CROWD4EMS: A CROWDSOURCING PLATFORM FOR GATHERING AND GEOLOCATING SOCIAL MEDIA CONTENT IN DISASTER RESPONSE

A. Ravi Shankar[1],*, J.L. Fernandez-Marquez[1], B. Pernici[2], G. Scalia[2], M.R. Mondardini[3], G. Serugendo[1]

[1] Centre Universitaire d'Informatics, University of Geneva, Switzerland
(amudha.ravishankar, joseluis.fernandez, giovanna.dimarzo)@unige.ch
[2] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
(barbara.pernici, gabriele.scalia)@polimi.it
[3] Citizen Science Centre, ETH - UZH, Zurich, Switzerland
maria.mondardini@uzh.ch

Commission VI, WG VI/4

KEY WORDS: Crowdsourcing, Social Media, Machine Learning Geolocation, Disaster Response

ABSTRACT:

Increase in access to mobile phone devices and social media networks has changed the way people report and respond to disasters. Community-driven initiatives such as Stand By Task Force (SBTF) or GISCorps have shown great potential by crowdsourcing the acquisition, analysis, and geolocation of social media data for disaster responders. These initiatives face two main challenges: (1) most of social media content such as photos and videos are not geolocated, thus preventing the information to be used by emergency responders, and (2) they lack tools to manage volunteers contributions and aggregate them in order to ensure high quality and reliable results. This paper illustrates the use of a crowdsourcing platform that combines automatic methods for gathering information from social media and crowdsourcing techniques, in order to manage and aggregate volunteers contributions. High precision geolocation is achieved by combining data mining techniques for estimating the location of photos and videos from social media, and crowdsourcing for the validation and/or improvement of the estimated location. The evaluation of the proposed approach is carried out using data related to the Amatrice Earthquake in 2016, coming from Flickr, Twitter and Youtube. A common data set is analyzed and geolocated by both the volunteers using the proposed platform and a group of experts. Data quality and data reliability is assessed by comparing volunteers versus experts results. Final results are shown in a web map service providing a global view of the information social media provided about the Amatrice Earthquake event.

1. INTRODUCTION

The second generation of the world wide web, Web2.0 or the *the participatory web*, as they call it, has changed the way people report, respond or share information across the world. People have started using the web not just as a source of information, but also as a platform where they share, create and contribute (Blank, Reisdorf, 2012). The increase in social media usage and the giant leap in the technological advancement has facilitated the speed, quantity and quality of first hand information attained from the ground. Collective wisdom has made its strong presence not just in scientific and humanitarian domain but also in our day to day activities . In 2010, Haiti earthquake showed the world how collective effort materialised in the form of a *Ushahidi*[1] crisis map, mapping the severity of the event and damages caused, crowdsourced from SMS reporting and Social media Posts. In the context of disaster response, Crowdsourced information, combined with technical data collected about a given emergency, can be used to improve situational awareness, decrease response time, and assess the severity of the damage. Crowdsourcing can boost preparedness for the times when we least expect to need it.

Increase in crowdsourcing demanded exclusive communities or focus groups that contributed in different aspects of crowdsourcing, by pooling the social media contributions of the many to pinpoint trouble areas, to gather critical information, and perform time-sensitive tasks and other such micro-tasks. Digital Humanitarian Network[2], Stand By Task Force[3], Virtual Operations Support Team[4], Crisis Mappers[5], Humanitarian Open Street Map Team[6], Crowd rescue[7] and GISCorps[8] are few of the well known crowdsourcing communities, who have contributed a great deal during various disasters. While crowdsourcing is an organisational framework and corresponds to processes for procuring services from a large amount of people external to an organisation, for example communities like SBTF *Human based computation* is an information processing framework and corresponds to methods for incorporating human intelligence into an information processing system, for example, volunteer annotation of large sets of images captured during crisis (Castillo, 2016). Such emergent communities spend several man-hours, often with no or limited tools.Over the decade, Crowdsourcing has given raise to platforms or web based tools,like Zooniverse[9] and Humanitarian Openstreet map, that invite volunteers to contribute to projects developed by citizens, professionals or institutions, who are required to complete several challenging tasks that requires human intelligence alongside machine intelligence.

*Corresponding author
[1] https://www.ushahidi.com/

[2] http://digitalhumanitarians.com/
[3] https://www.standbytaskforce.org/
[4] https://www.epicentermediatraining.com/vost/
[5] https://crisismapping.ning.com/
[6] https://www.hotosm.org/
[7] https://crowdsourcerescue.com/
[8] https://www.giscorps.org/
[9] https://www.zooniverse.org/

https://doi.org/10.5194/isprs-archives-XLII-3-W8-331-2019

By using intelligent image retrieval techniques, new technologies are providing means to automatise the process and to handle larger quantity of data in shorter time. Yet, the quality and reliability of the result are still unclear and need to be evaluated against ground truth (Alam et al., 2018). This paper proposes a hybrid approach, combining automatic extraction of relevant social media content and crowdsouring. Automatic extraction is based on keyword matching and allows the automatic extraction of locations with a NER (Named Entity Recognition) based and disambiguation approach called CIME (Context based IMage Recognition)(Francalanci et al., 2018). Crowdsourcing, uses Crowd4EMS[10], a facilitating platform for coordinating volunteer contributions and creating reliable and actionable data for disaster response.

These approaches are part of a research project Evolution of Emergency Copernicus Services (The E2mC Project)[11], funded by the European Commission through its Horizon 2020 programme . The goal of the project is to demonstrate the technical and operational feasibility of integrating social media and crowdsourced data into the EMS Mapping and Early Warning components. Therefore, a new EMS service component (Copernicus Witness) designed to exploit social media analysis and crowdsourcing capabilities to generate a new information product, ensuring the availability of relevant information in near real time (Havas et al., 2017).

The E2mC project is a multi-component tool called Copernicus Witness. The components include

- *Data Acquisition* : When initiated, the data collection component uses automated web crawlers to funnel relevant social media posts to the crowdsourcing phase of the project.
- *Data Analysis* : *Crowd4EMS*, facilitates data enrichment and analysis. this constitutes crowd platform that facilitates validation of the social media content and refines geolocation and classifies precision information.
- *Data Visualisation* : The final component is a real-time, web-based spatial data visualizer, that helps the Copernicus Emergency managers to quickly sort, filter, map and extract data within the first critical hours of a crisis to make decisions and develop a response strategy.

