Multimed Tools Appl
DOI 10.1007/s11042-013-1804-2

Social media for crisis management: clustering approaches for sub-event detection

Daniela Pohl · Abdelhamid Bouchachia · Hermann Hellwagner

Abstract Social media is getting increasingly important for crisis management, as it enables the public to provide information in different forms: text, image and video which can be valuable for crisis management. Such information is usually spatial and time-oriented, useful for understanding the emergency needs, performing decision making and supporting learning/training after the emergency. Due to the huge amount of data gathered during a crisis, automatic processing of the data is needed to support crisis management. One way of automating the process is to uncover sub-events (i.e., special hotspots) in the data collected from social media to enable better understanding of the crisis. We propose in the present paper clustering approaches for sub-event detection that operate on Flickr and YouTube data since multimedia data is of particular importance to understand the situation. Different clustering algorithms are assessed using the textual annotations (i.e., title, tags and description) and additional metadata information, like time and location. The empirical study shows in particular that social multimedia combined with clustering in the context of crisis management is worth using for detecting sub-events. It serves to integrate social media into crisis management without cumbersome manual monitoring.

Keywords Social media · Sub-event detection · Clustering · Information search and retrieval · Crisis management

D. Pohl (✉) · H. Hellwagner
Institute of Information Technology (ITEC), Multimedia Communication (MMC),
Alpen-Adria-Universität Klagenfurt, Universitätsstrasse 65-67,
9020 Klagenfurt am Wörthersee, Austria
e-mail: daniela@itec.uni-klu.ac.at

H. Hellwagner
e-mail: hermann.hellwagner@aau.at

A. Bouchachia
Smart Technology Research Center, School of Design, Engineering & Computing,
Bournemouth University, Poole House Talbot Campus, Fern Barrow Poole BH12 5BB, UK
e-mail: abouchachia@bournemouth.ac.uk

Published online: 27 December 2013

1 Introduction

Emergency management, especially in relation to large-scale disasters, has to deal with hundreds of different parties. It usually consists of three major phases [31]: preparedness, response, and recovery. Preparedness deals with awareness and monitoring aspects. Response focuses on actions to be taken for stabilizing the situation when the emergency already occurred. Recovery aims at reestablishing the original infrastructure existing before the crisis. In each of these phases, any kind of information regarding the situation at hand is vital for understanding, decision making, and learning.

Social (multi-)media turns out to be an important source of information. The high acceptance of social media in the public and the introduction of mobile social media applications make it possible to document any kind of situation people are currently involved in. People's willingness to document such information (e.g., in Egypt [7]) makes social media an interesting communication medium.

Having access to information about a crisis is of high importance indeed; it enables the first responders to be on-site from the very beginning of a crisis. Various studies underline the importance of social media during different crises [18, 39]. Yates and Paquette [42] describe the introduction of special social media platforms as knowledge management tools for information sharing and collaboration between different aid organizations.

Depending on the scale of an emergency situation, a huge amount of information may be accumulated. Hence, it is necessary for the emergency management parties to use such information after purposeful analysis. Motivated by this idea, we introduced a *Media Exploration Framework* [26] and begun studying the suitability of clustering techniques for the detection of sub-events considering the content of social media. Sub-events are special hotspots arising during a crisis situation, which need special attention (e.g., a flood in a specific district of a city). The detected sub-events give the possibility to understand the crisis better and to be prepared for similar disasters (e.g., which is very important for training purposes). Because emergency management benefits from multimedia information, especially videos [4], to gain a better view of the situation, we have focused so far on social media sharing tools like Flickr and YouTube as bases for the detection.

In this paper, we examine different clustering algorithms based on an evaluation framework introduced using various metadata information (text, time, and location) of pictures and videos collected from Flickr and YouTube. We used algorithms from our previous work [24–26]. The algorithms combine traditional methods, i.e. Self-Organizing Maps (SOM) and Agglomerative Clustering (AC), but we use them in the new application context of emergency management. In addition, we extended one algorithm [24] by using location and time information. The evaluation shows the potential of clustering algorithms for identifying sub-events. The algorithms become part of our exploration application for crisis management and are discussed with ample details from different perspectives.

The present work is structured as follows. Section 2 gives an overview of related work. Section 3 describes our understanding of sub-events and their detection methodology. It focuses on different metadata aspects that are relevant for the detection process and describes the clustering algorithms applied. Sections 4 and 5 present the evaluation framework and the results respectively. Finally, Section 6 concludes the work.

2 Related work

The high acceptance of social media by the public has generated growing research interests. Much of it is related to topic and event detection for Twitter. For instance, Marcus et al. [20] describe a method for summarizing the events extracted from Twitter data. A similar idea is investigated by Mathioudakis and Koudas [21] aiming at trend detection from tweets. Petrović et al. [23] present an analysis approach for dealing with tweet streams.

The importance of Twitter in emergencies is stated by different researchers, e.g., Vieweg et al. [39]. In particular Terpstra et al. [35] show the realtime analysis of Twitter in a specific emergency situation using keyword-based data filters. Rogstadius and Kostakos [29] show how Twitter can be considered in emergency scenarios. Ireson et al. [14] introduce an approach for analyzing forums related to a specific city where crisis-related information is posted.

Moreover, Liu et al. [18] state the importance of social media sharing platforms, like Flickr and YouTube. Rattenbury et al. [28] present an approach to find tags which are related to specific events described in Flickr; while Fontugne et al. [12] analyze Flickr tags with respect to the Tohoku earthquake and the tsunami in Japan. Tucker et al. [37] discuss the introduction of collational interfaces, including a map-based visualization and different ideas on how to filter the raw data related to an emergency situation.

Additionally, geo-data has been used in various research areas and applications, since it serves as a clue for identifying events. For instance, Yin et al. [43] use geo-tagging for the visualization of analyzed Twitter information. Jaffe et al. [15] also use geo-data in combination with a hierarchical clustering algorithm to visualize and rank geo-annotated images. In general, the identification of relevant locations in particular situations can be easily done using geo-data [44].

On the other hand, the temporal aspect for event detection is considered in many studies. Becker et al. [3] describe a classification framework for event detection based on location and time information. Petkos et al. [22] also identify social events based on a combination of classification and multimodal clustering using the time stamp. In other studies [12, 28], not only time is used for event detection but also the location information.

There is also a relation to research regarding the use of topic models. Blei et al. [5] describe a topic model based on textual features, where documents can be assigned to several topics identified via training data. Additionally, Han et al. [13] describe a topic model for classifying images based on visual features using Latent Semantic Analysis.

Our goal in this paper is to identify sub-events that occur in a crisis context to support crisis management. We make use of different metadata information directly and apply various clustering algorithms. Clustering as a methodology is chosen as it does not require labeling media items as in classification investigated in the state-of-the-art approaches. Moreover, it is difficult to replicate the application of classifiers in other crisis situations unless the related data is labeled and the classifiers are retrained, because each crisis has its own characteristics (area of occurrence, scalability, impact, etc.) and people use different terminology to describe the events.

3 Sub-event detection

An emergency situation does not necessarily relate to a single large incident [42], but it is very often the case that in a large scale crisis, like a tsunami or earthquake, many side

Table 1 Data sets

Title	Period in 2011	No. of images, videos	Geo. data (#pics, #videos)
Mississippi Flood (MF)	04–19 May	2039, 442	573 (535, 38)
Oslo Bombing (OB)	22 July	31, 222	13 (2, 11)
UK Riots (UK)	06–10 Aug.	178, 274	19 (11, 8)
Hurricane Irene (HI)	23–29 Aug.	455, 700	200 (153, 47)

events take place at different locations at different times [41]. That is, an emergency can be seen as a *complex event* consisting of many incidents, called *sub-events* also known as *mini-crises* [42]. For example, a complex event (referred to simply as "event") could be the *UK riots in 2011* and one related sub-event would be, e.g., *looting in Hackney London*. Thus, a set of sub-events forms the overall event. Our goal is to identify such sub-events, as they need special attention from the emergency management team [26].

For our experiments on sub-event detection, we use data from Flickr and YouTube describing crises that took place in 2011 (see Table 1). We extracted information on videos and pictures via the APIs[1] provided for Flickr and YouTube. The platforms are queried based on user-supplied keywords, describing best the initial situation of the crisis (e.g., as it would be the case when one initially hears or gets informed about the crisis at hand).

The data sets contain media items (e.g., images and videos) with the corresponding metadata annotation. The size of data sets along with that of geo-referenced data[2] are depicted. The data sets describe major incidents in 2011 (e.g., UK riots or the Norway attacks) with different scales and origins. Also natural disasters, such as the Mississippi flood and the impact of the Hurricane Irene are covered. The data sets have been created from Flickr and YouTube using a keywords-based harvesting approach. Keywords are used that describe best the current crisis (e.g., *Oslo bomb explosion* or *UK riots*). Due to the extensive amount of evaluation results, we selected two representative data sets, namely UK riots (UK) and Hurricane Irene (HI).

3.1 Information dimensions

The structure of social media data, especially from Flickr and YouTube, consists of two parts: metadata (e.g., title, description, etc.) and content (e.g., image, video, etc.). Content is a valuable source of information, but due to time limitation it is often not possible to see an entire video to judge about the content, even less so in a stressful situation like an emergency. Thus, metadata (referred to as context as well) of data is a useful source [33]. For example, a fire shown on a picture does not necessarily tell much about the location in detail, unless additional metadata is used. For this reason, we focus in a first step on metadata to identify sub-events. Here we distinguish between three dimensions of information related to metadata, within which a sub-event detection algorithm can operate.

One dimension describes the *annotations* comprising all texts (*title, description*, and *tags*) attached to a media item. Annotations give a detailed context about what people find

[1] See http://flickrj.sourceforge.net/, https://developers.google.com/gdata/

[2] The precision of the geo-coordinates depends on the tagging behavior of the user and the platform offering the tagging mechanism, e.g., for Flickr when placing it on a map it is the sixth decimal.

important in the given situation. Additionally, there is *geo-referenced data* describing a location, showing a specific place. It describes, for example, where a picture was taken. The last dimension assigns the *time* aspect related to each item.

Because of the characteristics of social media, it is often the case that media items lack information in different dimensions. Thus, algorithms can operate on the individual dimensions or on a combination of them. We applied various algorithms operating on those different dimensions for detecting sub-events.

3.2 Sub-event detection approaches

For sub-event detection, we rely on clustering techniques because they are unsupervised learning techniques needing no excessive preparation phase before they can be implemented. The aim of the work is to show the potential of clustering for identifying sub-events using the different dimensions. Additionally, the presented approaches can be used for an after-the-fact analysis (e.g., for training purposes), giving an overview of the situation and identifying areas of important incidents.

3.2.1 Media item representation for clustering

For clustering, a suitable representation of the data for each dimension is needed. We use the well established *Vector Space Model (VSM)* for data representation going back to Salton et al. [32]. Each media item is a vector that contains the values of the representative attributes. Since the annotations of the media items are used in this study, the textual metadata fields (title, description, tags) are used to produce a weighted term-based representation by applying the *term frequency-inverse document frequency* (tf-idf) [19, 32]. This results in a so called term vector for each media item which is obtained by tagging, stemming and weighting of the terms. To further enhance the quality and the compactness of the representation, synonymy is used so that similar words (e.g., blast, explosion) are treated as one term. We used WordNet [11] to identify synonymic terms. We consider different relationships between terms, such as synonyms and derivated-from of WordNet.

