The Ianos Cyclone (September 2020, Greece) from Perspective of Utilizing Social Networks for DM

Stathis G. Arapostathis(✉)

Harokopio University, Kallithea, 17671 Athens, Greece

Abstract. Main purpose of current research is to present evolutions in previous presented approaches of the author for manipulating social media content for disaster management of natural events. Those innovations suggest the adoption of machine learning for classifying both photos and text posted in social networks along with hybrid geo-referencing. As case study the author chose the Ianos cyclone, occurred between Italy and Greece, during September 2020. The geographic focus of the research was in Greece where the cyclone caused 4 human losses and damages in the urban environment. A dataset consisted of 4655 photos, with their corresponding captions, timestamps and location information was crawled from Instagram. The main hashtag used was #Ianos. Two data samples, one for each type, were classified manually for calibrating the classification models. The classes regarding photos were initially: (i) related and (ii) not related to Ianos, while the general classification schema for photos and text was: (i) Ianos event identification, (ii) consequences, scaled according to the impact of each report, (iii) precaution, (iv) disaster management: announcements, measures, volunteered actions. Author's approach regarding classification suggests the use of convolutional neural networks and support vector machine algorithms for image and text classification respectively. The classified dataset, was geo-referenced by using commercial geocoding API and list-based geoparsing. The results of the research in current status are at an initial level, a subset of data though of automatically or manually processed information is presented in four related maps.

Keywords: Machine learning · Disaster management · Social media · Volunteered Geographic Information · Ianos cyclone

1 Introduction

Social network data is an important part of data science. The enormous rhythm of data produced by social media users, forced the scientific world to research on innovative methods for utilizing them in benefit of a plethora of scientific disciplines. In Geography, social network data analysis is strongly associated to the phenomenon of Volunteered Geographic Information (VGI) which is described as the "act of normal citizens, produce geographic information" [1]. Through that point of view, social networks are considered as an "unconventional" VGI source [2, 3] with many potentials in almost

Published by Springer Nature Switzerland AG 2021
Y. Murayama et al. (Eds.): ITDRR 2020, IFIP AICT 622, pp. 160–169, 2021.
https://doi.org/10.1007/978-3-030-81469-4_13

every spatial-related scientific discipline. Despite any point of view, there is published research regarding its' applications in medicine [4, 5], politics [6], sentiment/marketing [7] and natural disasters. Especially, regarding the latter, the emergence of new innovative tools for DM has brought up a plethora of related re-search that aspires to contribute to DM related to floods [8–11], fires [12], earthquakes [13, 14] and other disastrous events including hurricanes [15].

According to author's latest research findings [3], the main challenges regarding effective social network data manipulation for DM needs, are accumulated in four main groups. A. Classification B. Geo-referencing, C. Visualization, D. Automation. For each one of those groups there is ongoing research. In particular, various classification schemas have been proposed [9, 16] that aspire to extract all the valuable information available through social media. The classification schemas in general may vary depending on data types which can be images, text or video. Moreover, geo-referencing is an important aspect which still faces challenges mostly in terms of enhancing the precision [3, 17]. This research provides some advancement by utilizing a hybrid approach which makes use of both commercial map APIs (google maps geocoder) and conventional list-based geoparsing approaches [18].

In addition, visualization is still a challenge considering that the processed information should be delivered to all citizens, despite their level of literacy, who may potentially face a disastrous event. Finally, automation is a key component that can convert all of those theoretic approaches to operational at real time. Current research provides significant advancements in automation as well, by employing machine learning techniques for both image and text, minimizing thus the whole data processing time to very few minutes.

Case Study

The case study used, was the recent Mediterranean cyclone (medicane) entitled as Ianos. That cyclone occurred during middle of September initially in Italy, between Sicily and Ionian Sea. In Greece, strong storms and floods occurred while the wind's speed reached the 150 km/h. Main affected areas were initially the Ionian Islands: Zante, Kefallinia, and Ithaca. The cyclone affected, sequentially, the prefectures of West Greece, Peloponnese while in Thessaly caused among others 4 human losses.

