A review of visual analytic approaches used on spatio-temporal Twitter data, for the identification, management and analysis of disasters

Juan Cova, Heiko Maerz, Sergio Naval, Joaquin Perez, and Veronica Tuffrey
City, University of London

Abstract—Natural disasters demand situational awareness for identification and effective management. The growth of Twitter data provides an opportunity to complement traditional sensors and data sources available. In the last decade Data Science, and within this, Visual Analytics, have been proved useful to improve data-driven decision making in many domains. We report the findings of a rapid review of the application of Visual Analytics techniques to disaster management using Twitter data. Based on 10 case studies we find that VA methods are potentially effective in assisting human analysts to summarize and extract patterns from large volumes of rapidly changing data. During data processing, clustering is commonly applied to identify subtopics within tweets. For visualization, interactive maps and timelines are widely used to summarize spatio-temporal characteristics of the events. Challenges to the use of Twitter data in this domain include intrinsic characteristics of the data such as noise, high volume and potentially not being representative of the overall population. Except with respect to information overload, users' feedback was positive. Evidence is needed from application of this technology in real-time disasters.

Index Terms—Data, Emergencies, Review, Twitter, Visual Analytics, Visualization; Situational awareness

1 INTRODUCTION

Huge numbers of people are affected by natural disasters every year. In 2017 alone there were 324 natural disasters in 105 countries, with 204 million people affected [1]. Numbers affected are projected to increase, not only due to population growth but also to increasing concentrations of population in urban areas, thereby increasing vulnerability, and to climate change with its likely impacts on sea level and extreme weather events [2]. Improvements in power of computers coupled with the emergence of new sources of data mean more options exist for detecting and monitoring disasters compared to a decade ago. We define a disaster as a sudden event, such as an accident or natural catastrophe that causes great damage or loss of life [2]. Improvement in quality and quantity of information should plausibly contribute to rapid decision-making and planning during and after disasters to the benefit of the affected populations. Such matters are examined in an emerging research field termed "Crisis Informatics" which focuses on the use of ICT and social media before, during, or after crisis events [3].

This paper reviews the application of Visual Analytics (VA) techniques on spatio-temporal Twitter data. We examine the varied approaches used in VA of Twitter data in disasters and examine the extent to which the available methods enable their improved identification, management and analysis.

1.1 Data Science and Visual Analytics

During the last decade or so, the term "Data Science" has been used to refer to the study of the generalizable extraction of knowledge from data in the context of "emerging machine intelligence fuelled by data and state-of-the-art analytics" [4]. The increasing power of machine learning is enhanced by the increasing availability of high-quality human-curated data so that predictive models can be built and feed into decision-making in varied public and private applications.

Visual Analytics (VA) is one of the many tools within Data Science that are used to try to gain insight from large and complex datasets. VA "combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets" [5]. Thus the strengths of automated computing are harnessed to augment human cognitive and visual skills and thereby help create understanding and value [6].

One category of data increasingly collected and studied in diverse domains is "spatio-temporal" data for which in addition to the measurements/attributes of primary interest, spatial and temporal attributes are also available. Increasing volumes of geo-spatial data is now accessible to non-expert communities and new tools are needed to take advantage of this rich source of information [7]. This is because in spatio-temporal data, there are dependencies among measurements because instances are structurally related to each other in the context of space and time and show varying properties in different time periods and spatial regions. Thus the assumption underlying many of the widely used data mining techniques - that data instances are independent and identically distributed - is violated [8].

1.2 VA for spatio-temporal data

While the effectiveness of classical data mining algorithms is limited with respect to analysis of spatio-temporal data, the coupling of spatial and temporal information opens new analytic opportunities, for example for interpolation and extrapolation, and for integration of information of different types using references to common locations (spatial overlay) [7]. VA tools have been developed which both enable discovery of patterns in spatio-temporal data, and also construction of formal models to represent the patterns. The basic components of such systems are [9]:

- Cartographic map display, in which spatio-temporal data can be represented by map animation or by embedded diagrams;
- Time series display;
- Interactive tools for clustering;
- Time-series modelling, and
- An interactive visual interface.

As indicated in the definition above, automated analysis has a pivotal role in VA. An important application of VA of spatio-temporal data is the analysis of social media, one of the major sources of "Big Data" (the technological trend of storing and analysing vast amounts of data [6]).

1.3 Twitter data and disasters

Twitter is a micro-blog social media tool where registered users read and write messages called "tweets" and unregistered users can read the messages. When the system was created in 2006, the length of messages was limited to 140 characters and this limit was doubled to 280 characters in November 2017. The most recent statistics (Q2, 2018) show there are currently around 335 million active monthly users of which 20% reside in the US [10].

There is a growing literature around the use of social media in emergencies, summarised by Reuter and colleagues in 2018 [3]. Also Martinez - Rojas and colleagues undertook a systematic literature review of research into the specific use of Twitter data as a tool for the management and analysis of emergency situations [11]. The literature indicates that social media networks including Twitter can significantly support and improve disaster management due to the "real-time" nature of the data (inputs are available without significant temporal delays), and their in-situ nature (information can be obtained about the local situation). In this way, social media data can be regarded as real-time in-situ sensor data [12] and can be produced more quickly than conventional remote-sensing-based approaches (approx. 1–3 h). Clearly to be effective, a functioning wireless network (mobile phone networks, Wi-Fi networks, etc.) is needed.

While the strengths of Twitter include large potential audience, instant communication, and provision of real-time information, Twitter also presents limitations that make the use of these data difficult in the context of an emergency situation, including lack of verification of information; potential spread of rumours; imprecision in data; and provision of irrelevant information [11]. Other limitations of Twitter data are noise [13, 14], lack of georeferentiation [13, 15] high volume and continuous stream [16] and privacy and data protection concerns [17].

Potential applications of Twitter data have been described from all phases of disaster events, from the early-warning and planning phases via identifying the occurrence and location of the event [18-20], through identifying the affected-population's needs and focusing the response [21, 22], and for identifying priorities after the event [23].

In the remainder of the paper, we first describe the methods adopted for our review, before describing our findings with respect to aims, methods and visualization techniques. Finally, we describe gaps we identified in the existing research and discuss the limitations of our review.

