Artificial Intelligence Research and Development
I. Sanz et al. (Eds.)

doi:10.3233/FAIA230701

Improving Disaster Response by Combining Automated Text Information Extraction from Images and Text on Social Media

Hafiz Budi FIRMANSYAH [a,1], Carlo Alberto BONO [b], Valerio LORINI [d], Jesus CERQUIDES [c], Jose Luis FERNANDEZ-MARQUEZ [a]

[a] *University of Geneva, Switzerland*
[b] *Politecnico di Milano, DEIB, Italy*
[c] *Artificial Intelligence Research Institute IIIA-CSIC, Spain*
[d] *European Comission - Joint Research Centre, Ispra*

Abstract. During disasters, social media can serve as a valuable source of real-time information about the impacts on people and infrastructure. However, due to the lack of geographical information in most social media posts, this information is often underutilized by first responders. Previous research has attempted to estimate the location of individual social media posts using text and image analysis, but limitations still exist in fine-grained disaster area mapping. To address this issue, this paper analyses the performance of combining text from social media post with textual information from the images on improving the geolocation of social media information during a disaster.

Keywords. Social Media, Disaster, Text Extraction, Geolocation

1. Introduction

Social media users share details about damaged infrastructure, injured victims, damage severity, and real-time situation of disaster scenes [1]. On the platforms, the social media posts may contain geolocation information. Such information is important for disaster responders to map the impacted area and take action to the situation quickly. However, only 3% of social media posts contain structured geolocation information [2,3], challenging the use of social media information in disaster management.

The potential of using social media data to improve disaster response attracted researchers to combine natural language processing (NLP) techniques and gazetteers to retrieve geolocation information. A large part of work contributions to geolocate social media posts focus on text analysis, for instance analyzing the text associated with images to find the location where the images were taken. These approaches have been proven to be able to geolocate a significant percentage of social media posts at country and city

[1] Corresponding Author: Hafiz Budi Firmansyah, hafiz.firmansyah@unige.ch

level. Text-based geolocation approaches have also been combined with crowdsourcing to increase the precision of the location. However, this hybrid approach, while performing well in terms of precision, can delay the production of a disaster area map due to the manual geolocation process. Therefore, it is necessary to increase the number of automatically geolocated images in order to reduce the time needed to produce the disaster area map.

The text contained in images, such as street names, traffic signals, and names of businesses or points of interest, serves a crucial role in crowdsourcing the geolocation of images. However, to the best of our knowledge, the automatic extraction of text from images and combining the extracted text with social media posts to retrieve geolocation has never been explored.

This paper presents a pipeline that utilizes text information extracted from social media images to improve the creation of disaster maps. The study systematically analyses the value of text embedded in images attached to social media posts for improving social media information during a disaster.

The remainder of the paper is structured as follows: Section 2 discusses related work. Section 3 presents the proposed approach. Section 4 shows the experimental results using real social media data from four disaster events. Section 5 describes the discussion. Finally, Section 6 reports the conclusion and future works.

2. Related works

The studies on geolocation in social media have attracted many researchers. Moreover, the research on text analysis for inferring geolocation information has been growing steadily recently. Several approaches have been presented, for instance, named entity recognition (NER), location-indicative word (LIW), user location profile (ULP), and social relations [4].

Some previous works attempted to develop methodologies based on machine learning, crowdsourcing, or hybrid approaches. Lorenzo et al. [2] proposed a text-based machine learning approach. They combined NER and network theory to extract relevant location information from tweets. The approach has been applied to nearly 3 million tweets, managing to extract country-level geolocation with 89.8% accuracy.

In another work, the authors exploited a naïve Bayes probabilistic method combined with location indicative words techniques. Focusing on feature selection, their work aimed at city-level geolocation. The results demonstrated how selected features outperformed the complete feature set [5].