Eventually, the medicane moved south, to the prefectures of Attica, Cyclades and Crete. Apart from human losses, the port of Zante had serious damages, while in Thessaly, in the area of Mouzaki, 2 bridges were completely destroyed. The main rail network which connects Athens and Salonica was cut in central Greece. In some areas the total rainfall was over 140 mm/h. Hundreds of people were captured within the muds, others were trapped within their cars, many trees fell and thousands of houses were flooded, while few of them collapsed. About three thousand local businesses applied for receiving financial aid from the government. The fire brigade received more than a thousand requests for help. Many areas remained without electricity for days. The floods invoked the Copernicus emergency service in Thessaly (Fig. 1).

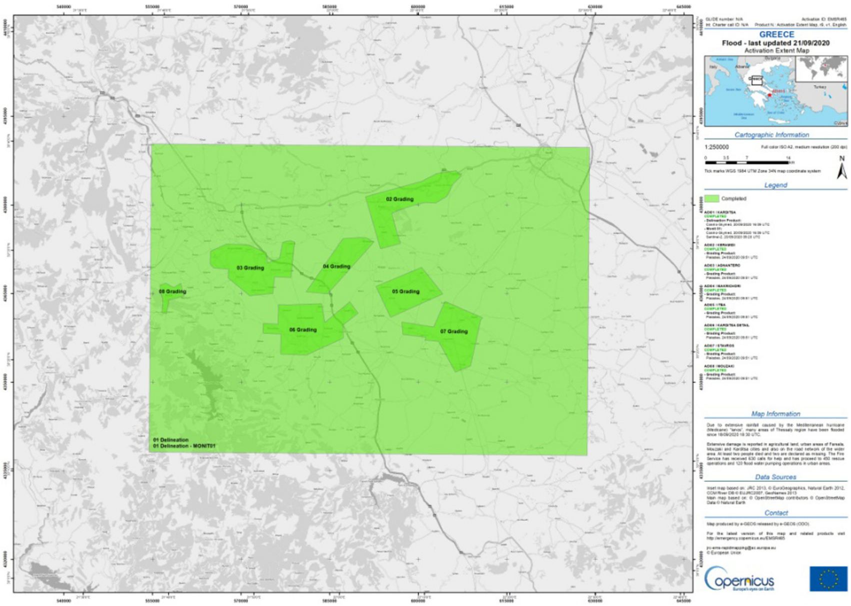

Fig. 1. Rapid mapping activation of the Copernicus emergency service [19].

2 Data and Material Used

An archive of 4649 Instagram photos along with their corresponding descriptive text-information consisted of captions, timestamps and location information (if available) was used. The data were scrapped through the use of a crawler [20].

Data processing was performed in R environment. Main packages used include the "mxnet" for image classification, "rtexttools" for classification of the captions, and "ggmap" along with various other scripts, developed by the author, for georeferencing. QGIS was used for creating the maps. Data manipulation approaches of the presented work can be found at author's github (https://www.github.com/stathisar).

3 Methodology

The main steps of the methodology were divided to the following parts: Initially, the data were crawled while all the metadata for each photo (caption, location, timestamp) were stored in a.json file-format. Sequentially, the json format was con-verted to data-frame and was exported as.csv file, while various pre-processing functions were applied to each text row for reasons of more effective processing.

The second part of the methodology was related to classification. In specific: a classification schema was defined for both image and text data, while the related training datasets, based on data samples were created. In specific, images were initially divided into two basic categories: related to Ianos cyclone and other photos, while the first of the two categories was further sub-classified manually. Regarding text information the categories included: (i) Ianos event identification, (ii) Consequences, quantified to a 5-degree range scale, (iii) Precaution info and (iv) Disaster management: announcements from authorities, measures, volunteered actions. The author suggests Convolutional Neural Networks(CNN) for image and Support Vector Machines (SVM) for text classification, provided through the Mxnet and Rtexttools R packages respectively. Classification especially of photos is at an early stage in current status of the research, limited to the creation of an initial CNN model (https://www.github.com/stathisar).

The next step was related to georeferencing (Fig. 2). The author had presented an innovative method [18] which provides credible results mostly at a municipality level. That method has been advanced to a hybrid approach which employs both geoparsing techniques and commercial geocoding APIs. In specific, all the photos that had location information were geo-referenced through the well known google maps API, while the rest of the text information was processed for detecting geolocations within each caption's text (Fig. 3). That approach increased, in some cases the accuracy; confronting thus with the general known problem of not accurate georeferencing of social media data [3, 17]. As the process is still at a development stage, manual checks and adjustments of the processed info were needed in order to verify the correctness and improve accuracy of the initial results.