The geo-references are given in longitude and latitude coordinates, hence no preprocessing is required as they are in a suitable format for clustering. On the other hand, time is transformed into milliseconds that passed since the 1st January 1970.

Depending on the applied algorithm and the information dimension, the representation vector consists in the general case of the 3 parts: term-based representation, the location and time representation.

3.2.2 Approaches

In this study, the application of four algorithms is investigated: *Self-Organizing Map* (SOM), *Agglomerative Clustering* (AC), *2PhaseGeo* (2PG) and *2PhaseGeoTime* (2PGT). In our first experiments [25, 26], we mainly focused on text-based annotation of media items for identifying sub-events using SOM and AC. Appendix A shows the general processing steps of SOM and AC. Hence, only textual metadata fields (title, tags, and description) of each media item are considered to compute the tf-idf representation which is then used for identifying sub-events by means of clustering. To capture the meaning and the semantical content of the resulting clusters, the most frequently occurring terms in each cluster are used to label them.

Geo-referenced data, when available, can be utilized to strengthen and support the sub-event detection. A natural extension of our previous work is to incorporate the geo-references and perform a clustering in two steps (2PhaseGeo) [24]. The 2PhaseGeo approach works as follows:

- **Pre-processing:**
  - Performing tf-idf analysis of the textual annotations (title, description, and tags) of the media items.
  - Localization of data containing geo-references.
- **Two-phase Clustering:**
  - Phase 1: Geo-referenced data is clustered using *Self-Organizing Maps (SOM)* [16].
  - Phase 2: The remaining data items (those lacking geo-reference) are assigned to the clusters produced in the previous step using *cosine similarity measure* [19] resulting in updated clusters.
- **Visualization:** Visualization of the geo-referenced data, clusters' centers, and the labels of the clusters in a map (OpenStreetMap was used http://www.openstreetmap.org)

In the first step of the clustering, media items tagged with geo-references are clustered together via SOM to identify crisis locations. For each resulting cluster given by its geo-referenced centroid (V_1), the textual annotations are then used to compute an additional term-based centroid (V_2). The final centroid of the geo-referenced data is thus a concatenation $V =< V_1, V_2 >$. V_1 is basically used as a starting point only. In the second step, the rest of the data items are assigned to the best fitting term-based centroids V_2's resulting from the first step using the cosine measure.

The results of the 2PhaseGeo algorithm are displayed on a map using the geo-referenced centroids V_1's (see Fig. 1 as an example). The compound labels of each cluster are displayed as well. They contain the highest tf-idf values of the term vector. For a better illustration, the locations of the geo-referenced items on the map are marked as red dots.

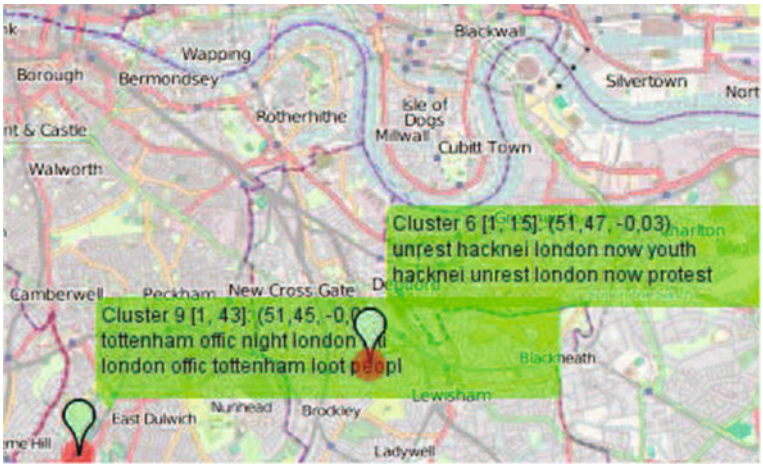

Fig. 1 Example for the 2PhaseGeo Visualization

We extended the 2PhaseGeo algorithm [24] to incorporate also the time dimension into the detection process resulting in 2PhaseGeoTime. This extension is motivated by the fact that people document the same event nearly at the same time (with similar terms). In the first step, the geo-reference and the time of the item are fed as a vector into the SOM algorithm producing location-time based clusters V_1. For each cluster, an additional term-based centroid for the second step $V_2 =< V_{2,1}, V_{2,2} >$ is calculated by averaging the term $V_{2,1}$ and time values $V_{2,2}$ of items assigned by similarity to this cluster.

In the second step, all non-geo-referenced media items are assigned to the previously calculated term-time based clusters. As each of the remaining data items is described also by a term-time vector, the best fitting/nearest centroid is determined. For finding the nearest centroid, we use a combined measure. For the term part $V_{2,1}$ the cosine measure between the tf-idf values of an item and a centroid is used, while for the time part $V_{2,2}$, a Gaussian membership function is applied. The variance for the Gaussian function is based on the data items already contained in the cluster (produced in the first phase). The farther away the time from the cluster, the higher/dissimilar is the value. Afterwards, both results are added and used as a combined distance measure. The data item is assigned to the cluster following the minimum of the combined measure.

The clusters are displayed on a map given by the longitude and latitude values of the location-time-based cluster similarly to the original 2PhaseGeo algorithm. Additionally, the evolution of the disaster can be tracked thanks to the temporal information.

4 Evaluation framework

For evaluating and comparing the different algorithms in a structured manner, we specified a framework focusing on different aspects and perspectives (algorithm, data, and results). The framework is characterized by four main criteria: scalability, metadata quality, ground truth, and clustering quality. Based on these criteria the best fitting algorithm for a given task can be identified. In the following the four criteria are introduced, whereas Section 5 describes the application of the criteria.

4.1 Scalability

Scalability compares different characteristics of the algorithms. The criterion *runtime complexity* shows a theoretical runtime estimate for the worst case. In addition, the number of *parameters* that must be set by the user is discussed in another subsection. Also, the *representations* of the clustering results are compared based on their clarity. This *representation* criterion discusses how the results are visualized for the user.

4.2 Metadata quality

This criterion is related to the quality of the metadata used as a basis for clustering. It allows to evaluate how complete the metadata information must be to reach meaningful results.

4.3 Ground truth

We also check the quality of the clusters (sub-events) regarding to their homogeneity/similarity with the real-world sub-events from the ground truth. Therefore, the produced clusters labeled with the most significant terms are compared against manually extracted

sub-events. We identify these sub-events based on publicly available descriptions of the treated incidents. For the comparison, we search if clusters recognized by the algorithms are labeled with similar keywords as extracted from the public descriptions. For the UK riots, the BBC timeline [2] is used, and for Hurricane Irene, the timelines from the federal Emergency Management Agency [27] and from the National Hurricane Center [1] are used.

The manually extracted sub-events are our ground truth (see Tables 2 and 3). The UK riots are spread over different cities in the UK and, hence, they contain many crisis-related sub-events (see Table 2). Mostly, the sub-events cover fire in buildings, cars, etc., and looting activities. The manually extracted sub-events from Hurricane Irene are described in Table 3. Almost all of the sub-events include struggling with wind, rain, and flooding issues.

4.4 Clustering quality

This criterion deals with different quality aspects of the clustering process. The quality of the resulting clustering depends on the number of clusters and terms. Therefore, we make

Table 2 UK riots sub-events as ground truth

Date	Time	Sub-event
06. Aug	08:45 p.m.	Start: London/Tottenham
		(till morning 07. Aug 4:30)
07. Aug	06:28 p.m.	London/Enfield
	06:30 p.m.	London/Brixton
	10:30 p.m.	London/Oxford Circus
08. Aug	00:45 a.m.	London/Brixton road
	02:20 a.m.	London/Enfield
	05:19 p.m.	London/Hackney
	06:42 p.m.	London/Peckham: bus on fire
	08:58 p.m.	London/Croydon
	11:05 p.m.	London/woolwhich high street
	11:23 p.m.	West London/Ealing
	Night	Birmingham, Bristol, Nottingham
09. Aug	00:45 a.m.	Birmingham: police station on fire
	00:59 a.m.	Liverpool
	01:05 a.m.	London/Ealing
	01:21 a.m.	London/Hackney/Newham/Coydon etc.
	03:06 a.m.	London/Enfield etc. divers fire
	05:46 p.m.	Manchester/Salford: attack police
	07:28 p.m.	Manchester/Salford: looting; Birmingham
	19:36 p.m.	London/Nottingham
10. Aug	Early hours	Liverpool/Birkenhead
	00:24 a.m.	Birmingham, London, Manchester
	00:24 a.m.	West Bromwich, Wolverhampton
	04:00 a.m.	Bristol
	05:37 a.m.	Birmingham: men killed
	09:13 a.m.	London/Croydon: fire
	06:19 p.m.	Manchester: men arrested

Table 3 Hurricane Irene sub-events as ground truth

Date	Time	Sub-event
22. Aug	01:35 a.m.	Puerto Rico: Virgin Island (flooding)
24. Aug	08:00 a.m.	Bahamas: Mayaguana, Grand Inagua
24. Aug	11:00 a.m.	Acklins and Crooked Islands
24. Aug	08:00 p.m.	Nassau/Long Island
25. Aug	02:00 p.m.	NW Bahamas
25/26. Aug		Evacuation Wilmington, Ocracoke Island
26. Aug		Evacuation: North Carolina (Dare, Hyde), regions of Virginia (Beach), Maryland and Delaware
26/27. Aug	evening/ morning	South Carolina: Charleston, Georgetown, Myrtle Beach
27. Aug	08:00 a.m.	North Carolina: Cape Lookout (power outage)
	07:10 p.m.	Virginia: Virginia Beach (road damage)
27. Aug		Evacuation New York City
28. Aug	05:35 a.m.	New Jersey (breaking crest on rivers)
	05:35 a.m.	Brigantine Island near Atlantic City
	09:00 a.m.	Coney Island, Brooklyn
	10:00 a.m.	Manhattan, New York City
	10:00 a.m.	(flooding and outage)
29. Aug	08:00 p.m.	New Hampshire/Vermont (flooding)
30. Aug	02:00 a.m.	North-eastern Canada

use of three indices [26], Dunn Index (Dunn) [36], Davies-Bouldin (DB) Index [9], and Silhouette (S) Index [30], to check different settings of the clustering algorithms. Appendix B shows the formulae of the indices. The indices help determine an initial setting of the parameters (i.e., the number of terms and clusters) of an algorithm. These quality indices allow to determine the optimal number of clusters that are compact and well-separated. While the Dunn and the Silhouette methods aim at finding the highest value of the performance index, the DB method targets the minimum value.

An index depends on the characteristics of the data (e.g., well separated data clusters). However, it is difficult to know about the characteristics of real-world data, hence the application of different indices is needed. The impact and scale of the crisis can sometimes give an indication about the number of clusters and terms too.

