Labelling Relevant Events to Support the Crisis Management Operator

Labelling Relevant Events to Support the Crisis Management Operator

Tommaso Zoppi[1], Andrea Ceccarelli[1], Francesco Lo Piccolo[2], Paolo Lollini[1], Gabriele Giunta[2], Vito Morreale[2], Andrea Bondavalli[1]

[1] *Department of Mathematics and Informatics - University of Florence, Viale Morgagni 65, Firenze, Italy*
[2] *Engineering Ingegneria Informatica, Viale Regione Siciliana, 7275, Palermo, Italy*

ABSTRACT

Thanks to the large availability of portable devices and the growing interest in the Internet of Things, during crises, social networks or alerts sent through mobile devices or sensor networks are available and can be matched each other to perform situational analysis. However, the inclusion of multiple heterogeneous sources in situational analysis leads to two main issues: i) a source could deliver (voluntarily or erroneously) wrong data damaging the integrity and the correctness of the analysis, and ii) a significant amount of heterogeneous data need to be processed. As a consequence, the crisis management operator faces a large amount of potentially unreliable data. In this paper we present a relevance labelling strategy to process information gathered from heterogeneous data streams to select the most relevant events. These are presented to the crisis management operator with the highest priority. Our strategy is evaluated using events collected by the Secure! crisis management system, considering three real crisis scenarios happened in Italy in 2015. Results show that our strategy is able to correctly identify sets of relevant events, supporting the activities of the crisis management operator.

KEYWORDS: Crisis Management System; Human Sensors; Heterogeneous Data; Data Filtering; Relevance Labelling; Twitter

1. INTRODUCTION

One of the main tasks performed by public authorities is to take care and ensure safety and security of infrastructures, society and citizens. The management of crises as for example earthquakes or terroristic attacks consists in “encompass the immediate response to a disaster, recovery efforts, mitigation, and preparedness efforts to reduce the impact of possible future crises” [1]. A strong support is provided by *Crisis Management Systems* (CMSs) that implement functionalities to sustain and support the different parts of the management process e.g., collection, filtering and visualization of data [16].

Recently, the interest in researching and developing CMSs is growing significantly, mainly due to an increasing number of available information [6], [27]. This includes information provided by physical sensors and humans, both citizens and trained personnel, which generate information (e.g., using their smartphones [2], [8]) accessible through social networks [15], [29]. On one side having multiple information sources is a benefit for the analysis of a crisis scenario, but on the other side the risk is that the crisis management operator is *overloaded of potentially unreliable information*, negatively affecting the response process. For this reason, the information needs to be processed and filtered to become readable and trusted for the human operator.

Crisis data from the citizens generate *Volunteered Geographic Information* (VGI [8]) that is shared for example through SMS [40], Social Media [38] or dedicated applications [22]. This activity involves crowd-sourcing [20] and crowd-sensing [21] techniques: *crowd-sourcing* is the process of obtaining needed services, ideas, or content by soliciting contributions from a crowd of people (e.g., online communities), while *crowd-sensing* refers to the involvement of a large group of participants in retrieving reliable data from a specific field.

Considering "human sensors" together with sensors dispatched in the infrastructures leads to an increased number of sources and consequently data, but introduces quality issues that cannot be ignored [2], ultimately requiring the definition and the implementation of complex data filtering and aggregation techniques to ensure a satisfactory credibility confidence [22].

Previous studies and frameworks (such as [12], [14], [32]) tackle the integration of crisis data collected from different heterogeneous sources, figuring out common features and merging them into a unique format. Still, examining all the individual events may not be feasible for authorities or operators that want to i) interpret these data at runtime, or ii) store them for a-posteriori analysis. Thus, software supports are needed to realize advanced filtering techniques, discarding untrusted information and showing with highest priority the information labelled as most relevant. Such *relevance labelling strategy* plays a key role in supporting the crisis management operators to focus on the relevant events.

In this paper we present a *relevance labelling strategy* for CMSs whose streams are from heterogeneous sources. After presenting the motivations, the structure and the details of this strategy, we show its application in the 'Secure!' [13] CMS. Secure! is able to collect and aggregate heterogeneous data from sources like webcams, gyroscopic sensors, social media or linked applications providing data to the crisis coordination office supporting response management. The evaluation of our strategy is performed showing three real crisis scenarios happened in Italy, where data is collected and elaborated in Secure!.

This paper is an extended version of [34], where a preliminary version of the relevance labelling strategy is presented. The main novel contributions of this paper are the following. The strategy was revised, and enriched with additional statistical indexes. Further, novel labels as the severity score were considered. The assessment now includes a detailed sensitivity analysis of the parameters. Finally, guidelines for profiling the application in typical scenarios were included. The paper is structured as follows. Section II presents the motivations of our work and the state of the art on crisis management systems that manage data from heterogeneous sources including humans. Section III discusses the goal and the structure of the relevance labelling strategy, while Section IV presents details of our instantiation. Section V describes the implementation of Secure! CMS, which is then used in Section VI to exercise our strategy on real crisis scenarios, together with a sensitivity analysis of the main parameters of our strategy. In Section VII possible improvements of the strategy and different usage profiles are discussed to tune the relevance labelling strategy depending on the target scenario. Finally, Section VIII concludes the paper.

2. BACKGROUND, RELATED WORKS AND MOTIVATIONS

Gathering data from crowds. There are several works on crisis management based on lessons learned facing past emergencies and natural disasters. It is possible to find detailed reports about *Tahiti* earthquake [5], *Katrina* hurricane [6], fire episodes happened in Russia [7] in the summer 2010 or tsunami raged in Japan in the last decade [39] (especially the *Great Tohoku Tsunami* in 2011). In most of these contexts, crisis management was significantly supported by the collection of data from sensors or citizens, which provided VGI through telephone alerts or posts on social media. For example, regarding the earthquake that struck Port-au-Prince in January 2010 [7], [40], a live crisis map of Haiti was launched using the *Ushahidi* [11] platform. Information on the impact of the disaster was initially collected from online sources, social media e.g., *Facebook* and *Twitter*, and SMS sent by citizens that wanted to signal their most urgent needs and locations. Information coming from all sources was geo-located to build a crisis map that ten days after the earthquake was recognized from the head of the *US Federal Emergency Management Association* as the most comprehensive and up to date map available to the humanitarian community.

