COMBINING SOCIAL MEDIA AND AUTHORITATIVE DATA FOR CRISIS MAPPING: A CASE STUDY OF A WILDFIRE REACHING CROATIAN CITY OF SPLIT

M. Tavra 1*, I. Racetin 1, J. Peroš 1

1 Dept. of Geodesy and Geoinformatics, Faculty of Civil Engineering, Architecture and Geodesy, University of Split, Split, Croatia - (mtavra, racetin, iperos) @gradst.hr

KEY WORDS: Geosocial data mining, VGI, Social media, Wildfire, Croatia

ABSTRACT:

Due to climate changes, wildfire breakout get more frequent and difficult to control. In the mid-July 2017, the wildfire spread from wildland to the city of Split, the second-largest city in Croatia. This unpredictable spread almost caused the collapse of emergency response systems. Fortunately, a greater tragedy was avoided with the composure of the responsible services and the help of the citizens. The citizens helped in extinguishing the fire and timely provided the significant amount of disaster-related information on different platforms and through social media. In this paper, we address the problem of identifying useful Volunteered Geographic Information (VGI) and georeferenced social media, for improving situation awareness while the wildfire was reaching the Croatian city of Split. Additionally, we combine social media with other external data sources (e.g. Sentinel-2 satellite images) and authoritative data (e.g. Croatian National Protection and Rescue Directorate official data and Public Fire Department of Split data) to establish the geographical relations between the wildfire phenomena and social media messages. In this manner, we seek to leverage the existing knowledge and data about the spatiotemporal characteristics of the Split wildfire in order to improve the identification of useful information from georeferenced social media with other integrated data sources that can be valuable for improving situation awareness in wildfire events.

1. INTRODUCTION

1.1 Background

Emergency response and risk management of natural hazards is getting more and more interaction with social impact through crowdsourcing, in a case of fire and wildfire (Nayebi et al., 2017; Daly, Thom, 2016; Williams, 2013; Becken, Hughly, 2013; De Longueville et al., 2009), earthquakes (Han, Wang 2019; Xu, Nyerges 2017; Hewitt, 2014; Alexander, 2014; Xu et al., 2013) and floods (Chan, 2015; Begg et al., 2015; Tingsanchali, 2012; Merz et al., 2010; Schanze, 2006). Except for purposes of innovative practices based on the disaster type, existing theories and solutions based on different technical backgrounds, data collection from social networks (Ryabchenko et al., 2016; Xu et al., 2015), classification of social media messages (mitigation, preparedness, response, and recovery) (Xiao, Huang, 2015), analytical models from different sources like videos (To et al., 2015), geographical approach to analyse social media to indicate message usefulness (De Albuquerque et al., 2015), real-time data mining tools (Zhu et al., 2019; Zhong et al., 2016) or predictions based on Twitter events that belong geographical analysis of spatio-temporal Big Data (Shi et al., 2016).

Castillo (2016) pointed out that immediacy is key to the relevance of social media information. People on the ground collect and share information before mainstream media or disaster management systems can even react. From a plethora of general and specific emergency management theories and services, the specific field related to the crowdsourcing data and their application in wildfire response and rescue systems emerged. For example, Oliveira et al. (2017) presented a fire alert service FDWithoutFire, which improved Forest Fire emergency system with crowdsourcing data. Villela et al. (2018) used crowdsourcing as a ground map for a decision support system for emergency and crisis management called RESCUEr. They used mobile crowdsourcing data to recognise and respond to an incident in the industrial area. According to Castillo (2016), there are several emergency management systems which incorporate different sources of data, and some of them are crowdsourced or social media. SaferCity (Berlingerio et al., 2013 to Castillo, 2016) integrate social media and news. STED (Hua et al., 2013 to Castillo, 2016) uses traditional news media over social media messages. LITMUS (Musaev et al., 2014 to Castillo, 2016) create alerts about landslides using information collected from social networks and official data from the U.S. Geological Survey (USGS) and rainfall data from NASA’s Tropical Rainfall Measuring Mission (TRMM).

1.2 Motivation and study area

The motivation for this research is a frequent occurrence of wildfire breakout in Croatia during the summer days. Due to the high summer heats, strong winds and human factors, as in countries with similar climate (Spain, Portugal, Greece, Italy), Croatia is under constant fire risk. The Dalmatian fire brigade unit’s records a minimum of ten interventions per day in the summer months (EU, 2018). For this research, we have selected a wildfire incident that happened in July of 2017 in the outskirts of the city Split as the case study (Figure 1). In this event, wildfire, driven by heavy wind, reached several populated places and city of Split suburbs as well as residential districts in a short period. During and after the disaster, many citizens wanted to help and to know what exactly had happened, so it was our motivation to use authoritative data of this event and merge it with available sources of crowdsourcing data. The focus of this study is data collection through crowdsourcing from a variety of sources and a different context.