Once the optimal setting of each algorithm is found (relying on the 3 indices), we can compare the clustering algorithms using the *Normalized Mutual Information* (NMI) criterion. The calculation of the NMI is based on the original mutual information (MI) (1) and the entropy (H) (3) [19]. MI is expressed in terms of entropies (see (2)) [34]. The higher the NMI value, the more similar are the clusterings. The formula of the NMI for two clusterings $A = \{a_1, a_2, ...\}$ and $C = \{c_1, c_2, ...\}$ containing N data points is given by (4) [8].

$$MI(A;C) = \sum_i \sum_j \frac{|a_i \cap c_j|}{N} log \frac{N|a_i \cap c_j|}{|a_i||c_j|} \tag{1}$$

$$MI(A;C) = H(A) + H(C) - H(A;C) \tag{2}$$

$$H(A) = -\sum_i \frac{|a_i|}{N} log \frac{|a_i|}{N} \tag{3}$$

$$NMI(A;C) = \frac{MI(A;C)}{\sqrt{[H(A)H(C)]}} \tag{4}$$

We evaluate the clustering quality at two levels:

- At the *topic level* to check whether the topic can be identified by the clustering algorithms (see description above).
- At the *item level* to check how good the documents (items) are assigned to pure clusters, that is, whether the items are correctly associated with the desired sub-events/clusters (addressed in Section 5.5 for the UK riots).

5 Evaluation results

Based on the described evaluation framework, the following four algorithms are evaluated: *SOM*, *AC*, *2PG*, and *2PGT*.

5.1 Scalability

This section gives detailed information on runtime complexities, features/parameters, and representations of the discussed algorithms.

5.1.1 Runtime complexity

Algorithms are often compared based on their runtime. Therefore, we make a runtime estimation of the presented algorithms, where n describes the number of media items, f is the number of textual metadata fields used, and d is the dimension of the input vectors for clustering. We separate the estimation into two steps: the preprocessing of the media items and the algorithm itself.

In the *preprocessing* stage, we have to go through the metadata fields (title, description, and tags) of each media item to calculate the tf-idf of terms. We choose the terms with the highest tf-idf. In addition, we apply a synonym check of terms based on the number of terms already seen to get a compact representation (e.g., *blast* or *explosion*, based on WordNet dependencies). This is performed during the tf-idf analysis and results in a complexity of $O(nf)$. The preprocessing affects all algorithms, as each algorithm requires the textual features.

Note that the topology of the map used in SOM is given by a 2-dim array of m map-units (see Appendix A.1 for a general description). The SOM algorithm [38] is run over a number of epochs, where each epoch corresponds to one training round over all media items. The number of epochs is equal to the ratio of map units to the number of media items (m/n) [38]. For each media item, the weights (the number of weights is equal to the input vector dimension d) of each map unit must be updated during a specific epoch. The number of map units is approximately $m = 5\sqrt{n}$ in the default settings [38]. Therefore the SOM complexity is $n(md)(m/n)$. Depending on the selection of the map units as $m = a * n$ (where $a << 1$) rather than $\sqrt{n}$, the complexity increases to $O(n^2 d)$ [38], which we are considering as the worst case for the batch mode.

On the other hand, in each step of the *AC*, two clusters are merged (see Appendix A.2 for a general description). At the beginning, each media item represents one cluster. Hence,

$n-1$ steps are needed until only one cluster remains. The similarity matrix ($n \times n$) must be updated and traversed for identifying the best fitting cluster. For a naive implementation of agglomerative clustering, the time complexity is of $O(n^3)$, in some cases it can be improved to $O(n^2)$ [19].

The two algorithms *2PhaseGeo* and *2PhaseGeoTime* are based on SOM. Therefore, the runtime complexity for the worst case is $O(n^2 d)$, where all used media items n are geo-referenced. Compared to the other algorithms the dimension d of the input vector is smaller.

5.1.2 Features/parameters

In SOM, different parameters (shape of map units, neighborhood function, linear/batch mode, and linear/random initialization) can be specified [38]. The predefined parameters (rectangular shape, Gaussian neighborhood function, linear initialization, and batch mode) show good performance along the presented algorithm in identifying the sub-events. The user has only to define the number of clusters created by the algorithm describing the degree of accuracy. The number of clusters depends on the size of the crisis, the dominant sub-events and on the number of considered information dimensions (SOM vs. 2PG vs. 2PGT).

In AC, different parameters must be set, e.g., the distance measure (defining how similar two clusters are) and the agglomerative method/link measure [10] (defining which specific media items should be used for the distance measure). The Euclidean distance combined with the WARD link measure [40] shows good performance for our data sets. The AC algorithm creates clusters until only one cluster remains [26]. So, the user can traverse through a hierarchy of clusters; it is not necessary to define the exact number of clusters.

Additionally, in all approaches the number of terms to consider must be defined and this depends on the scale of the crisis.

5.1.3 Representation

The SOM and AC algorithms work only on the textual annotations and show their results in tabular form (see Section 5.3.1, Table 4, for an example). The terms are ordered based on the tf-idf value and displayed in their stem form (also for the other representations) considering the 4 most important terms. Each row in the table describes a specific cluster. In addition, the amount of media items belonging to a cluster is also visualized. Hence, this representation shows the mostly discussed issues during the disaster.

The 2PG and 2PGT algorithms, operating on the geo-referenced data, visualize their results in a map. Each cluster is represented by a centroid which is equal to the averaged coordinates over all data items belonging to this specific cluster. Each cluster is visualized as a bubble containing the labels (highest tf-idf terms) for describing the related content (see Section 5.3.1, Fig. 2). The first row in the bubble shows the cluster name. It contains the amount of media items used to create the initial base cluster and the number of non-geo-referenced media items assigned to the base cluster. The coordinates are also given in the first row. The next row shows the labels (highest tf-idf) for the base cluster, whereas the last row shows the labels after the assignment of the non-geo-referenced data.

The 2PhaseGeoTime results contain a color coding for the cluster bubbles; lighter colors display the first sub-events (see Section 5.3.1, Fig. 3, for an example). We use a simple intuitive representation for the timescale that can be easily understood by a user. The two additional lines compared to the 2PG show the date of the base cluster and the date after

Table 4 UK riots 2011: SOM results (20 terms, 22 clusters)

Cluster	Top 4 words
Cluster 1(129)	riotpolic, demo, urban, life
Cluster 2(44)	peopl, birmingham, london, riot
Cluster 3(31)	england, birmingham, london, burn
Cluster 4(26)	london, riot, polic, peopl
Cluster 5(8)	manchest, uk, peopl, birmingham
Cluster 6(20)	riotpolic, demo, urban, life
Cluster 7(16)	birmingham, england riot, london
Cluster 8(16)	birmingham, england, riot, london
Cluster 9(16)	england, birmingham, london, loot
Cluster 10(15)	manchest, street, uk, loot
Cluster 11(15)	london, peopl, polic, riot
Cluster 12(13)	polic, uk, loot, peopl
Cluster 13(13)	manchest, loot, uk, salford
Cluster 14(12)	uk, loot, street, polic
Cluster 15(12)	london, riot, polic, fire
Cluster 16(12)	manchest, riotpolic, demo, urban
Cluster 17(11)	england, birmingham, london, riot
Cluster 18(11)	london, england, loot, peopl
Cluster 19(10)	loot, poepl, london, polic
Cluster 20(8)	loot, peopl, london, polic
Cluster 21(7)	street, uk, loot, machest
Cluster 22(7)	peopl, uk, birmingham, street

the non-geo-referenced data points are assigned. It is obvious that geo-referenced cluster information displayed on a map gives a better situational awareness.

For each algorithm the parameters “number of clusters” and “number of terms” must be defined. These parameters depend on the used information dimension and the scale of the disaster (indices help in identifying initial settings see Section 5.4). Based on the performance, the SOM algorithm shows the best runtime behavior. The 2PG and 2PGT results are presented in a more user-friendly manner.

5.2 Metadata quality

For social media, it is more likely that textual annotations for pictures or videos are given, e.g., titles, tags, etc. The portion of data containing geo-referenced information is smaller [28]. This can also be seen in the data sets above (see Table 1). The upload time of a picture or video is set during the upload process to the platform.

The first algorithms, SOM and AC, operate only on the textual annotations. It is obvious that the results depend on the quality and amount of textual annotations given by the social media users. Nevertheless, it is the most widely available information on Flickr and YouTube.

Another algorithm, 2PG, is working on the textual annotation and the geo-referenced information. It performs quite well, even if there are only few geo-referenced media items, as long as all events are annotated/mentioned [24]. In the future, it is possible to improve

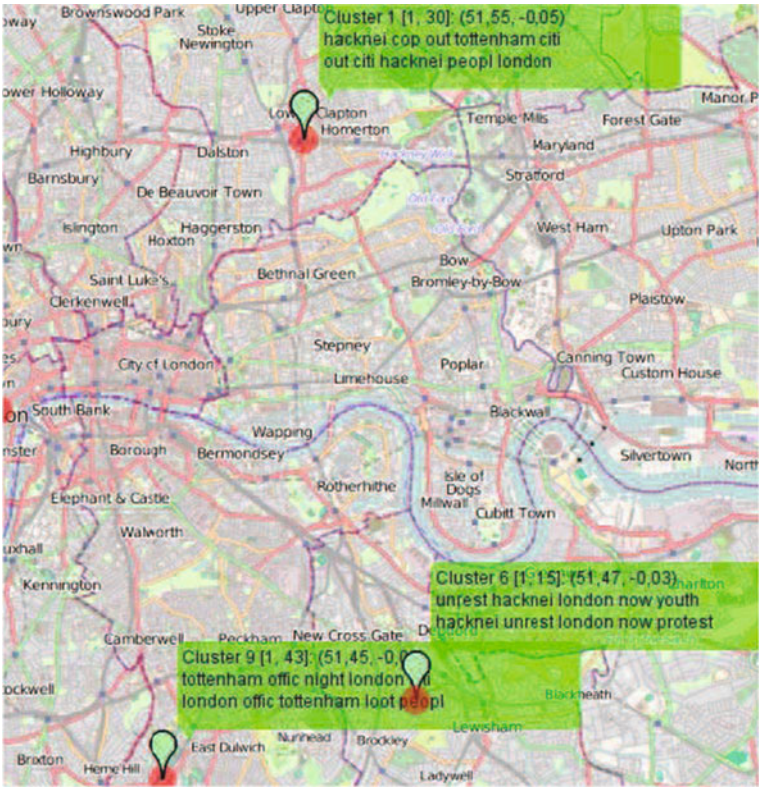

Fig. 2 UK 2PhaseGeo snapshot of the results (40 terms, 9 clusters)

this by including other sources to enrich the amount of geo-referenced data (e.g., Twitter messages). Additionally, the more social media platforms are used, the more discussion about sub-events can be covered.

The 2PGT algorithm works on all information dimensions. Hence, it is necessary that all sub-events are described based on the time and location dimension. This is not always the case, especially, due to some missing geographical coordinates. An extension of the preprocessing for the algorithm with additional geo-tagging of media items with missing coordinates could overcome this issue (Section 5.3.1). The 2PGT shows good results separating the data sets in timely related information. Additionally, it considers also all other information dimensions.

5.3 Ground truth

Based on the manually identified sub-events of the crisis (see Section 4.3), it is possible to judge the quality of the clustering results against this set of true sub-events. We tried different parameter settings (related to the indices; see Sections 4.4 and 5.4) to evaluate the results based on these extracted sub-events. In summary, we made the following observations:

– The most sensible parameters for influencing the clustering results are the numbers of terms and clusters.

