to the participants, in order to be fully compliant with the Twitter Developer Agreement & Policy[5].

[2]https://github.com/multimediaeval/2022-DisasterMM/blob/main/DisasterMM2022_RCTP_keywords.json
[3]https://github.com/multimediaeval/2022-DisasterMM/blob/main/DisasterMM2022_LETT_keywords.json
[4]With scikit-learn's train_test_split(): https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
[5]https://developer.twitter.com/en/developer-terms/agreement-and-policy

Table 1
Annotation of a sentence with regards to locations

Allagamento	in	via	Prati	della	Farnesina
O	O	B-LOC	I-LOC	I-LOC	I-LOC

4. Evaluation

In RCTP, F1-score[6] is selected as the official evaluation metric for the binary classification of tweets as relevant (1) or not relevant (0) on the test set.

In LETT, F1-score will be used too, not in sentence level, but in word level. To further explain, if a given label for a word matches the label of the annotator for this particular word, then it is considered as true (true positive if "B-LOC"/"I-LOC", true negative if "O"). Two scores will be measured per each run: the exact F1-score, where labels have to fully match, and the partial F1-score, where either "B-LOC" or "I-LOC" can be considered as true as long as the annotator's label concerns location.

Participants are also greatly encouraged to carry out a failure analysis of their results in order to gain insight in the mistakes that their models make.

Acknowledgements

This work has been supported by the EU's Horizon 2020 research and innovation programme under grant agreements H2020-833435 INGENIOUS, H2020-101004152 CALLISTO, H2020-832876 aqua3S.

References

[1] W. Du, G. J. FitzGerald, M. Clark, X.-Y. Hou, Health impacts of floods, Prehospital and disaster medicine 25 (2010) 265–272.
[2] N. W. Arnell, S. N. Gosling, The impacts of climate change on river flood risk at the global scale, Climatic Change 134 (2016) 387–401.
[3] D. L. Paterson, H. Wright, P. N. Harris, Health risks of flood disasters, Clinical Infectious Diseases 67 (2018) 1450–1454.
[4] L. Li, M. Bensi, Q. Cui, G. B. Baecher, Y. Huang, Social media crowdsourcing for rapid damage assessment following a sudden-onset natural hazard event, International Journal of Information Management 60 (2021) 102378.
[5] A. Goncalves, C. Silva, P. Morreale, J. Bonafide, Crowdsourcing for public safety, in: 2014 IEEE International Systems Conference Proceedings, IEEE, 2014, pp. 50–56.
[6] D. E. Alexander, Social media in disaster risk reduction and crisis management, Science and engineering ethics 20 (2014) 717–733.
[7] Y. Jin, B. F. Liu, L. L. Austin, Examining the role of social media in effective crisis management: The effects of crisis origin, information form, and source on publics' crisis responses, Communication research 41 (2014) 74–94.
[8] A. Moumtzidou, S. Andreadis, I. Gialampoukidis, A. Karakostas, S. Vrochidis, I. Kompatsiaris, Flood relevance estimation from visual and textual content in social media streams, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 1621–1627.
[9] M. Imran, F. Ofli, D. Caragea, A. Torralba, Using AI and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions, 2020.

[6]https://en.wikipedia.org/wiki/F-score

[10] A. Kruspe, M. Häberle, E. J. Hoffmann, S. Rode-Hasinger, K. Abdulahhad, X. X. Zhu, Changes in twitter geolocations: Insights and suggestions for future usage, arXiv preprint arXiv:2108.12251 (2021).
[11] S. Andreadis, I. Gialampoukidis, A. Bozas, A. Moumtzidou, R. Fiorin, D. Norbiato, M. Ferri, S. Vrochidis, I. Kompatsiaris, WaterMM: Water Quality in Social Multimedia Task at MediaEval 2021, in: Proceedings of the MediaEval 2021 Workshop, Online, 2021.
[12] B. Bischke, P. Helber, C. Schulze, V. Srinivasan, A. Dengel, D. Borth, The Multimedia Satellite Task at MediaEval 2017., in: MediaEval, 2017.
[13] B. Benjamin, H. Patrick, Z. Zhengyu, B. Damian, et al., The Multimedia Satellite Task at MediaEval 2018: Emergency response for flooding events (2018).
[14] B. Bischke, P. Helber, S. Brugman, E. Basar, Z. Zhao, M. Larson, K. Pogorelov, The Multimedia Satellite Task at MediaEval 2019: Estimation of Flood Severity, in: Proc. of the MediaEval 2019 Workshop, Sophia Antipolis, France, 2019, p. 1621–1627.
[15] S. Andreadis, I. Gialampoukidis, A. Karakostas, S. Vrochidis, I. Kompatsiaris, R. Fiorin, D. Norbiato, M. Ferri, The Flood-related Multimedia Task at MediaEval 2020, in: Proceedings of the MediaEval 2020 Workshop, Online, 2020, pp. 14–15.