* Corresponding author
1.3 Case study and results

In the post-event phase, social media can provide and receive information about disaster response and recovery (Houston et al., 2015). For this case study, we reconstruct spatiotemporal social media and other relevant data for period of 24h from the start of a wildfire. In this research, we identify sources of data from social media (Twitter, Facebook), news portals and merge it with other external data sources to develop capabilities for emergency response based on social media information. Therefore, we developed the methodology workflow for aggregating data from different sources and the procedure for data mining based on the existing knowledge. In comparison to other studies, our approach integrates several sources of data, including the theoretical background. There are identified components between multiple sources in time and after the disaster to help to map disaster with spatiotemporal data. The presented results could help to develop new capabilities for emergency response based on combining social media information to improve efficiency and analysis for disaster information extraction.

2. METHODOLOGY

2.1 Study area

City of Split is situated mainly on a peninsula, encircled by the Kozjak and Mosor mountains (Figure 2). The Split is the second largest city in Croatia (The Croatian Bureau of Statistics, 2018) with more than 200,000 citizens (including surrounding settlements). During the summer touristic season, the city population increases for more than 10%. The rapid city development in the second part of the 20th century resulted in a massive build-up without proper urban planning. This problem is particularly emphasized in the suburban area and negatively impacts the quality of life. Life quality problems mainly arise from inadequate infrastructure: too narrow roads, water shortages and unsolved sewerage. The issue of over build-up of can also cause lower safety of citizens, for example, evacuation in the case of emergencies would be exceptionally complicated due to the insufficient road width.

In the time while the fire was raging, there was no electricity in most of the city districts, and the situation was on the edge of evacuation. For that night, there was a Facebook Safety Check – Crisis response as social network disaster response. Also, there were many tweets about the fire on the Twitter social network. Additionally, the threat of great catastrophe emerged as the fire spread to the landfill of Karepovac (around 20 ha) situated at the city border (Figure 3).

Figure 1. Fire in the Split surrounding (source/author: Damira Kalajdzic)

Figure 2. Study area - City of Split

A Mediterranean climate characterises the area with hot, dry summers and mild, wet winters. The mean annual precipitation (1971-2000) is 782.8 mm, mostly in the period from October to April and the average annual air temperature is 16.1 °C. In the July monthly extremes (1971-2000) of precipitation and temperature occur, the lowest mean precipitation of 25.5 mm and the highest monthly air temperature of 25.7 °C (Croatian Meteorological and Hydrological Service, 2018). Uncultivated neighbouring rural and mountain areas are covered by shrub lands and forests. The Aleppo pine (Pinus halepensis) is the most common species present there, while the significantly smaller area is covered by black pine (Pinus nigra) and pubescent oak (Quercus pubescens). These areas are foremost fire danger areas as they provide easily flammable fuel, particularly during the dry season. Long-lasting drought and heavy winds which started blowing in mid-July created ideal preconditions for the fast and horrifying spread of wildfire. The terrain itself was often unreachable for the firefighters because of the local relief characteristics and clogged access roads and fire trails.

In the time while the fire was raging, there was no electricity in most of the city districts, and the situation was on the edge of evacuation. For that night, there was a Facebook Safety Check – Crisis response as social network disaster response. Also, there were many tweets about the fire on the Twitter social network. Additionally, the threat of great catastrophe emerged as the fire spread to the landfill of Karepovac (around 20 ha) situated at the city border (Figure 3).

Figure 3. Landfill of Karepovac after fire (source: SD)
The threat lasted for two days (Figure 3), and in one moment, there were not enough firefighters and Canadairs could not fly because of the strong winds. At that moment, citizens organised and started to defend their houses. They improvised and combated the fire with the watering hoses, washbowls and branches.

2.2 Identifying sources for data collection

This study is mainly based on data that is collected by crowdsourcing that refers to the voluntary participation of citizens and organisations in the process of data collection, to an already prepared platform called Crowdmap. Because it is an organised event (with explicit instructions to volunteers) that part of collecting data refer to an activity of volunteered geographic information (VGI). Other crowdsourcing data is collected using some of the apps that approve the use of location and position information (by social networks platform) for gathering data from the crowd, for example, on mobile phones. In this way, citizens become a kind of sensors (Goodchild, 2007).