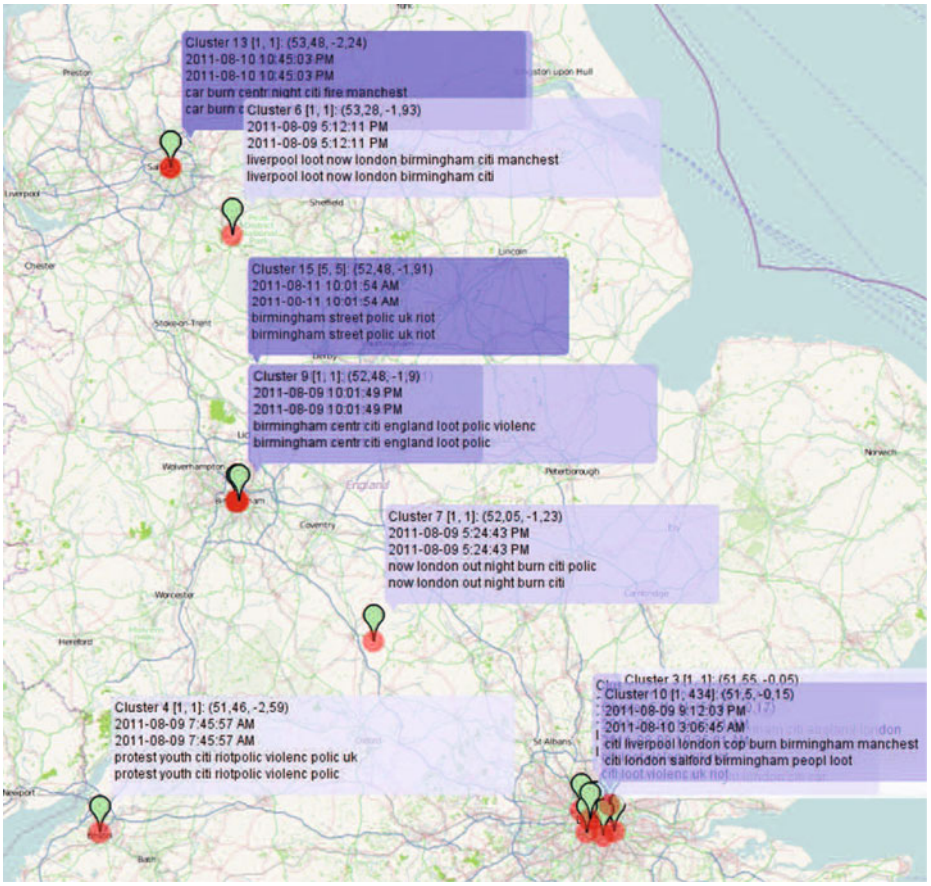

Fig. 3 UK 2PhaseGeoTime results overview (40 terms, 15 clusters)

- The number of terms and clusters depend on the scale and type of an emergency.
- The number of clusters depends on the algorithms used.
- The presented algorithms are suitable for different tasks: general overview and situational awareness (discussed later).
- Geo-tagging improves the results of clustering, this is exemplified by the 2PhaseGeoTime algorithm (see Section 5.3.1).

Obviously the number of terms and clusters influence the clustering results. Depending on the complexity and the nature of an emergency, different numbers of clusters and terms may be needed to describe the whole situation accurately. Furthermore, the number of clusters depends also on the clustering algorithm. If the algorithm operates in all three information dimensions, it is clear that the number of clusters should be higher than that for algorithms operating in fewer dimensions to represent all facets. In the 2PGT the clusters must represent all sub-events on all time aspects contained in the data sets.

This is also true for the terms. We believe that there is an evolution of sub-events, where in each step other terms are used for describing the situation. People writing about a sub-event directly after its occurrence typically use different terms compared to people who

write when time passes by, because they have different knowledge and information of the situation at hand. This implies that in future activities it is important to find algorithms to represent/reproduce these dynamical aspects.

5.3.1 UK riots

The UK riots are a large-scale emergency covering many cities and districts in UK resulting in high number of clusters. The 2PhaseGeo approach shows sufficiently good results for the UK data set. The number of clusters found in the first phase is limited by the amount of geo-information included in the data set. Hence, most clusters are found in the area of London, as most media items are located there. Sub-events are identified in the area of London (Hackney) (see Fig. 2). This is also true for the 2PhaseGeoTime algorithm (see Fig. 3; please be aware that clusters overlap on the zoom-level chosen for the overview).

It is possible to improve the results (see Fig. 4) with an additional geo-tagging approach considering the most important locations in the terms; the geo-data is assigned using a web service.[3] Then, also the data concerning Liverpool is correctly assigned and displayed. Although, there could be some outliers (see Cluster 18) when people assign geo-coordinates.

The identified sub-events are closely related to the manually identified ones. For example, riots are noticed in Hackney London on the 8th of August and in the night from 8th to 9th of August (see Fig. 5).

Results of SOM on the UK data set considering the results of the indices can be found in Table 4. Additional results for SOM and AC are given in [26]. Changing the settings, e.g., increasing the number of terms (to 40 terms) yields more city names (with fewer media items), which indeed are an important information (e.g., a cluster with following labels: tottenham, car, offic, london). Hence, it is possible to "sharpen" the results by additionally extending the number of clusters.

5.3.2 Hurricane Irene

Hurricane Irene affected several islands and cities along the east coast of the US (see Table 3). Therefore, the number of clusters must be higher than for other data sets to capture the full range of locations. Considering a higher number of clusters and terms (e.g., 50 terms and 43 clusters for the 2PhaseGeo clustering; see indices) results in detailed areas of, e.g., Bahamas, Florida, Virginia, New Jersey and New York. It is possible to follow the course of the hurricane (see Fig. 6). For the 2PhaseGeoTime we used 40 words and 40 clusters as suggested by the indices. It shows the most important sub-events.

The data set contains some outliers (see Fig. 6 with data points in the middle of US). This happens because it is not unusual that people of other areas write about the occurrences. For example, a very early cluster (23rd of August 2011) is labeled with terms regarding the path of the storm (Cluster 1 in Fig. 7: updat path storm east cost beach). Additionally, Cluster 11 shows facts on evacuation (term: evacu) on the 26th of August, where indeed evacuations in North Carolina happened. For New York city, also information about the performed evacuation can be found (see Fig. 8). The 2PhaseGeoTime algorithm shows the first affected areas in the region of Nassau and the Bahamas (see Fig. 9). Compared with sub-events extracted regarding the landfalls of the hurricane (see Table 3) it shows a sub-event located near Nassau at the 24th of August 2011 where indeed the hurricane raged.

[3] http://www.geonames.org/

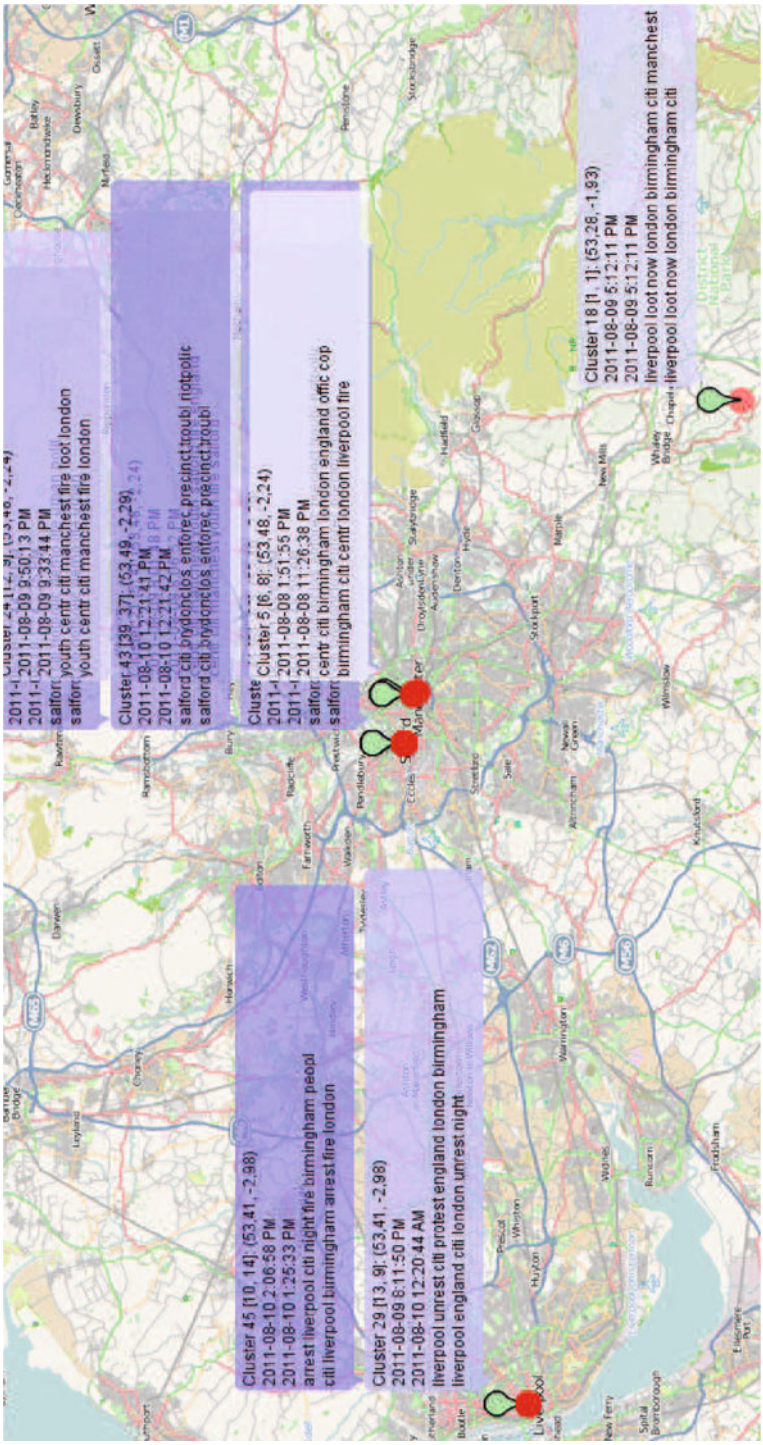

Fig. 4 UK 2PhaseGeoTime snapshot of the results (40 terms, 15 clusters, geotagged)