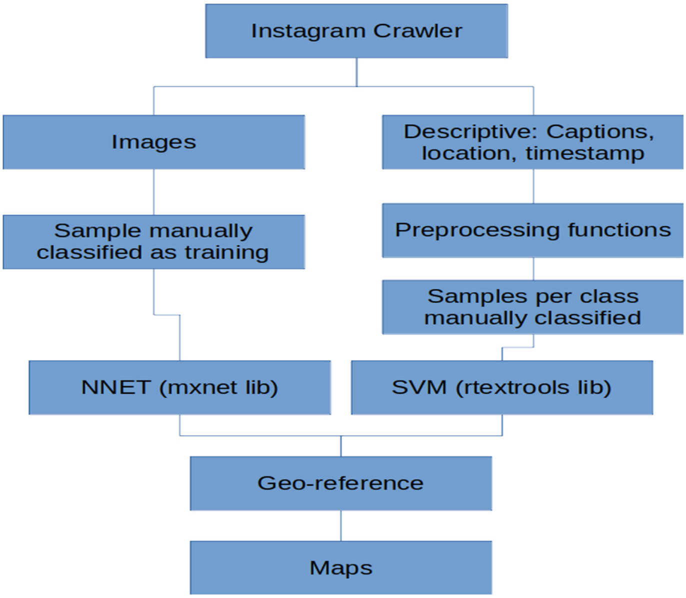

Fig. 2. General schema of the methodology applied

Eventually, the final part of the methodology was related to visualizing the processed output. In current status, the author presents three maps, with information related to: Ianos identification (Fig. 5) in various geographic areas, tracked consequences (Fig. 4), and DM information (Fig. 6). Moreover, a fourth map, displays the geographic positions of three indicative related to Ianos photos (Fig. 7).

Table 1. Basic description of consequence ranged scale used.

Value	Basic description
I	Simple identification, almost zero consequences
II	Big rainfall, measurements of the rainfall without any damages reported
III	Small damages in premises
IV	Big damages, destroyed property and elements of the urban and physical environment. People missing or injured
V	Human loss

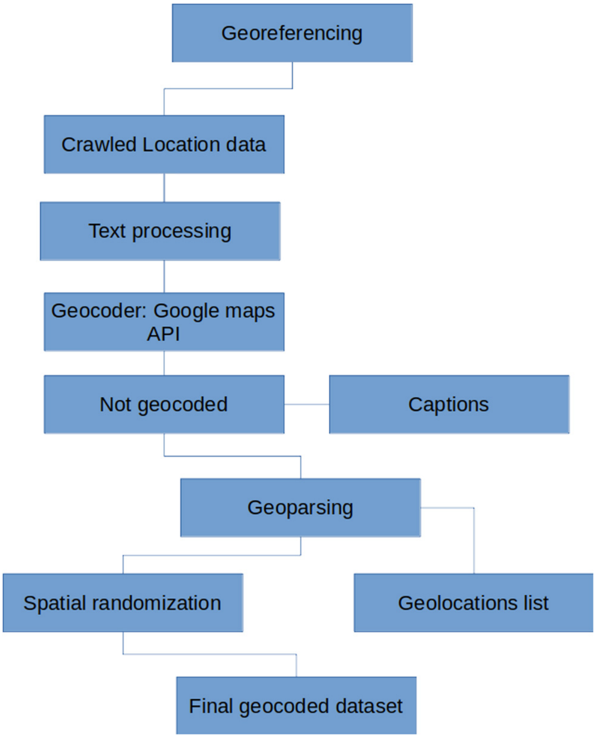

Fig. 3. Hybrid georeferencing method applied

4 Results

Figure 4 displays the location of point observations that represent the consequences, as those reported in Instagram posts. The point values and colors are scaled: The more close to red and bigger each point is the more serious the impact, while the red is related to human loss (Table 1).