2 METHODS

In this section we outline the methodological approach used for the review. As will be described below, our analysis and conclusions are based on a selection of case-studies.

2.1 Objectives

Research questions addressed by the review were as follows:

1. What are the methods/ approaches used for VA of Twitter data related to disaster situations?
2. What are the applications of the findings of this VA?
3. What are the limitations of the methods / approaches being used?
4. What are the limitations of current research into the methods?

To address questions 1 and 2 we examined 10 selected papers in turn to determine the techniques applied by the authors and the questions asked of the data. To address questions 3 and 4 we synthesised our findings from these individual studies together with contextual information from other sources, to identify both the strengths and limitations of the current VA methods being applied, and also of the research into these methods.

2.2 Approach used

Phase 1: Search

To choose 10 papers, we adopted the following inclusion criteria:

- in English language;
- in peer-reviewed journals;
- describes a single approach to analysis of disaster-related Twitter data;
- describes use of Visual Analytics (though did not need to be explicitly identified as such, it could be our judgment).

The databases searched included IEEE Explore (provided by the Institute of Electrical and Electronics Engineers), and the ACM (Association for Computing Machinery) portal. Once key papers had been identified, the citation lists of those papers were scanned, and Google Scholar was used to identify papers that had cited the key paper. Keywords used for the search included "Visual Analytics", "Twitter", "Disaster", "Emergencies", and "Critical event".

Our searches returned numerous papers, but many methods described did not simultaneously investigate all the three domains: spatial, temporal and semantic. Papers that related to one or two domains were rejected as potential case-studies, but some were retained for their contextual information and citations e.g. Onorati & Diaz [24], together with relevant reviews e.g. Wanner et al. [14].

Initially we found more than 10 papers that met our criteria to be case studies and made the final selection with the aim of including a range of methodological approaches and types of disasters. However, on detailed review we discovered that several of these papers described visualization rather than VA. We therefore relaxed the third inclusion criteria. The final selection of 10 papers includes one that describes a method identified as potentially of use in disasters [25] but has not been tested using disaster-related data.

Phase 2: Paper review

We used a checklist to extract and record relevant information from the 10 chosen papers in order to address research questions 1 and 2.

Phase 3: Synthesis of findings

We reviewed the findings from phase 2 and identified the common and divergent features of the various approaches. We used a table to extract the elements of the ten selected applications (see Table 1 below, which was loosely adapted from the framework of Wanner and colleagues [14]) and we examined the use of these elements in turn. Next, we used an iterative process to derive a framework to characterise the process of VA of disaster-related spatio-temporal Twitter data (see Figure 6 below). This framework facilitated the final stage of our review, that of an overall appraisal of the existing approaches and identification of the main challenges to their use.

2.3 Strengths and limitations of our review

As described above, our analysis and conclusions are based on a selection of case-studies. This approach is less rigorous than a systematic review, in which all available documented research that meets specified criteria is evaluated. Key papers may have been missed during our rapid search process, with the consequence that our conclusions may differ slightly from those reached if our review had been systematic. However, our approach enabled a quick and effective exploration of current methods, and our review provides a useful foundation for a future detailed exploration of the topic.

3 FINDINGS

Source [reference]		Application	Aims			Data collection & processing				Visualization						
			Identify	Manage	Analysis	Filtering	Clustering	Classification	Anomaly detection	Map	Timeline	Glyph	Node-link	Word cloud	Bar chart	Other
MacEachren et 2011	[15]	SensePlace2		X	X	X		X		X	X					
Marcus et al 2011	[26]	TwitInfo	X		X	X	X	X	X	X	X	X				
Cao et al 2012	[27]	Whisper			X	X		X		X	X	X	X			X
Thom et al 2012	[28]	ScatterBlogs	X	X			X		X	X	X			X		
Yin et al 2012	[29]	CCC	X	X		X	X	X	X	X	X			X		
Bosch et al 2013	[17]	ScatterBlogs2	X	X		X	X		X	X	X	X		X		
Morstatter et al 2013	[30]	TweetXplorer		X	X	X				X	X		X	X		X
Middleton et al 2014	[31]	TweetComP1	X	X		X	X		X	X	X	X			X	X
Steed et al 2015	[32]	Matisse			X	X		X		X	X					
Choi et al 2018	[25]	TopicOnTiles	X		X	X	X		X	X	X	X		X	X	

Table 1: Aims, data collection and processing methods, and visualization methods used in the 10 chosen case-studies.

Table 1 provides summary information about each of the applications described in our 10 case-studies. For each of the applications, the table shows the aims, and the methods for data collection, processing and visualization. This table facilitates identification of the common and divergent characteristics of the applications.

3.1 Aims of the applications

The main goal of papers that research on spatio-temporal events is trying to answer, in the best possible way, the questions of "where" and "when" an event is generated. This is the case in our review, however, it is possible to distinguish nuances in the objectives of each application (see Table 1).

Thus, we identified 3 main objectives: *Identify*, *Manage* and *Analyse* disastrous events using Twitter data. When we classify a document with the aim of *identify*, we refer to documents that focus their efforts on trying to distinguish where and when an event is generated, proposing tools that can be used for real-time detection. At the same time, when we talk about *manage*, we refer instead to applications designed to understand how an event evolves once it has already occurred, providing valuable information to mitigate a crisis. Finally, with *analyse*, we refer to documents which main goal is trying to explain, in detail, an event a posteriori, getting insights for future events that could have similar characteristics.