Analyzing VGI. Consequently to the massive usage of the social networks, several studies are conducted to fetch data from social network sources for crisis management purposes [1], [2]. The focus is especially on analyzing data coming from tweets [38], [42], a compact source of information that can be easily indexed using the hashtags. For example, *Twitcident* [29] started with the purpose of filtering, searching and analyzing information about real-world incidents or crises using *Twitter* as a unique data source. In the last years this project conveyed into *PublicSonar* [30], providing near real-time intelligence from social media in general (not only *Twitter*) and supporting most of the critical activities conducted by public administrations.

However, data obtained from social networks can contain *noise*, *misinformation* and *bias*, which can get amplified by the viral nature of social media. These data will require advanced forms of filtering and verification techniques that are generally not needed. A filtering algorithm is reported in [9], where experiments are conducted to assess the capability to detect anomalies, allowing the system to discard untrusted data. Also, it is possible to find open source tools which enable the filtering and verification of real-time data from independent channels such as SMS, Email, *Twitter* and RSS feeds, such as *SwiftRiver* [10].

Heterogeneous data streams. In general, the availability of multiple heterogeneous data streams makes it possible to retrieve information about a specific event from different points of view. Thus, the data streams coming from the considered sensors should be integrated in a unique framework. For example, *CodeBlue* [31] provides a common software layer to integrate known sensors and other wireless devices into a disaster response module managing i) ad hoc network formation, ii) resource naming and discovery, iii) security, and iv) aggregation of sensor-produced data.

Reviewing CMSs. Several works describing solutions for crisis management exist integrating VGI obtained through crowd-sourcing and - where available - other heterogeneous data streams. To the best of the authors' knowledge, the most relevant contributions to the topic addressed in our work are [3], [4], [15], [32], [33], which we survey with the aid of Table I at the end of the paper. The authors in [3] describe a *QoS-aware Service Oriented Architecture* for an environmental information management system that uses real-time geospatial datasets and complex presentation tools. The authors in [4] present a framework to deliver a reliable tsunami warning message in the Asiatic south-east. SIADEX [32], instead, defines a framework that integrates several *Artificial Intelligence* techniques and is able to design fighting plans against forest fires. The authors in [15] define a framework that supports authorities using the social network feeds, trying to geolocate and categorize tweets following specific *crisis* trends. Tweets are input sources also in the work [33], where a system is presented that produces situation awareness reports on social media activity during large-scale events, such as natural disasters.

In Table I we summarize the common characteristics of the CMS presented above to compare how they face the key issues addressed in our work. We structure the crisis management process in three blocks:

- *Data Collection and Integration*. The operations to collect, organize and control the significance and the integrity of data acquired from sensors.
- *Information processing*. Once the data is structured and collected, operations to transform raw data into actionable information are conducted.
- *Human Interface*. It includes the techniques to present the outcomes of the process to the crisis management operator.

Only the works [3], [4], [32] tackle the problems generated by the simultaneous usage of *heterogeneous* sources. This improves the accuracy of the information but at the cost of the enhanced integration effort and the inclusion of advanced filtering techniques. Such filtering addresses the data collected from the whole set of sources, and should not be applied to individual data flows.

Motivation for Relevance Labelling. In the surveyed frameworks, the analysis of the input data, the integrity check, and the integration are carefully performed by dedicated software modules. However, it appears difficult for a human operator to examine these sets of correlated data in a reasonable time. For example, the authors in [4] deal with several available data streams, which can be read from sensors (e.g., water height control, meteorological feeds), but the data analysis needed to understand relations amongst events originated by the different streams is demanded to the human operator.

Further, we observe that the works reported in Table I (see *Event Relevance Labelling* category) do not select the most relevant information to be shown to the operator, and also assume that data-level integrity and filtering mechanisms have already checked their quality and trustworthiness. Noteworthy, it is difficult to understand the relevance of a single value, because it depends on the scenario the operator is interested in.

When different events happen in different places at the same time, the operator may be interested in giving precedence to one or more sets of events (e.g., tackling the adverse effects of an earthquake which damaged the power grid before considering to recover food supply infrastructures). Thus, he must be able to *rapidly* choose sets saving key time to initiate the response. In this paper we present a relevance labelling strategy that achieves this goal; the solution we propose can be integrated in crisis management systems, closing the gaps we identified surveying the above works. We remark that our *relevance labelling strategy* we devise does not discard events, but it gives priority to certain events with respect to others for visualization to the operator.

3. EVENT RELEVANCE LABELLING

We consider a *human operator* that analyses the data collected by the crisis management system in two ways: i) *runtime analysis*: the operator analyses events that happen at runtime to manage the occurring crisis [15], and ii) *historical studies*: the operator analyses past data to identify critical events for e.g., a-posteriori analysis of the crisis or to identify patterns and scenarios that may repeat in the future [39]. In both approaches, the operator retrieves events from a data collection system and is interested in detecting the main *sets of related events* that refer to a specific crisis situation. As shown in Figure 1, the operator searches for the available events, from e.g., an *Event Database.* The event database provides a set of events matching the search query. Due to possible inaccuracies of the query or noise in the events, only a part of the events returned to the operator may be relevant to the scope of his analysis. Consequently, these events are sent to the *Relevance Labelling Strategy*, which marks each event with a *relevance score*. Finally, the events are delivered to the operator, which analyses them following the order of the relevance score.

Noteworthy, the relevance labelling strategy is intended for integration in already existing CMSs. The only change it requires in the hosting CMS is an update of its interface to visualize the relevance scores provided by the relevance labelling strategy. In the rest of the section we will formulate the criterion and function used to decide whether an event is relevant or not.