Data was collected from several different sources with different context:
- Crowdsourcing;
- Collecting data from geosocial networks;
- Extracting data from the news;
- Croatian National Protection and Rescue Directorate (NPRD) official data;
- Satellite images (Sentinel-2);
- Public Fire Department of Split (PFDS) data.

Crowdsourcing data is collected through the Crowdmap (Herbert, Crowdmap Team, 2017) platform. Crowdmap is a free and open source tool based on Ushahidi. Ushahidi is a tool or a concept that is developed by Kenyan civil activists in 2008 to track and prevent ethical clashes using the geographic data (Mäkinen et al., 2008). Data from geosocial networks (Twitter and Facebook) helped with filling the gap between other data sets.

On the other side, we used multispectral satellite data (Copernicus Sentinel-2), field observations, Natural protection and rescue directorate (NPRD) polygon of burned areas for quality assurance (accuracy assessment) of collected data.

The important source of data was data from Public Fire Department of Split (PFDS) call centre, over 4000 calls of citizens were interpreted, and one part of them was possible to geocode.

2.3 Data mining

After we identified relevant sources of data to integrate presented by this proposal, we began with the data mining of collected data.

As the event happened on micro-locations, that did not involve many people, and everything went relatively fast (around 24h), we did most of the data mining work manually. Another advantage in manual data mining is that we can monitor the information relevance and based on that create methodology procedure later.

Crowdsourcing data from the Crowdmap was later exported in tabular format. Although users can enter the location and time of the event on Crowdmap application, some of the posts didn’t have that information. The users described the sites and typed the timestamps in the description box. So, we have manually geocoded that kind of entries and placed it in the right timeline. Geosocial network data was mostly manually mined with the little help of the Octoparse software. In this case, the Internet and social platforms also helped to mobilise volunteers for action, and later for collecting data. Octoparse software was used to point out interesting posts and threads. The first step was to find the right keywords and hashtags (Murzinteic, Cheng, 2017), on which we used Google tools to show the popularity of different keywords on the Croatian language that, are related to theme wildfire. The second group of keywords was related to the location, city of Split, nearby places and Croatia. The third group of keywords was related to warnings and dangerous situation. Irrelevant data to this event we manually removed from data collection.

NPRD official data was already georeferenced vector polygon data and didn’t need to be specially arranged. Same as Satellite images from Copernicus Sentinel-2 that were used calculated to accentuate the burned area.

PFDS deliver raw data in the tabular format that contained three columns: timestamp, telephone number and description of the call. In the 4000 call records from citizens to fire department call centre there was some official communication between firefighting units and police due to lack of availability of the communication equipment. Records were manually selected to ones that have location description and those that were called from the fix telephone lines that were related with the address. The result was about 100 records that were geocoded.

2.4 Data processing

We used multispectral satellite data (Copernicus Sentinel-2), field observations, Natural protection and rescue directorate (NPRD) polygon of burned areas for quality assurance (accuracy assessment) of collected data.

Copernicus Sentinel-2 images from 18 May 2017 and 6 August 2018 were selected and differenced Normalized Burn Ratio (Garcia, Caselles, 1991) was calculated to accentuate the burned area (Figure 4).

The most important source of data was data from Public Fire Department of Split (PFDS) call centre, over 4000 calls of citizens were interpreted (Figure 5), and only small part of them was possible to geocode.
3. RESULTS AND DISCUSSION

3.1 Results

On the workflow of Split wildfire crowdsourcing (Figure 6) a methodology for aggregating data from different sources is presented. The methodology was derived by conducting a small search of our own data and following the process of using a different approach. Base data framework is formed from several data sources: crowdsourcing data, data from geosocial networks and PFDS data. These sources of data conducted into the design of geodatabase.

Once data were combined and prepared for visual interpretation and validation, it was loaded into ArcGIS software for further processing. After enabling time and preparing time format we created the wildfire spreading map – raw version for visual interpretation. Also, at this validation phase, we left the possibility to add different layers as meteorological data, data about firefighters or even types of vegetation to make the future map more informative. Data quality assurance is made by multispectral satellite data (Copernicus Sentinel-2), field observations to filling gaps, and NPRD polygon of burned areas. These data sets have the capacity to tell us about the processes of wildfires in different places. Change in fire patterns is still being studied by different scale analyses of the data (Chas-Amil et al., 2015, Huan et al., 2012), but there are several geographical relations as will be indicated by the observations of this study.