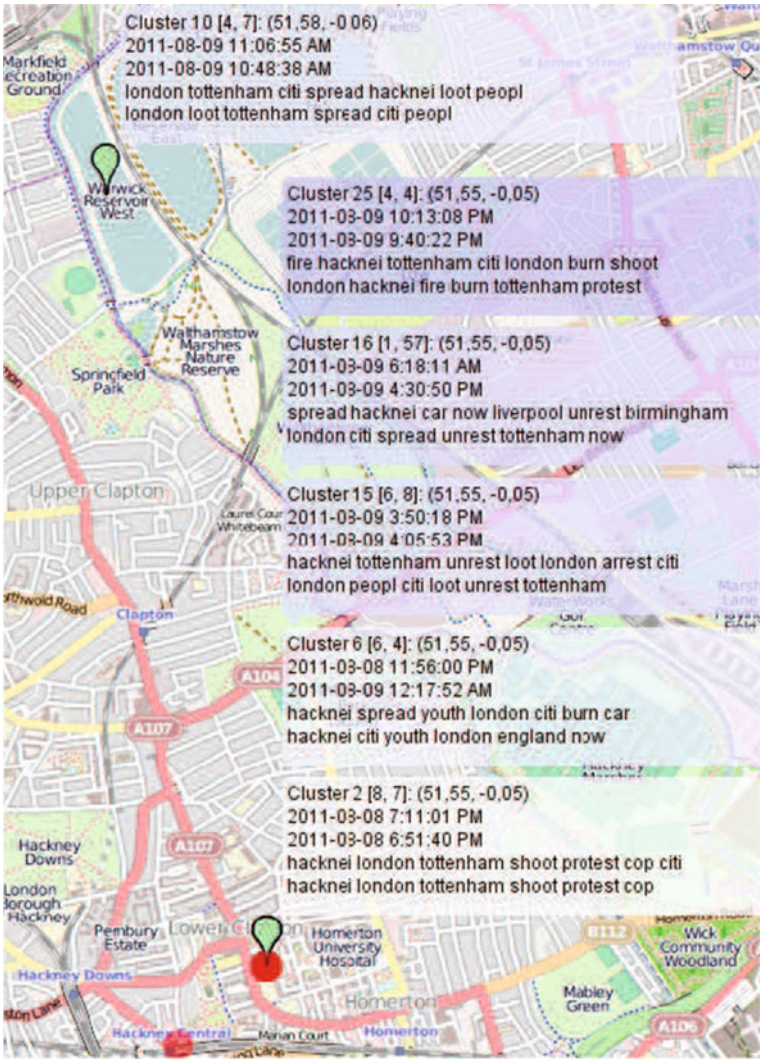

Fig. 5 UK 2PhaseGeoTime results part of London (40 terms, 15 clusters)

The SOM and AC results also show related sub-events based on the settings of the indices. It shows for example the power outages in New York. The results depend on the quality of the available metadata, but they give an overview of the situation and what the public is concerned about. In summary, the algorithms working with geo-referenced data help to easier understand the situation, especially due to the fact that the results can be visualized on a map. Hence, the 2Phase clustering algorithms need a high number of clusters to cover the time and geo-data aspects contained in the data. A streaming (online) approach would be suitable to overcome this issue, by uncovering the situation/clusters directly when new information arrives. In addition, it would allow to track the age of clusters.

Also for other data sets (MF and OB) the presented algorithms identify related sub-events. For example, for the MF data set, sub-events related to specific cities along

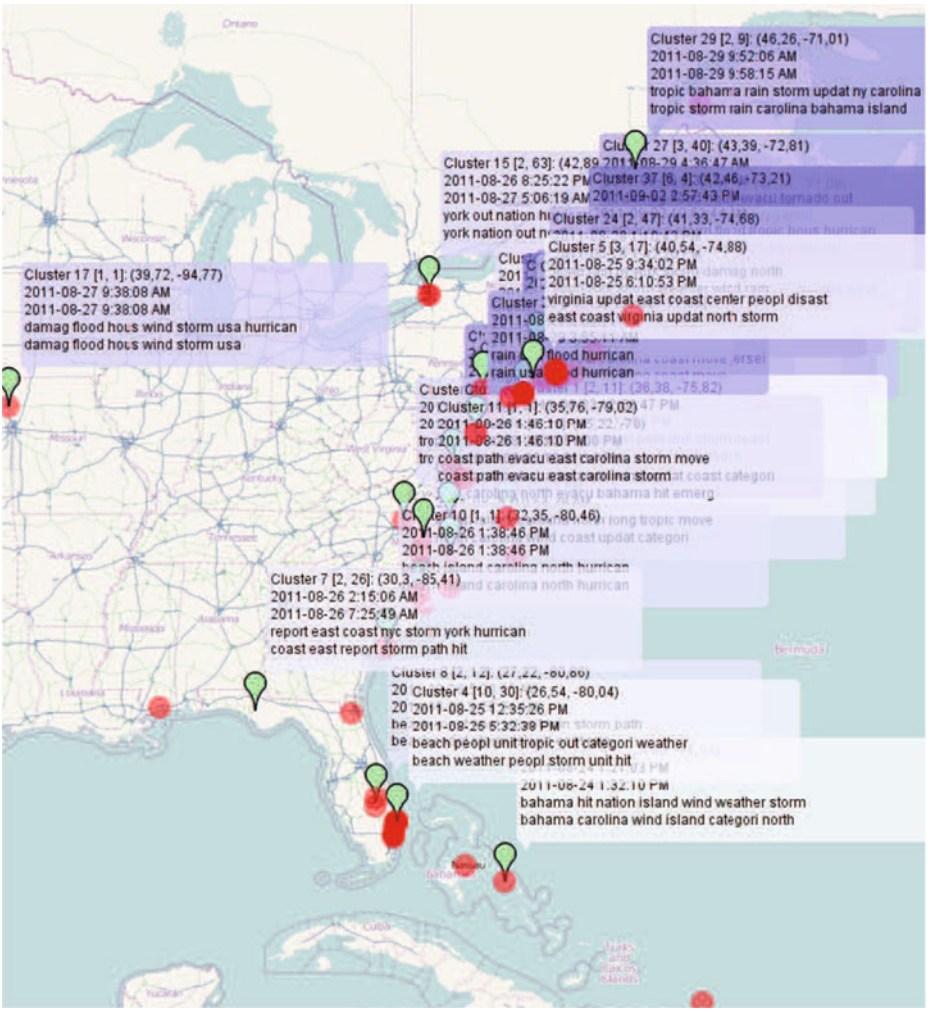

Fig. 6 HI 2PhaseGeoTime Overview (40 terms, 40 clusters)

the Mississippi river are of interests. The algorithms identify clusters related to cities like Memphis, Vicksburg, Greenville or New Orleans. Additionally, it is possible with the 2PGT to follow the flood from the origin to the delta. For the OB data set, the algorithm also identifies sub-events related to the bombing in the center of Oslo and the shooting at the Utøya island.

All algorithms show promising results in identifying sub-events. Hence, 2PG and 2PGT give a more detailed insight in what is going on during a disaster. Especially, 2PGT shows the local and timely circumstance of the disaster.

5.4 Clustering quality

We calculated the indices (Dunn, Davies Bouldin, Silhouette) for each data set with different numbers of words and clusters. For each plotted index, the maxima/peaks have to

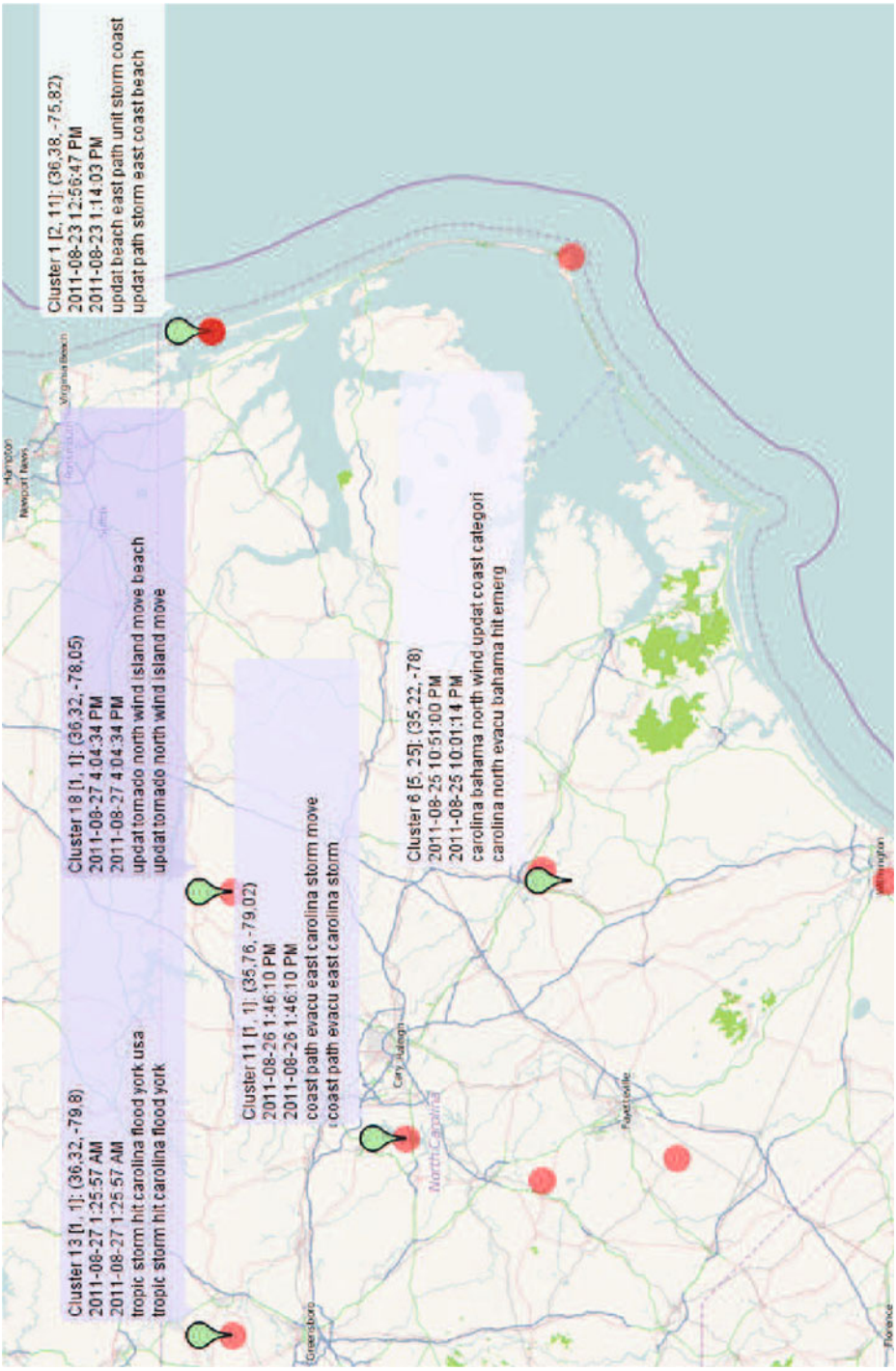

Fig. 7 HI 2PhaseGeoTime detail (40 terms, 40 clusters)