Fig. 1. Overall interactions between the operator and the system

The Relevance Criterion. A *relevance criterion* establishes how to judge each event, labelling it with a *relevance score*. The criterion is also required to tune the relevance labelling algorithm, and it must be selected to penalize events that are less relevant for the crisis management operator.

We consider a data collection system which uses heterogeneous sources. We assume that during a crisis the considered data sources generate a volume of data higher than usual and with common values for specific features (e.g., a similar latitude and longitude). For example, during an earthquake the number of alerts from gyroscopic sensors and tweets referred to that area may raise, especially being significantly higher than the alerts that come from surrounding regions. This assumption has been proven reasonable in several works that analysed tweets and alerts during past crises [39], [40]. The operator is interested to explore connected events, analysing them and eventually activating response strategies (alerting the authorities, dispatching and guiding intervention teams). However, response strategies are outside of the scope of this paper and will not be further elaborated here.

In general, it is not always true that large groups of connected events are more relevant than others composed by few events. It follows, that, where possible, each event should include a *severity score*, which helps distinguishing between large groups of events with low severity (e.g., quiet crowd of football supporters that go to the stadium) and others composed by few severe events (e.g., people damaging infrastructures or threatening the citizens).

The Relevance Labelling function. The operator is interested in understanding if in a large set of events it is possible to detect smaller subsets composed by events sharing common characteristics. Hence, we are looking for a relevance labelling function $rlf\,(CE, sf, \Omega) = S$ which takes as input i) a set CE of critical events, ii) a severity function sf: $CE \rightarrow \mathbb{R}$ which matches each event in CE with the severity score mentioned above, and iii) a set of parameters Ω related to the chosen implementation strategy. The function defines a set S of scores that indicates what is the relevance in the context of the user query for each event in the starting set CE. This is a general formulation of the problem that can be adapted depending on the characteristics of the context, as it is described in the following section.

4. SPECIFICATION OF THE RELEVANCE LABELLING STRATEGY

We explain our relevance labelling strategy that realizes the *rlf* function presented above. In our implementation the general purpose *rlf* function is adapted to a function *lf* such that:

$$[R, NR, W] = lf(CE, A, \delta)$$

Function *lf* takes as inputs i) a set *CE* of critical events, ii) a set of acceptability thresholds A, and iii) a tolerance parameter set δ. The function *lf* outputs the triple [*R*, *NR*, *W*] where *R*, *NR* and *W* are three disjoint sets whose union is *CE*. In details, each event in CE is labelled as relevant (R), non-relevant (NR) or wrong (W). The events set W contains the events with erroneous or incomplete values for one or more features, e.g., missing timestamp, or wrong spatial coordinates. R is composed of distinguished sets of events, where each individual set includes connected events with common characteristics. The remaining events are labelled as non-relevant (NR); these are intended to be shown to the operator only after the relevant (R) ones. Finally, we define *CS* as the set of valid CE events, obtained without considering the wrong events (CS = CE \ W).

This implementation does not take into account the *sf* parameter of *rlf*, which assigns the severity scores to each event in CE. Our strategy is designed to be integrated in a wide range of CMSs without requiring specific data and massive changes in the implementation of the hosting framework. Therefore, we choose to not consider this parameter because the *sf* function may be difficult to obtain, as most of the CMSs do not provide sufficient inputs to compute it. In Section 7.2, a possible modification to the relevance labelling strategy for CMSs is discussed that makes the *sf* parameter available.

4.1 Involved Techniques

Here we report the main techniques used to realize our solution. It is important to highlight that we are considering events from heterogeneous sources which have common features: in particular, we assume that each event is *at least* geo-located and time-referenced, thus we define a feature set *F* = {*time, latitude, longitude*}. We chose this minimum set of features to correlate events because it is usually easy to collect; other common features, including semantic descriptions, could be included with trivial updates of the techniques discussed below.

Integrity Check. The target of this step is to detect events that have incomplete or wrong values and to place them in the *W* set. This allows avoiding pollution of data due to events that are not valid because of malfunctions of the sensors, the transmission channels, the database or due to adversarial activities e.g., attacks corrupting the data streams. Events with these problems are stored in the W set and not considered for further analysis, which will take into account only the events in the CS set defined above.

Statistical Check. This step runs a simple and fast statistical check to verify the statistical dispersion of the events. It calculates the deviation of values of the features from a specific *statistical reference index*. This check is applied to all the event features f ϵ F (time, latitude, longitude), obtaining the *index(f)* value.

For the statistical reference indexes, possible options we considered are the *mean*, *median,* and *mode* values, because they summarize different properties of the investigated event set. The operator can select the preferred *statistical reference index.* More formally, let CS be the set of valid events with common features set F, and let f_e be the value of the feature f ϵ F for the event e ϵ CS. For each event, we will check the statement

$$\forall_{f \in F} \ |index(f) - f_e| < \delta_f$$

where $\delta = \{\delta_f \in \mathbb{R} | f \in \mathrm{F}\}$ contains the tolerance values for each feature f ϵ F. It follows that an event is relevant for the scenario *if and only if* the values of all its features are sufficiently close to the reference index considering the chosen δ_f value.

Clustering. The CS events may be clustered in more sets, where the events of each set have similar features. In this case the statistical check is not able to identify such sets, because it is effective only in detecting a single group of related events. Thus, we introduce a clustering step to detect clusters of events with similar features values. Amongst the many existing clustering algorithms, we selected the *K-Medoids* clustering algorithm [17]. The algorithm partitions the CS set into *k* clusters by selecting *k* events as leaders (*medoids*) and assigning all the other events to the neighbourhood of the closest leader. The *K-Medoids* is an evolution of *K-Means* [18] algorithm: this choice of a leader event (medoid) instead of leader coordinates (centroid) makes *K-Medoids* less sensitive than *K-Means* to outliers, keeping a similar efficiency in terms of computational time [19]. Noteworthy, the choice of the medoids leader, which may have an impact on the results, is left to the *K-Medoids* algorithm which selects its reference medoids depending on the investigated dataset. The clustering algorithm is intuitive to implement, execute and configure, and it presents an adequate tradeoff between performance, false positive and false negative, as it will be shown in Section 6 and Section 7. Thus, we favor it although most likely its performance is not optimal if compared with novel state-of-the-art algorithms [44].