Of course, these objectives are not mutually exclusive, and in many cases they are intertwined. Thus, four of our applications reviewed [17, 28, 29, 31, 32] have as their main objective *identify* and *manage*, given that they are oriented not only to the detection of anomalous events, but also to facilitate situational awareness in emergencies. The authors base the effectiveness of their applications through case studies that include natural disasters such as earthquakes [28, 29, 31, 32], tsunamis [31, 32], hurricanes [28], among other events [17, 29]. Two applications have a greater focus on *manage* and *analyse* [15, 30], which also concentrate their efforts on supporting situational awareness for crisis events, however, they do not provide solutions for event detections. What these interfaces pursue is to provide various graphic tools to explore, characterize and compare information. As in the previous cases, they show examples applied to earthquakes [15] and hurricanes [30]. On the other hand, two applications are mainly aimed at *identify* and *analyse*. One [25] seeks to detect anomalous events from social media data, as well as to facilitate in-depth analysis on this kind of events, and another [26] proposes several methods to synthesize a coherent timeline of events and sentiment on a topic. The kind of case studies that those papers provide are more related on sports events [25, 26] and protests [25] rather than natural disasters. Finally, two applications concentrate only on the focus of *analyse*. [27, 32]. In the first case, the fundamental target is to understand how opinions spread away in a particular time and space dimension, using "retweet" information. In the second one, pursuing "to reveal key trends and associations in complex social media streams", incorporating a powerful data-processing operation, an automated sentiment and emotion analytics, as well as interactive visualization tool aimed to do exhaustive analysis. Real examples that they use are an application to political campaigns [27], an earthquake [27] and a terrorist attack [32].

As a final point, it should be noted that none papers reviewed present evidence of real-time use, in other words, applications that have been used effectively to *identify*, *manage* and/or *analyse* a catastrophic event while the disaster is actually evolving. Nevertheless, the authors base their potential use in concrete examples of case studies.

In the next section we describe methods, and distinguish which are common to all applications, and which are application-specific.

3.2 Methods: Automatic Processing

Given the challenges inherent to Twitter data, many papers propose automatic processing techniques to summarize and identify patterns in the data that are later presented visually to the human analyst. Different machine learning methods are being used depending on the specific use case, that is, the specific type of pattern that the analyst might be interested. For example, clustering methods are being used to automatically identify topics being discussed and aggregate tweets in those topics. Note that given the large continuous volume of tweets this task would be very difficult to handle by an unaided analyst. This section describes the methods and applications in the papers analysed.

Additionally, this section starts by briefly describing data capturing and pre-processing steps required before automatic processing steps.

3.2.1 Data capture and pre-process (Filtering)

Twitter API provides a continuous stream of tweets which unfiltered quickly become a huge dataset that in a very vast proportion will not be used for the task at hand. Twitter API allows to specify several filters to the stream, most common used are by location or keywords. When the solution proposed includes functionality to identify potential incidents based on frequency anomalies the stream is generally only filtered based on locations. For example, Yin and colleagues [29] designed their application for the Australian Crisis Coordination Centre (CCC) and consequently they are only interested in events in Australia (this application is hereafter referred to as CCC). The applications TopicOnTiles [25] and ScatterBlogs2 [17] have similar approaches.

Solutions proposed that do not include incident detection functionality generally filter further the data stream using user-specified keywords. For example, SensePlace2 [15] and TweetTracker [22] prefilter the dataset. This approach reduces data volume and could simplify architecture and technical design although it will limit the ability for the analyst to switch keywords and be able explore historical trends.

When automatic methods use the actual text of the tweet to extract patterns additional pre-processing steps are required. These generally include removing stop words, links and tokenise text into unigrams or n-grams (split texts into individual words). Additionally, after tokenisation texts are generally encoded into bag-of-words [for example, 25, 29]. By encoding in bag-of-words, each tweet is represented with a vector with length of vocabulary where the terms present in the text have scores and all non-present terms are 0.

3.2.2 Anomaly detection

Several papers propose automatic methods to detect potential incidents that require the analyst attention. For example, the method used in the CCC application is based on comparing long-term word or expression frequencies with relative frequencies in 5 minutes sliding windows [29]. Frequencies are modelled as the probability in a binomial model, when the relative frequency in the current window is abnormally higher than the baseline probability the event is identified as a bursty feature and an alert is shown to the analyst. Proposed burst-detection method successfully identified several emergencies in reference dataset. The authors report a true positive rate of 0.72 and false positive rate of 0.014 [29].

In TweetComP1, several key words related to specific incidents (e.g. earthquake) are tracked [31]. A user defines thresholds for each of the keywords, the alert system is triggered when in a sliding time window the frequency of the keywords surpass the predefined threshold.

The application TopicOnTiles [25] uses a method for anomaly detection that considers also spatial density of the word frequency and not only the time dimensions. Maps are split into fixed width and height tiles, the tool implemented compares density of words in the different tiles and identifies as anomalies tiles / words with higher density than rest of the tiles.

3.2.3 Clustering

Many of the papers analysed propose to use clustering methods as a means to aggregate and summarize the large volume of tweets for the analyst. Methods differ in the dimensions considered by the clustering algorithm, all of them do take into account time dimension because they are used to identify time constrained events. It is also possible to appreciate that the methods proposed are adapted for the continuous data stream so they all include mechanism to update and depreciate clusters with the newly received tweets. Another common factor identified is that authors favour methods that do not require a number of clusters as hyperparameters, and we consider that this is consistent with the nature of the data analysed.

The TwitInfo application [26] uses a method to analyse a posteriori events and be able to reconstruct its subtopic sequences. The method identifies peaks in a given timeline and assign labels to peaks automatically using word frequencies. Although it could be argued that the approach is not a formal machine learning clustering method, we include this method in this section as it is being used to aggregate/categorise tweets in time delimited subtopics. Labelling of peaks utilizes IDF normalization to remove noisy query terms.

The CCC application [29] uses an online clustering technique based on Term Frequency and Inverse Document Frequency. Method proposed does not require a predefined number of clusters, it forms clusters based on similarity in words contained in tweets. Two similarity measures are tried, Jaccard and cosine, and Jaccard obtains denser clusters. The authors also propose to adjust similarity measures by including a time factor, so clusters are time delimited.

The ScatterBlogs application [28] uses a method that is slightly different from those described above. Rather than assigning categories to tweets, it clusters the occurrence of individual words in the space-time. The centroids of the clusters are being used to assign locations in a map to the words in the word cloud.

Finally, TweetComP1 [31] uses hierarchical clustering considering only spatial and temporal dimensions on a subset of tweets prefiltered using keyword search. In contrast to the previous methods discussed, the use case for this method is not to identify subtopics but to locate in time-space the occurrence of user defined event.

3.2.4 Classification

In situation response and situation awareness it is necessary to be able to filter information relevant to broad topic, for example infrastructures. Given that such broad topics could be searched by many different key words, Yin and colleagues [29] propose utilizing SVM and Naïve Bayes classifiers to assist human analyst identify tweets talking about infrastructure. Methods seem to achieve better accuracy when including word bigrams and additional features as number of retweets. SVM achieves classification accuracy of 0.87 [29].