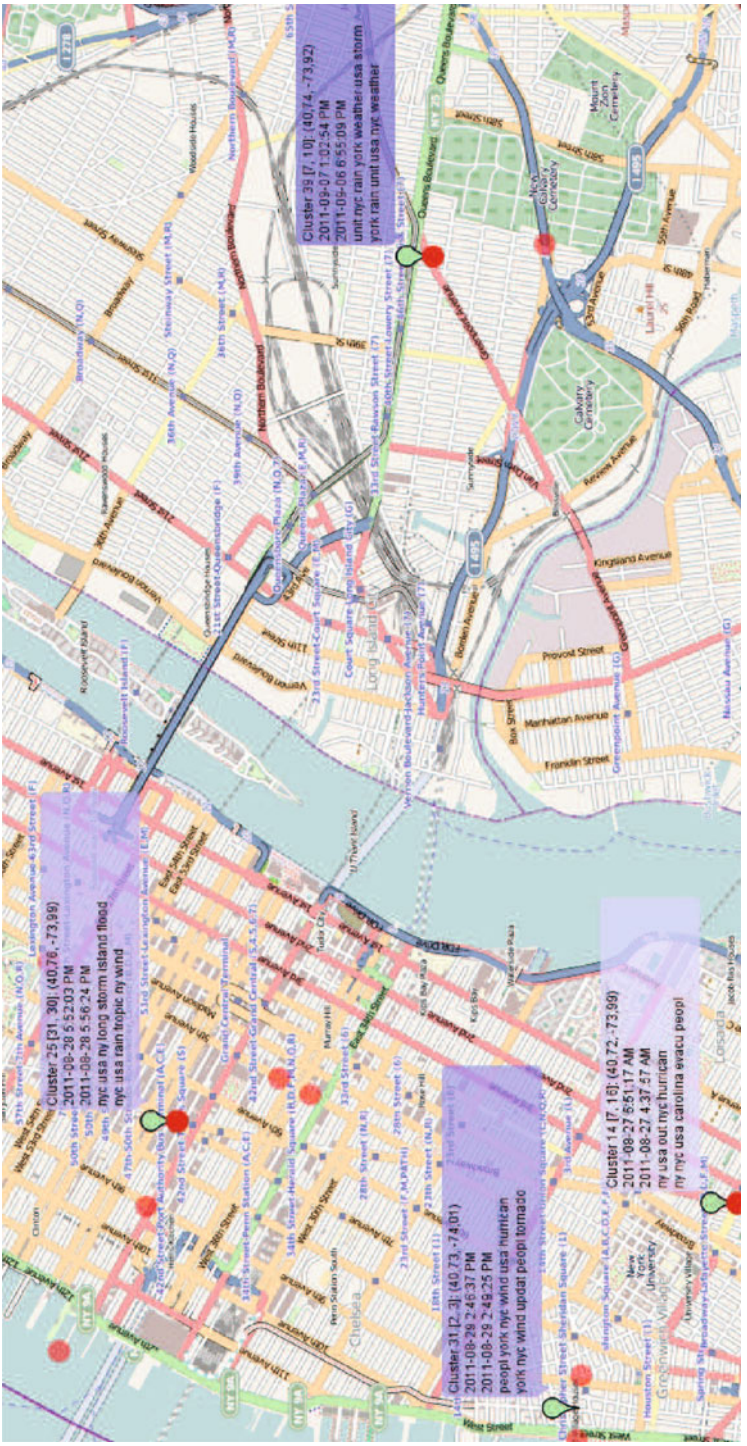

Fig. 8 HI 2PhaseGeoTime detail New York (40 terms, 40 clusters)

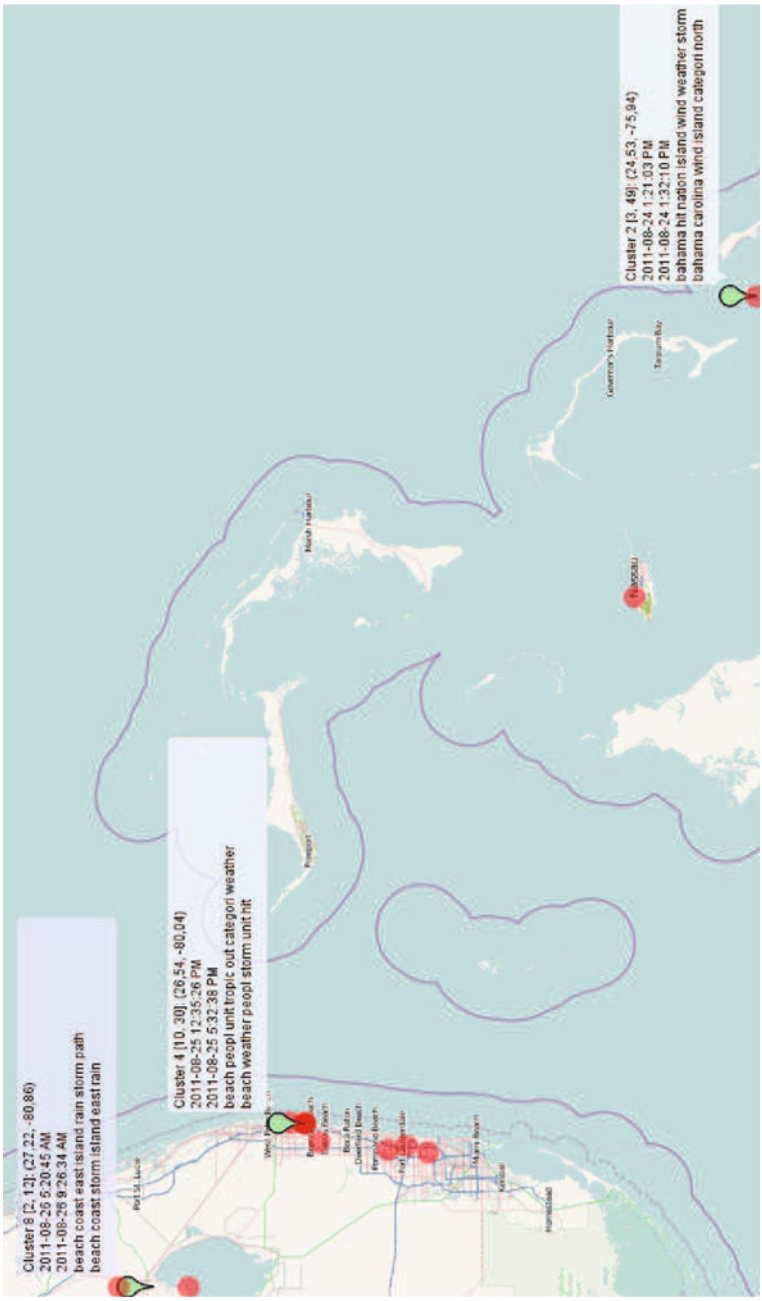

Fig. 9 HI 2PhaseGeoTime overview on first affected areas (40 terms, 40 clusters)

be considered. This is also true for the DB as it is plotted inversely for the representation in the diagram. For example, the higher the value for the Dunn Index the more compact and separated the clusters are. The indices do not always agree on the same number of clusters

and term values, but they give a hint where to look when establishing an after-the-fact analysis. For example, it is possible to take the intersection/agreement of two (or more) indices for an initial setting for the sub-event identification process.

The indices support emergency managers. Based on their experiences with emergencies of different scale, they have a sensibility to adjust parameters accordingly. Both aspects—indices and experiences—give a hint on how the number of clusters should be set. It can be seen as establishing a semi-supervised process, where the user brings in experience in setting the parameter.

For our ground truth evaluation, we make use of the indices and take the most appropriate term-cluster setting (see for example Fig. 10). Figure 10 describes the three indices for the AC algorithm based on the UK riots data set. The indices are calculated for each algorithm and data set pair for different numbers of clusters and terms). It is possible to decide on settings where most of the indices agree to find a suitable clustering result. Details about the evaluation can be found in Section 5.3.

Based on the settings gained from the indices for each algorithm, we also computed the NMI per data set (see Tables 5 and 6). The first line shows the algorithms, whereas the second line shows the used cluster and term settings for the evaluation. The higher the NMI value the more similar are the clusterings (1 means perfect match).

2PG, AC and SOM are more related, since they cluster items in a similar way. They have a good correlation in the NMI tables showing higher values across algorithms and data sets. This indicates that people describe based on the location the things that happened (resulting in sub-events). The 2PGT algorithm has different characteristics, since the clustering depends on the terms, geo and time information. The HI data set has a good term-to-time correlation (e.g., specific terms are used for specific timely-relevant sub-events) as the 2PGT results also correlate with the 2PG, AC and SOM. Although the algorithms work on different information/metadata dimensions, they show an acceptable correlation.

The 2PGT approach shows appropriate results for huge amount of clusters, as they are needed to cover all time aspects in the data sets. This results in a higher separation of the data sets which affects the correlation with other clustering results. For our future work, we plan to perform an online algorithm to analyze data (i) in real-time, and (ii) include aging aspects of clusters for showing the dynamic evolution of the situation. Additionally, no previously defined number of clusters should be needed.

5.5 Clustering quality in the presence of labeled data

Having discussed so far the clustering quality of the topic level, this section will focus on the item level evaluation. We examine the quality of clustering in terms of how good documents (items) are correctly assigned to the clusters (sub-events) using the metrics *purity* and *entropy*. This evaluation however requires the documents to be labeled and the sub-events to be known. Therefore, as a first step, the UK riot data set was labeled to define the ground truth assignment pairs (document, sub-event).

5.5.1 Manual labeling process

The labeling task is conducted using the textual metadata annotations associated with the picture and video items. The sub-events have been identified along with the items assigned to them. A sub-event is described by means of three labeling elements. The first two elements indicate the city and the district, respectively, where the sub-event happened, whereas the third element indicates the specific incident (like riot, police intervention, etc.). A

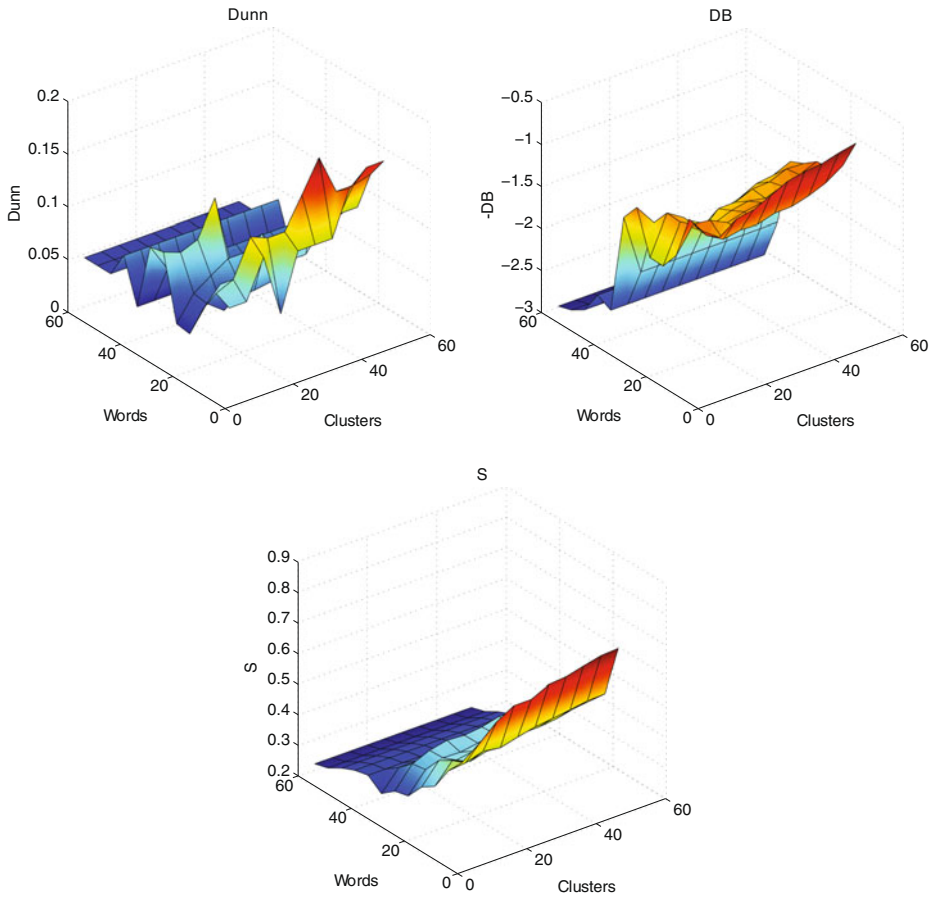

Fig. 10 Exemplified indices for UK riots; for the AC algorithm

sub-event is therefore presented as City - District - Incident. However, it is not always possible to identify all these elements and hence less specific sub-events (topics/clusters) can be obtained by this manual labeling.

