A Web Application to Map Disaster Impacts from Text

Lydia Price [1], Kristin Stock [1], Sophie Francis [1] and Kalana Wijegunarathna[1]

[1] *Massey Geoinformatics Collaboratory, Massey University, Auckland, 0745, New Zealand*

Abstract
Social media allows users to connect and share information, news, and content with a large network. Following a natural disaster, people use social media to share information relevant to recent disasters, often faster than official channels. This information is extremely valuable for emergency responders to plan and implement relief operations. While web mapping applications and satellite images are commonly used to visualize disaster impacts, they either lack specific information about individual impacts or require manual intervention. Unlike previous work, we are developing an application to map fine-grained, individual impacts that are automatically extracted from tweets, along with an accurate and informative open-source mapping interface. This paper discusses the development of this web mapping application that displays accurate disaster impact details from social media posts.

Keywords
Disaster impacts, natural language processing, web mapping application

1. Introduction

Social media enables users to connect and share information, news, and content with a large network of people. It contains a wealth of knowledge, being updated constantly in real-time. Following a natural disaster, people turn to social media to share information regarding the damage and injuries sustained. The information provided on social media is often updated faster and more often than official channels [1]. For example, even though seismologists monitor earthquakes and provide information about them, this alone is insufficient for an effective and timely response because it lacks detailed information regarding the specifics of the earthquake impacts. Emergency responders would greatly benefit from this information in their efforts to plan and implement relief operations more effectively [2].

In this paper, we present a web mapping application that displays disaster impact details from posts extracted from social media and plots them by place name. Unlike previous work, we are developing an application to map fine-grained, individual impacts that are automatically extracted from tweets, along with an accurate and informative open-source mapping interface. We present a description of the method we are using to extract the information from the tweets and display them as well as the application's functionality. We will also outline our future plans for advancing this application, as this paper describes work in progress.

2. Related Work

Many past research works have used classification techniques to identify useful content from social media following disaster events. A range of automated methods following this approach have been developed including: filtering messages through type and categories [3], informativeness [2], using search functions with keywords, spatial boundaries, or dates [4] and the relevance of the post [5]. However, these methods are often only able to extract very course-grained content (for example, they may classify posts into categories such as damage to utilities, casualties, caution and advice, donations, etc.) [6], [7], and are not able to explicitly identify the key information needed by disaster responders. Work that has addressed the task of extracting explicit, fine-grained disaster impacts includes [8], which uses an ontology and rules-based approach to extract specific items of information, but does not evaluate the performance against ground truth data, and [3], which uses Part of Speech (POS) Tagging to define sequences of parts of speech that are expected to contain impact information, achieving variable results with a small data set. Another research direction combines multiple data sources to describe a single event, using clusters to reduce the risk of noise or unreliable information [9], [10]. However, this means that many individual reports of impacts at the detailed level are missed. Our project will go beyond previous research by applying state of the art deep learning methods to extract fine-grained, individual impact information to gather and map crucial and pertinent information that can be utilized by emergency responders in real-life situations.

Our second aim is to combine place names and spatial indicators (terms that pinpoint location relative to a place name) to precisely determine the area affected by a disaster event. For example, in the tweet "*new island erupts off the coast of #pakistan*", the phrase "*off the coast of*" gives a lot more information about the location of the new island than the place name Pakistan alone. Currently, many methods use only the place names from the text or determine the location based on the coordinates of the post. Both methods can be unreliable as georeferencing based on place names can result in a broad area, not the specific location of the impact, the location where the post was made may not be accurate, and the post may not be made at the same location as the event described [11]. Making use of location and spatial indicators, if present, can narrow down the location and provide more accurate information for emergency responders. We will extract terms from complex location descriptions to map disaster impacts more accurately than has been done previously.

Web mapping applications for disaster impacts are not new, with recent examples of mapping interfaces utilizing posts from social media include mapping geographic distribution of events using Named Entity Recognition (NER) and the Google Maps Geocoding API [12], location and probability of a tweet being relevant [13] and mapping the location solely based on names mentioned within a tweet [14]. Crowd sourcing mapping applications for disaster mapping, including Ushahidi and Sahana Eden, are widely recognized as valuable crisis response tools, providing access to both posts made directly to the platform, as well as social media analytics. While these and many other tools display social media content in maps and other formats using methods such as sentiment analysis, trend/keyword analysis and social network analysis, they do not extract specific information about individual impacts [15], [16]. Our plans are to make the information plotted more informative and accurate compared with current methods. Furthermore, the interfaces do not compare with what we plan to develop - see future work for more information.