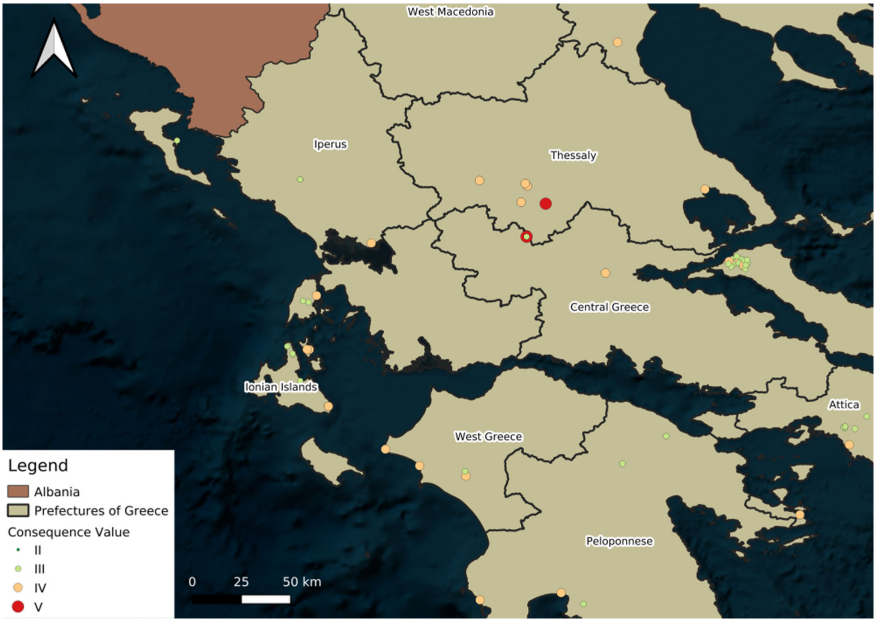

Fig. 4. Point consequence scaled information, extracted though Instagram.

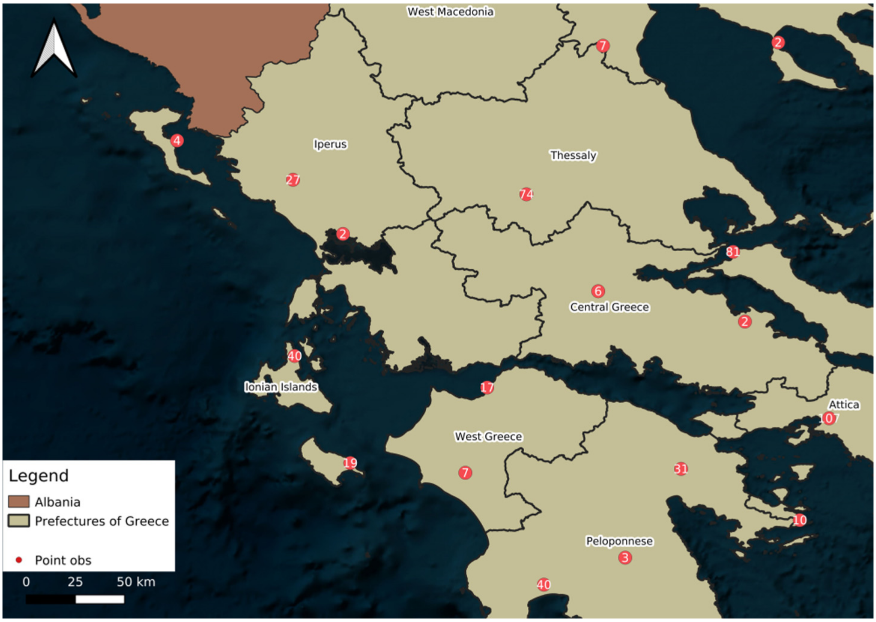

Fig. 5. Ianos identification in various areas: Info extracted through Instagram.

Figure 5 displays clusters of geo-referenced posts regarding the identification of Ianos. Even data processing, in current stage, is still ongoing it can be easily seen that there was information posted about the presence of Ianos, in many areas.

Moreover, the map in Fig. 6 demonstrates point observations that are related to: Precaution and DM: instructions from the authorities, volunteered actions etc.