This paper discusses the data acquisition and data analysis aspect of the witness component and in detail the facilitating crowd platform and its components. The paper is structured as follows. In section 2 the paper discusses the challenges faced in data acquisition and data analysis, while outlining some related work. In section 3 there is a detailed description of the components of the crowd4ems platform, the experimental framework and its application to a case study are presented. The evaluation of the crowd contribution and the results are discussed in section 4.Finally, concluding remarks and future research are discussed in Section 5.

2. RELATED WORK

2.1 Crowdsourcing platforms in disaster response

Crowdsourcing is not a new concept, but the improvement in technology and the increase in access to internet and the spread

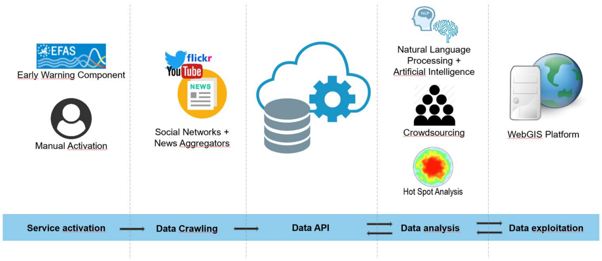

Figure 1. E2mC Framework

of social media networks across various walks of life, makes it easier to combine technology and the available human resource to address the information scarcity during any disasters. Manual geolocation of social media imagery has been proven useful in several situations. In investigative journalism, it has helped to verify the relevance of an image to a specific event. During natural disasters, manual geolocation has been used to mark accessibility features in specific areas and to enhance the situational awareness of first responders and emergency managers[12].

Exisitng crowdsourcing platforms include Zooniverse, Crowdcrafting, Sahana and ArcGIS platforms which has been proven useful in disaster response and recovery during various humanitarian crisis(Poblet et al., 2013) Several online crowdsourcing communities have taken up the challenge to geolocate Social Media content. In the field of humanitarian response, communities such as Stand by Task Force and GISCorps have collated, aggregated and geolocated social media data during several natural disaster events such as Hurricane Harvey, Hurricane Maria and Hurricane Florence. But manual geolocation is a task very difficult to perform and hard to scale.

2.2 Social media geolocation

Geolocation refers generically to the activity of assigning a location to an object. Automatic geolocation has long been investigated for traditional documents and web pages, often relying on Natural Language Processing (NLP) techniques. More recently, it emerged as a fundamental topic in the context of social media analysis, given the profound impact and the new challenges brought by this domain (Ajao et al., 2015, Zheng et al., 2018). Indeed, many kinds of social media analyses require some geographical knowledge, which is often very scarce, but, at the same time, social media texts are usually more difficult to analyse than traditional well-structured texts. However, social media objects are typically characterised by other kinds of (meta)data that can be exploited by automatic tools, such as social, temporal and contextual information.

Research on social media geolocation has focused on different targets, ranging from users, posting locations and posts' content locations. A platform such as Crowd4EMS needs to assign (or, at least, estimate) a coherent location for posted contents, and this is often achieved by recognising and disambiguating locations mentioned in the text (Zheng et al., 2018).

Several approaches have been proposed for this goal. Supervised techniques have been described both for location recognition and disambiguation (Zhang, Gelernter, 2014, Liu et al., 2014, Ji et al., 2016, Inkpen et al., 2015). Among the main

[10] `https://crowd4ems.org/`
[11] https://www.e2mc-project.eu/
[12] `https://www.giscorps.org/napsg_243/`

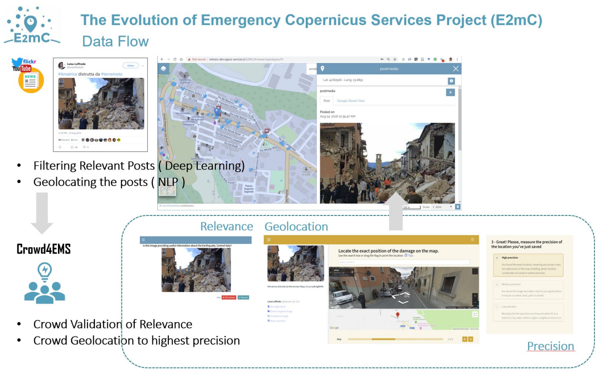

Figure 2. Data Flow in E2mC

limitations of supervised techniques there is (1) the need of labeled examples and (2) resulting models which do not easily generalise on events different with respect to those seen during training. This is particularly relevant in disaster response, since each event is a *unicum* characterised by complex specificities. Another class of techniques exploiting gazetteer insights has demonstrated state-of-the-art geolocation performance in terms of precision and recall (Middleton et al., 2014, Middleton et al., 2018). However, these approaches require knowing and preloading the area of the target event in advance, which is often not feasible for unexpected emergency events. While many methods in this way reach a good precision in identifying names of localities, associating precise geographical coordinates to a post is still a challenge and requires further research.

2.3 Challenges faced

A major limitation of existing algorithms is reaching high precision when estimating geolocations within the range of one kilometer or less. To overcome this limit, a growing body of research seeks to use computer vision techniques to automatically geolocate images (Weyand et al., 2016). Advanced computer vision technologies makes it possible to *guesstimate* the location of an image (Hays, Efros, 2008). However, the cost of wrongly guesstimated locations in case of a damage assessment or disaster response could be huge. The research question that the paper attempts to answer are the reliability and actionability of the information processed by crowdsourcing communities. How does a facilitating platform minimizes the time taken on each micro task and how such platform assists in aggregating and acquiring necessary information from the social media. Geolocation of social media information has often resulted in creating reliable and actionable data source, which can increase the productivity of the emergency responders especially during the early hours of a disaster. Crowd4EMS facilitates the volunteering communities and enables them to geolocate faster by providing assitive features like a street view, translating component and precision options.

3. CROWD4EMS

Crowd4EMS is the crowdsourcing platform of the E2mC project. Social media content extracted by the CIME algorithm is presented to the digital volunteering community via the crowd4ems platform. The communities enrich the social media content by validating the relevance and geolocating the social media imagery.

