

Floods Detection in Twitter Text and Images

Naina Said¹, Kashif Ahmad², Asma Gul³, Nasir Ahmad¹, Ala Al-Fuqaha²

¹CSE, University of Engineering and Technology, Peshawar, Pakistan ²Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.

³Department of Statistics, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan

{kahmad,aalfuqaha}@hbku.edu.qa,asmagul@sbbwu.edu.pk

{nainasaid,n.ahmad}@uetpeshawar.edu.pk

ABSTRACT

In this paper, we present our methods for the MediaEval 2020 Flood Related Multimedia task, which aims to analyze and combine textual and visual content from social media for the detection of real-world flooding events. The task mainly focuses on identifying floods related tweets relevant to a specific area. We propose several schemes to address the challenge. For text-based flood events detection, we use three different methods, relying on Bag of Words (BOW) and an Italian Version of Bert individually and in combination, achieving an F1-score of 0.77%, 0.68%, and 0.70% on the development set, respectively. For the visual analysis, we rely on features extracted via multiple state-of-the-art deep models pre-trained on ImageNet. The extracted features are then used to train multiple individual classifiers whose scores are then combined in a late fusion manner achieving an F1-score of 0.75%. For our mandatory multi-modal run, we combine the classification scores obtained with the best textual and visual schemes in a late fusion manner. Overall, better results are obtained with the multimodal scheme achieving an F1-score of 0.80% on the development set.

1 INTRODUCTION

Social media outlets, such as Facebook, Twitter, and Instagram, allow users to create, obtain and share instant information. Being an instant source of information, social media outlets especially Twitter has been widely exploited for information gathering and dissemination especially in adverse events, where instant access to relevant information is more crucial [2, 3]. The literature reports several situations where social media content has been mined to get timely access to information in the events of natural and man made disasters [14].

Being one of the most frequently occurred natural disasters, flood events detection in social media has been the focus of the research community over the last few years. For instance, the research work presented in [12] assesses the informativeness of a tweet in the event of an earthquake using machine learning techniques. Similarly, in another study [13], the authors analyze domain adaptation classifiers by utilizing the labeled data from a past disaster event and unlabelled data from a current event. Flood events detection has been also part of the MediaEval challenge for the last four years where each time a different aspect of flood events has been targeted. This year, the task is focused on the detection of flood events relevant to a specific area [6]. In the task, the participants

are provided with a collection of Twitter data containing a large number of tweets' text and associated images, and are asked to develop a multi-modal system capable of automatically detecting flood related events that occurred in a particular area in Italy. It is to be noted that the tweets are provided in the Italian language.

This paper provides a detailed description of the methods proposed by team UEHBKU for the task. In total, we submitted five runs including a late fusion based multimodal, one image-based, and three textual information-based solutions as detailed in the next section.

2 PROPOSED APPROACH

2.1 Image-based Floods Detection

For image-based floods detection, based on our experience on a similar type of task [1, 4, 15], we rely on multiple pre-trained CNNs to extract the object-level features from the images, which are then used to train multiple SVM classifiers. The SVM classifiers provide the results in terms of posterior probabilities, which are then combined in a late fusion manner by aggregating the scores. A label with the highest aggregate is selected as the final outcome of the framework. In total, we used three different models namely (i) DenseNet [11], (ii) VggNet-19 [16], and ResNet [10]. It is to be noted that all the models are pre-trained on ImageNet [8], and expected to extract object-level features.

Moreover, to deal with the class imbalance problem, we use Synthetic Minority Oversampling Technique (SMOTE) [7] to synthesize new examples of the rare class. The SMOTE technique is based on under-sampling the majority class and oversampling the rare class and is found to be more effective in improving the classifier performance in contrast to only under-sampling the majority class. During the oversampling process, the rare class has been increased by a factor 3 to have an equal number of samples in both classes.

2.2 Text-based Floods Detection

For text-based analysis of the tweets, two different methods, namely (i) BoW, and (ii) state of the art BERT model [9], are used to obtain feature vectors from the tweets. The BoW model represents text by describing the occurrence of words within a document where each word count is considered as a feature. BERT, on the other hand, applies bidirectional training of Transformer, which is a popular attention model, to language modeling. It is to be noted that different variations of the BERT model are available. Since the tweets provided for the task are in the Italian language so we utilize the Italian version of the model. Before training the models, the text is cleaned by removing punctuation keys, such as commas,full-stops,

Table 1: Evaluation of our proposed approaches in terms of F1-scores. The first and second columns represent our results on the development and test sets, respectively. For comparison against the other teams, we also provide the average score of all the teams on each run in the third column.

Run	Method	Dev. Set	Test Set	Teams' Average
1	Textual (BoW)	0.77	0.437	0.357
2	Textual (BERT)	0.68	0.276	0.156
3	Textual (BoW & BERT)	0.70	0.343	0.199
4	Visual	0.75	0.129	0.131
5	Multi-modal	0.80	0.093	0.221

emojis, URLs, and stop words from the tweets. Similar to the image-based solution, we relied on SMOTE to tackle the class imbalance problem in the textual data.

The feature vector obtained with BoW is used to train a Naive Bayes classifier where as a logistic regression model is trained on the BERT features. The classification scores obtained with the both individual models are then combined in a late fusion manner by aggregating the probabilities obtained with both models for the final decision.

