Deep Learning models for passability detection of flooded roads

Laura Lopez-Fuentes[1, 2, 3], Alessandro Farasin[4, 5], Harald Skinnemoen[3], Paolo Garza[4]
[1]University of the Balearic Islands, Spain, [2]Autonomous University of Barcelona, Spain
[3]AnsuR Technologies, Norway, [4]Politecnico di Torino, Italy [5]Istituto Superiore Mario Boella, Italy
l.lopez@uib.es,alessandro.farasin@ismb.it,harald@ansur.no,paolo.garza@polito.it

ABSTRACT

In this paper we study and compare several approaches to detect floods and evidence for passability of roads by conventional means in Twitter. We focus on tweets containing both visual information (a picture shared by the user) and metadata, a combination of text and related extra information intrinsic to the Twitter API. This work has been done in the context of the MediaEval 2018 Multimedia Satellite Task.

1 INTRODUCTION

Social media are becoming, year by year, ever more popular and used for sharing people daily activities. This massive adoption brought to a large availability of contents, such as texts and pictures, in the most varied sectors, making the social media a great source of information. The large availability of data is precious for extracting knowledge and it lays the basis for several applications. One of them falls in the context of emergency management related to natural disasters, in which computer vision and machine learning techniques are investigated [2, 3, 11, 16] to extract key information and help first responders in their activities.

In this work we focus on analyzing social media posts in order to extract valuable information about roads affected by flood. In more detail, we propose a multi-modal deep learning network which processes flood-related social media pictures and related metadata (e.g., Twitter, Flickr, YFCC100M), both to provide (a) evidence of roads and (b) whether they are also passable. The approach presented in this paper takes his inspiration (and it wants to be an extension) from the work [10] proposed in the MediaEval challenge held in 2017.

2 RELATED WORKS

Recent literature approaches leverage on satellite [8, 12, 15] or ground acquisitions [5] to identify flood events. Other works focus more on urban elements detection such as roads [6, 9]. To the best of our knowledge there are no existing works to determine road passability evidence during flood events.

3 DATA

The dataset used in this work was distributed by MediaEval 2018 Multimedia Satellite Task [1, 4]. It consists of 5820 Twitter images with its related metadata, from which ~36% of the images present flooded regions with evidence of roads. Only the images belonging to the earlier class are considered for the second task evaluation: among them, the ~45% present passable roads. Furthermore, for each image, metadata is available. Metadata is a set of information concerning the tweet itself and the user who wrote it. For instance, the text message and the respective language, the number of retweets and likes, the number of replies, the coordinates (and whether it is geo-located) and the user's followers number are only a subset of properties associated to a tweet.

MediaEval'18, 29-31 October 2018, Sophia Antipolis, France

4 APPROACH

To properly deal with heterogeneous data, we opted for a "divide-et-impera" approach: we created a model for each kind of data. In detail we developed: (i) a classification model using only the metadata information, (ii) three classification models using only the images, (iii) a model which combines the metadata and the visual information. In this section we will briefly describe the three different systems.

4.1 Metadata only

We processed metadata (i) filtering out properties not available in the whole dataset, (ii) studying the correlation among the remaining ones and selecting the most relevant ones. As a result we kept the text written by the user, the language in which the tweet was originally written, the number of retweets and the number of persons who had favourited the tweet. As a pre-processing step we have translated all texts to English, removed emojis, urls and special characters; then we used lemmatization and tokenization techniques. We represented the text features using a word embedding initialized with Glove [7]. After that, we normalized between 0 and 1 the number of retweets and the number of times the tweet was favourited. Then, we binarized the original language information by assigning 0 to English and 1 to any other language. Finally, we defined a neural network composed by a bidirectional Long Short-Term Memory (LSTM) network. The result of the LSTM has been concatenated to the normalized extra fields and passed through a two parallel fully-connected (FC) layers with a softmax classifier, one per task.

Both tasks have been trained in parallel. Initially, we set all the images which had no evidence of flood as having no passability issues either. However, this strategy introduced a big imbalance to an already imbalanced dataset which made the training more difficult, so we finally decided to use only the images containing evidence of flood to train the passability classifier, while still doing the training in parallel.

4.2 Visual Information only

As a pre-processing step for the images we applied several data augmentation techniques: image rotation, width and height shifts, horizontal flip and zoom. We designed two different systems to process the images, which we will compare.