Two main goals to be addressed to provide responders with appropriate information extracted from social media are:

- *Selecting relevant information :* A social media data is considered *relevant*, if it provides useful information about the given disaster. In this case, the need of enriched social media data is for damage assessment. Any image creating a situational awareness of the disaster, where the impact of the disaster or the damage caused are clearly visible, is considered relevant. Any image that shows no damage, news article, weather report or even any misinformation or a rumour is considered *irrelevant*.

- *Geolocating :* *Geolocation* is defined as the process of identifying the geographical location of the information contained in a social media post, and it usually refers to the position of the person when the image or video was captured or the position of the damage.

While several approaches have been proposed in the literature to analyze and geolocalize social media in the context of emergencies, the obtained results from the automatic analysis still need to be analyzed with respect to relevance, as many posts may be off-topic or only weakly related to the event in progress, moreover as illustrated in (Middleton et al., 2018), current approaches still do not provide a precise localisation of the extracted information.

However, often the text and the media associated to posts often provide some clues that, even if they are hard to use in an automatic system, can help a volunteer to assess the relevance and the location. We propose therefore an enrichment of information extracted from social media using crowdsourcing in two directions: (1) increasing the relevance of data by filtering data considered irrelevant by volunteers, and (2) Geolocating each of the relevant social media imagery (Photo & Video).

The information to be analyzed by the crowd is extracted from social media as illustrated in Figure 1 which presents the main components in the E2mC framework for extracting, filtering and enriching social media information. Figure **??** indicates the flow of data within the E2mC framework.

3.1 Automated data enrichment

In this section we briefly illustrate the social media initial analysis steps, while in the following section we focus on the crowdsourcing components.

The goal of the system is to extract relevant and geolocalized images and videos from social media. First, crawlers on social media sources, and in particular Twitter, Flickr, and YouTube, select potentially interesting posts with specific queries specifying the period, with a 15 minutes moving window approach, keywords specific for the type of event being considered (e.g. floods or earthquake), and event-specific keywords that can be added by the operators. Only posts containing visual evidence are retained. Multilingual posts are analyzed using the Natural Language Processing feature of Polyglot for Named Entity Recognition (Al-Rfou et al., 2015) to extract Named Entities. Currently, 40 different languages are supported. Extracted posts are memorized in a data store (currently supported by Postgres), from which other services can extract stored information and add evaluation tags based on microservices through data APIs.

As a basis of this work we focus here on the description of the results of the geolocalization service called CIME (Context-based Media Extractor) which has been developed for the project. The CIME algorithm is briefly illustrated in (Francalanci et al., 2018) and its presentation is not the focus of this work. Many services are also provided as a support to filtering relevant images, to eliminate duplicates and to identify images with a poor quality (e.g. too dark), as described in detail in (Barozzi et al., 2019).

As most of the social media posts do not carry a native geolocation tag, the goal of CIME is to associate geographical coordinates to a post. As in Crowd4EMS we have the goal of retrieving from social media visual material, such as pictures and videos, to get awareness of the ongoing situation, the focus of CIME is to extract geographical coordinates for the locations mentioned in the posts. CIME extracts Named Entities as potential names of locations from the text of the post using NLP Stanford Core (Manning et al., 2014) and extracts potential locations also from the metadata associated to the post, such as hashtags and places. The potential names of locations and the geographical coordinates for the extracted localities and places are retrieved using Nominatim on Open Street Map (OSM) (Haklay, Weber, 2008). The hiercarchical organization of locations in OSM and distance measures are used to disambiguate locations, selecting the most fine grained candidate localities based on the proximity with other candidates. When the location can not be disambiguated using a single post, the location of related posts such as replies and mentions are also taken into account. The result is one (or a set of) location name, represented with a string which mentions the location of the place with a hierarchical structure, such as for instance “Avenida de Machupichu, Canillas, Hortaleza, Madrid, rea metropolitana de Madrid y Corredor del Henares, Community of Madrid, 28001, Spain”, its geoJSON representation and the coordinates of its center point (40.4639234,-3.6353076).

The precision of this result depends on how the location is mentioned in the post, and supporting evidence available for disambiguation. In the following we discuss how this result can be used in crowdsourcing to improve the precision of the location obtained using the CIME algorithm.

3.2 Crowdsourced data enrichment

It is to be noted that the Crowd4EMS federates existing online volunteering communities instead of creating a new set of volunteers. Previous beta-testers of GeoTAGX , young digital beta testers of Goodwall [13] and existing members of digital humanitarian communities were involved during organised online activations. Timed-activations as part of student workshops were also conducted as an alternative approach. Availability and sustaining the interest of the community is a crucial factor in determining the outcome of any crowdsourcing task.

3.2.1 Crowdsourced Relevance As mentioned before, the data extracted from social media is presented to the volunteers through the Crowd4EMS component to assess its relevance . 3 shows a screenshot of the CROWD4EMS relevance task for evaluating the relevance of a tweet coming from the crawlers (i.e. automatically processed information gathering from social media based on keyword matching).

A tweet is atleast classified by 5 different volunteers to consider the task as completed. There are two main approaches to validate volunteers contributions in a crowd sourcing project: The first method uses the role of a validator (i.e. experienced volunteer) to ensure data quality coming from new volunteers (E.g. HOTOSM). A second approach is to ask different volunteers to do the same task a certain number of times. This number could range from 5 to 20 evaluations E.g. Zooniverse or Crowdcrafting. The quality of the data would be estimated according to the agreement between the different volunteers. The Crowd4EMS component uses the multi-evaluation approach to ensure the data quality coming from the crowd. The agreement between the users is computed using the Fleiss Kappa Algorithm(Fleiss et al., 1971).