Labeling a picture considering this description of a sub-event is straightforward; a video instead can contain/show more sub-events. This means that different sub-events could be assigned to one video item. We transform this into a single-labeling problem and take the first label which was assigned to a video since this can be seen as the most important one. In the future, we also want to examine the multi-labeling aspect in more detail.

The identified sub-events and the label information enable the evaluation of the different algorithms (SOM, ACM, and 2PG) considering different granularities (e.g., on the district or city level). In addition, to evaluate the time aspect of the 2PGT algorithm, each identified sub-event also gets a fourth labeling element appended, that is the date of occurrence (e.g., Birmingham - City Centre - Store Closure - 10th of August). Based on the amount of data in the UK data set we decided that the occurrence date is a sufficient time frame to consider for the evaluation. In addition, the date allows to define different granularities (e.g., all sub-events of one specific district at a specific date) for the 2PGT algorithm.

Table 5 Hurricane Irene—Normalized Mutual Information

HI	2PGT	2PGT	2PG	AC	SOM
(terms, clusters)	40,54	40,40	50,43	20,32	20,32
2PGT	1	0.6649	0.3920	0.3638	0.3314
2PGT	0.6649	1	0.3489	0.3201	0.2800
2PG	0.3920	0.3489	1	0.5081	0.4300
AC	0.3638	0.3201	0.5081	1	0.5673
SOM	0.3314	0.2800	0.4300	0.5673	1

The manual labeling results in 109 sub-events that involve the following cities in the UK: *Birmingham, Bristol, Gloucester, Leicester, Liverpool, London, Manchester, Nottingham, West Bromwich* and *Wolverhampton*. Additionally, the related districts are extracted, e.g., for Birmingham the following districts are uncovered: *City Centre, Handsworth, Harborne, Lozells* or *Winson Green*. In case it is not possible to identify a sub-event for an item, it is assigned to a class describing "unrelated items". The main incidents that could be identified were *Looting, Criminal Damage, Clean-up, Representation of Emergency Services, Fatalities* and *Store Closures*.

5.5.2 Evaluation

After the labeling process, the quality of the clustering can be judged using standard metrics [6]: *purity* (5) and *entropy* (6). This allows to judge how good the automatic assignment of items to the clusters (sub-events/topics) using the ground truth is.

$$\pi_i = \frac{1}{n_i} Max_j(n_{i,j}) \tag{5}$$

In (5) n_i is the total number of data items in cluster i and $n_{i,j}$ is the number of items of class j in cluster i. The purity measures how pure a cluster is (that is the proportion of classes in the cluster). The higher (the closer to 1) the purity value, the better the clustering quality (i.e., each cluster tends to correspond to a class).

The entropy is calculated as follows:

$$E_i = -\frac{1}{log(H)} \sum_{j=1}^{H} \frac{n_{i,j}}{n_i} log \frac{n_{i,j}}{n_i} \tag{6}$$

Table 6 UK riots—Normalized Mutual Information

UK	2PGT	2PG	AC	AC	SOM	SOM
(terms, clusters)	40,15	40,9	20,22	40,18	20,22	40,18
2PGT	1	0.1918	0.1710	0.1584	0.1527	0.1518
2PG	0.1918	1	0.4452	0.4348	0.3703	0.3936
AC	0.1710	0.4452	1	0.9641	0.5951	0.5525
AC	0.1584	0.4348	0.9641	1	0.5807	0.5432
SOM	0.1527	0.3703	0.5951	0.5807	1	0.5720
SOM	0.1518	0.3936	0.5525	0.5432	0.5720	1

where H represents the number of classes. The smaller the entropy value, the better the clustering quality. An entropy close to 1 indicates bad clustering.

5.5.3 Results and discussion

We developed three sets of experiments corresponding to the levels of the labeling:

- Using the original format City - District - Incident - Date (*CDID*) as described above.
- Using the combination City - District (*CD*): All items related to the same district and the same city are considered to be of the same class. This is motivated by the fact that we have a small number of items for each sub-event.
- Using city (*C*): All items related to the same city are considered to be of the same class.

For the 2PGT algorithm, also the time aspect was considered in the CD and C options (i.e., only sub-events with the same date are combined). Table 7 shows the purity and entropy values for the different granularities CDID, CD and C using the same setting for the algorithms (number of terms, number of clusters) as used in Section 5.4 for the topic level evaluation.

The CDID option evaluates each algorithm based on the basic labeling results. SOM and AC show similar behavior regarding the purity and entropy metrics. The values are improved by including more descriptive terms into the calculation (e.g., SOM/AC with 20 or 40 terms). This is also true for the CD and C options. 2PG produces the lowest values of purity and the highest entropy values. The geo-coordinates given in the data set do not cover the full spectrum of City - District - Incident - Date as identified in the labeling tasks. The best result is given by the 2PGT algorithm.

With the CD option, we obtain higher purity and lower entropy values compared to the CDID granularity. SOM and AC are operating solely on the textual features. For describing very specific sub-events too few related items are available to get related terms/features from them for clustering. Hence, on a coarser granularity the items are assigned in a better way. This is also true for 2PG regarding the selection of keywords. In addition, also the small amount of inherent geo-coordinates in the data set leads to better results using the coarser granularity.

Using the C option, we obtain the best values for purity and entropy, especially for the 2PG algorithm which depends on the geo-coordinates given in the data set. The values of the purity increase (and the entropy values decrease) for all options as the number of terms used in the index increases (SOM and AC with 20 and 40 terms) as expected because more discriminating terms may be included. This can also be observed when having a close look at

Table 7 UK riots 2011: Purity and Entropy

Algorithm	CDID		CD		C	
(terms, clusters)	Purity	Entropy	Purity	Entropy	Purity	Entropy
SOM (20,22)	0.5046	0.3429	0.5709	0.3548	0.6483	0.3883
SOM (40,18)	0.5275	0.3425	0.5944	0.3399	0.6519	0.3770
AC (20,22)	0.5332	0.3041	0.6355	0.2935	0.7160	0.3001
AC (40,18)	0.5526	0.3070	0.6581	0.2876	0.7228	0.2920
2PG (40,9)	0.4568	0.4047	0.5509	0.3982	0.6449	0.4124
2PGT (40,15)	0.9296	0.0514	0.9309	0.0525	0.9611	0.0393

the results of the manual labeling. The labeling process results in a larger portion of specific sub-events containing less than five items. The overall amount of data and the amount of data per sub-event is too small to identify all discriminating key features.

The best results using the CDID, CD and C options are obtained when using the 2PGT algorithm. This shows that time is a good indicator for distinguishing sub-events. Given that only few items describe a specific sub-event within this data set, the time is a discriminating feature independently of the indexing keywords used. Therefore, there are two ways to improve the results: (i) by selecting additional and more different terms, or (ii) by using the time as an additional distinguishing feature.

Although we have not performed any optimization regarding the assignments of items, the results are quite good, especially for the AC and 2PGT algorithms. Indeed, additional studies and experiments are needed for the item-based evaluation to cross-check the results for other data sets. Also multi-labeling needs special attention in future investigations.

5.6 Summary and future work

A summary of the algorithms' characteristics can be found in Table 8. Depending on the criteria highlighted in the table, one can make a choice of the algorithm to be used. The evaluation shows the appropriateness of clustering techniques for identifying sub-events of crisis-related data. It also shows the different characteristics of algorithms and data sets, which must be considered when doing the analysis.

In the present investigation, we used besides geo-references and time information mainly textual information. As a next step, we also want to examine the potential of visual information for our detection framework. One idea would be to better visualize the results, e.g., improve the arrangement of the pictures in the visualization considering visual information like color-coding.

In the future, we will focus on real-time analysis of the incoming multimedia data besides the current after-the-fact analysis. We intend to develop a streaming (online) approach so that the outcome can be used in real-time during emergency response activities. Specifically, we plan to use Twitter messages to further understand the situation. The online approach is crucially important in order to gain an overview of the situation during the emergency. Real-world scenarios and direct involvement of end-users of the proposed framework are two additional aspects we will consider in our future work.

Table 8 Short summary of the evaluation

	SOM	AC	2PG	2PGT
Performance	$O(n^2)$	$O(n^3)$	$O(n^2)$	
Parameter	number of clusters and terms			
Representation	table with text rows		map display	
Metadata	text		text+geo	text+geo +time
Ground Truth	identifies most important sub-events 2PG and 2PGT easier to read			
Cluster Quality	similar in assigning data items			future extension with streaming

6 Conclusion

In this paper, we show the suitability of clustering algorithms to identify crisis-related sub-events via social media data. To provide a full picture, different algorithms are applied: Self-Organizing Maps (SOM), Agglomerative Clustering (AC), 2PhaseGeo (2PG), and 2PhaseGeoTime (2PGT). For comparison, an evaluation framework has been designed. It takes the characteristics of the algorithms, data, and results into account. The experimental results show in particular that AC and SOM can be used for a fast overview, whereas 2PG and 2PGT have a more user-friendly visualization and, hence, simplify the understanding of the situation. The application of clustering for detecting sub-events from social media data is successful. It allows the integration of social media into the crisis management as an additional source of information, relieving the response personnel from having to manually analyze social media information. As outlined in Section 5.6, the present framework requires additional development, especially with respect to the online and real-time operation.

Acknowledgments The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°261817 and was partly performed in the Lakeside Labs research cluster at Alpen-Adria-Universität Klagenfurt. Thanks are due to Alexander Pask-Hughes for helping in labeling the data set.

Appendix A: Algorithm descriptions

In the following, it is assumed that there are n input vectors with m features ($\{x_i = x_{i,1}, x_{i,2} \ldots x_{i,m}\}, i = 1, \ldots, n$).

A.1 Self-Organizing Maps (SOM)

SOM can be seen as a neural network consisting of an input and an output layer without any hidden layers [17]. It maps inputs from the input layer to a lower dimensional output layer (i.e., the map) and clusters them accordingly. The output layer is described with so called map units. Closely related inputs are also closely related in the output layer, this means, that they are mapped to the same map unit. This mapping results in the corresponding clustering. The amount of map units N is given by the user. A map unit u is described by m weights (i.e., vector of $\{u_j = u_{j,1}, u_{j,2} \ldots u_{j,m}\}, j = 1, ..., N$) allowing the adaptation based on incoming inputs. Algorithm 1 shows the general processing steps, where α describes the learning rate and γ represents the neighborhood function. The neighborhood function ensures that neighboring units are updated correspondingly. We used SOM in [25] and also in [26].

A.2 Agglomerative clustering (AC)

Agglomerative clustering is hierarchical clustering, where in each step the most similar clusters are merged [10]. At the beginning, each input (i.e., vector) is an individual cluster c. The clustering ends when only one cluster remains or another stopping criterion is reached (e.g., a specific number of clusters N is reached). The merge is based on a distance measure (e.g., average, center, complete-linkage, etc.) identifying the most similar pair of clusters.