Satellite images and drones are also commonly used to visualize what is happening on the ground during a disaster. Although they can map disaster impacts, they do not automatically highlight impacts, so often require manual effort to be effective when responding to disasters. Machine learning methods are being developed to automate these methods to gather available information in near-real-time [17], however these are used mainly for the prevention of hazards and damage, and not emergency response in the event of a disaster.

3. Method

We extracted data from original tweets using Named Entity Recognition (NER) with several BERT-based models, identifying the following entities: type of impact, object affected, place names, quantity or severity and location modifier[18]. We then applied relation extraction to identify the correct syntactic links between the entities [19]. The entities data was provided in a csv format, and we then developed a program to extract relevant information and plot this on a map. Figure 1 illustrates the program pipeline.

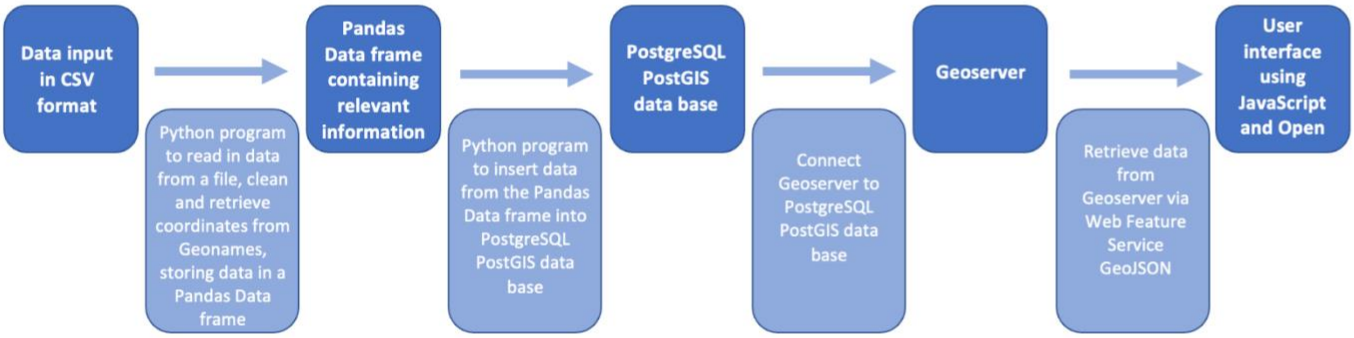

Figure 1: Pipeline

3.1. Reading in Data

In this step, comparisons were done between different csv readers. Using the csv reader from the Pandas library [20], the data was read by a python program. We extracted the 4 key columns to be used with the mapping interface: label, instance, tweetId and tweetText and created a Pandas Data Frame[20]. The label specifies the type of instance, for example, place name, location modifier and type of impact. The tweetID is the ID of the Tweet, and tweetText is simply the text contained by the tweet, for example, "*Napa 6.0 Earthquake Caused 'Significant' Damage in County: USGS: Cleanup efforts were underway in the Bay Area...*". The program also scans a directory containing the data files, each file being marked when read, to cater for a situation in which large volumes of new tweets are being processed quickly in a disaster event, with entities marked and created as new csv files in batches (e.g. every 30 seconds) for input into the mapping application.

3.2. Retrieving Coordinates

In this step, we determine coordinates for place names that were extracted from the csv file. This was done by connecting to GeoNames[21] using GeoCoder [22], and passing each place name label to GeoNames [21]. If coordinates for the place name were found, this data was added to the Pandas [20] data frame. The coordinates are converted into point geometries and stored in a GeoPandas[23] data frame. At present, the coordinates are obtained from the data provided by the first place name that the Geocoder[22] finds. We will further improve this in the future, by incorporating toponym resolution methods such as [24].

3.3. Database and Server

In this step, the data is inserted into a PostGIS database and then a connection is established between the database and Geoserver [5]. This is done on the assumption that the data files do not contain duplicate information. Geoserver[25] connects to the PostGIS database and creates a GeoJSON file to be used with Openlayers [26].