Another use-case of classifiers identified in several papers is sentiment analysis. In TwitInfo [26] a Naïve Bayes classifier is used to assign positive or negative sentiment to topics. The authors propose a method to normalize algorithm outputs so they can be aggregated given that classifiers have different recall for positive and negative sentiments. It is important to note that the user group evaluating the application seemed to agree that sentiment analysis was not useful for their task. Also the applications Whisper [27] and Matisse [32] use sentiment analysis. It is important to note that the three applications using sentiment analysis are not specifically designed to support emergency management but to monitor general real-time events.

3.3 Methods: Visualization and user interaction

In this section, a brief comparison between papers for each visualization tool will be presented with an emphasis on the applications that stand out above the others for each case.

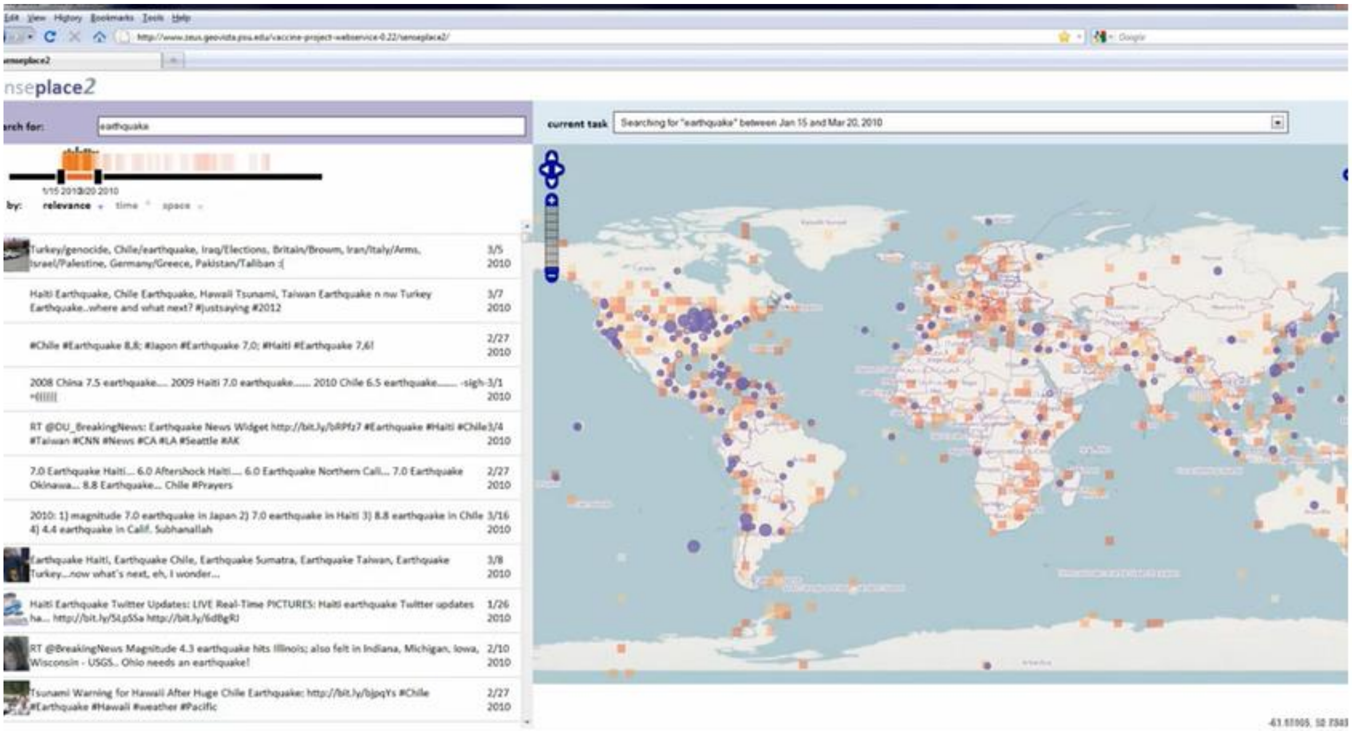

Figure 1: Map visualization of tweets, SensePlace2
(from https://www.youtube.com/watch?v=Nx9Cqcl_SAs)

3.3.1 Map

Given the context of this literature review, maps represent the most effective way to show geographic locations where crisis events occur. For this reason, each one of the applications reviewed use this tool to display the information about tweets collected and processed in the previous stages.

Some applications that stand out in terms of the use of maps as visualization tools are SensePlace2 [15] and TweetXplorer [30]. In both, a map of the world showing the distribution on tweets is available to the user, as seen in Figure 1. These applications also give analysts the ability to interact easily by scrolling over the map, selecting groups of tweets and zooming in and out of locations.

TweetXplorer [30] and TopicOnTiles [30] add a heatmap functionality based on kernel density estimations (an algorithm that assigns a higher probability to elements that are closer to each other). These regions with a higher density of tweets are then assigned to a more intense hue, whereas a softer palette of colours is used to less dense locations. The use of this technique facilitates the visualization of geographical locations where tweets tend to densify, and it is especially useful to analyse the displacement of tweets in different time windows.

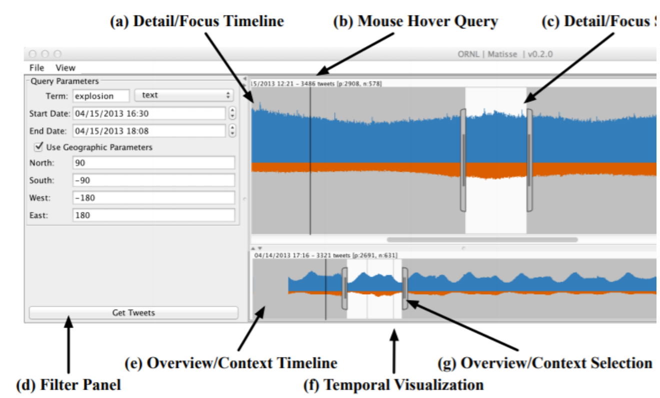

Figure 2: Timeline for event detection and sentiment analysis, Matisse [32]

3.3.2 Timeline

Some version of a timeline is also included in all applications. This kind of visualization tool is generally used to identify peaks and visualize frequencies of tweets over time.