[13]https://www.goodwall.io/

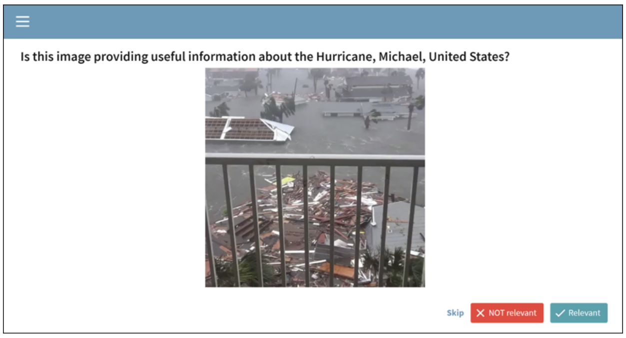

Figure 3. Crowd4EMS Relevance Validation

The Fleiss Kappa algorithm provides an agreement rate between 0 and 1 for each analysed tweet. The agreement is *1* when all volunteers provide the same classification and is *0* if they provide completely different classifications. A total of 238 tweets concerning the UK Floods were evaluated by experts providing the ground truth, and the same set of tweets were evaluated by the crowd. Each tweet was evaluated 5 times. The volunteers performed 1190 classifications. Out of the 236 classified tweets, the volunteers fully agreed on 146 classifications (61.35%), reached a high agreement on 57 classifications (23.95%), and did not reach an agreement in 33 classifications (13.8%) Based on the volunteers agreement, for each task computed with the Fleiss Kappa algorithm,relevance was computed.37.3% are considered as fully relevant, and 24.6% as completely irrelevant. The rest of tweets (38%) are classified with a relevance between 0.2 and 0.8. As a summary of this analysis, 57.2% of the tweets analyzed by the crowd represent a high degree of relevance (i.e. equal to or higher than 0.6) . Every post that is validated using rater-agreement as *Relevant* is presented again to the crowd for Geolocation.

3.2.2 Crowdsourced Geolocation Each post that is classified as relevant by the crowd and and are with pre-suggested geolocation tags by the CIME algorithm is presented to a minimum number of volunteers using Crowd4EMS in order to validate, correct or improve its location. While the automated technique facilitates faster means to crawl, consolidate, and extract relevant basic information, crowdsourcing is used to validate and improve the reliability of the retrieved information. Location hints provided by CIME for each post include :

- A textual string for the location (point of interest, street or location), including the names of the area in which the data belong - such as “Pescara del Tronto, Province of Ascoli Piceno, Italy” shown in Figure 4.
- The geographical coordinates (latitude and longitude) of its centroid for the positioning on a map.

Figure 4 shows a screenshot of the interface of Crowd4EMS, that allows volunteers to geolocate images and videos from social media, starting from the geolocation information presented by CIME (both textual and geolocated as a point on a map). The user interface combines all resources that are commonly used in the process of geolocating social media content manually, namely Google maps, Google street view, direct access to Google image search, and the link to the original post.

The users improve the precision of the geolocation based on the visual and textual cues present in the social media post. The users are also facilitated by additional features wherein they can *Translate* the tweet if in other languages. The users can also *search google* for similiar images posted or can look up the *original tweet* or the social media post

4. EVALUATION

This section will discuss the activation of Amatrice Earthquake . It will list out the need for geolocation of social media information.Two aggregation algorithms are proposed and then compared with an expert evaluation in this section.

4.1 Aggregation algorithms for geolocation

Asking a number of volunteers to geolocate the same media content helps improving the data quality and reliability. However, it also introduces a new challenge: how to choose the right answer based on the volunteer contribution? This question is particularly relevant when time constraints (imposed by the need to obtain results in the shorter possible time) limit the statistics to a few contributors for each post. Traditional methods for inter-rater agreement such as Fleiss' Kappa algorithm (Fleiss et al., 1971), widely used in aggregating crowdsourced contribution, can hardly be adapted here, because geospatial data is not discrete but rather continuous - in other words, it cannot be classified in a finite number of categories.

In this study, each post has been presented to a minimum of 3 volunteers in order to validate or improve/correct the approximate location provided by CIME. Next, two aggregation algorithms have been tested to calculate the location based on the multiple answers from the volunteers: Highest precision and Agreement aggregation.

Highest precision: this algorithm chooses the answer with highest precision. For instance, in the case:

1. Volunteer 1 answer: Madrid
2. Volunteer 2 answer: Calle Ramon Power, Madrid
3. Volunteer 3 answer: Calle Ramon Power, 3, Madrid.

the algorithm chooses the 3rd answer, because it is the one with highest precision. There are two main motivations behind the algorithm. First, information geolocated with high precision can be used in a wider range of cases. Second, if volunteers claim to geolocate a photo or video with high precision, we trust that either they really know the place, or the place can be identified precisely on the basis of the text associated with the post, or the image can be compared with images extracted from available sources such as Google Map's street view[14].

It is worth noticing the importance of trust in this first approach, which is motivated by knowing the experience of the volunteers often involved in this kind of first-response efforts (see examples in Section 2.1). However, the main limit of the highest precision algorithm is the robustness of the system. If one user introduces a wrong answer claiming high precision, this will be chosen as the location of the social media content.

[14] https://mapstreetview.com/

Agreement aggregation aims to provide a more robust geolocation. In this case, social media content is geolocated only if there is a minimum agreement between at least 2 users. The agreement can happen at different levels of precision. To assess the agreement the following rules are applied:

- *Agreement on low level precision*: At least 2 users agreed on the name of the city or town. An agreement at low level precision happens in the example above with the volunteer 1 and volunteer 2. The location in this case would be "Madrid".
- *Agreement on medium level precision*: At least 2 volunteers agree on the name of the street, river, area of the social media content. In the example above this agreement occurs between the volunteers 2 and 3.
- *Agreement at high precision level* happen just when at least 2 volunteers agree upon high precision levels, i.e., the distance between the 2 suggested location coordinates (i.e., lat, lon) is less than 100 meters.

In the following section both the algorithms are compared against the expert evaluation of the same dataset.

4.2 Experimental Framework

This section describes how the studies have been carried out and the main tools used to assess the quality of the geolocation of social media content by combining automatic and crowdsourcing techniques.

4.2.1 Amatrice Earthquake case study To evaluate the geolocation quality an original dataset of 1,153 posts with images or videos - geolocated with the CIME algorithm - was collected about the Amatrice Earthquake, coming from Twitter, Flickr, and Youtube. A total of 106 posts were validated as highly relevant by the crowd, and used for the geolocation evaluation. The datasets is available from the authors on request. The Amatrice Earthquake[15] hit Central Italy on August 24, 2016. Severe damages were reported in the towns of Amatrice, Accumuli, and Pescara del Tronto. It was the largest earthquake in Italy since 2009, 299 people died, and the economic loss was estimated between 1 and 11 Billions.