2.3 Multi-modal Analysis

For the mandatory multi-modal run, the visual and textual information are combined in a late fusion scheme by aggregating the probabilities obtained with the individual models trained on visual and textual features. It is to be noted that in the current implementation, all the models are treated equally by assigning them equal weights. In the future, we aim to use more sophisticated fusion methods by assigning merit-based weights to the models.

3 RESULTS AND ANALYSIS

We submitted five different runs for the task where the first three runs are based on the text while Run 4 and Run 5 are based on visual and multi-modal information, respectively. In Run 1, we used BoW for text representation to differentiate between flooded and non-flooded events on Twitter. In Run 2, we relied on a multilingual BERT model to obtain a feature vector for the tweets, and a logistic regression model is then trained on the generated word embeddings.

The variation in the performance of the models motivated us for the joint use of the models in a late fusion manner for our third run. However, lower than the best individual model' performance (i.e., BoW) is observed for the joint use of the models indicating that BERT is not suited well in our case.

Our Run 4, which is based on visual information only, is motivated by our previous experience [5], where we combined multiple state-of-the-art pre-trained models in a late fusion manner by aggregating the posterior probabilities obtained with the individual models.

In our final run, we enrich the textual information with the images associated with each tweet for accurate classification of the tweets. Again, a late fusion method by aggregating the posterior probabilities is utilized to combining the complementary information obtained with text and associated images.

As can be seen in Table 1, overall, better results have been obtained with the multimodal approach, which indicates the superiority of the joint use of textual and visual information for the task. In

the case of individual models trained on a single type of feature (i.e., textual or visual), better results have been observed for BoW on the textual features. However, comparable results have been observed for the joint use of the different deep models on visual information.

Moreover, as can be seen in Table 1, the average score of all the teams is very low, which shows the complexity of the task. However, the scores on the development set are reasonably good, which indicates the issues with the test set especially because the teams' average scores for most of the runs are below 20%.

4 CONCLUSIONS AND FUTURE WORK

The 2020 Mediaeval flood-related multimedia task was concerned with analyzing Twitter data for flood detection. The goal of the task was to combine the textual and visual information from Twitter in order to develop an automatic classification system to indicate whether a particular tweet's text and the associated image is relevant to an actual flooding event or not. We proposed five different solutions including a multimodal, a visual information based solution, and three text-based methods. We observed that both types of data complement each other, and indeed improves the overall accuracy. In the present study, we performed late fusion using equal weights for all the models. However, in the future, we would investigate different optimization techniques, such as Particle Swarm Optimization and Genetic Algorithm, for assigning merit-based weights to the models in fusion. In addition, we will also explore other text-based models specifically for the Italian language to improve the accuracy of Italian text classification.

REFERENCES

- [1] Kashif Ahmad, Mohamed Lamine Mekhalfi, Nicola Conci, Giulia Boato, Farid Melgani, and Francesco GB De Natale. 2017. A pool of deep models for event recognition. In *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2886–2890.
- [2] Kashif Ahmad, Konstantin Pogorelov, Michael Riegler, Nicola Conci, and Pål Halvorsen. 2018. Social media and satellites. *Multimedia Tools and Applications* (2018), 1–39.
- [3] Kashif Ahmad, Michael Riegler, Konstantin Pogorelov, Nicola Conci, Pål Halvorsen, and Francesco De Natale. 2017. Jord: a system for collecting information and monitoring natural disasters by linking social media with satellite imagery. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*. ACM, 12.
- [4] Kashif Ahmad, Amir Sohail, Nicola Conci, and Francesco De Natale. 2018. A Comparative study of Global and Deep Features for the analysis of user-generated natural disaster related images. In *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 1–5.
- [5] Sheharyar Ahmad, Kashif Ahmad, Nasir Ahmad, and Nicola Conci. 2017. Convolutional neural networks for disaster images retrieval. In *Proceedings of the MediaEval 2017 Workshop (Sept. 13–15, 2017)*. Dublin, Ireland.
- [6] Stelios Andreadis, Ilias Gialampoukidis, Anastasios Karakostas, Stefanos Vrochidis, Ioannis Kompatsiaris, Roberto Fiorin, Daniele Norbiato, and Michele Ferri. 2020. The Flood-related Multimedia Task at MediaEval 2020. (2020).
- [7] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

- [8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. Ieee, 248–255.
- [9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [11] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.
- [12] Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. 2014. AIDR: Artificial intelligence for disaster response. In *Proceedings of the 23rd International Conference on World Wide Web*. 159–162.
- [13] Hongmin Li, Nicolais Guevara, Nic Herndon, Doina Caragea, Kishore Neppalli, Cornelia Caragea, Anna Cinzia Squicciarini, and Andrea H Tapia. 2015. Twitter Mining for Disaster Response: A Domain Adaptation Approach.. In *ISCRAM*.
- [14] Naina Said, Kashif Ahmad, Michael Riegler, Konstantin Pogorelov, Laiq Hassan, Nasir Ahmad, and Nicola Conci. 2019. Natural disasters detection in social media and satellite imagery: a survey. *Multimedia Tools and Applications* (17 Jul 2019). <https://doi.org/10.1007/s11042-019-07942-1>
- [15] Naina Said, Konstantin Pogorelov, Kashif Ahmad, Michael Riegler, Nasir Ahmad, Olga Ostroukhova, Pål Halvorsen, and Nicola Conci. 2018. Deep Learning Approaches for Flood Classification and Flood Aftermath Detection.. In *MediaEval*.
- [16] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).