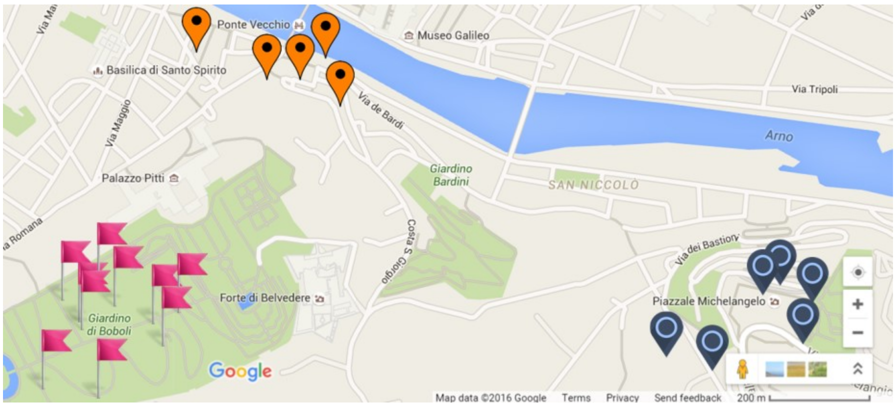


Fig. 2. Artificial sample scenario related to an area of Firenze (Italy)

To increase the chance of success of this strategy, we select *k* out of a set *K* of possible number of clusters the algorithm must look for. Moreover, the selected clusters - which could compose the R set - are distinguished through indexes matched to the constituent events.

4.2 Influence of A and δ Parameters

For clarity, we discuss the influence of A and δ parameters of our implementation of the relevance labelling strategy.

As mentioned above, the δ parameter is used in the statistical check to define a circular (considering |F| dimensions) tolerance interval around the statistical index in which the value of each event feature e_f must fall. The wider this interval is, the more events are considered sufficiently close to the statistical reference index. In the sample scenario of Figure 2, we refer to an area of Firenze city (Italy), where the set CS includes 21 events located: i) 48% near *Giardino di Boboli* (flags in the left-bottom corner of the figure), ii) 29% near *Piazzale Michelangelo* (markers in the right-bottom corner), and iii) 23% near *Ponte Vecchio* (markers in the upper part of the figure). We assume as feature set F = {latitude, longitude}, and the *mean* as reference statistical index. In this case, high δ values (e.g., δ = {10.000m, 10.000m}, see the scale on the map) lead to label all the events as relevant, because they are within the range of the chosen statistical index. Low δ values shrink the range around the statistical index, defining a subset of similar elements.

The parameter α ϵ A defines an acceptability threshold that must be reached to build the relevant events set *R*. Consider the scenario in Figure 2 and δ = {200m, 200m}: with α = 50% we are not able to identify any relevant subset, because the biggest cluster - the one in the area *Giardino di Boboli* - is composed of only 48% of the events in CS. In this case, the relevance labelling fails and it is not possible to identify relevant subsets. Taking α = 40%, instead, allows identifying a set of linked events (left-bottom corner of Figure 2) that contains more than 40% of the events in CS. Thus, these constitute the *R* set for this specific analysis. We obtain the same result with α = 30%; instead, considering α = 20% we identify the 3 different relevant subsets in Figure 2. Each of them contains at least 20% of the events of the starting CS set, and represents a relevant group of events that need to go to the attention of the crisis management operator. While this is an intuitive explanation on α and δ, a detailed sensitivity analysis and instructions on the tuning process are presented in Section 6.4 and Section 7.

4.3 Building the Process

We describe our relevance labelling strategy with the aid of Figure 3. The first step is the *integrity check*: events that have incomplete or wrong values (e.g., latitude of 1000 degrees), are removed. In addition, a low-priority notification is sent to the database administrator, which

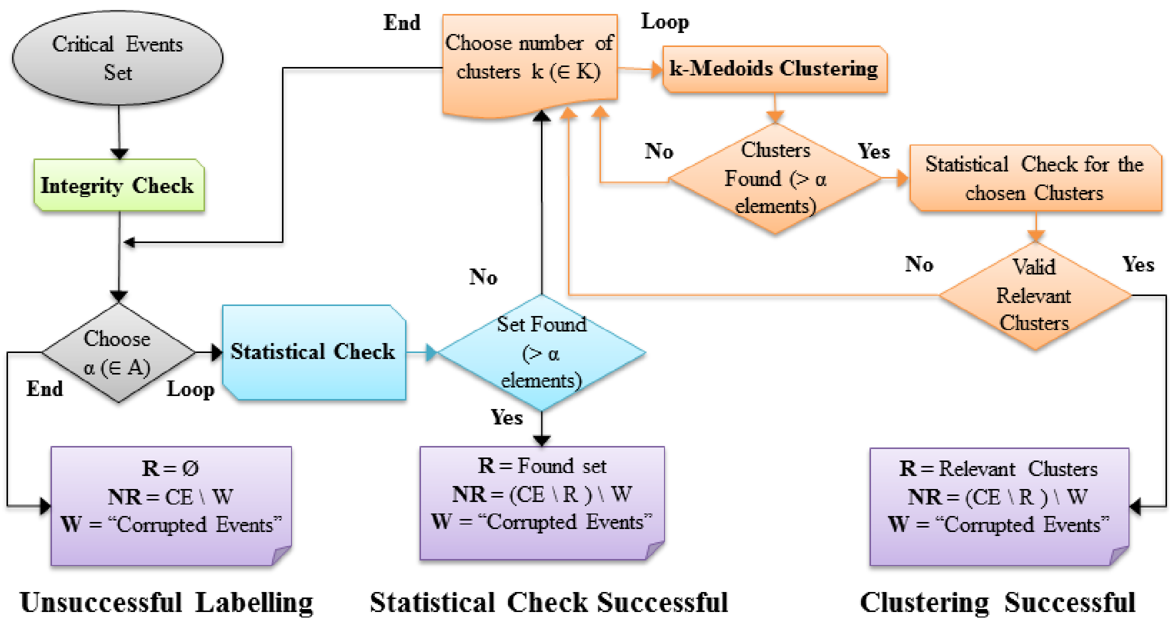


Fig. 3. Relevance labelling strategy: our *lf* function.