Any information denoting the severity of the disaster and its location is key for first responders. Especially during the first hours of the information blackout, Social media reporting could be a great source of information. However, as most tweets/posts are merely comments and contain very little or no information related to the location of the disaster, analyzing the large volume of information remains a challenging task in contrast to the ease of acquiring it from social media platforms (Nguyen et al., 2017).

[15] https://en.wikipedia.org/wiki/August_2016_Central_Italy_earthquake

https://doi.org/10.5194/isprs-archives-XLII-3-W8-331-2019

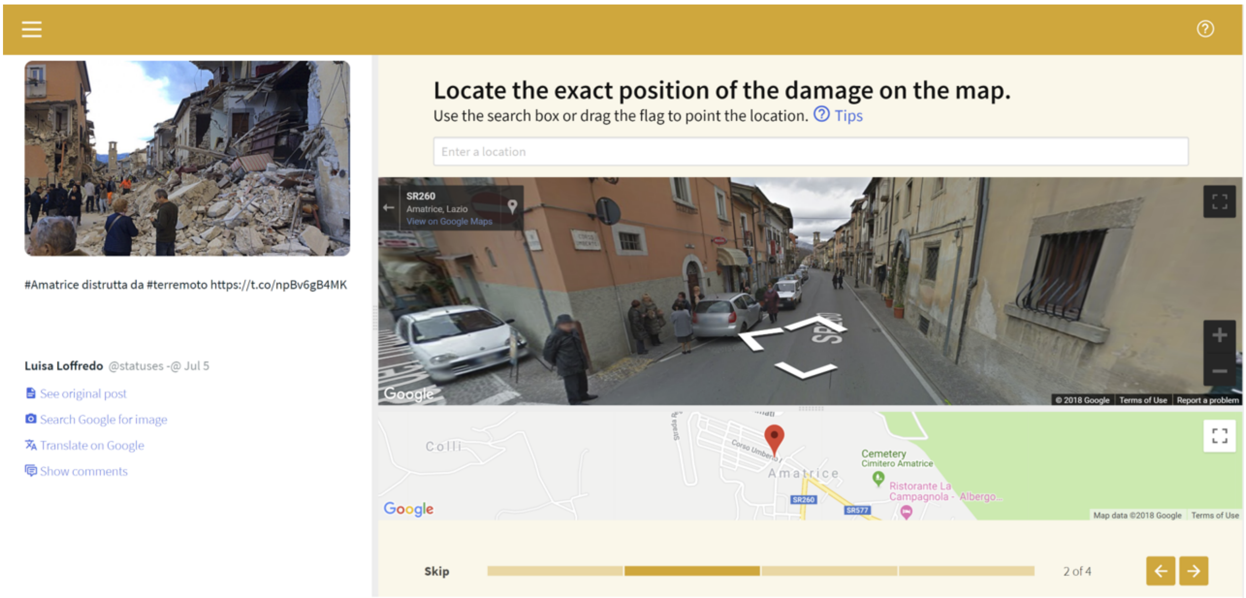

Figure 4. Crowd4EMS Geolocation

4.2.2 The ground truth To evaluate the quality of geolocation provided by the combined approach presented in this paper, results need to be compared with the ground truth. However, the ground truth is unknown, as the precise locations of many posts in social media are unknown and they will probably remain so. In this experimental framework, the geolocation of social media post provided by a team of 5 experts, i.e., professionals working on disaster management, is considered as the ground truth.

To carry out the evaluation, the data set was split between the experts to be geolocated once. They were asked to provide for each picture, the location (lat, lon) and the precision (high, medium, low). The different precision levels are defined as follows:

- *High*: the exact location is found, meaning the post is located as a very accurate point on the map: building, street number, coordinates of a road or similar precision.
- *Medium*: the image is taken close to the provided geolocation. It may be a street, road, path or similar.
- *Low*: Meaning that the precision provided is at the level of a city, town, district, region, neighbourhood and so on.

Figure 5 shows four examples of pictures from the dataset, that have been geolocated with different precision levels by the experts.

4.2.3 Precision and Accuracy The evaluation of geolocation in this experimental framework is based on two metrics: precision and accuracy. In this paper, we define *precision* as a qualitative metric to indicate how exact a geolocation information is, according to the high/medium/low classification mentioned above. Whether a photo is geolocated at the level of the city, street, or at the level of a point in the map. The precision can be high, however the information can be wrong. A photo can be classified with high precision in Rue de Lausanne, 1, Geneva. However, the actual location of the picture could be in a different city. *Accuracy* refers to the reliability of the geolocation provided, defined as a Boolean value. Independent of the precision provided, accuracy measures whether the information about the location is true. Based on the comparison with the expert geolocation, we considered a data point as accurate, considering separately the three precision levels, according to the following definitions:

- A photo or video geolocated with *high precision* is accurate if it is within 100m distance for the expert geolocation.
- A photo or video geolocated with *low precision* is accurate if the name of the city/town/village matches with the expert evaluation.
- A photo or video geolocated with *medium precision* is accurate if the name of the street is the same than the expert geolocation or the distance with the expert geolocation is less or equal to 100m.

Medium precision points have an additional complication, as indicating a different road than the expert does not necessarily mean the answer is “ wrong”. For instance, if we consider that the image shows a road intersection, either road would be an acceptable answer. In this case we also consider the geolocation right if it is located within 100m from the expert geolocation.

4.3 Experimental Results

This section analyses the geolocation quality of social media content by using the proposed method for images and videos from the Amatrice Earthquake case study.