3.4. Mapping Interface

In this step, the Openlayers [6] library connects from the JavaScript application to Geoserver [5] which retrieves the GeoJSON data and creates a vector layer on the map. After comparison between Openlayers[26] and Leaflet, both popular reliable mapping libraries, we chose Openlayers [26]. Openlayers offers more functionality than Leaflet which is mainly used solely for developing a map, not a GIS application [27] .

Figure 2: User Interface

4. Mapping Interface

Figure 3: Popup with multiple tweets

From the data obtained from the machine learning process, the cleaned data is plotted on the world map as displayed in Figure 2. The interface is simple and easy to use. For each placename identified in earlier stages, the location, coordinates, and the tweet contents are displayed within a popup. If there are multiple tweets with the same location and coordinates, then they are all displayed within the same popup with a scroll feature displayed in Figure 3.

5. Conclusions and Future Work

Social media allows users to communicate and share information with a large network of people, and it is a source of relevant knowledge, which is updated constantly in real-time. Map visualization of spatially accurate and specific information related to different impacts will enable disaster responders to direct their response more effectively than is currently possible.

As a next step, we will refine the geolocations of the tweets on the map by incorporating the location modifiers by creating models for specific spatial terms using a similar approach to [28], but applying the latest deep learning methods. Such modifiers describe location more accurately than the place name alone, and their inclusion will improve the accuracy of the points depicted on the map. To make it as easy as possible for disaster responders, we intend to display different impacts with different icons. We will also consider alternative presentations to the current pop ups for locations in which multiple tweets are clustered.

As part of our efforts to ensure that communication between responders remain effective, we are also considering adding the ability to add comments to tweets on the interface. Adding temporality to the application will also provide the ability to view selected data on the mapping interface within specific time frames, to remove confusion as impacts change and are resolved during live disaster events.

6. References

[1] Z. Xing *et al.*, “Crowdsourced social media and mobile phone signaling data for disaster impact assessment: A case study of the 8.8 Jiuzhaigou earthquake,” *Int. J. Disaster Risk Reduct.*, vol. 58, p. 102200, May 2021, doi: 10.1016/j.ijdrr.2021.102200.

[2] F. Alam, F. Ofli, M. Imran, T. Alam, and U. Qazi, “Deep Learning Benchmarks and Datasets for Social Media Image Classification for Disaster Response,” in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Dec. 2020, pp. 151–158. doi: 10.1109/ASONAM49781.2020.9381294.

[3] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, “Extracting Information Nuggets from Disaster- Related Messages in Social Media,” p. 10, 2013.

[4] M.-H. Tsou *et al.*, “Building a Real-Time Geo-Targeted Event Observation (Geo) Viewer for Disaster Management and Situation Awareness,” in *Advances in Cartography and GIScience*, Cham, 2017, pp. 85–98. doi: 10.1007/978-3-319-57336-6_7.

[5] A. Olteanu, S. Vieweg, and C. Castillo, “What to Expect When the Unexpected Happens: Social Media Communications Across Crises,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA, Feb. 2015, pp. 994–1009. doi: 10.1145/2675133.2675242.

[6] F. Alam, F. Ofli, M. Imran, T. Alam, and U. Qazi, “Deep Learning Benchmarks and Datasets for Social Media Image Classification for Disaster Response,” in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Dec. 2020, pp. 151–158. doi: 10.1109/ASONAM49781.2020.9381294.

[7] P. K. Roy, A. Kumar, J. P. Singh, Y. K. Dwivedi, N. P. Rana, and R. Raman, “Disaster related social media content processing for sustainable cities,” *Sustain. Cities Soc.*, p. 103363, Sep. 2021, doi: 10.1016/j.scs.2021.103363.

[8] S. Traverso, V. Cerutti, K. Stock, and M. Jackson, “EDIT: A Methodology for the Treatment of Non-authoritative Data in the Reconstruction of Disaster Scenarios,” in *Information Systems for Crisis Response and Management in Mediterranean Countries*, Cham, 2014, pp. 32–45. doi: 10.1007/978-3-319-11818-5_4.

[9] F. Hamborg, C. Breitinger, and B. Gipp, “Giveme5W1H: A Universal System for Extracting Main Events from News Articles,” *ArXiv190902766 Cs*, Sep. 2019, Accessed: Jan. 20, 2021. [Online]. Available: http://arxiv.org/abs/1909.02766

[10] N. Algiriyage, R. Prasanna, K. Stock, E. E. H. Doyle, and D. Johnston, “DEES: a real-time system for event extraction from disaster-related web text,” *Soc. Netw. Anal. Min.*, vol. 13, no. 1, p. 6, Dec. 2022, doi: 10.1007/s13278-022-01007-2.