For example, TweetXplorer [30] allows users to query groups of keywords for different time periods such as before, during and after the occurrence of a crisis event. Then, the timeline provides them a fast way to observe peaks in frequency of tweets for each group of keywords, which is very useful to understand the development of the event over time.

Another interesting use of timeline as a visualization tool is presented in Matisse [32]. In this case, the timeline is used in a river format to show both frequency of tweets and positive/negative categorisation of the message using sentiment analysis, as seen in

Figure 2. This allow users to identify not only peaks on general frequency but also for each class of sentiment over time, which can be useful for the posterior analysis of a crisis event.

3.3.3 Node link

Just Whisper [27] and TweetXplorer [30] use node links as visualization tools. The reason for this is their interest in how the message is spread when an event take place. In both, this tool is used to represent the retweet network, i.e. the relationship between original tweets and the ones that are just replications.

In the case of TweetXplorer [30], the retweet network and its node link representation are proposed to be an auxiliary tool for analyst to identify and characterize prominent users for a better understanding of tweet content.

On the other hand, the way information diffusion take place is the main interest in Whisper [27]. This application uses node links on top of a circular map of the world showing the nodes at the centre. When a retweet happens, arrows connect the nodes with retweets in different parts of the world showing how the message spread in a time window.

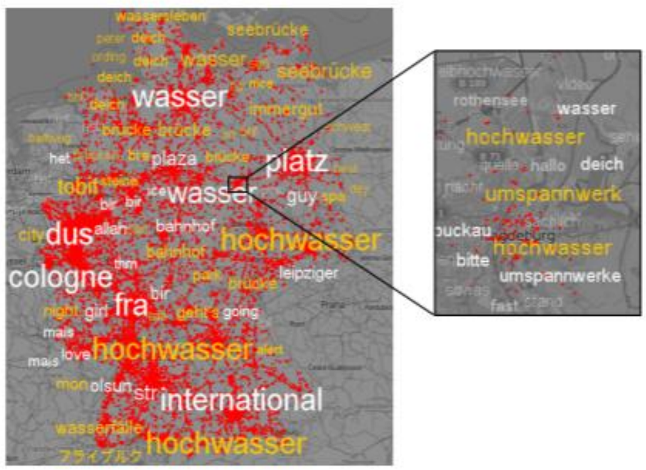

Figure 3: Word cloud in a geographical context, ScatterBlogs [28].

3.3.4 Word cloud

A word cloud is a very useful technique to show the most relevant words that are being tweeted in a particular area. In fact, this tool takes an important role in ScatterBlogs [17, 28], CCC [29] and TopicOnTiles [25], meanwhile for TweetXplorer it is used in particular cases [30].

The CCC application [29] uses word clouds that differentiate by colours and font size the words that are more frequently used for a cluster that has been obtained in previous phases of the process, without displaying them in a geographical position.

Both ScatterBlogs [17, 28] and TopicOnTiles [25] use a similar approach in terms of the use of a hue scale and different font sizes for most frequent words, but they also show them in their corresponding geographical position. The example of the first one is shown in Figure 3. This last characteristic makes them practical for users either in the context of identification, management or analysis of any crisis events.

Lastly, TweetXplorer [30] incorporates the Word Cloud tool in the context of an analysis of a particular Twitter user, showing with bigger font sizes the most important words and hashtags that has been used by the individual in the time period specified.

3.3.5 Glyph

Glyphs are used in TopicOnTiles [25] to represent the relative frequency of occurrences of each topic in a tile (i.e. zone of the map which has been subdivided using a regular grid into equal-sized areas).

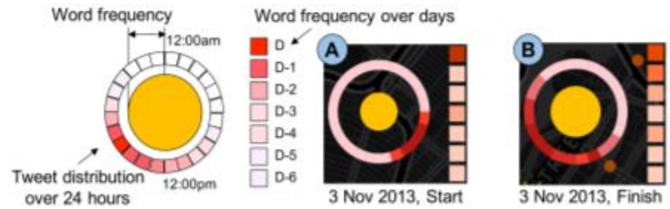

Figure 4: Glyph visualization of temporal frequency for words Start and Finish, TopicOnTiles [25]

In this case, it takes the form of a "donut chart-like glyph (see Figure 4). There are two layers - the radius of the inner layer represents the total count of tweets with the user-selected keyword within the tile, and the outer shows the hourly distribution of tweets containing the keyword over 24 hours. The colour becomes darker as the frequency of the keyword increases. The outer ring indicates the particular time when the event occurs. User feedback indicated that more information may be provided than can be effectively assimilated, see section 4.3.

3.3.6 Other visualization tools

In TopicOnTiles [25], bar charts and pie charts are used as secondary means to support their geographical analysis. Also tiles in which anomalous events have been detected are highlighted using borders with coloured thick edges. In TweetComP1 [31] bar charts are used to show frequency of tweet clusters classified by their relevancy score.

3.3.7 User interaction in general

All applications offer an interactive map with the capability to pan and zoom.

With the exception of CCC [29] all applications provide an interactive timeline to select a time period. A slider is used to track the temporal development of an event.

All applications offer filters for keywords or sets of keywords, time periods and geolocations. These filters can be applied at time of data capture as well as during the process of event monitoring and managing.

Applications developed for event detection offer the option to drill down by selecting the event. This is done via an interactive map, the selection of a time interval or a set of keywords. This sets the focus on the terms, time, geolocation, or a combination thereof. In all applications the analyst can perform this operation down to the detail level of an individual tweet.

Of the ten applications reviewed, ScatterBlogs2 [17] offers the most comprehensive tools for user interaction, and this case-study is therefore described in more detail below:

Bosch and colleagues identified two challenges: the large volume of data might require keyword filtering, but a too narrow set of keywords might miss valuable data. On the other hand, low occurrence data with high relevance to the anomaly might be missed. They propose a two-step solution with ScatterBlogs2 [17].