4.3.1 Expert evaluation First, we analyze the results of the geolocation carried out by the experts, showing the extent to which the selected social media content can be geolocated.Figure 6 shows that 49% of posts have been geolocated with a high precision, i.e., experts claimed to geolocate the social media content with a precision of one point in the map. 10% of the posts were located at the level of the street, road, path or similar (medium). 38% were located at the city level, e.g. a picture was taken in Amatrice, but the experts are not able to geolocate it inside Amatrice (low). Finally, 3% of the content was not possible to geolocate.

precision examples.JPG

Figure 5. Geolocation Precision Examples

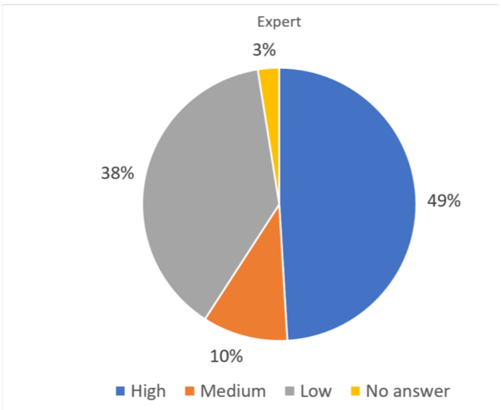

Figure 6. Expert Precision

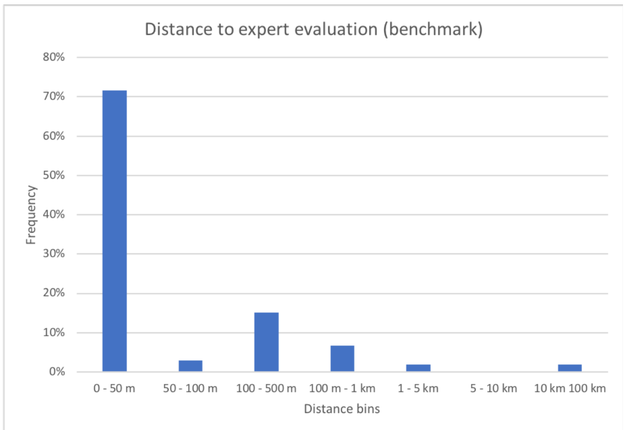

Figure 7. Crowd Vs Expert Optimal

Notice that in a real scenario the expert evaluation is most probably not feasible due to the unavailability of experts to carry out the geolocation tasks in a limited amount of time, and the time required to geolocate the data. The experts evaluation are used in this paper as ground truth to compare the results obtained with the combination of automatic and crowdsourcing techniques, comparing the two aggregation strategies.

4.3.2 Crowdsourcing evaluation This section compares the expert evaluation to the evaluation coming from the crowd.The same dataset has been presented to a crowd of volunteers through the Crowd4EMS platform. Each post has been geolocated by three different volunteers. This redundancy aims to improve data quality and make the geolocation more robust against errors, however it adds an additional challenge when choosing which of the three geolocations is the best. In this section we compare the two proposed approaches to the aggregation of volunteer answers presented in Section 4.1 , i.e.,highest precision and agreement, to an additional offline method called *optimal*, where the geolocation retained is the closest to the expert evaluation. While this optimal aggregation cannot be applied in a live scenario, it is used here as a benchmark to evaluate the quality of the aggregation approaches presented in this paper. The optimal aggregation answers the question: for each task what is the best answer that could be provided by the volunteers?

Figure 7 shows the distance between the locations provided by the volunteers and the experts. Most volunteer contributions are very close to the expert geolocations, with over 70% of the volunteers' points within 50 m from the expert ones. Figure 8 shows the precision that can be reached with the optimal aggregation benchmark, with 55% of the images classified with high precision. Figure 9 shows the accuracy of the optimal aggregation. It shows more than 90% of social media content geolocation is considered right, following the evaluation criteria presented in Section 4.2.3. Regarding the 9% of wrong geolocations, Figure 10 shows that most of the wrong answers

https://doi.org/10.5194/isprs-archives-XLII-3-W8-331-2019

are coming from geolocations claiming high precision, i.e., the volunteer claims to have geolocated the photo or video at the level of the point. These geolocations are considered wrong in our evaluation framework when the distance between the expert's evaluation and the volunteer's evaluation is more than 100 meters, which might be a very constrained threshold, in particular for photos with a wide area view.

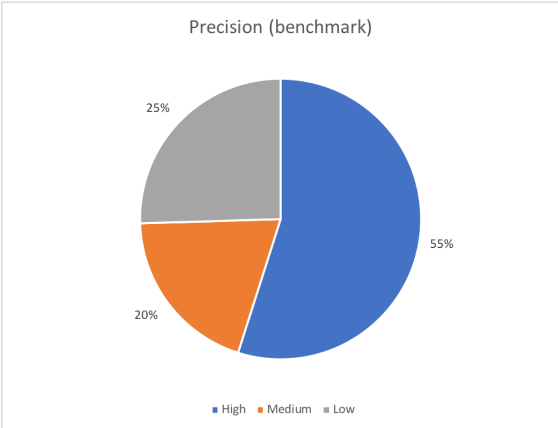

Figure 8. Precision Optimal

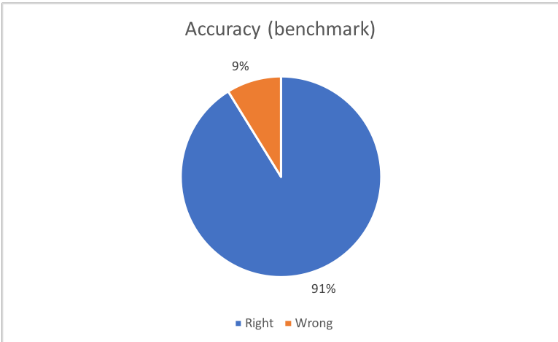

Figure 9. Accuracy Optimal

4.3.3 Highest precision aggregation This method selects the volunteer contribution with highest precision, ideally at the level of one point in the map. Although this approach provides data with higher precision (65% high precision, see Figure 11), it does so at the expense of accuracy, as it returns 25% of incorrect data (see Figure 12). This can be explained as the method assumes that a high precision answer is likely correct. In reality, however, a user could select the wrong level of accuracy by mistake, or they could be marking a very similar yet different place than the one in the image. Unless there is a very trusted community of volunteers, relying on a single answer is not a good approach in terms of accuracy, because the probability of having a wrong answer dramatically increases.Notwithstanding this high percentage of wrong data, more than 80% of the social media content is geolocated within 500 meters of error, and more than 56% within 50 meters (see Figure 17).