can check these events at his convenience and eventually fix the detected integrity errors. Then, the resulting CS becomes the input to the *statistical check*, which is executed using the selected statistical reference index. The statistical check identifies which events are sufficiently close to the reference index. If this number is above the chosen acceptability threshold α, these events constitute the relevant set R, while the remaining events are placed in the NR set (*Statistical Check Successful*) and the algorithm stops.

Otherwise, the *clustering* step is executed. It attempts to divide CS into *k* clusters. Depending on the chosen *k* value from *K* set, the *k-medoids clustering* is executed. If it succeeds, we check if any of the resulting clusters has at least α elements. Since the clustering does not take into account A and δ, some events may be included in the same cluster even if they do not respect these constraints. For this reason, the *statistical check* is run on each identified cluster. If it succeeds for one or more clusters (*Clustering Successful*), the strategy terminates and the elements in the clusters are labelled as R, while the others compose the NR set.

If both the *statistical check* and the *clustering* step are not able to identify an R set, we can choose among two options:

- Consider all the events in CS as non-relevant i.e., NR = CS;
- Lower the value of α and execute *lf* again, to search smaller sets of relevant events.

In Figure 3 we apply the second option. We first define the A set of possible α values for the current setup of the relevance labelling strategy. Then we attempt to label events taking as input the greater α value in the A set: if the strategy fails, we attempt to label the events considering a lower α value. If the algorithm is not able to find relevant subsets with any α value in A, the process terminates with *Unsuccessful Labelling*.

5. INTEGRATION IN A CRISIS MANAGEMENT SYSTEM

In this section we present Secure! [13], a *Decision Support System* (DSS, [14]) for crisis management in which we integrated the relevance labelling strategy described in Section 4.

Secure! exploits information retrieved from a large quantity of heterogeneous data sources available in a target geographical area. It aims at detecting critical situations and commanding the corresponding reactions e.g., guiding rescue teams or delivering emergency information to the population. Sample sources are social media (e.g., *Twitter*), surveillance camera, proximity sensors for suspicious people movements detection and vibration sensors for detecting events such as explosions or earthquake.

5.1 Micro and Macro-events

Data are collected, homogenized and aggregated in order to produce a situation for the DSS system that is ultimately shown to operators in a control room. First, data are collected from the heterogeneous set of sources and processed to extract basic information called *micro-events* (e.g., a gun recognized in a video from a surveillance camera). The information that constitutes a *micro event* is i) the textual description of the generic event, ii) the time when it happened, iii) its location, iv) a category assigned following an ontology [23], [44] and v) the source that generated it. In practice, micro-events are usually generated at the end of the features and information extraction process, leveraging on the acquired data processing (e.g., image, video and audio analysis, text mining, social network analysis) from the considered sources.

Macro-events, also called Secure! events, are the aggregation results of the information contained in a set of micro events which are correlated by spatial, temporal and causal relations. A macro-event contains more detailed and complete information than the related micro-events; it aims at describing a critical situation at a certain time and in a certain place. Macro-events are intended to describe the situation to the operator: for example, they can describe the status of a demonstration in a part of the city. During a crisis, micro events are generated, and correlated producing a set of macro-events. Macro-events are structured accordingly to the ontology presented in [23], [44]. Each macro-event is characterized by a set of features extracted from the micro events that compose it. The main information contained in a macro-event are i) spatial (i.e., *latitude*, *longitude*) and temporal (i.e., *time*) information, ii) details on the involved entities e.g., *the individuals or groups involved*, and iii) its category, assigned taking into account the main disaster/crisis risk typologies, natural and man-made, to be faced (e.g., natural disaster, damage, terrorist attack, violence [44]). As the macro-events describe a situation, the operator can visualize them and plan response accordingly. A complete discussion on the structure of micro-events and macro-events is reported in [44]; further this paper details the process to build macro-events starting from the micro events acquired from the field.

5.2 Data collection and Storage

Secure! collects data from heterogeneous streams through an *Integration Framework*, which is composed of services managing the sources and the extraction of data and metadata. The integration framework is composed of four low-level modules that work on a multi-modal data source; for example, *Crawler* modules extract data to identify persons, companies, cities and other types of entities from HTML documents or Web contents in general.

Moreover, for sensor network and mobile applications, *Interface* modules represent an important bridge between data sources and Secure! framework. The heterogeneous data sources are managed using a program that allows decoupling I/O modules and normalising output using predefined schemas e.g., *RDF, XML*. The data stream extracted from information sources is automatically analysed by another process whose goal is to manage analysis services.

5.3 Secure! Control Panel

As mentioned in Section 5.1, when a critical situation happens, the generated macro-events are displayed on the operator dashboard. In Figure 4, a snapshot of the Secure! dashboard is reported where events in the area of Rome are shown. Here macro-events are represented in an intuitive view through circles: the dimension of the circle is set depending on the number of macro-events in the area. This gives to the operator an immediate view on the amount of macro-events in a specific area, and helps defining the level of risk for the area considered. Further, the control panel offers the possibility to set the colour of the circle depending on the severity of the event, although a feature to compute severity is not implemented in the current version of Secure! engine (see also Section 7.2). Here becomes evident that a relevance labelling strategy integrated in the framework can help to rate this relevance, supporting the operator in understanding which are the more relevant events. Thanks to the relevance labelling strategy, the NR events can be painted with different colours or ignored, while the R events can be grouped and visualized as clusters with different colours or markers.