Another popular approach to acquiring geolocation is extracting the location from the user's profile. In various works, authors processed the information in the user profile using a NER approach to obtain users' locations. However, the location quality could be limited, e.g., the users might write a fictitious location in their profile [6,7,8].

In [8] and [9], the authors proposed social relation techniques to obtain geolocation prediction. Specifically, they estimated the location based on the friendship network. However, also in this case limitations could arise. For example, users could be linked with someone they do not actually know (e.g., public figures). Furthermore, the approach demands high computational resources for processing.

In terms of text-based geolocation prediction algorithm, Scalia et al. proposed an approach called CIME, a technique based on NER and disambiguation of candidate locations [10]. An application of this methodology to large-scale flood characterization is proposed in [11]. Shankar et al. developed Crowd4EMS, a platform to geolocate a social media post, has successfully combined crowdsourcing techniques and an automatic approach for gathering information. Figure 1 illustrates how crowdsourcing can be leveraged to validate automated information extraction. The picture on the right, collected from Twitter, represents damaged buildings due to an earthquake. The picture on the left shows the same location on Google Street View. Using a crowdsourcing platform, crowd workers were asked to identify whether the image is located in a specific location. The approach was evaluated using data related to the Amatrice 2016 earthquake [12].

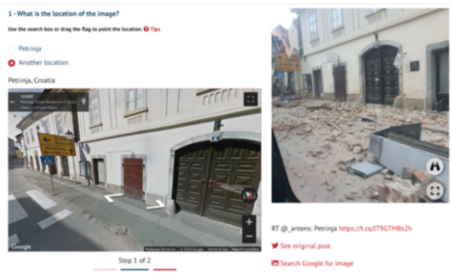

Figure 1. The crowdsourcing approach to validate geolocation information of a social media image

Besides text-based geolocation prediction, research on image-based geolocation prediction also exists. Murgese et al. address the problem of estimating images' locations by introducing Focal Modulation Network. They used geolocated images from Flickr and Mapillary [2] to evaluate their approach. However, the approach was not performing well with Flickr data which is more similar to the one used in disaster situations [13].

In a more recent study, Firmansyah et al. proposed a text extraction from images pipeline. This pipeline enables the integration of the extracted text into several geolocation prediction algorithms. The result shows that the geolocation prediction is reasonably good at the country level [14].

Despite extensive work on post-based and initial works on text extraction from image geolocation prediction, the approach proposed in this paper builds on the pipeline proposed by [15] and [14] which focuses on the geolocation of social media information by analyzing the text extracted from the image. Instead of focusing on a country level, this research aims to evaluate the number of more specific impacted disaster areas.

3. Our approach

This research evaluates the potential of using text extraction from social media images to find their locations. An image often comes with relevant text information such as the name of a street, a shop name, traffic signals directions, a hotel or a hospital name, which are key clues for deriving the location of an image.

The proposed approach aims to combine the extracted text from the images with social media post and predict the geolocation information using state-of-the-art algorithms from the literature.

[2]https://www.mapillary.com

Our approach proposes a pipeline for extracting relevant information from social media, related to a given disaster, and locating them in a disaster map to allow emergency responders to take better-informed decisions. Figure 2 illustrates the proposed pipeline, which is composed of five processes:

1. **Social media crawling** process fetches informative posts from social media by selecting appropriate keywords which depend on the type of disaster.
2. **Cleaning and Filtering** process retains posts that are relevant to the event, removing duplicates, memes, not safe to work, and images which do not provide information related to the disaster event (i.e. not informative images).
3. **Text extraction from images combined with text from post** process automatically extracts any text in the social media images. Then, we combined the extracted text with text from social media posts. They will be used later on by text-based geolocation algorithms.
4. **Geolocation prediction** process retrieves the candidate locations for each text fragment using text-based geolocation algorithms.
5. **Map visualization** process implements a disaster map where each dot represents a social media image.