Table 1: F-Score (%) evaluated on the test set for the two sub-tasks. Firstly, it is computed on a subset of 50 tweets from the training set and manually annotated by 4 persons. Then, it is computed on the test set for each developed approach: Metadata-only Approach (MA), Visual Approach 1 (VA1, Double-Ended Classifier with Compact Loss), Visual Approach 2 (VA2, Network Stacking with average aggregation), Visual Approach 3 (VA3, Network Stacking with average and voting aggregation).

Approach \Data	EVIDENCE [%]					PASSABILITY [%]				
	Metadata	Images			Meta + Imgs	Metadata	Images			Meta + Imgs
Human annotation	51.48		87.32		-	18.18		47.71		-
Metadata only [MA]	43.88		-		-	19.3		-		-
Image only [VA1, VA2, VA3]	-	85.6	86.43	87.79	-	-	24.09	67.13	68.38	-
Metadata and Image [VA1+MA]	-		-		83.12	-		-		28.34

Double-ended classifier with compact loss: We used the Inception V3 [14] network pre-trained on ImageNet with two FC layers and a softmax classifier at the end. Each end of the network was trained for each task. These tasks can be subsumed as two separate *One-class classification* problems, in which the single class is the flood event for the first case, its passability condition for the second one. We took inspiration from [13] and we customized the InceptionV3 optimization function, as: $\hat{g} = \max_g \mathcal{D}(g(t)) + \lambda C(g(t))$, where: (a) g is the deep feature representation for the training data t, (b) λ is a positive constant, (c) $\mathcal{D}$ is the *Descriptive loss function* (within this approach, we used the cross-entropy) and (d) C is the *Compactness loss function*, which evaluates the batch intra-class deep feature distance to derive objects from the same class.

Network staking: We used 9 state-of-the art networks (InceptionV3, Xception, VGG16, VGG19, InceptionResNetV2, MobileNet, DenseNet121, DenseNet201, NASNetLarge) all of them pre-trained on ImageNet. They were separately trained for both problems in 5 different train-validation folds, which generated 90 networks (45 per task). The output of each network is a number between 0 and 1 which represents the probability of the picture containing evidence of roads in flooded regions and evidence of passable roads, respectively. Being n the number of networks and p_i the probability of the picture corresponding to class 1, we define $\overline{p}$ as the average of p_i for all $1 \leq i \leq n$. We define $voting(p_1, ..., p_n) = |\{i/p_i > 0.5, 1 \leq i \leq n\}|$, where $|.|$ is the set cardinality. We used two different methods to aggregate the results from the networks: (i) *Average aggregation*: $pred(p_1, ..., p_n) = (\overline{p} > 0.5)$, (ii) *Average and voting aggregation*:

$$pred(p_1, ..., p_n) = \begin{cases} 1 & \text{if } \left(\overline{p} > 0.5 \text{ and } voting(p_1, ..., p_n) > \frac{n}{2} - 2\right) \text{ or} \\ & \left(\overline{p} > 0.45 \text{ and } voting(p_1, ..., p_n) \geq \frac{n}{2}\right), \\ 0 & \text{otherwise.} \end{cases}$$

4.3 Metadata and Visual Information

To combine the metadata and the images, we merged the *Metadata-only* and the *Double-ended classifier with compact loss* approaches. The two networks were taken without their respective double-ended fully-connected (FC) layers and merged with two newer FC layers (one per task) with a softmax classifier.

5 RESULTS

To get a first idea of the upper-bound for our task we asked 4 persons to do the task on a subset of 50 images, the results are given in Table 1. In the subsequent rows we have included the results for the 5 different approaches introduced in this paper. As it can be seen, our approach for flood evidence classification using metadata obtains very poor results but close to the results obtained by human annotators, which means that the metadata was not very discriminative for this task. Since the error is cumulative, the results of both the humans and the metadata classifier drop significantly for the passability detection, being the F-score again very close to one another. All the image classification approaches achieve similar results on the first task, while the *network stacking* achieves a small improvement with respect to the *double-ended classifier with compact-loss* network. Furthermore, the aggregation of the networks using *average and voting* slightly improves the aggregation compared to using only the average. However, there is a big gap between the performance of the *double-ended classifier with compact loss* and the network stacking approaches. When we decided to use the same network body to perform both tasks we thought that since the tasks were very related, one task could benefit from the others knowledge. However, since one task was more difficult than the other, the double-ended network seems to have specialized on the easy task while leaving aside the difficult task. It is also remarkable that the network stacking algorithms achieve significant better results than the human annotators, probably because: (i) road passability is subjective in several cases and (ii) while the network learnt over the whole training set, the human annotators were not given any examples about the task.
Finally, combining metadata with images does not provide much improvement or it even worsens the results due to the lack of discriminative features in the metadata.