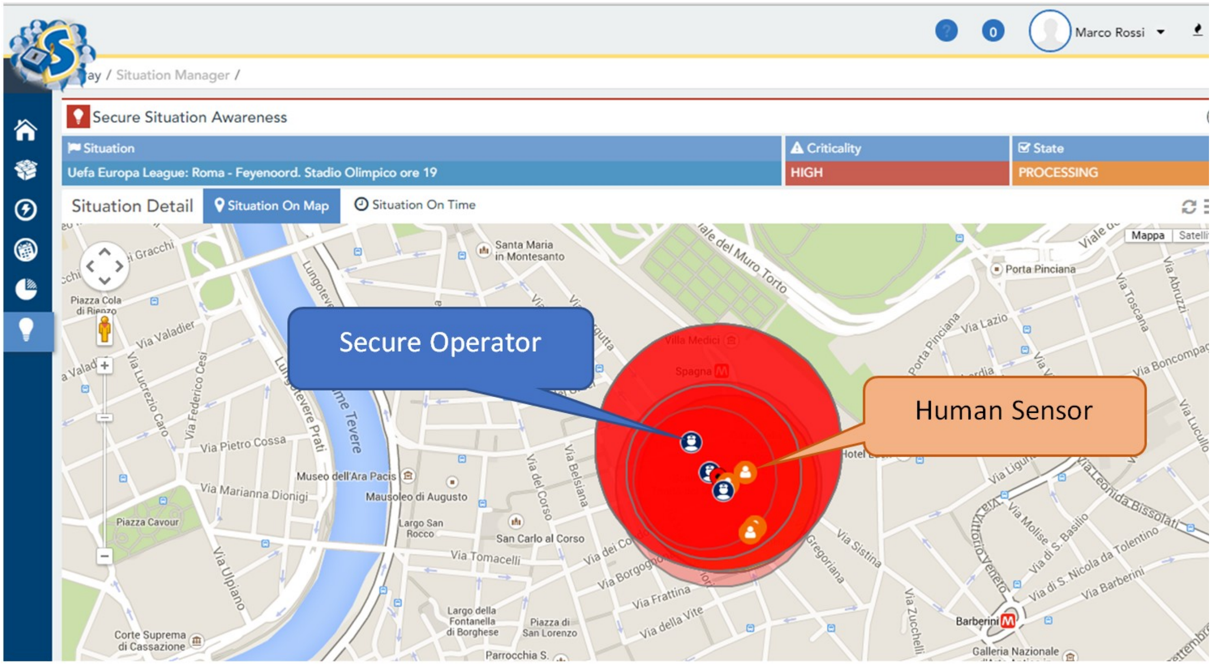


Fig. 4. Secure! GUI: events related to a crisis scenario in Rome.

Moreover, the Secure! framework allows defining profiles for *Secure Operator* and *Human Sensor*, that are shown on the map together with the macro-events differentiated by specific icons. Both profiles are tracked using GPS coordinates providing information on the presence of trained personnel in the investigated area. The map is interactive, allowing the operator to zoom on specific area of events and coordinate rescue operations communicating with the personnel on the field.

5.4 Integration of our solution in Secure!

The architecture of the integration framework is based on the SOA paradigm, with a *RESTFul* web service approach that allows defining a set of unambiguous identifiers supporting I/O Interface (XML, JSON) and the canonical HTTP operations. The most important web service is the *getEventWithAndCondition*, which allows advanced search for *macro-events* stored in the Secure! database, offering a JSON output of retrieved events. The operator can query the dataset adding filters on the features of the events: for example, it is possible to query for events generated in a restricted interval of time [t_1, t_2] adding the clauses *fromDate*=t_1 and *toDate*=t_2. All the communications are encrypted using a standard SSL certificate owned by the operator and checked before each invocation of the Secure! services.

Our relevance labelling strategy, described in Section 4.3, is implemented in *Java* as a *RESTful* service: when the operator performs a query on the Secure! database using the *getEventWithAndCondition* service, the extracted events are transmitted to the module implementing the *Relevance Labelling Strategy*. It executes the strategy described in Section 4 and returns to the operator the three sets R, NR, W.

6. EXERCISING OUR SOLUTION IN SECURE!

We discuss the experimental evaluation of the relevance labelling strategy exercised within the Secure! system described in Section 5. As target scenarios, we rely on three crises happened in Italy in 2015 whose data are stored in the Secure! database [28]. All data and results concerning the experiments are publicly accessible from [35]. In particular, Section 6.1 presents the scenarios and the configuration of the relevance labelling strategy, while in Section 6.2, the three scenarios are tested a-posteriori for the purpose of *historical studies*. Consequently, the macro-events generated from information related to the crises are stored in a database. *Runtime analysis,* i.e., the relevance labelling of a live stream of events, is addressed in Section 6.3. In Section 6.4 a sensitivity analysis for the different parameters of the relevance labelling strategy is carried forward. Finally, the performance analysis presented in Section 6.5 lists both the computational complexity and the computation time required for the relevance labelling in the investigated scenarios considering the current setup of the Secure! framework.

For the sake of simplicity, in the following of the discussion we will use the term *event* instead of *macro-event* when there are no ambiguities.