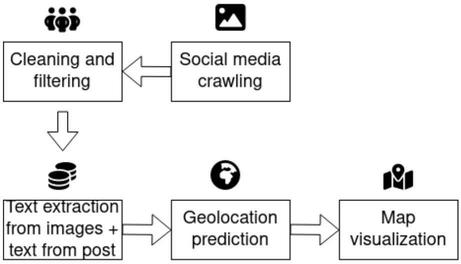

Figure 2. Pipeline using text extraction from images to geolocate social media data

4. Experimental settings and results

4.1. Dataset assessment

The dataset used for the validation of the presented approach is composed of Twitter data from two types of disasters, earthquakes and floods. The dataset consists of two modalities: text and images. The dataset has been created using the Social Media for Disaster Risk Management (SMDRM) Platform[16] gathered by Copernicus Emergency Management Service[3].

The text dataset contains 5,430 posts. From that number, we found that about 4,714 posts were unique. Furthermore, the images dataset contains 1,752 images from four disasters, Catania floods, 2021 (70 images), Central European floods, 2021 (214 images), Croatia earthquake, 2020 (549 images), and Haiti earthquake, 2010 (961 images).

[3] https://emergency.copernicus.eu/

To evaluate which images contain text, a crowdsourcing project was created using the Citizen Science project builder [4] as shown in Figure 3. This analysis is used later on as a ground truth to evaluate the automatic extraction of text. Each image is classified into three categories:(1) “yes”, (2) “no”, and (3) “yes, but not legible”, as shown in Figure 3.

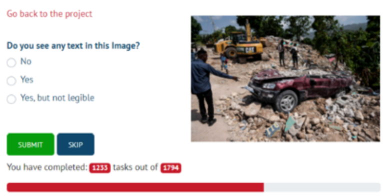

Figure 3. Classifying the images using Citizen Science Project Builder

From 1,752 images, we observed 852 images (49%) having text, 281 (16%) categorized as “yes but not legible” and 619 (35%) images not containing text. The class “yes but not legible” was defined as having small or blurry text which cannot be readable (see Figure 4).

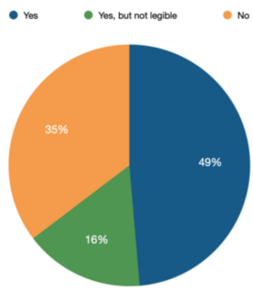

Figure 4. Images category distribution

Out of the 1,133 images containing text, we found 3 images were no longer accessible after these operations. Then, we finally considered 1,130 images.

4.2. Experimental settings

Recent advances in computer vision enable highly accurate extraction of text from images. In this research, we incorporated several existing technologies. We used the Amazon Rekognition service to extract text from images. The service implements a Convolutional Neural Network (CNN) pre-trained on a large labeled dataset. Previous research demonstrated that Convolutional Neural Networks show excellent performance when dealing with image data.

In terms of development environment, programming language, and library, we relied on Python and Google Colab as a development environment. The folium[5] library was used to visualize the predicted geolocation in a map.

[4] https://lab.citizenscience.ch/en/
[5] https://python-visualization.github.io/folium/

A geolocation algorithm, CIME, was implemented into the experiment. The CIME algorithm is deployed in two different configurations, with or without providing the detected language. The language was estimated directly from the text. The estimation was performed using the langdetect[6] library, which supports 55 different languages.

4.3. Experimental results

The Amazon Rekognition[7] service extracted the text from 1,130 images. From 1,130 images, Amazon Rekognition managed to detect the text of 1,029 images. From 1,029 images, we selected 532 images containing real text by excluding images with banners, news logos, images caption, and watermarks. To observe the improvement, we concatenated the extracted text from images into the post. Finally, we had 532 data that were concatenating extracted text from images and social media posts. The text from the post and text from the images were concatenated using a space as the separator. The text concatenated text was then fed to the CIME algorithm for inferring geolocation information. We aim to measure the delta improvement between text from the post and the combination of text from the post and text from images.