Based on the presented retrospective data analysis, we suggest procedure and methodology for crowdsourcing data integration. Procedure is based on activities which we have taken to reconstruct an event by combining crowdsourcing data. So, we choose the inverted pyramid shape, depicting the invested effort, to describe this approach in an effective way (Figure 7).

![Figure 6. Methodology workflow for aggregating data from different sources](image)

**Figure 5. Overview of calls to Public Fire Department call centre**

Figure 5 shows an overview of all calls to PFDS call centre by a number of calls and call duration. There is a visible peak of wildfire around 18h. The peak of calls has a good overlap with geo-coded spatial-temporal geosocial data because, at that moment, the fire was the most turbulent and closer to the urban and suburban area. This also suggests that people call because increased risk during the night period.

**Figure 7. Activity pyramid of combining crowdsourcing data**

The procedure consists of five phases:

1. At first, sources with relevant data is identified. In this phase, Crowdmap campaign is created and promoted.
2. In the second phase, raw data is collected from identified sources and requests for official data are sent.
3. The third phase consists of data mining and data pre-processing from combined sources.
4. In four-phase, the data is processed.
5. At last, information is prepared and analysed, also validated through design of test maps and others visualisation.

3.2 Discussion

From a disaster management perspective, key issues are the unpredictability of human behaviour and prediction of disaster accompanying hazards. The first problem can be solved by providing precise and timely information to the citizens. In the case of a natural disaster, citizens can be informed in three phases: before, during and after the event (Houston et al., 2015). The second problem can be solved by developing better technical response systems based on a theoretical framework that is often developed on the reconstruction of past events.

This kind of analysis helps us to more understand cause and the event flow. As many studies pointed out, there is importance for VGI volunteer’s education as well as the need for increased motivation (Fritz et al., 2017, Mooney et al., 2016, Fonte et al., 2015, Sui et al., 2012). Educated volunteers provide and collect more relevant data, even participate in the data quality process.

4. CONCLUSIONS

This research presents an innovative approach to data collection about the disaster shown in the example of the wildfire event in
Split in 2017. By collecting data from various sources that occurred during the event, we developed an approach for better disaster management in case of a real-time catastrophe. This paper focuses on the integration of different data sources after a disaster. We combine the data from the identified sources that were available at the time after the disaster accident to reconstruct the event. Based on the presented retrospective data analysis, we suggest procedure and methodology for integration. This research opens new horizons to organisations, whose main activity is fire protection. The achievements shown in this article, generally can be applied to other disasters management organisations. This article highlights the importance of using geoinformations from geosocial networks, providing a different perspective on disaster management through the formulation of data combined from multiple sources.

4.1 Future work

Next phase of this research is to find and visualise geographical relations between the wildfire phenomena with combined data sources as we identified as issues. It will be addressed in future work such as a better geographical representation of cartographic elements from crowdsourced spatiotemporal data. This approach can also help developing disaster response and analytic system that collect spatial data from numerous crowdsourcing data sources.

ACKNOWLEDGEMENTS

We would like to thank all the people and organisations that help us with collecting data: Public Fire Department of Split, National Protection and Rescue Directorate, Croatian Mountain Rescue Service, Tomislav Mamić, Jelena Kilić, Our dear college Marina Vučemilović-Vranjić who had help us with data processing. And brave Damira Kalajdžić who took photos on the site of fire. We would like to thank ESRI company for software support which led us to uninterrupted work.

REFERENCES


Castillo, C., 2016: Big crisis data: social media in disasters and time-critical situations. Cambridge University Press.


Daly, S., Thom, J. A., 2016: Mining and Classifying Image Posts on Social Media to Analyse Fires. In ISCRAM.


De Longueville, B., Smith, R. S., Luraschi, G., 2009: Omg, from here, i can see the flames: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. Proceedings of the 2009 international workshop on location based social networks (73-80). ACM. doi.org/10.1145/1629890.1629907


Han, X., Wang, J., 2019: Earthquake Information Extraction and Comparison from Different Sources Based on Web Text. ISPRS International Journal of Geo-Information, 8(6), 252. doi.org/10.3390/ijgi8060252


To, H., Kim, S. H., Shahabi, C., 2015: Effectively crowdsourcing the acquisition and analysis of visual data for disaster response. In *Big Data (Big Data), 2015 IEEE International Conference*, 697-706. doi.org/10.1109/bigdata.2015.7363814


Zhong, X., Duckham, M., Chong, D., Tolhurst, K., 2016: Real-time estimation of wildfire perimeters from curated crowdsourcing. *Scientific reports*, 6, 24206. doi.org/10.1038/srep24206