4.3.4 Agreement aggregation The agreement aggregation only provides an answer if there is an agreement between at least two volunteers. The Agreement aggregation improves the accuracy with 95% of correct answers (see Figure 15), albeit at the expense of precision, as 78% of the crowd points have low precision (see Figure 16). When answers have different levels of precision for a given post, the agreement needs to be evaluated based on the lowest precision level, i.e.: if the an-

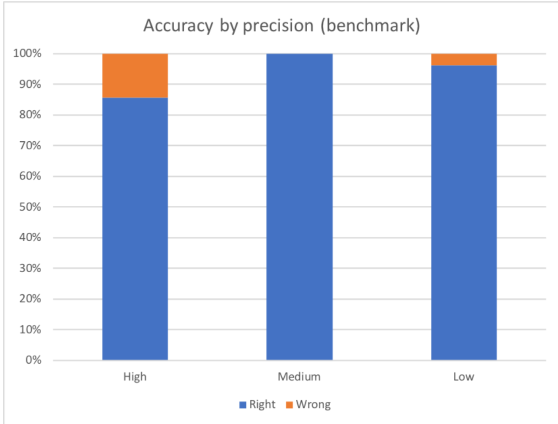

Figure 10. Accuracy by Precision Level Optimal

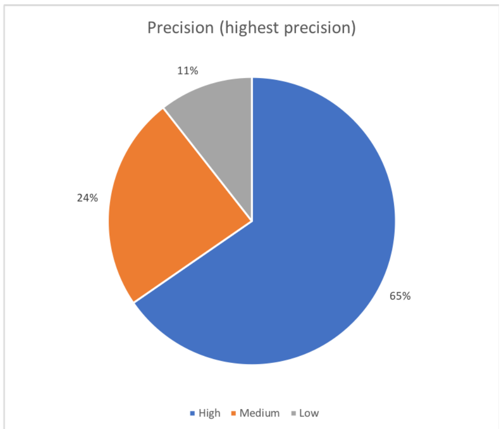

Figure 11. Level of precision achieved with highest precision aggregation

swers precision levels are low, medium, high, then the agreement is evaluated at the city level (low). Nevertheless, evaluating answers that have an agreement among more than one person, reduces outliers in data, tightening the distance distribution curve, and improving the accuracy. This is made more apparent when observing the distance to expert points frequency distribution (Figure 13): due to the fact that the aggregated points are centroids of the points of agreement, less than 25% of the data is within 50 meters from the expert points, however, this method also returns the smallest mean (0.73 km) and max (18.12 km) distances, as well as the smallest standard deviation (2.5 km) of all three aggregation methods. Figure 17 summarizes the main statistics of all three aggregations, as well as those of the CIME algorithm results, as compared to the expert evaluation. Figure 17 shows a dramatic improvement compared with the CIME algorithm. Namely, the highest precision algorithm is able to geolocate more than 60% of post within 100 meters of error (Compared with the expert evaluation), while the Agreement algorithm provide 37% of post geolocated within 100 meters of error. Combining CIME with crowdsourcing techniques has the potential to geolocate more than 70% of the social media content within 50 meters in comparison with expert evaluation.

This evaluation shows that the crowd can be very effective in filtering out data that would be considered irrelevant, which is very important, considering the large volumes of data that are crawled from social media. Additionally, combining automatic algorithms with the volunteers input; this data can be enriched by being very precisely geolocated. Using the appropri-

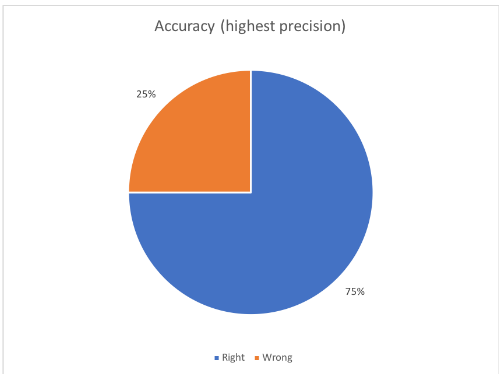

Figure 12. Accuracy achieved with highest precision aggregation

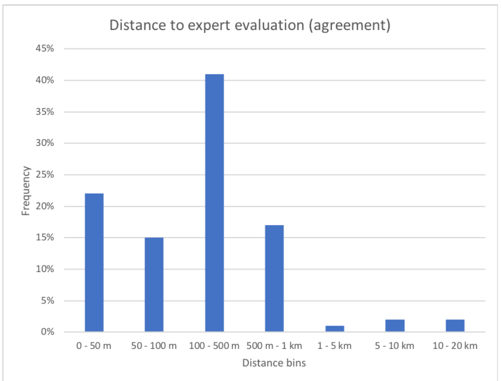

Figure 13. Crowdsourcing vs Expert geolocation - Agreement aggregation

ate data aggregation method, the crowd can locate over to 80% of posts from social media within 500 m of the expert established ground truth, and over 50% can be located within 50 m. The crowd validated and geolocated social media content can be visualised real time in a web map , that helps the Emergency managers to quickly sort, filter, map and extract data within the first critical hours of a crisis to make decisions and develop a response strategy.(see Figure 14)

5. CONCLUSIONS

This paper presents a new approach to improve the geolocation of social media content, combining automated social media text analysis (to filter and geolocate relevant information from social media) with crowdsourcing (to improve the social media content geolocation). The proposed approach is evaluated using data from the Amatrice Earthquake that hit central Italy in 2016.The paper presents two different aggregation algorithms in order to estimate the content's location from a given number of volunteers. The two algorithms behave differently in terms of accuracy and precision of the results, and the best one will depend on the use case requirements. The volunteers participated on the geolocation process through an online crowd sourcing platform, called Crowd4EMS.This paper also presents how right tools and approaches can better facilitate the crowd sourcing communities in providing high quality actionable data in disaster response.

ACKNOWLEDGEMENTS

This work was funded by the European Commission H2020 project E^2mC "Evolution of Emergency Copernicus services".