Algorithm 1 Pseudocode of the SOM Algorithm (adapted from [17])

1: Randomly initialize the N map units with the corresponding m weights for each unit
2: $i = 1$
3: **for** $i <= n$ **do**
4: Given the input x_i compute the best-matching-unit (BMU) with minimum distance given by the measure *dist*

$$BMU = argmin_j(dist(x_i, u_j)) \tag{7}$$

5: The weights of the BMU and the corresponding topological neighbors are updated based on γ

$$u_{j,(t+1)} = u_{j,t} + \alpha\ \gamma_{BMU,j}(x_i - u_{j,t}) \tag{8}$$

6: $i = i + 1$
7: **end for**

Algorithm 2 shows the processing steps for agglomerative clustering. For our settings, we used the WARD distance measure [26].

Algorithm 2 Pseudocode of the AC Algorithm (adapted from [10])

1: Initialize the individual clusters with the inputs ($c_i = x_i$); $\forall c_i \in \omega_0$
2: $t = 1$
3: **for** $t < N$ **do**
4: Identify the closest-related pair of clusters (c_1, c_2) in ω_{t-1}

$$(c_1, c_2) = min_{i,j}(c_i, c_j),\ i \neq j \tag{9}$$

5: Define a new cluster $c_{new} = c_1 \cup c_2$

$$\omega_t = (\omega_{t-1} - \{c_1, c_2\}) \cup c_{new} \tag{10}$$

6: $t = t + 1$
7: **end for**

Appendix B: Indices

Equation (11) from [36] describes the Dunn Index for m clusters where d describes a dissimilarity measure and $diam$ the diameter of a cluster.

$$d(C_i, C_j) = \min_{x \in C_i \text{ and } y \in C_j} d(x, y)$$

$$diam(C) = \max_{x,y \in C} d(x, y)$$

$$Dunn = \min_{i,\ldots,m} \left\{ \min_{j=i+1,\ldots,m} \left(\frac{d(C_i, C_j)}{max_{k=1,\ldots,m}\ diam(Ck)} \right) \right\} \tag{11}$$

Equation (12) from [36] describes the DB index based on the s_i dispersion of a cluster C_i and d is again the dissimilarity measure between two clusters.

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$
$$R_i = \max_{j=1,\ldots,m; j\neq i} R_{ij}, \quad i = 1, \ldots, m$$
$$DB = \frac{1}{m}\sum_{i=1}^{m} R_i \tag{12}$$

Equation (13) from [30] shows the Silhouette value of an item i of the cluster A. $a(i)$ is the average dissimilarity of the item i to all other items in A. $d(i, C)$ is the average dissimilarity of all objects of C to i. The Silhouette value of a cluster is the average $s(i)$ from each item i in the cluster.

$$b(i) = \min_{C\neq A} d(i, C)$$
$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{13}$$

References

1. Avila LA, Cangialosi J (2012) National Hurricane Center. http://www.nhc.noaa.gov/data/tcr/AL092011_Irene.pdf
2. BBC News Europe (2012) England Riots: Maps and Timeline. http://www.bbc.co.uk/news/uk-14436499
3. Becker H, Naaman M, Gravano L (2010) Learning similarity metrics for event identification in social media. In: Proceedings of the 3rd ACM international conference on web search and data mining, WSDM '10. ACM, New York, pp 291–300
4. Bergstrand F, Landgren J (2009) Information sharing using live video in emergency response work. In: Proceedings of the 6th international conference on information systems for crisis response and management
5. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022
6. Bouchachia A, Pedrycz W (2006) Enhancement of fuzzy clustering by mechanisms of partial supervision. Fuzzy Sets Syst 157(13):1733–1759. doi:10.1016/j.fss.2006.02.015. http://www.sciencedirect.com/science/article/pii/S0165011406000960
7. Choudhary A, Hendrix W, Lee K, Palsetia D, Liao WK (2012) Social media evolution of the Egyptian revolution. Commun ACM 55(5):74–80
8. Cover TM, Thomas JA (2006) Entropy, relative entropy and mutual information. In: Elements of information theory. Wiley, New Jersey
9. Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Trans Pattern Anal Mach Intell PAMI-1(2):224–227. doi:10.1109/TPAMI.1979.4766909
10. Duda P, Hart E, Stork D (2001) Pattern classification. Wiley, New York
11. Fellbaum C (1998) WordNet: an electronic lexical database. MIT Press, Cambridge
12. Fontugne R, Cho K, Won Y, Fukuda K (2011) Disasters seen through Flickr Cameras. In: Proceedings of the special workshop on internet and disasters, SWID '11. ACM, New York, pp 5:1–5:10
13. Han D, Li W, Li Z (2008) Semantic image classification using statistical local spatial relations model. Multimed Tools Appl 39(2):169–188
14. Ireson N (2009) Local community situational awareness during an emergency. In: 3rd IEEE international conference on digital ecosystems and technologies (DEST '09), pp 49–54
15. Jaffe A, Naaman M, Tassa T, Davis M (2006) Generating summaries and visualization for large collections of geo-referenced photographs. In: Proceedings of the 8th ACM int'l workshop on multimedia information retrieval. ACM, New York, pp 89–98
16. Kohonen T (1990) The self-organizing map. Proc IEEE 78(9):1464–1480
17. Larose DT (2005) Discovering knowledge in data: an introduction to data mining. Wiley, Hoboken

18. Liu S, Palen L, Sutton J, Hughes A, Vieweg S (2008) In search of the bigger picture: the emergent role of on-line photo-sharing in times of disaster. In: Proceedings of the information systems for crisis response and management conf
19. Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University Press, Cambridge
20. Marcus A, Bernstein MS, Badar O, Karger DR, Madden S, Miller RC (2011) Twitinfo: aggregating and visualizing microblogs for event exploration. In: Proceedings of annual conference on human factors in computing systems. ACM, New York, pp 227–236
21. Mathioudakis M, Koudas N (2010) TwitterMonitor: trend detection over the Twitter stream. In: Proceedings of the international conference on management of data, SIGMOD '10. ACM, New York, pp 1155–1158
22. Petkos G, Papadopoulos S, Kompatsiaris Y (2012) Social event detection using multimodal clustering and integrating supervisory signals. In: Proceedings of the 2nd ACM international conference on multimedia retrieval, ICMR '12. ACM, New York, pp 23:1–23:8
23. Petrović S, Osborne M, Lavrenko V (2010) Streaming first story detection with application to Twitter. In: The 2010 annual conferenece of the North American chapter of the association for computational linguistics, HLT '10. Association for Computational Linguistics, Stroudsburg, pp 181–189
24. Pohl D, Bouchachia A, Hellwagner H (2012) Automatic identification of crisis-related sub-events using clustering. In: International conference on machine learning and applications (ICMLA). Florida
25. Pohl D, Bouchachia A, Hellwagner H (2012) Automatic sub-event detection in emergency management using social media. In: First international workshop on social web for disaster management (SWDM), in conjunction with WWW'12. Lyon
26. Pohl D, Bouchachia A, Hellwagner H (2012) Supporting crisis management via sub-event detection in social networks. In: International conference on collaboration technologies and infrastructures. Toulouse
27. Public Health Emergency: Hurricane Irene 2011 (2012) http://www.phe.gov/emergency/news/sitreps/Pages/irene-2011.aspx
28. Rattenbury T, Good N, Naaman M (2007) Towards automatic extraction of event and place semantics from Flickr tags. In: Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval, SIGIR '07. ACM, New York, pp 103–110
29. Rogstadius J, Kostakos V (2011) Towards real-time emergency response using crowd supported analysis of social media. In: Proceedings CHI workshop on crowdsourcing and human computation: systems, studies and platforms
30. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65
31. Sagun A (2010) Advanced ICTs for disaster management and threat detection: collaborative and distributed frameworks, chap. Efficient deployment of ICT tools in disaster management process. Peremier Reference Source, pp 95–107
32. Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. Commun ACM 18(11):613–620
33. Slaney M (2011) Web-scale multimedia analysis: does content matter? IEEE Multimed 18(2):12–15. doi:10.1109/MMUL.2011.34
34. Strehl A, Ghosh J (2003) Cluster ensembles—a knowledge reuse framework for combining multiple partitions. J Mach Learn Res 3:583–617
35. Terpstra T, de Vries A, Stronkman R, Paradies GL (2012) Towards a realtime Twitter analysis during crises for operational crisis management. In: Proceedings of the 9th international ISCRAM conference Vancouver
36. Theodoridis S, Koutroumbas K (2006) Pattern recognition. Elsevier/Academic Press, San Diego
37. Tucker S, Lanfranchi V, Ireson N, Sosa A, Burel G, Ciravegna F (2012) "Straight to the information I need": assessing collational interfaces for emergency response. In: Proceedings of the 9th international ISCRAM conference. Vancouver
38. Vesanto J (2000) Neural network tool for data mining: SOM toolbox. In: TOOLMET 2000 symposium-tool environments and development methods for intelligent systems. Oulu
39. Vieweg S, Hughes AL, Starbird K, Palen L (2010) Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In: Proceedings of the 28th international conference on human factors in computing systems, CHI '10. ACM, New York, pp 1079–1088

40. Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244
41. Yang Y, Carbonell JG, Brown RD, Pierce T, Archibald BT, Liu X (1999) Learning approaches for detecting and tracking news events. IEEE Intell Syst 14(4):32–43
42. Yates D, Paquette S (2011) Emergency knowledge management and social media technologies: a case study of the 2010 Haitian earthquake. Int J Inf Manag 31(1):6–13
43. Yin J, Lampert A, Cameron M, Robinson B, Power R (2012) Using social media to enhance emergency situation awareness. IEEE Intell Syst PP(99):1
44. Zhou C, Frankowski D, Ludford P, Shekhar S, Terveen L (2007) Discovering personally meaningful places: an interactive clustering approach. ACM Trans Inf Syst 25(3):65–95

Daniela Pohl is a research assistant and PhD candidate at the Institute of Information Technology (ITEC) Alpen-Adria-Universität Klagenfurt, Austria. She works in the scope of the EU-funded FP7 project BRIDGE (www.bridgeproject.eu). Her research interests include social media, information retrieval, data mining, and machine learning.

Abdelhamid Bouchachia is currently an Associate Professor at the Bournemouth University, Department of Smart Technology Research Center, UK. He obtained his Doctorate in Computer Science from the Alpen-Adria-Universität Klagenfurt in 2001. He then spent one year as a post-doc at the University of Alberta, Department of Computer and Electrical Engineering, Edmonton, Canada. His major research interests include Soft Computing and Machine Learning encompassing nature-inspired computing, neurocomputing, fuzzy systems, incremental learning, semi-supervised learning, prediction systems, and uncertainty modeling. He is the general chair of the International Conference on Adaptive and Intelligent Systems (ICAIS) and serves as a program committee member for several conferences.

Hermann Hellwagner is a full professor of Informatics in the Institute of Information Technology (ITEC), Alpen-Adria-Universität Klagenfurt, Austria, leading the Multimedia Communications group. His current research areas are distributed multimedia systems, multimedia communications, and quality of service. He has received many research grants from national (Austria, Germany) and European funding agencies as well as from industry, is the editor of several books, and has published more than 100 scientific papers on parallel computer architecture, parallel programming, and multimedia communications and adaptation. He is a senior member of the IEEE, member of the ACM, GI (German Informatics Society) and OCG (Austrian Computer Society), and a member of the Scientific Board of the Austrian Science Fund (FWF).