BMC@MediaEval 2017 Multimedia Satellite Task Via Regression Random Forest

Xiyao Fu, Yi Bin, Liang Peng, Jie Zhou
Yang Yang, Heng Tao Shen
Center for Future Media and School of Computer Science and Engineering
University of Electronic Science and Technology of China
fu.xiyao.gm@gmail.com,yi.bin@hotmail.com,pliang951125@outlook.com,jiezhou0714@gmail.com
dlyyang@gmail.com,shenhengtao@hotmail.com

ABSTRACT

In the *MediaEval 2017 Multimedia Satellite Task*, we propose an approach based on regression random forest which can extract valuable information from a few images and their corresponding metadata. The experimental results show that when processing social media images, the proposed method can be high-performance in circumstances where the images features are low-level and the training samples are relatively small of number. Additionally, when the low-level color features of satellite images are too ambiguous to analyze, random forest is also a effective way to detect flooding area.

1 INTRODUCTION

The outburst of social media provides us with an opportunity to deal with specific tasks, *e.g.,* disaster prediction and specific scene identification. Such problems can be crucial in agriculture, urbanization and environment monitoring. The *MediaEval 2017 Multimedia Satellite Task* consists of two subtasks: Disaster Image Retrieval from Social Media (DIRSM) task and Flood-Detection in Satellite Images(FDSI) task. The former one requires the prediction system to identify flooding circumstances in social media pictures, while the latter task aims to judge that which district in a certain area of a satellite image is suffering from flooding.

As to the theoretical basis of the task, existing work such as [1], which using Twitter as main data source and analyze associated geographical, textual, temporal and social media information. They split the task into four events (*metadata analysis, text analysis, image analysis and temporal aggregation*), each of which represents an utilizations of the data from tweets. In the processing of the satellite images in FDSI subtask, Chaouch et al. [3] exploited and combined different low-level color selectors to identify flooding areas on different satellite pictures. However, they used RGB color map to detect flooding area by predicting the water level, which means that this method requires the images to have strong diversity. In other conditions such as this subtask when the colors of the images is dim, the method may have difficulty processing them. In [6], the authors employed SVMs and low-level features descriptors (*e.g.*, SIFT descriptor) to detect fire scenes. In [8], the authors aim to generate spatial variants of satellite images in order to map the flooding areas. Same as aforementioned, the histogram they used to map the variance is time-consuming, and the data format is not fit for FDSI task to deal with.

In this paper, we propose to employ regression random forests to rank the relevance of flooding in both social media images taken by cameras and satellite images describing the overall situation of a certain district. It is shown that our method can well balance the efficiency and effectivity. In other words, our method achieves compatible performance with extraordinary short time-consuming. More specifically, we design two prediction systems based on regression random forest method, which has been proven effective in handling high-dimensional data and preventing overfitting when training set is comparatively small [7]. In the rest of this paper, we mainly discuss the approach developed for our systems and the evaluation of experimental results.

2 APPROACH DESCRIPTION

2.1 DIRSM Subtask

The goal of DIRSM subtask is to retrieve all images which show direct evidence of a flooding event from social media streams. The details of this subtask are described in [2]. The main challenge of this task is in two folds: (a) discrimination of the water levels in different areas, and (b) consideration of different types of flooding events. In many cases the images can be confusing to classify (e.g., to tell images showing a flushing river or a rainforest from the real flooding ones such as a flooded park).

2.1.1 Feature Extraction. In recent years, Convolutional Neural Networks (CNNs) has been dominating in the field of computer vision, such as recognition and detection. Therefore, except for the baseline features provided by the organizers, we also extract robust CNN feature to improve the performance of our system. Specifically, we apply *ResNet-152* [4] as the extraction network, and employ *Caffe* toolbox [5] to extract features from the training set. Each image is extracted from the bottom conv layer of ResNet. The dimension of each image feature vector is 2048.

2.1.2 Models Definition. In order to improve the performance of our system, the learning algorithm we choose must satisfy several requirements: (a) it should remain high-performance under circumstances where data is restricted; (b) it should excel in accuracy among current algorithms; (c) it can handle thousands of input variables without variable deletion; (d) its speed should be high enough. Due to the consideration above, we use regression random forest for the ranking of the 5 runs.

MediaEval'17, 13-15 September 2017, Dublin, Ireland

As an important application in ensemble methods, random forest is a high-performance method both for classification and regression. When the number of training set images is not large enough to utilize other learning methods (e.g., deep learning), using random forest can prevent overfitting and unbalance of features in datasets [7]. A random forest consists of many classification(or regression) trees and uses bagging mechanism to learn base estimators. As one of the main contributions in ensemble methods, bagging requires randomly allocating training data (including features learned) to each classifier(regressor) to train a base estimator.

Theoretically, the results will improve with the number of trees in a random forest increasing. However, since the computation cost increases as well as the advancement decreases when new trees are added to a larger forest. This phenomenon indicates that the number of trees should be limited.

As for the details of the parameters setting, we set the bagging percent of the forest as 0.9 [9], the minimal leaf size as 10 (when the data points get down the value, stop splitting the data), we set the number of regression trees in the random forest as 500. During training, we split the development set into training set and validation set with 80 and 20 percent, respectively.