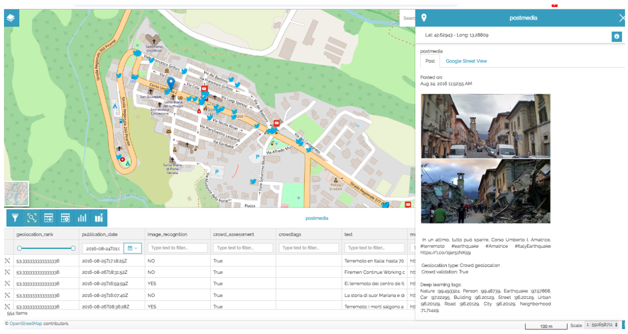

Figure 14. Copernicus Witness : Real time Visualisation of Social media data

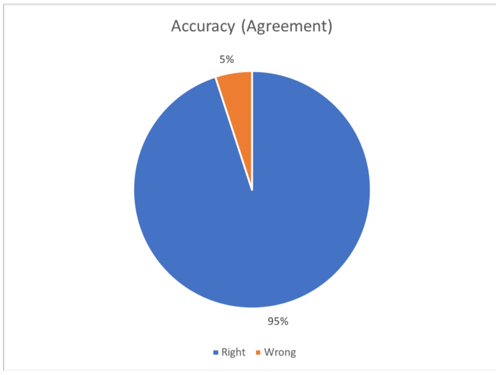

Figure 15. Accuracy achieved with Agreement aggregation

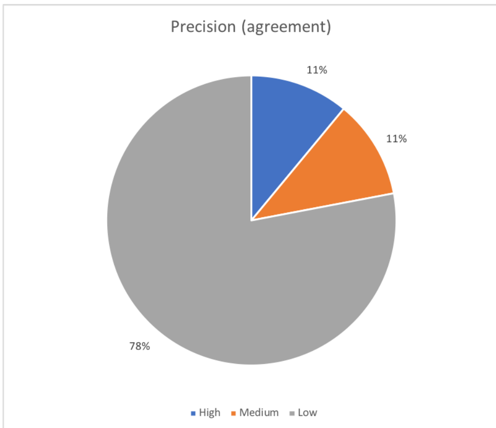

Figure 16. Level of precision achieved with highest precision aggregation

	CIME	Optimal	Highest Precision	Agreement
Distance (Km) to expert point summary statistics				
Mean	88.19	0.82	9.23	0.73
Min	0.00	0.00	0.00	0.00
Max	7304.51	53.40	394.75	18.12
Standard deviation	710.28	5.46	46.49	2.48
Percentage of points per distance to expert point				
Within 50 m	12.26%	71.70%	56.60%	22.00%
Within 100 m	14.15%	74.53%	61.32%	37.00%
Within 500 m	63.21%	89.62%	80.19%	78.00%
Within 1 km	80.19%	96.23%	88.68%	95.00%

Figure 17. Comparison of CIME and hybrid approaches with the expert valuation

Grant Agreement No. 730082. 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.

REFERENCES

Ajao, O., Hong, J., Liu, W., 2015. A survey of location inference techniques on Twitter. *Journal of Information Science*, 41(6), 855–864.

Al-Rfou, R., Kulkarni, V., Perozzi, B., Skiena, S., 2015. Polyglot-ner: Massive multilingual named entity recognition. *Proceedings of the 2015 SIAM International Conference on Data Mining*, SIAM, 586–594.

Alam, F., Ofli, F., Imran, M., 2018. Processing Social Media Images by Combining Human and Machine Computing during Crises. *International Journal of Human–Computer Interaction*, 34(4), 311–327.

Barozzi, S., Marquez, J. L. F., Ravi Shankar, A., Pernici, B. et al., 2019. Filtering images extracted from social media in the response phase of emergency events. *16th Conference on Information Systems for Crisis Response and Management*, 1–12.

Blank, G., Reisdorf, B. C., 2012. The participatory web: A user perspective on Web 2.0. *Information, Communication & Society*, 15(4), 537–554.

Castillo, C., 2016. *Big Crisis Data*. Cambridge University Press.

Fleiss, J. et al., 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382.

Francalanci, C., Pernici, B., Scalia, G., Zeug, G., 2018. Talking about places: Considering context in geolocation of images extracted from tweets. *GI-Forum 2018, Issue 1, Salzburg, July 2018, Short paper*, 243–250.

Haklay, M. M., Weber, P., 2008. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, 7(4), 12–18.

Havas, C., Resch, B., Francalanci, C., Pernici, B., Scalia, G., Fernandez-Marquez, J. L., Achte, T. V., Zeug, G., Mondardini, M. R. R., Grandoni, D., Kirsch, B., Kalas, M., Lorini, V., Rüping, S., 2017. E2mC: Improving Emergency Management Service Practice through Social Media and Crowdsourcing Analysis in Near Real Time. *Sensors*, 17(12), 2766.

Hays, J., Efros, A. A., 2008. Im2gps: estimating geographic information from a single image. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, IEEE, 1–8.

Inkpen, D., Liu, J., Farzindar, A., Kazemi, F., Ghazi, D., 2015. Detecting and disambiguating locations mentioned in twitter messages. *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, 321–332.

Ji, Z., Sun, A., Cong, G., Han, J., 2016. Joint recognition and linking of fine-grained locations from tweets. *Proceedings of the 25th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee, 1271–1281.

Liu, F., Vasardani, M., Baldwin, T., 2014. Automatic identification of locative expressions from social media text: A comparative analysis. *Proceedings of the 4th International Workshop on Location and the Web*, ACM, 9–16.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., McClosky, D., 2014. The stanford corenlp natural language processing toolkit. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, System Demonstrations*, 55–60.

Middleton, S. E., Kordopatis-Zilos, G., Papadopoulos, S., Kompatsiaris, Y., 2018. Location Extraction from Social Media: Geoparsing, Location Disambiguation, and Geotagging. *ACM Trans. Inf. Syst.*, 36(4), 40:1–40:27.

Middleton, S. E., Middleton, L., Modafferi, S., 2014. Real-Time Crisis Mapping of Natural Disasters Using Social Media. *IEEE Intelligent Systems*, 29(2), 9–17.

Nguyen, D. T., Ofli, F., Imran, M., Mitra, P., 2017. Damage assessment from social media imagery data during disasters. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, ASONAM ’17, ACM, New York, NY, USA, 569–576.

Poblet, M., García-Cuesta, E., Casanovas, P., 2013. Crowdsourcing tools for disaster management: A review of platforms and methods. *International Workshop on AI Approaches to the Complexity of Legal Systems*, Springer, 261–274.

Weyand, T., Kostrikov, I., Philbin, J., 2016. PlaNet - Photo Geolocation with Convolutional Neural Networks. *CoRR*, abs/1602.05314. http://arxiv.org/abs/1602.05314.

Zhang, W., Gelernter, J., 2014. Geocoding location expressions in Twitter messages: A preference learning method. *Journal of Spatial Information Science*, 2014(9), 37–70.

Zheng, X., Han, J., Sun, A., 2018. A Survey of Location Prediction on Twitter. *IEEE Trans. Knowl. Data Eng.*, 30(9), 1652–1671.