[11] K. Stock, “Mining location from social media: A systematic review,” *Comput. Environ. Urban Syst.*, vol. 71, pp. 209–240, Sep. 2018, doi: 10.1016/j.compenvurbsys.2018.05.007.

[12] C. Fan, F. Wu, and A. Mostafavi, “A Hybrid Machine Learning Pipeline for Automated Mapping of Events and Locations From Social Media in Disasters,” *IEEE Access*, vol. 8, pp. 10478–10490, 2020, doi: 10.1109/ACCESS.2020.2965550.

[13] Kulranjan and S. Ojha, “Use of Social Media/Microblogging Platform along with GIS for Mapping Disaster,” in *2019 International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Nov. 2019, pp. 1–4. doi: 10.1109/UPCON47278.2019.8980246.

[14] B. Anbalagan and C. Valliyammai, “#ChennaiFloods: Leveraging Human and Machine Learning for Crisis Mapping during Disasters Using Social Media,” in *2016 IEEE 23rd International Conference on High Performance Computing Workshops (HiPCW)*, Dec. 2016, pp. 50–59. doi: 10.1109/HiPCW.2016.016.

[15] H. Gao, G. Barbier, and R. Goolsby, “Harnessing the Crowdsourcing Power of Social Media for Disaster Relief,” *IEEE Intell. Syst.*, vol. 26, no. 3, pp. 10–14, May 2011, doi: 10.1109/MIS.2011.52.

[16] L. Ngamassi, A. Malik, J. Zhang, and D. S. Ebert, “Social Media Visual Analytic Toolkits for Disaster Management: A Review of the Literature.,” *ISCRAM*, 2017.

[17] G. Antzoulatos *et al.*, “Flood Hazard and Risk Mapping by Applying an Explainable Machine Learning Framework Using Satellite Imagery and GIS Data,” *Sustainability*, vol. 14, no. 6, Art. no. 6, Jan. 2022, doi: 10.3390/su14063251.

[18] S. Francis, K. Stock, and S. Hameed, "Annotating and Extracting Disaster Impacts from Social Media with Named Entity Recognition. Poster presented at ISCRAM Asia Pacific, 7-9 November, Melbourne."
[19] K. Wijegunarathna, K. Stock, S. Francis, C. B. Jones, R. Prasanna, and E. Hudson-Doyle, "Relation extraction to identify locations of disaster impacts from social media text. Poster presented at ISCRAM Asia Pacific, 7-9 November, Melbourne."
[20] pandas - Python Data Analysis Library. URL: https://pandas.pydata.org/
[21] GeoNames. URL: https://www.geonames.org/
[22] Geocoder 1.38.1 documentation. URL: https://geocoder.readthedocs.io/ (accessed Feb. 06, 2023).
[23] Documentation — GeoPandas 0.12.2. URL: https://geopandas.org/en/stable/docs.html (accessed Feb. 07, 2023).
[24] M. Karimzadeh, S. Pezanowski, A. M. MacEachren, and J. O. Wallgrün, "GeoTxt: A scalable geoparsing system for unstructured text geolocation," *Trans. GIS*, vol. 23, no. 1, pp. 118–136, 2019, doi: 10.1111/tgis.12510.
[25] GeoServer Documentation. URL: https://docs.geoserver.org/ (accessed Feb. 06, 2023).
[26] OpenLayers - Welcome. URL: https://openlayers.org/ (accessed Feb. 06, 2023).
[27] "Leaflet vs OpenLayers. Pros and cons of both libraries," *Geoapify*, Apr. 12, 2019. https://www.geoapify.com/leaflet-vs-openlayers/ (accessed Feb. 06, 2023).
[28] R. Liao, P. P. Das, C. B. Jones, N. Aflaki, and K. Stock, "Predicting Distance and Direction from Text Locality Descriptions for Biological Specimen Collections," in *15th International Conference on Spatial Information Theory (COSIT 2022)*, Dagstuhl, Germany, 2022, vol. 240, p. 4:1-4:15. doi: 10.4230/LIPIcs.COSIT.2022.4.