6 CONCLUSIONS

In this paper we studied several approaches to perform flood and road passability detection. We proposed several approaches to deal with textual and visual information. According to our tests, we discovered that when a network tries to accomplish several tasks with different difficulties, even if they are related, it focuses on one of them (presumably the simplest one), achieving good performance in one case, but bad in the latter one.

ACKNOWLEDGMENTS

This work was partially supported by the Spanish Grants TIN2016-75404-P AEI/FEDER, UE, TIN2014-52072-P, TIN2016-79717-R, TIN2013-42795-P and the European Commission H2020 I-REACT project no. 700256. Laura Lopez-Fuentes benefits from the NAERINGSPHD fellowship of the Norwegian Research Council under the collaboration agreement Ref.3114 with the UIB.

REFERENCES

[1] 2018. MediaEval 2018 Multimedia Satellite Task. http://www.multimediaeval.org/mediaeval2018/multimediasatellite/. (2018). Data released: 31 May 2018.

[2] Flavia Sofia Acerbo and Claudio Rossi. 2017. Filtering informative tweets during emergencies: a machine learning approach. In *Proceedings of the First CoNEXT Workshop on ICT Tools for Emergency Networks and DisastEr Relief*. ACM, 1–6.

[3] Federico Angaramo and Claudio Rossi. 2017. Online clustering and classification for real-time event detection in Twitter. (2017).

[4] Benjamin Bischke, Patrick Helber, Zhengyu Zhao, Jens de Bruijn, and Damian Borth. The Multimedia Satellite Task at MediaEval 2018: Emergency Response for Flooding Events. In *Proc. of the MediaEval 2018 Workshop* (Oct. 29-31, 2018). Sophia-Antipolis, France.

[5] Tom Brouwer, Dirk Eilander, Arnejan Van Loenen, Martijn J Booij, Kathelijne M Wijnberg, Jan S Verkade, and Jurjen Wagemaker. 2017. Probabilistic flood extent estimates from social media flood observations. *Natural Hazards & Earth System Sciences* 17, 5 (2017).

[6] Yinghua He, Hong Wang, and Bo Zhang. 2004. Color-based road detection in urban traffic scenes. *IEEE Transactions on intelligent transportation systems* 5, 4 (2004), 309–318.

[7] Christopher D. Manning Jeffrey Pennington, Richard Socher. 2018. GloVe: Global Vectors for Word Representation. (2018). https://nlp.stanford.edu/projects/glove/

[8] Victor Klemas. 2014. Remote sensing of floods and flood-prone areas: an overview. *Journal of Coastal Research* 31, 4 (2014), 1005–1013.

[9] Hui Kong, Jean-Yves Audibert, and Jean Ponce. 2010. General road detection from a single image. *IEEE Transactions on Image Processing* 19, 8 (2010), 2211–2220.

[10] Laura Lopez-Fuentes, Joost van de Weijer, Marc Bolanos, and Harald Skinnemoen. 2017. Multi-modal deep learning approach for flood detection. In *Proc. of the MediaEval 2017 Workshop (Sept. 13–15, 2017). Dublin, Ireland.*

[11] Laura Lopez-Fuentes, Joost van de Weijer, Manuel González-Hidalgo, Harald Skinnemoen, and Andrew D Bagdanov. 2018. Review on computer vision techniques in emergency situations. *Multimedia Tools and Applications* 77, 13 (2018), 17069–17107.

[12] Igor Ogashawara, Marcelo Pedroso Curtarelli, and Celso M Ferreira. 2013. The use of optical remote sensing for mapping flooded areas. *Int. J. Eng. Res. Appl* 3, 5 (2013), 1956–1960.

[13] Pramuditha Perera and Vishal M Patel. 2018. Learning Deep Features for One-Class Classification. *arXiv preprint arXiv:1801.05365* (2018).

[14] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2818–2826.

[15] CJ Ticehurst, P Dyce, and JP Guerschman. 2009. Using passive microwave and optical remote sensing to monitor flood inundation in support of hydrologic modelling. In *Interfacing modelling and simulation with mathematical and computational sciences, 18th World IMACS/MODSIM Congress*. 13–17.

[16] Luca Venturini and Evelina Di Corso. 2017. Analyzing spatial data from twitter during a disaster. In *Big Data (Big Data), 2017 IEEE International Conference on*. IEEE, 3779–3783.