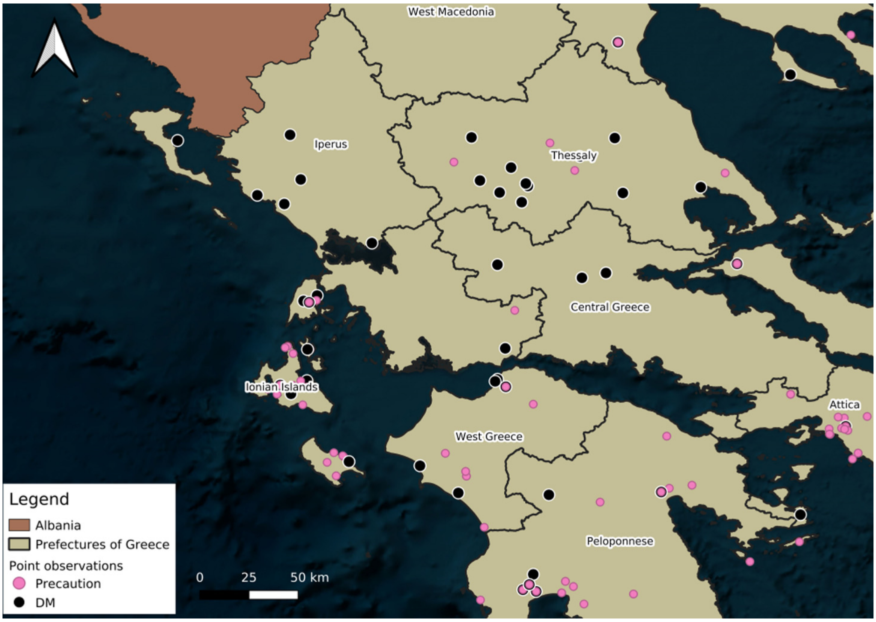

Fig. 6. Point observations of precaution and DM info, extracted through Instagram

Finally Fig. 7 displays the location of three indicative Instagram photos, classified as related to Ianos. Each image's location was mentioned within their corresponding captions.

5 Discussion, Conclusions

In general, automated classification is quite a complex process. There is a wide conversation in international research regarding the capabilities of various machine learning algorithms, while their classification seem to vary according to topic, classification type (binary/multiclass) and data types and properties (i.e. text length) [21–23]. Moreover, specific machine learning algorithms have been emerged as more capable of "generalizing" in a sense of using a general model for all similar classification tasks. Those cases and options need further research in order to evaluate the most optimum algorithms for social network data processing for DM needs.

In addition, initially the author, tried to calibrate the SVM model by using training dataset, created from other natural disastrous event of floods and storms, without any success. The correct classification percentage was less than 40%. By re-feeding the algorithm though, accuracy increased radically. Calibration of the models is still on-going, while the author performed manual corrections in order to have a credible result presented in current status. The initial findings though are promising.

Moreover, even the time needed for training the algorithms and finally perform the classification is, comparatively to manual procedures, significantly less, further research is needed for having models that can operate effectively for similar events.

Regarding georeferencing, even there are promising results by using the hybrid approach there are still quite a few challenges ahead. Initially, there is always the very logical assumption in [24] according to which, the location of a post does not necessarily reflect the post's information. Machine learning for clearing ambiguities could be an effective way, thinking though a real-time system there are a lot of steps towards a desired level of credibility and accuracy. Credibility after all, is vital for the applicability of the methods in DM.

Fig. 7. Images related to Ianos cyclone, posted in Instagram along with their location

The future steps of current research are related to calibrating both image and text classification models and completing the classification automatically. Moreover the optimization of georeferencing is also a mandatory future task. Eventually the general disciplines of author's research are always aligned to generalizing and providing real-time applicability.