6.1 Scenarios and configuration

The Secure! system has been subject of an experimentation process whose goal is to endorse the framework and its components in a real context. Macro-events resulting from the aggregation of micro events collected using i) Twitter trends, ii) authorities feeds, and iii) web site of online news, constitute the following scenarios:

- *Europa League Match*: collisions between supporters and police and vandalism in Rome occurred [24], [25] before the Roma - Feyenoord European football match. 68 macro-events were collected, starting from 845 micro events, and stored in the Secure! database regarding this scenario (18^{th} - 19^{th} February 2015).
- *Political Manifestation*: clashes [36] between police and members of two Italian political factions (*Lega Nord*, *Casapound*) occurred during a manifestation in Rome on the 27^{th} and 28^{th} February 2015. The whole scenario is composed of 55 macro-events generated from 630 input micro events.
- *Weather Warnings*: intense atmospheric events (strong winds, [41]) raged in Tuscany on the 4^{th} and 5^{th} of March 2015. The sensors and the micro event correlation techniques in Secure! made 65 macro-events available in the abovementioned time window. These were generated from a significant number of micro events, that comprised tweets trends describing weather conditions in locations of Tuscany, events from the local civil protection agency (it periodically dispatches updates on weather status in case of bad weather) and web site of online news.

In Table II details on the scenarios are reported. We highlight the temporal clauses that the operator can use to explore the Secure! database, in addition to a textual geographic description (i.e., Rome and Tuscany). We also report the user query that was used to retrieve the macro-events, the chosen tolerance $\delta = \{\delta_{lat}, \delta_{lon}, \Delta t\}$ (δ_{lat} refers to the tolerance of the latitude feature, δ_{lon} the longitude and Δt represents the tolerance regarding the time value for each event) and the selected statistical reference index. We select α out of a set A = {50%, 40%, 30%, 20%} and the cluster number out of a set K = {2, 3, 4, 5}. We use the same setup for each scenario: in [34] the strategy was not able to provide a relevance labelling score for all the scenarios, but with the current modifications, this setup of the parameters is suitable for all the listed scenarios. Lastly, we remark that due to the geographical dimension of the considered areas, localization uncertainty of the events (which may be up to several meters [37]) is not significant for our analysis and consequently not considered in this work.

6.2 Relevance Labelling Strategy for Historical Data

We applied our strategy to the three scenarios, aiming at understanding how our solution is able to help the human operator identifying blocks of related critical events. Results are summarized with the graphical support of *Google Maps* in Figure 5, Figure 6 and Figure 7. The red pin markers locate the R events, while the yellow circles locate the NR ones. We also considered the temporal dimension of analysis, but it is not represented in the figures. With the support of Table III we discuss the results.

Europa League Match in Rome. As depicted in Figure 5, the macro-events for this scenario refer to different areas of Rome: i) nearby the stadium (upper part of the figure), ii) in *Campo de' Fiori* district and iii) around the *Barcaccia* fountain (bottom of Figure 5). Despite the highest aggregation of people was in the stadium, most of the macro-events generated by the Secure! framework refer to other areas of the city, where the Italian police and *Feyenoord* supporters had several collisions.

The relevance labelling strategy considers all macro-events as valid, without detecting any integrity problem (CE = CS, and W = Ø). Then it executes the statistical check, which labels 24% of the events in CS as relevant i.e., the candidate R contains 24% of the events. At the beginning we chose α = 50%, thus this first result is discarded because the R set is too small. According to our strategy, the clustering step becomes necessary. The clustering process identifies four distinct subsets of events, where the biggest contains 63% of the events in CS. These events refer to the area of *Campo de' Fiori*, where the most significant collisions among policemen and Feyenoord supporters occurred [25]. The other three clusters contain

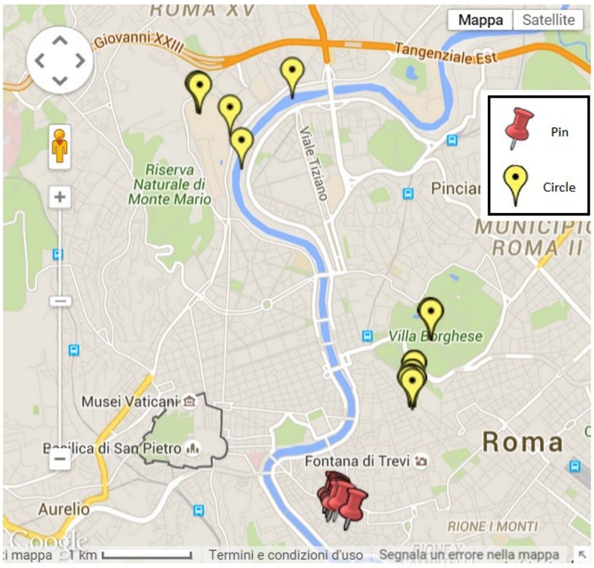


Fig. 5. Geo-location detail for events in "Europa League" scenario.

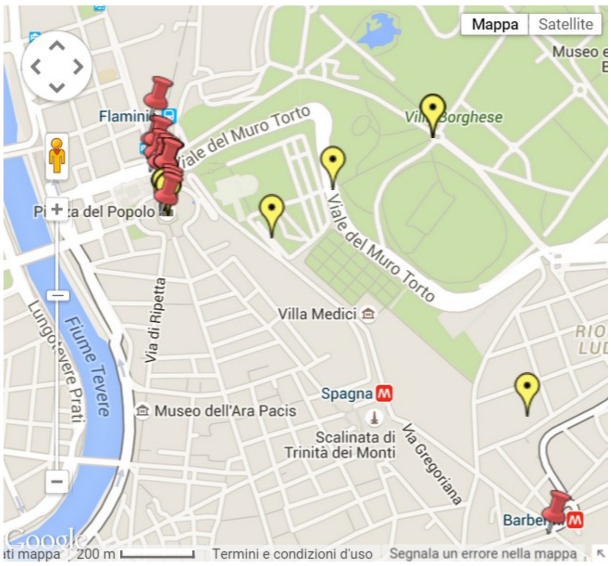


Fig. 6. Geo-location detail for events in "Political Manifestation" scenario.

events i) generated by alerts (both Tweets and authorities alerts) from the stadium while the match was being played (13% of events in CS), ii) related to the authorities alerts delivered in the morning of the 27th of February concerning the most probably dangerous zones (6% of events in CS) and iii) Tweets and website news linked to the vandalisms to the *Barcaccia* fountain [24] (18% of events in CS). The *Campo de' Fiori* cluster is the only cluster containing more than α events: only the events in this cluster (the red pins in Figure 5) are part of R and thus visualized to the human operator with the highest priority.