4.4. Assessing the relevance of texts from images for geolocating the image

This subsection shows the comparison between different settings in the experiment and measures their enhancement. The goal of this subsection is to contrast the results when incorporating the extracted text from images into social media posts and having only text from social media posts.

	CIME without language prediction			**CIME with language prediction**		
Macro average	Post	Post+Image	Delta	Post	Post+Image	Delta
Precision	0.91	0.96	0.05	0.91	0.96	0.05
Recall	0.97	0.97	0	0.97	0.97	0
F1 score	0.93	0.96	0.03	0.93	0.96	0.03

Table 1. Delta improvement in the two different settings of geolocation algorithm

Table 1 demonstrates the macro average score for each CIME algorithm setting. Macro average is the arithmetic mean of the scores on individual classes. The macro average includes precision, recall, and F1 score. The term post refers to text from post. The term post+image indicates to the concatenated text (text from post and text from images). Overall, both settings produce better performance for the combination of text from the post and text from images compared to text from the post. The improvement is ranging from 3 to 5 % for each performance metric. Furthermore, we found that the different settings (with or without language prediction) indicated no difference result.

In the CIME algorithm without language setting, the number of candidates were decreasing from 310 (only post) to 257 (post + extracted text). Despite the lower number on the post and extracted text, we found that this combination provided better precision, recall, and F1 score. In the CIME algorithm with language settings, the total number of

[6]https://pypi.org/project/langdetect/
[7]https://aws.amazon.com/id/rekognition/

candidates were equally declining from 307 (only post) to 258 (post + extracted text). Similar to previous settings, the result demonstrates better performance.

4.5. Evaluation of distribution of administrative level boundaries

This experiment analyses the administrative level boundaries obtained by the CIME algorithm using just posts and CIME algorithm using posts and texts extracted from the images. The administrative level represent the details of a feature within a government hierarchy. [8]. Intuitively, using texts from images and text from the post would better inform the CIME algorithm and the distribution of administrative level boundaries will be more precise, i.e. using the text extracted from images we are able to geolocate more images at the level of the street and number of the building.

Figure 5 illustrates the distribution of candidates based on the administrative level. In general, the x-axis represents the detail of the administrative level while the y-axis shows the candidate percentage. We can observe that the distribution is concentrated in administrative level 15 (equivalent to quarters, neighborhood). For the experiment without a language detection setting, the percentage is improving from 63% (see green bar) to 65% (see blue bar). For the experiment with the language detection setting, the percentage is increasing from 62 % (see orange bar) to 64 % (see grey bar).

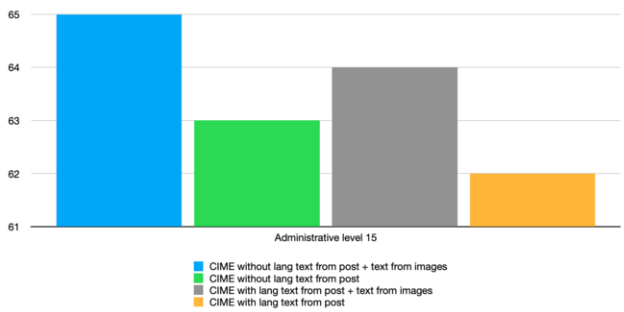

Figure 5. Admin level discrete distribution

4.6. Evaluation of candidate within bounding box and map visualization

This experiment analyses the ratio of geolocation candidates predicted within the affected area for each of the four disasters presented in the database. Each affected area is represented by a bounding box as shown in Figure 6. Each dot represents a geolocation candidate and the geolocation algorithm setting (without language or with language detection). The blue bounding box was manually created and represented the impacted area. The rationale for creating this bounding box is to provide a middle point and trade-off between ”too general” and ”too tight”. The term ”too general” refers to the wide area for instance country level, while ”too tight” means that it only concentrated on the impacted area based on Copernicus activation [9].