References

1. Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. GeoJournal **69**(4), 211–221 (2007)
2. Annis, A., Nardi, F.: Integrating VGI and 2D hydraulic models into a data assimilation framework for real time flood forecasting and mapping. Geo Spat. Inf. Sci. **22**(4), 223 (2019)
3. Arapostathis, S.G.: Fundamentals of volunteered geographic information in disaster management related to floods. In: Flood Impact Mitigation and Resilience Enhancement. IntechOpen (2020)
4. Gorayeb, A., et al.: Volunteered geographic information generates new spatial understandings of covid-19 in Fortaleza. J. Lat. Am. Geogr. **19**(3), 260–271 (2020)
5. Depoux, A., Martin, S., Karafillakis, E., Preet, R., Wilder-Smith, A., Larson, H.: The pandemic of social media panic travels faster than the COVID-19 outbreak (2020)
6. Asghar, M.Z., RahmanUllah, A.B., Khan, A., Ahmad, S., Nawaz, I.U.: Political miner: opinion extraction from user generated political reviews. Sci. Int. (Lahore) **26**(1), 385–389 (2014)
7. Stojanovski, D., Chorbev, I., Dimitrovski, I., Madjarov, G.: Social networks VGI: Twitter sentiment analysis of social hotspots. In: European Handbook of Crowdsourced Geographic Information, p. 223 (2016)
8. Li, Z., Wang, C., Emrich, C.T., Guo, D.: A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. Cartogr. Geogr. Inf. Sci. **45**(2), 97–110 (2018)
9. Arapostathis, S.G.: Tweeting about floods of Messinia (Greece, September 2016) - towards a credible methodology for disaster management purposes. In: Murayama, Y., Velev, D., Zlateva, P. (eds.) ITDRR 2018. IAICT, vol. 550, pp. 142–154. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32169-7_11
10. Feng, Y., Sester, M.: Extraction of pluvial flood relevant volunteered geographic information (VGI) by deep learning from user generated texts and photos. ISPRS Int. J. Geo Inf. **7**(2), 39 (2018)
11. Kankanamge, N., Yigitcanlar, T., Goonetilleke, A., Kamruzzaman, M.: Determining disaster severity through social media analysis: testing the methodology with South East Queensland Flood tweets. Int. J. disaster Risk Reduct. **42**, 101360 (2020)
12. De Longueville, B., Smith, R.S., Luraschi, G.: "OMG, from here, I can see the flames!" A use case of mining location based social networks to acquire spatiotemporal data on forest fires. In: Proceedings of the 2009 International Workshop on Location Based Social Networks, pp. 73–80, November 2009
13. Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J.: #Earthquake: Twitter as a distributed sensor system. Trans. GIS **17**(1), 124–147 (2013)
14. Yang, C., Tian, W.: Social media geo-sensing services for EO missions under sensor web environment: users sensing information about the Ya'an earthquake from Sina Weibo. In: 6th International Conference on Agro-Geoinformatics, pp. 1–6. IEEE, August 2017
15. Feng, Y., Brenner, C., Sester, M.: Flood severity mapping from Volunteered Geographic Information by interpreting water level from images containing people: a case study of Hurricane Harvey. arXiv preprint arXiv:2006.11802 (2020)
16. Gründer-Fahrer, S., Schlaf, A., Wustmann, S.: How social media text analysis can inform disaster management. In: Rehm, G., Declerck, T. (eds.) GSCL 2017. LNCS (LNAI), vol. 10713, pp. 199–207. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73706-5_17
17. de Bruijn, J.A., de Moel, H., Jongman, B., Wagemaker, J., Aerts, J.C.: TAGGS: grouping tweets to improve global geoparsing for disaster response. J. Geovis. Spat. Anal. **2**(1), 2 (2018)

18. Arapostathis, S.G.: Automated methods for effective geo-referencing of tweets related to disaster management. In: Proceedings of GeoMapplica International Conference 2k18, 23–29 June 2018, Syros, Mykonos (2018)
19. Copermicus Emergency Homepage. https://emergency.copernicus.eu/. Accessed 3 Oct 2020
20. Instagram crawler Homepage. Accessed 3 Oct 2020
21. Suliman, A., Nazri, N., Othman, M., Abdul, M., Ku-Mahamud, K.R.: Artificial neural network and support vector machine in flood forecasting: a review. In: Proceedings of the 4th International Conference on Computing and Informatics, ICOCI, pp. 28–30, August 2013
22. Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., Gupta, B.: Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. J. Comput. Sci. **27**, 386–393 (2018)
23. Hernandez-Suarez, A., et al.: Using Twitter data to monitor natural disaster social dynamics: a recurrent neural network approach with word embeddings and kernel density estimation. Sensors **19**(7), 1746 (2019)
24. Huiji, G., Barbier, G.: Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intell. Syst. **26**(3), 1541–1672 (2011)