The first step is the creation of filters and classifiers. A domain expert or analyst will use a historic event and tweets collected for that time period in the "Classifier Creation Environment". To broaden the filter all keywords are extracted from the subset of data the analyst identifies as relevant. The terms or term co-occurrences with the highest frequency are identified by the application. The analyst can evaluate the list and add or delete additional terms [17].

Classifiers offer better results for low occurrence events. The domain expert again chooses messages that are relevant and messages that are not relevant to train a classifier based on the support vector machine framework. Both filters and classifiers are saved for later use by the system [17].

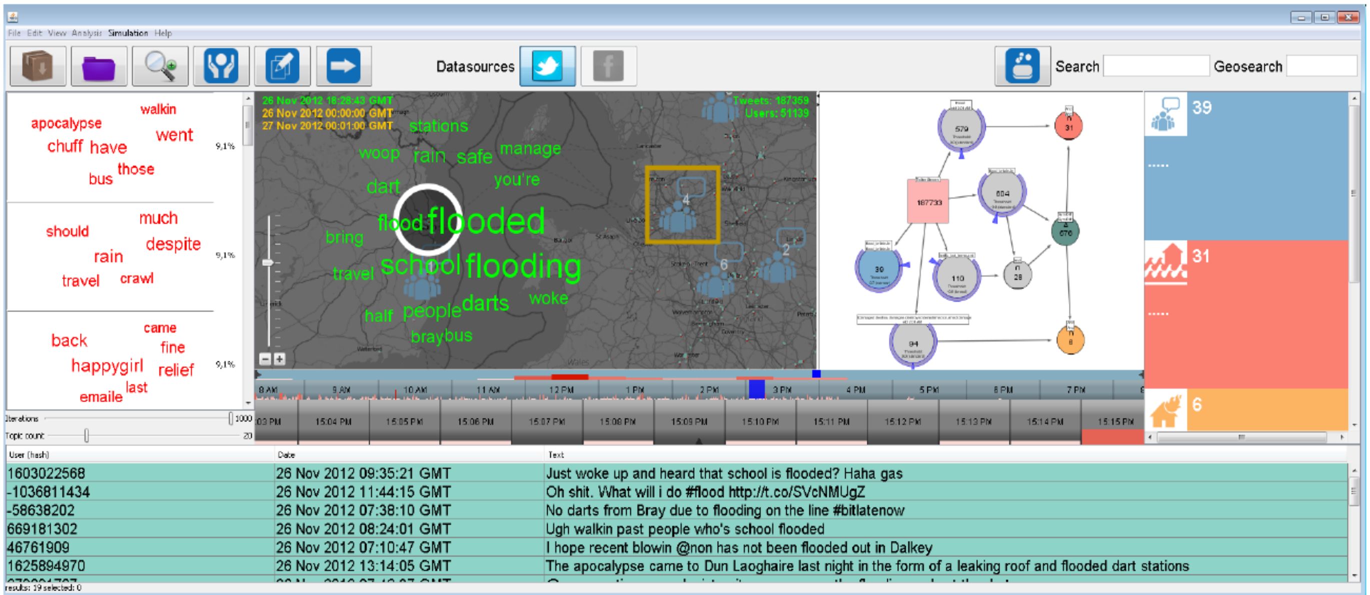

Figure 5: "Visual Filter Orchestration", ScatterBlogs2 [17]

The second step is the actual process of event monitoring. The authors propose "Visual Filter Orchestration". A graphic user interface enables the analyst to add filters and classifiers to the unfiltered data stream. The output of such a filter node can serve as the input for another filter, "combination nodes" allow the aggregation of outputs. The graphic user interface displays the number of resulting messages, allowing the analyst to adjust thresholds. The resulting data can be assigned a colour, a name, and a symbol for display on the interactive map [17].

4 Discussion

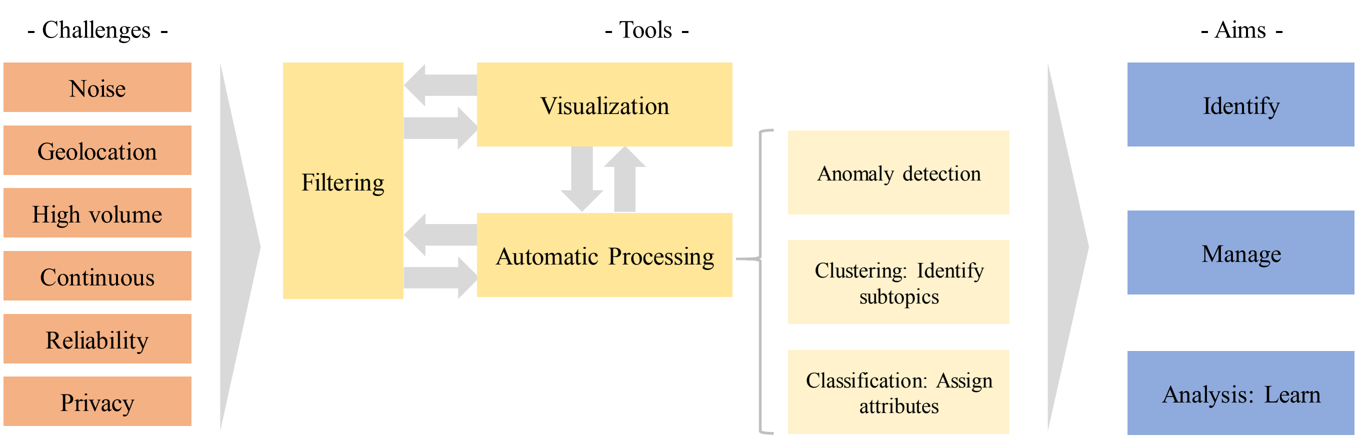

Figure 6: Framework representing analysis and utilization of disaster-related tweets in the context of data-related challenges

In the previous section we presented findings from our review of 10 selected papers, in which we identified the methods used in each of the applications and the questions asked of the data. Figure 6 shows the framework we derived based on these findings, which summarises the process of VA of disaster-related spatio-temporal Twitter data. As outlined in the left-most column, the framework introduces one important new element of our review – not discussed earlier because it is pertinent to all the applications – that of challenges. There are various characteristics of Twitter data that constrain the potential value of such data in the context of disasters, and these are discussed below, together with other limitations on the potential effectiveness of the approaches used, and of research into this domain.