[8]https://wiki.openstreetmap.org/wiki/Key:admin_level
[9]https://emergency.copernicus.eu/mapping/list-of-activations-rapid

For CIME without language setting, about 42% (310/736) of text from posts were within the bounding box. While 43.2% (257/594) of concatenated text was inside the bounding box. There was a 1.2% improvement in the number of candidates predicted in the bounding box. For CIME with language setting, About (307/758) 40.8 % text from posts and (258/630) 40.9 % post + extracted text were within the bounding box.

Figure 6. Maps with predicted geolocation

5. Discussion

Results from the experiment show that adding extracted text from images to text from social media post increases the performance. The result demonstrates consistent improvement in both language prediction and without language prediction settings. Practically speaking, the operation team, public information team and the first reponders could benefit the practical contribution from the research as the end users [17].

As the additional practical contribution, we also proposed a pipeline that could be deployed as a feeder for a text-based geolocation prediction algorithm. The evaluation was done by measuring the performance of the extracted text from images into the contents of the posts with the text from posts.

However, we identified that there are four limitation on the use of social media data in the context of disaster response. First, the risk of having false geolocation information should be also considered carefully before adapting the approach to the production. Second, the issue of inclusiveness and accessibility needs to be anticipated for instance by providing data from various social media platforms (Facebook, Instagram, Flickr, and Foursquare). Third, the privacy aspect needs to be reviewed regularly in accordance with privacy policy. For instance, by removing personal information such as name, identification number, age, exact address in the social media data. Finally, the semantic meaning on Twitter might be more difficult to interpret compared to Facebook. This could happen because the character limitation in Twitter platform [18].

Even tough social media offers some advantages, it also comes up with the downside. As an example, the data might be irrelevant and noisy. One approach to address this problem is by incorporating the combination of machine learning and human (expert or crowd) to filter out the misinformation or noise.

6. Conclusion and Future Works

Geoinformation in social media posts serves a significant role for the first responders in the event of a disaster. However, most social media content does not contain geolocation information originally. Despite abundant research on text-based location prediction conducted by researchers in recent years, the research on combining text extraction-based geolocation from images and social media posts was still not discovered.

This paper analyses the potential of automated text extraction from images to enhance geolocation prediction. The proposed approach builds on the state-of-the-art text-based geolocation algorithms, by providing them a concatenation of extracted text from the images and the text from the post. The experimental result shows that the extracted text from images better infers the text-based geolocation algorithm thus improving the quality of geolocation candidate predictions.

As a future works, we plan to improve the geolocation prediction by filtering process by favoring the clear location indication in the social media images, might augment the quality of geolocation prediction.

Measuring multiple dimensions in order to establish a confidence score, for instance lightning condition, text direction in the images, extracted text confidence, number of times that the image appears on social media (also possibly in the past), geographical distribution of the predictions, and scoring derived from the disambiguation process, might be meaningful to investigate and corroborate the efficacy of the predictions. This measurement could be also considered as future direction of the works.

Acknowledgements

This work was funded by LPDP (Indonesia Endowment Fund for Education), the European Commission H2020 project Crowd4SDG "Citizen Science for Monitoring Climate Impacts and Achieving Climate Resilience", #872944. This work expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this work. The authors thank all the Crowd4SDG partners for the discussions and common work during the project.

References

[1] F. Alam, F. Ofli, and M. Imran, "CrisisMMD: Multimodal Twitter datasets from natural disasters," *Proceedings of the 12th International AAAI Conference on Web and Social Media (ICWSM)*, 2018.

[2] J. Alonso-Lorenzo, E. Costa-Montenegro, and M. Fernández-Gavilanes, "Language independent big-data system for the prediction of user location on twitter," in *2016 IEEE International Conference on Big Data (Big Data)*, 2016, pp. 2437–2446.

