Floods Relevancy and Identification of Location from Twitter Posts using NLP Techniques

Muhammad **Suleman**[1,†], Muhammad **Asif**[1,†], Tayyab **Zamir**[2,†], Ayaz Mehmood[1,†], Jebran **Khan**[3], Nasir Ahmad[1] and Kashif Ahmad[4]

[1]*DCSE, University of Engineering and Technology, Peshawar, Pakistan*

[2]*Abasyn University Islamabad Campus, Pakistan*

[3]*Department of AI, AJOU University, South Korea*

[4]*Department of Computer Science, Munster Technological University, Cork, Ireland*

Abstract

This paper presents our solutions for the MediaEval 2022 task on DisasterMM. The task is composed of two subtasks, namely (i) Relevance Classification of Twitter Posts (RCTP), and (ii) Location Extraction from Twitter Texts (LETT). The RCTP subtask aims at differentiating flood-related and non-relevant social posts while LETT is a Named Entity Recognition (NER) task and aims at the extraction of location information from the text. For RCTP, we proposed four different solutions based on BERT, RoBERTa, Distil BERT, and ALBERT obtaining an F1-score of 0.7934, 0.7970, 0.7613, and 0.7924, respectively. For LETT, we used three models namely BERT, RoBERTa, and Distil BERTA obtaining an F1-score of 0.6256, 0.6744, and 0.6723, respectively.

1. Introduction

Natural disasters represent hazardous events that are generally caused by geophysical, hydrological, climatological, and meteorological elements. These hazardous events may have an adverse impact on human lives and infrastructure. Floods are one such event and it frequently occurs in different parts of the world. Similar to other natural disasters, floods may have a significant impact on public health and infrastructure. For instance, it has been noticed on numerous occasions that roads and communication infrastructure are badly damaged during floods [1].

A rapid and effective response to disasters may help in mitigating their adverse impact. Access to relevant and timely information is critical for an effective response. The literature demonstrates several situations where access to relevant information may be possible due to several reasons, such as the unavailability of reporters in the area and damage to communication [2]. Recently social media and crowdsourcing have been explored as a source of communication, information collection, and dissemination in emergency situations. To this aim, several interesting solutions have been proposed to collect, analyze, and extract meaningful insights from social media content [2]. However, social media content also comes with several limitations. For instance, social media content is generally noisy, thus, making access to relevant information very challenging. Similarly, geolocation information, which is critical for the relevance of the content, is not necessarily available for all the relevant posts.

---

*Corresponding author.

† These authors contributed equally.

✉ kashif.ahmad@mtu.ie (K. Ahmad)

CEUR Workshop Proceedings (CEUR-WS.org)

Considering the importance and applications of social media content in disaster analytics floods detection in social media content has been also included in the MediaEval benchmark competition as a shared task for several years. This paper presents a solution for the MMDisaster task presented in MediaEval 2022 [3]. The challenge aims to solve two key challenges to disaster analytics in social media. The first subtask aims at reducing social media noise by automatically filtering social media content to obtain relevant content. The second subtask aims at extracting location information from social media text, allowing automatic positioning of a potential incident due to floods. For both subtasks, we proposed several interesting solutions as described in Section 3.

2. Related Work

In recent years, the potential of social media has been widely explored in different application domains [4, 5]. Some of the key applications where social media content has been already proven very effective include public health [6], education [7], and public resource management [8]. Social media outlets have also been widely explored for a diversified set of applications in disasters and emergency situations [2]. For instance, Hao et al. [9] proposed a multi-modal framework utilizing multi-social media imagery and textual information for damage assessment in disaster-hit areas. The key factors analyzed in the work include hazard/disaster type, severity, and damage type. Wu et al. [10] also utilized social media data and the associated geo-location information for generating early warnings and damage assessment analysis after disasters. Ahmad et al. [11], on the other hand, used social media imagery for the analysis of road conditions after the floods. More specifically, the authors proposed an early and late-fusion framework to identify passable roads in flooded regions. Alam et al. [12] explored the potential of social media content in another relevant task of assessing flood severity. To this aim, the authors collected a large-scale benchmark dataset namely CrisisMD. The dataset provides a large collection of Twitter posts including textual and visual content. Hassan et al. [13] explored a slightly different aspect of natural disasters by extracting sentiments and emotions from visual content shared in social media outlets. The authors detailed how visual sentiment analysis of disaster-related social media visual content can be utilized by different stakeholders, such as news agencies, public authorities, and humanitarian organizations.