2.2 FDSI Subtask

The aim of the FDSI subtask is to develop a model that is able to identify regions in satellite imagery which are affected by a flooding. Same as the DIRSM subtask, the details can be find in [2]. The main challenge relies on defining flooding area based on conjoint area's situation. For example, in a satellite image, a lake has bounds and belongs to the area without flooding, while a river does not have intact bounds and partly belongs to flooding area.

The same as before, we use ResNet and caffe to extract the features of the satellite images. We still use the 2048-dim vectors of bottom conv layer in ResNet. Meanwhile, we utilize random forest to process the images in the development set due to the same reason, which is that the number of features to learn in the satellite images is small and the development set is even smaller than the DIRSM subtask. We set the bagging percent of the forest at 0.8, the minimal leaf size as 20, and the number of trees at 400. During training, we use the first four development set folders as training set, the other two as validation set.

3 EXPERIMENTS AND RESULTS

3.1 DIRSM Subtask

Because of different requirements in the 5 runs, we train features with slightly different setting. In run 1, 4 and 5, we only utilize the images features in training. To augment the dataset, we randomly crop and flip horizontally the original images in the development set, and obtain 5320 more images for run 4 and 5 additionally. In run 2, we utilize the text given in associated metadata of the development set. We process each word in every image description as GloVe vector, the dimension of each vector 300, and constrict each sentence to its maximum length to generate a matrix including all the sentences. In run 3, we use both the text and the development set images to train the random forest.

The mean average precision (mAP) scores we get from the 5 runs is shown in Table 1. The mAP scores listed are the mean of average precision at the top 50, 100, 200, 300, 400 and 500 rankings of each run.

Table 1: Performance on Testset of DIRSM Subtask

	run1	run2	run3	run4	run5
MAP	19.21	12.84	18.30	17.24	17.72

Table 2: Performan on Testset of FDSI Subtask

	run1	run2	run3	run4	run5
location 01	0.3657	0.368	0.3525	0.3617	0.3678
location 02	0.3286	0.3125	0.3226	0.3221	0.3224
location 03	0.3408	0.3359	0.342	0.3486	0.3421
location 04	0.3107	0.32	0.3155	0.3129	0.3106
location 05	0.426	0.427	0.424	0.433	0.4341
new location	0.402	0.401	0.402	0.403	0.401

Experimental results of DIRSM are shown in Table 1. As we can see, metadata only (run 2) perform much worse than visual information only model (run 1), which indicates that associated descriptions are much noisy than visual information for flooding prediction. Intuitively, more feature bring more information, and gain better performance. However, run 3 (combination of visual and textual feature) performs a little worse than run 1. This also demonstrates that textual description introduces much noise, even induce to decrease the performance of original images.

3.2 FDSI Subtask

For the run 1, 2 and 3, we only utilize the original satellite images and their ground truth masks in training. For run 4 and 5, we use the cropped satellite images and horizontally flipped images as well.

Table 2 exhibits the intersection of union (IoU) of experimental results. the best performances lie in the location 05 and the new location provided by the task organizers. Possible reasons may be that the number of the images in these location is relatively small, reducing the possibility of overfitting. Besides, the mean performance of the last 2 runs is better than the first 3 ones, claiming that the general run cast the better results. It is possible that the performance in the first 3 ones suffer from the variance of more images, but generally the results are at large satisfying.

4 CONCLUSION

In this paper, we illustrated our approach for the *MediaEval 2017 Multimedia Satellite Task*. In both subtasks, combining random forests and CNN features enhanced the performance of the detection. In the DIRSM subtask, combining the features learnt from text and images improved the regression performance of labeling, but our methods still suffer from noise. As to FDSI subtask, the performance of the proposed method could be better when the number of test images fewer. The best result remained in the location 05 and the new location. Overall, the proposed method got promising performance in processing both social media stream and satellite images.

REFERENCES

[1] Benjamin Bischke, Damian Borth, Christian Schulze, and Andreas Dengel. Contextual enrichment of remote-sensed events with social media streams. In *Proceedings of the 2016 ACM on Multimedia Conference*, 2016.

[2] Benjamin Bischke, Patrick Helber, Christian Schulze, Srinivasan Venkat, Andreas Dengel, and Damian Borth. The multimedia satellite task at mediaeval 2017: Emergence response for flooding events. In *Proc. of the MediaEval 2017 Workshop*, Sept. 13-15, 2017.

[3] Naira Chaouch, Marouane Temimi, Scott Hagen, John Weishampel, Stephen Medeiros, and Reza Khanbilvardi. A synergetic use of satellite imagery from sar and optical sensors to improve coastal flood mapping in the gulf of mexico. In *Hydrological processes*, 2012.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[5] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, 2014.

[6] Ryan Lagerstrom, Yulia Arzhaeva, Piotr Szul, Oliver Obst, Robert Power, Bella Robinson, and Tomasz Bednarz. Image classification to support emergency situation awareness. In *Frontiers in Robotics and AI*, 2016.

[7] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. In *R News*, 2002.

[8] Igor Ogashawara, Marcelo Pedroso Curtarelli, and Celso M Ferreira. The use of optical remote sensing for mapping flooded areas. In *International Journal of Engineering Research and Application*, 2013.

[9] Vladimir Svetnik, Andy Liaw, Christopher Tong, J Christopher Culberson, Robert P Sheridan, and Bradley P Feuston. Random forest: a classification and regression tool for compound classification and qsar modeling. In *Journal of chemical information and computer sciences*, 2003.