4.1 Challenges

Noise: User generated Twitter posts are inherently noisy due to the limitation of 140 (later 280) characters, but also due to abbreviations, orthographic mistakes, and the colloquial use of language [13, 14]. This poses a challenge to natural language processing tasks.

Geolocation: Twitter allows to add geolocation meta-data to posts, but users may disable this feature due to privacy concerns. Fuchs and colleagues conducted an experiment using a known event and geolocated Tweets for the relevant time period [13]. The experiment was conducted in Germany, where privacy concerns result in a low proportion of geolocated tweets. The authors concluded that "spatio-temporal visual analysis of the data still allows to detect significant events with reasonable accuracy" [13]. Other authors concur that an event will generate a sufficient amount of geolocated posts [15]. There exist several alternative approaches to assign geolocation to tweets with missing geolocation such as named entity recognition [14].

High volume and continuous data streams: Two challenges have been identified with respect to volume, the first is technical, and requires efficient algorithms and data handling that enable processing the large volume of data generated. The second one is the "temporal context", that is "new data has to be shown in the context of older data" [16].

Reliability: Twitter users have been likened to "social sensors" which are less reliable than physical sensors due to their unpredictability [20]. Domain experts have questioned the credibility of Twitter data, and emphasise that human interaction is still needed [33]. In particular, emergency managers can be doubtful about the quality of social media data [3]. Thus resistance to the inclusion and use of social media data in the emergency control room itself [31] must be taken into account by designers of applications.

Privacy and Data Protection: Only a small proportion of papers reviewed address the aspect of data rights and privacy in their discussion. Bosch and colleagues question whether users are conscious that microblog posts are public, and suggest a solution of anonymising user and message IDs [17].

The lack of consideration of data policies, ownership, data rights, privacy, and ethics has been noted by others reviewing the literature on use of geo-social media for disaster management [34]. Improved data governance to clarify who owns the data and to define limits of use is needed [34].

4.2 Lack of evidence from real-time disasters

The case-studies papers we reviewed only described analysis of historical data. Some experts suggest this is because "real-time analysis of social media data aggravates many of the challenges", and "there is still a lack of widely available real-time analytic tools to handle and process in real time vast amounts of data streams" [34], Given the recent development of powerful Data Science tools, we would challenge this view, however we do not have evidence to do so since the literature does not include accounts of application use in actual disasters.

We know that one application has gone into private production so this might account for the lack of recent literature (ScatterBlogs: https://www.scatterblogs.com/)

The lack of published studies from actual rather than historical disasters makes it difficult for us to properly critique the existing techniques.

4.3 End-users missing in design

Several of the papers reviewed did not explicitly indicate an end-user or stakeholder. Those that do, indicate emergency managers [for example, 31], but frequently the needs of emergency operators do not appear to have been kept in mind. Not many examples of collaboration with emergency operators at the design stage can be found in the literature analysed, although when there is collaboration it is possible to find clearer objectives for the applications of visualization or algorithms [for example, 28, 29].

The lack of input by end-users in the design phase might translate in some cases in information overload as reported by some users [25, 27, 32]. For example one participant in a case-study stated '*there exist too many visual components, and sometimes it is confusing which one I should be looking at*.' [25].

It is also important to restate that we have not identified any literature describing how these applications are being used and the value that they add in a real-time disaster as described in section 4.2 above.

4.4 Performance evaluation

Evaluation of the applications is not always included in the case-studies reviewed. With respect to user feedback analysed as part of the case-study, with the exception of the issue of information overload mentioned in the previous section, feedback was uniformly positive [17, 26, 27, 32]. This supports the idea of applications being effective in summarizing Twitter data when this task is generally challenging for a human analyst. In general, applications analysed seem to be helpful to extract additional understanding in crisis events but input of final human criteria is always needed in all case-studies.

Regarding the performance of statistical learning methods, we noted that scores for clustering algorithms are very rarely reported so readers are unable to judge how well the clustering algorithms are performing. Classification and anomaly detection scores are most commonly included, including false positive rate for anomaly detection applications [31].

4.5 Bias in situation awareness inference

Tweets, as social sensors for natural events or impact assessment in specific location, can be biased because they rely on telecommunication infrastructure that could be damaged in natural disasters. Also, Twitter users might not be uniformly distributed in geographical location. Consequently, situation reflected by the Twitter data might have a strong bias and mislead crisis management efforts. We have not identified approaches to adjust or understand these possible biases.

On the other hand, lack of social media activity could potentially be used as evidence of major destruction as long as activity is detected in nearby regions [31].

5 Conclusion

This paper reports a rapid review of the literature concerning visual analytic approaches used on spatio-temporal Twitter data, for the identification, management and analysis of disasters.

We have not identified a clear evolution of the methods over the period when these periods have existed in the use-cases analysed. Twitter data is a use case for Natural Language Processing (NLP) and several papers use NLP models in their applications. But the NLP field has clearly evolved in recent years due to broader Deep Learning method, and we did not see these new approaches being incorporated by the authors.

We found that VA methods are potentially effective in assisting human analysts to summarize and extract patterns from large volumes of rapidly changing data. Nevertheless, given the specific context of emergencies, where rapid decision-making is essential, we found that the issue of information overload is not sufficiently addressed by the papers analysed.

Our review is only based on 10 case studies. Key papers may have been missed during the rapid search process, so that our conclusions may differ slightly from those reached if the review had been systematic. However, this review does provide a useful overview of existing techniques which could be easily built on by other researchers to provide more robust findings.

Acknowledgments

The authors wish to thank Professor Gennady Andrienko for his advice.

References

[1] UNOCHA (United Nations Office for the Coordination of Humanitarian Affairs). *World Humanitarian Data and Trends 2017.* Policy Development and Studies Branch of OCHA; 2018.

[2] Gencer EA. *Natural Disasters, Urban Vulnerability, and Risk Management: A Theoretical Overview.* In The Interplay between Urban Development, Vulnerability, and Risk Management: A Case Study of the Istanbul Metropolitan Area. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013: 7-43

[3] Reuter C, Hughes AL, Kaufhold M-A. Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research. *International Journal of Human–Computer Interaction* 2018; 34:280-294.