Despite being proven very effective in different tasks of disaster analytics, social media content has several limitations, such as noisy data and the unavailability of geolocation information. In this paper, we propose a solution to overcome such challenges.

3. Approach

3.1. Relevance Classification of Twitter Posts (RCTP)

As a first step, we analyzed the available multimedia content. During the analysis, we observed that most of the posts missing visual content. Moreover, most of the images were irrelevant. Thus, we decided to use textual information only in our solution. Our framework for the RCTP subtask is composed of two steps. In the first step, we performed some pre-processing techniques to clean the data by removing unnecessary information, such as usernames, URLs, emojis, and stop words.

After pre-processing, several state-of-the-art NLP algorithms including BERT [14], Roberta [15], Distil BERT [16], and ALBERT [17] are used for the classification of the text. Since its a

Table 1
Experimental results of RCTP task on the development set.

Model	F1-Score on the Clean Dataset	F1-Score on the Un-clean Dataset
BERT	0.95	0.94
RoBERTa	0.94	0.93
Distil BERT	0.93	0.93
ALBERT	0.92	0.92

binary classification task, in all methods, our cost function is based on binary crossentropy. Moreover, we used Adam optimizer with a mini batch size of 32 for 20 epochs.

3.2. Location Extraction from Twitter Texts (LETT)

LETT subtask is treated as Named Entity Recognition (NER) task. NER involves locating and classifying named entities in text into pre-defined categories [18]. In this task, we are interested in the identification of words representing the starting and subsequent words of a text sequence referring to a location. In LETT, annotations are provided at the word level. Similar to the RCTP task, in this task, we rely on multiple state-of-the-art algorithms including BERT, Roberta, Distil BERT, and ALBERT. We note that in this task, since annotations are provided at the word level, we did not use any pre-processing technique before training our models.

3.3. Dataset

For both subtasks, separate datasets are released. The dataset for RCTP subtask contains data from a total of 8,000 tweets. The tweets are collected between May 25, 2020, and June 12, 2020, using flood-related keywords in the Italian Language, such as "alluvione", "allagamento", and "esondazione". The dataset is provided in two different sets namely the development set and the test set. The development set is composed of 5,337 tweets while the test set contains a total of 1,315 tweets.

The dataset for the LETT subtask is composed of around 6,000 tweets collected between March 25, 2017, and August 1, 2018, using flood-related Italian keywords. The annotations for this subtask are available per word in the tweets.

4. Results and Analysis

4.1. Runs Description of RCTP Subtask

Table 1 shows the experimental results of the proposed solutions on the development set. We note that during the experiments on the development set, we used 70% samples of the development set for training, 20% for testing, and 10% samples for validation. As can be seen in the table, no significant differences can be observed in the performance of the models on the clean and un-clean datasets. As far as the performance of the individual models is concerned, slightly better results are obtained with BERT compared to the other models. Table 2 provides the official results of the proposed solutions on the test set. We note that for the experiments on the test set the models are trained on the complete development set. In total, 4 different runs are submitted for the task. Our first, second, and fourth runs are based on BERT, RoBERTa, and Distil Bert models trained on the un-cleaned dataset, respectively. Our third run is based on the BERT model trained on the cleaned dataset. The performance of the models trained on the

Table 2

Experimental results of the RCTP task on the test set.

Run	Precision	Recall	F1-Score
1 (BERT on Un-clean Dataset)	0.6949	0.9251	0.7934
2 (RoBERTa on Un-lean Dataset)	0.6947	0.9347	0.7970
3 (BERT on Clean Dataset)	0.6486	0.9213	0.7613
4 (Distil BERT on Un-clean Dataset)	0.6940	0.9232	0.7924

Table 3

Experimental results of the LETT task on the development set.

Model	F1-Score
BERT	0.7752
RoBERTa	0.8014
Distil BERT	0.7658
ALBERT	0.6827

Table 4

Experimental results of LETT task on the test set.

Run	Exact Results			Partial Results		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
1 (BERT)	0.596	0.522	0.556	0.628	0.622	0.625
2 (RoBERTa)	0.540	0.676	0.600	0.577	0.810	0.674
3 (Distil BERT)	0.563	0.604	0.583	0.610	0.760	0.677

un-cleaned dataset is higher than the models trained on the cleaned dataset. This indicates that the pre-processing information resulted in the removal of some relevant features and thus has a negative impact on the results.