[3] C. Francalanci, P. Guglielmino, M. Montalcini, G. Scalia, and B. Pernici, "IMEXT: A method and system to extract geolocated images from tweets — analysis of a case study," *2017 11th International Conference on Research Challenges in Information Science (RCIS)*, 2017.

[4] M. N. Utomo, T. B. Adji, and I. Ardiyanto, "Geolocation prediction in social media data using text analysis: A review," *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018.

[5] H. Bo, E. Cook, and T. Baldwin, "Geolocation prediction in social media data by finding location indicative words," in *24th International Conference on Computational Linguistics (COLINGS 2012)*, 2012, pp. 1045–1062.
[6] M. Maghrebi, A. Abbasi, T. H. Rashidi, and S. T. Waller, "Complementing travel diary surveys with twitter data: Application of text mining techniques on activity location, type and time," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 208–213.
[7] H. Shekhar, S. Setty, and U. Mudenagudi, "Vehicular traffic analysis from social media data," in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2016, pp. 1628–1634.
[8] E. Williams, "Geocontext: Discovering geographical topics from social media," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016, pp. 1342–1346.
[9] D. Xu, P. Cui, W. Zhu, and S. Yang, "Find you from your friends: Graph-based residence location prediction for users in social media," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014, pp. 1–6.
[10] G. Scalia, C. Francalanci, and B. Pernici, "CIME: Context-aware geolocation of emergency-related posts," *GeoInformatica*, vol. 26, no. 1, pp. 125—-157, 2022.
[11] C. Bono, B. Pernici, M. O. Mulayim, J. L. Fernandez-Marquez, A. R. Shankar, and E. Nemni, "TriggerCit: Early flood alerting using twitter and geolocation - a comparison with alternative sources," in *19th ISCRAM Conference*, 2022.
[12] A. Ravi Shankar, J. L. Fernandez-Marquez, B. Pernici, G. Scalia, M. R. Mondardini, and G. Serugendo, "Crowd4EMS: A crowdsourcing platform for gathering and geolocating social media content in disaster response," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-3/W8, p. 331–340, 2019.
[13] F. Murgese, G. Costa-Alcaina, M. O. Mulayim, J. Cerquides, and J. L. Fernandez-Marquez, "Automatic outdoor image geolocation with focal modulation networks," in *Artificial Intelligence Research and Development: Proceedings of the 24th International Conference of the Catalan Association for Artificial Intelligence*. IOS Press, 2022.
[14] H. B. Firmansyah, J.-L. Fernandez Marquez, J. Cerquides, C. Bono, V. Lorini, and B. Pernici, "Enhancing disaster response with automated text information extraction from social media images," in *IEEE Big Data Service*, 2023.
[15] C. Bono, M. O. Mülâyim, C. Cappiello, M. J. Carman, J. Cerquides, J. L. Fernandez-Marquez, M. R. Mondardini, E. Ramalli, and B. Pernici, "A citizen science approach for analyzing social media with crowdsourcing," *IEEE Access*, vol. 11, pp. 15 329–15 347, 2023.
[16] V. Lorini, E. Panizio, and C. Carlos, "Smdrm: A platform to analyze social media for disaster risk management in near real time," in *Workshop Proceedings of the 16th International AAAI Conference on Web and Social Media ICWSM*, 2022.
[17] D. P. Freitas, S. B. Marcos R., and d. C. Paulo Victor R., "A conceptual framework for developing solutions that organise social media information for emergency response teams," *Behaviour & Information Technology*, vol. 39, no. 3, pp. 360–378, 2020.
[18] C. Havas, B. Resch, C. Francalanci, B. Pernici, G. Scalia, J. L. Fernandez-Marquez, T. Van Achte, G. Zeug, M. R. R. Mondardini, D. Grandoni, B. Kirsch, M. Kalas, V. Lorini, and S. Rüping, "E2mc: Improving emergency management service practice through social media and crowdsourcing analysis in near real time," *Sensors*, vol. 17, no. 12, 2017.