[4] Dhar V. Data science and prediction. *Communications of the ACM* 2013; 56:64-73.

[5] Keim D, Andrienko G, Fekete J-D, Görg C, Kohlhammer J, Melançon G. *Visual analytics: Definition, process, and challenges.* In Information visualization. Springer; 2008: 154-175

[6] Blytt M. *Big challenges for visual analytics: Assisting sensemaking of big data with visual analytics.* Norwegian University of Science and Technology; 2014.

[7] Keim D, Kohlhammer J, Ellis G, Mansmann F (Eds.): *Mastering the information age solving problems with visual analytics*. Goslar, Germany: Eurographics Association; 2010.

[8] Atluri G, Karpatne A, Kumar V. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)* 2018; 51:83.

[9] Andrienko N, Andrienko G. A visual analytics framework for spatio-temporal analysis and modelling. *Data Mining and Knowledge Discovery* 2013; 27:55-83.

[10] Twitter. *Investor Fact Sheet, Q2, 2018.* San Francisco, CA: Twitter, Inc.; 2018.

[11] Martínez-Rojas M, del Carmen Pardo-Ferreira M, Rubio-Romero JC. Twitter as a tool for the management and analysis of emergency situations: A systematic literature review. *International Journal of Information Management* 2018; 43:196-208.

[12] Resch B, Usländer F, Havas C. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science* 2018; 45:362-376.

[13] Fuchs G, Andrienko N, Andrienko G, Bothe S, Stange H. *Tracing the German centennial flood in the stream of tweets: first lessons learned.* In Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information. pp. 31-38. Orlando, Florida: ACM; 2013:31-38.

[14] Wanner F, Stoffel A, Jäckle D, Kwon BC, Weiler A, Keim DA. *State-of-the-Art Report of Visual Analysis for Event Detection in Text Data Streams.* In *EuroVis*. 2014

[15] MacEachren AM, Jaiswal A, Robinson AC, Pezanowski S, Savelyev A, Mitra P, et al. *SensePlace2: GeoTwitter analytics support for situational awareness.* In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST); 23-28 Oct. 2011*. 2011: 181-190.

[16] Rohrdantz C, Oelke D, Krstajic M, Fischer F. *Real-time visualization of streaming text data: Tasks and challenges.* In *VIS-Week*. 2011

[17] Bosch H, Thom D, Heimerl F, Püttmann E, Koch S, Krüger R, et al. Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *IEEE Transactions on Visualization and Computer Graphics* 2013; 19:2022-2031.

[18] Carley KM, Malik M, Landwehr PM, Pfeffer J, Kowalchuck M. Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. *Safety science* 2016; 90:48-61.

[19] Landwehr PM, Wei W, Kowalchuck M, Carley KM. Using tweets to support disaster planning, warning and response. *Safety science* 2016; 90:33-47.

[20] Sakaki T, Okazaki M, Matsuo Y. *Earthquake shakes Twitter users: real-time event detection by social sensors.* In Proceedings of the 19th international conference on World wide web. pp. 851-860. Raleigh, North Carolina, USA: ACM; 2010:851-860.

[21] Kejriwal M, Gilley D, Szekely P, Crisman J. *THOR: Text-enabled Analytics for Humanitarian Operations.* In Companion Proceedings of the The Web Conference 2018. pp. 147-150. Lyon, France: International World Wide Web Conferences Steering Committee; 2018:147-150.

[22] Kumar S, Barbier G, Abbasi M, Liu H. *TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief.* In *Fifth International AAAI Conference on Weblogs and Social Media*. 2-11

[23] Jamali M, Nejat A, Ghosh S, Jin F, Cao G. Social media data and post-disaster recovery. *International Journal of Information Management* 2019; 44:25-37.

[24] Onorati T, Díaz P. Giving meaning to tweets in emergency situations: a semantic approach for filtering and visualizing social data. *Springerplus* 2016; 5.

[25] Choi M, Shin S, Choi J, Langevin S, Bethune C, Horne P, et al. *TopicOnTiles: Tile-Based Spatio-Temporal Event Analytics via Exclusive Topic Modeling on Social Media.* In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1-11. Montreal QC, Canada: ACM; 2018:1-11.

[26] Marcus A, Bernstein M, Badar O, Karger D, Madden S, Miller R. *Twitinfo: aggregating and visualizing microblogs for event exploration.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 227-236. Vancouver, BC, Canada: ACM; 2011:227-236.

[27] Cao N, Lin Y, Sun X, Lazer D, Liu S, Qu H. Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time. *IEEE Transactions on Visualization and Computer Graphics* 2012; 18:2649-2658.

[28] Thom D, Bosch H, Koch S, Wörner M, Ertl T. *Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages.* In *2012 IEEE Pacific Visualization Symposium; 28 Feb.-2 March 2012*. 2012: 41-48.

[29] Yin J, Lampert A, Cameron M, Robinson B, Power R. Using Social Media to Enhance Emergency Situation Awareness. *IEEE Intelligent Systems* 2012; 27:52-59.

[30] Morstatter F, Kumar S, Liu H, Maciejewski R. *Understanding Twitter data with TweetXplorer.* In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1482-1485. Chicago, Illinois, USA: ACM; 2013:1482-1485.

[31] Middleton SE, Zielinski A, Necmioğlu Ö, Hammitzsch M. *Spatio-Temporal Decision Support System for Natural Crisis Management with TweetComP1.* In Decision Support Systems III - Impact of Decision Support Systems for Global Environments. pp. 11-21. Cham: Springer International Publishing; 2014:11-21.

[32] Steed CA, Drouhard M, Beaver J, Pyle J, Bogen PL. *Matisse: A visual analytics system for exploring emotion trends in social media text streams.* In *2015 IEEE International Conference on Big Data (Big Data); 29 Oct.-1 Nov. 2015*. 2015: 807-814.

[33] Thom D, Krueger R, Ertl T. *Can Twitter Save Lives? A Broad-Scale Study on Visual Social Media Analytics for Public Safety.* In *IEEE Transactions on Visualization and Computer Graphics*. 2015: 1-1.

[34] Granell C, Ostermann FO. Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management. *Computers, Environment and Urban Systems* 2016; 59:231-243.