4.2. Runs Description of LETT Subtask

Table 3 provides the experimental results of the proposed solutions on the development set for the LETT subtask. Similar to RCTP, we used 70% samples of the development set for training, 20% for testing, and 10% samples for validation. A significant variation can be observed in the results of the models on the development set. Overall, better results are obtained for RoBerta with a significant improvement of 3% over the second-highest results obtained with the BERT model. Table 4 provides the official results for the LETT subtask. In the experiments on the test set, the models are trained on the complete development set. We note that in the partial results the omitted samples are counted as false while in the partial results the omitted samples are completely ignored without any penalty. As can be seen in the table, overall, better results are obtained with Roberta and Distil BERT compared to the original implementation of the BERT model.

5. Conclusions

In this paper, we presented our solutions for the DisasterMM challenge posted in MediaEval 2022. For both subtasks, multiple state-of-the-art NLP algorithms are employed. In the current implementation, all the models are used individually, however, we believe these models can complement each other if jointly utilized in a merit-based fusion method. In the future, we aim to employ different merit-based fusion methods to jointly utilize the capabilities of the individual models in both subtasks.

References

[1] D. T. Nguyen, F. Ofli, M. Imran, P. Mitra, Damage assessment from social media imagery data during disasters, in: Proceedings of international conference on advances in social networks analysis and mining 2017, 2017, pp. 569–576.

[2] N. Said, K. Ahmad, M. Riegler, K. Pogorelov, L. Hassan, N. Ahmad, N. Conci, Natural disasters detection in social media and satellite imagery: a survey, Multimedia Tools and Applications 78 (2019) 31267–31302.

[3] S. Andreadis, A. Bozas, I. Gialampoukidis, A. Moumtzidou, R. Fiorin, F. Lombardo, T. Mavropoulos, D. Norbiato, S. Vrochidis, M. Ferri, I. Kompatsiaris, DisasterMM: Multimedia Analysis of Disaster-Related Social Media Data Task at MediaEval 2022, in: Proceedings of the MediaEval 2022 Workshop, Bergen, Norway and Online, 2023.

[4] K. Ahmad, K. Pogorelov, M. Riegler, N. Conci, P. Halvorsen, Social media and satellites, Multimedia Tools and Applications 78 (2019) 2837–2875.

[5] I. Alsmadi, K. Ahmad, M. Nazzal, F. Alam, A. Al-Fuqaha, A. Khreishah, A. Algosaibi, Adversarial nlp for social network applications: Attacks, defenses, and research directions, IEEE Transactions on Computational Social Systems (2022).

[6] S. Andreadis, G. Antzoulatos, T. Mavropoulos, P. Giannakeris, G. Tzionis, N. Pantelidis, K. Ioannidis, A. Karakostas, I. Gialampoukidis, S. Vrochidis, et al., A social media analytics platform visualising the spread of covid-19 in italy via exploitation of automatically geotagged tweets, Online Social Networks and Media 23 (2021) 100134.

[7] A.-R. et al., The influence of information system success and technology acceptance model on social media factors in education, Sustainability 13 (2021) 7770.

[8] K. Ahmad, M. Ayub, J. Khan, N. Ahmad, A. Al-Fuqaha, Social media as an instant source of feedback on water quality, IEEE Transactions on Technology and Society (2022).

[9] H. Hao, Y. Wang, Leveraging multimodal social media data for rapid disaster damage assessment, International Journal of Disaster Risk Reduction 51 (2020) 101760.

[10] D. Wu, Y. Cui, Disaster early warning and damage assessment analysis using social media data and geo-location information, Decision support systems 111 (2018) 48–59.

[11] K. Ahmad, K. Pogorelov, M. Riegler, O. Ostroukhova, P. Halvorsen, N. Conci, R. Dahyot, Automatic detection of passable roads after floods in remote sensed and social media data, Signal Processing: Image Communication 74 (2019) 110–118.

[12] F. Alam, F. Ofli, M. Imran, Crisismmd: Multimodal twitter datasets from natural disasters, in: Twelfth international AAAI conference on web and social media, 2018.

[13] S. Z. Hassan, K. Ahmad, S. Hicks, P. Halvorsen, A. Al-Fuqaha, N. Conci, M. Riegler, Visual sentiment analysis from disaster images in social media, Sensors 22 (2022) 3628.

[14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[15] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692 (2019).

[16] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, arXiv preprint arXiv:1910.01108 (2019).

[17] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, Albert: A lite bert for self-supervised learning of language representations, arXiv preprint arXiv:1909.11942 (2019).

[18] J. Li, A. Sun, J. Han, C. Li, A survey on deep learning for named entity recognition, IEEE Transactions on Knowledge and Data Engineering 34 (2020) 50–70.