Summarizing, in this scenario the relevance labelling can help the operator to focus the attention on the area of *Campo de' Fiori*, e.g., restricting the query parameters to zoom on the specified area and to identify the areas that need intervention.

Political Manifestation. In this scenario the relevance labelling strategy (Figure 6) follows a different flow with respect to the previous case study. Five macro-events generated from the aggregation of tweets are classified as not valid (W set). An in-depth view of the raw data shows us that their latitude and longitude values are set to *null*, meaning that the geo-localization of those macro-events generated aggregating *tweets* failed. The statistical check applied to events in CS finds 59% of events in the range of the mean value index for all the three considered features (the red pins in Figure 6). While most of the red pins in Figure 6 are grouped together, one is located at a significant distance. This event is considered relevant even if it is located in a different position than the others, because its time value is very close to the one of the events created during clashes with the police. The values of the tolerance parameters δ_{lat} and δ_{lon} are both set to 0.06, identifying an area of 14 kilometres of diameter around the mean value. All the events are in the range of the mean value considering only spatial coordinates. It follows that the events labelled as non-relevant are not in the range of the mean only because of their time value. These events are related to the political demonstration and the successive political meeting, which did not create problems to the authorities and happened before the clashes between demonstrators and police officers. In this case, the relevance labelling strategy constitutes an R set of events that are temporally close, meaning that in the whole *Political Manifestation* scenario a smaller group of linked events is identified and needs further analysis.

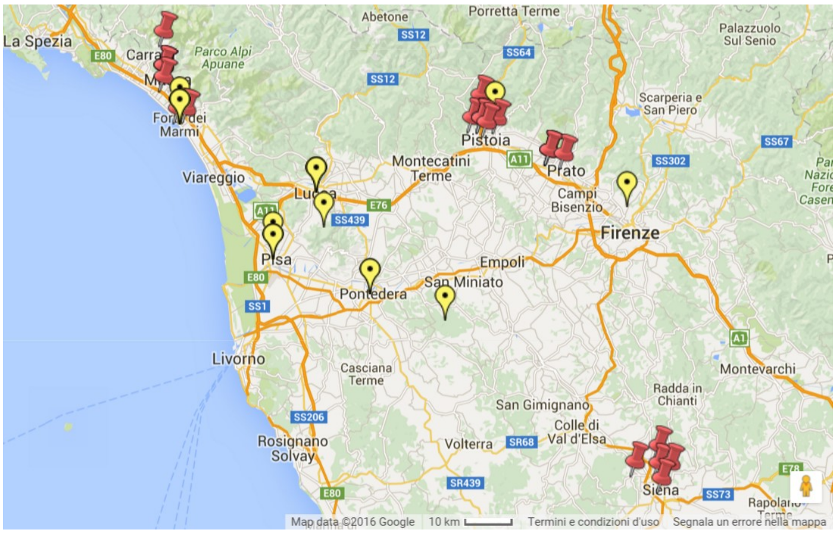


Fig. 7. Geo-location detail for events in *Weather Warnings* scenario (α=20%)

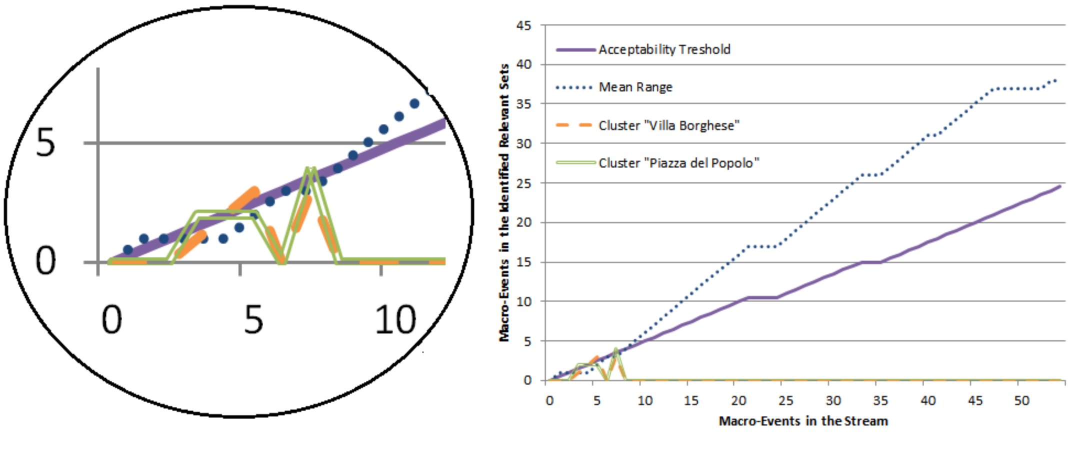
Weather Warnings in Tuscany. The last scenario regards a violent windstorm [41] which occurred in Tuscany: differently from the previous two scenarios, the investigated area is wider comprising an entire Italian region of 23.000 km^2, and macro-events are spread on the whole territory. This scenario allows exploring the case in which our strategy is not able to identify the R set considering the biggest α value in A. 21 macro-events are classified as wrong and put in the W set. 44 macro-events related to *Weather Warnings* constitute the CS set. These are geographically scattered and generated in different time slots of the two days. Therefore neither the statistical check nor the clustering are able to find a relevant group with the settings that worked with the other two scenarios. The dispersion of the events does not make possible to identify a candidate R set composed by at least 50% of the collected events considering the chosen δ setup. Consequently, the strategy is repeated with lower α values from the A set.

With α = 20% (Figure 7), the clustering step finds 3 clusters c_1, c_2, c_3 containing more than 20% of events of the starting set. Therefore, the strategy terminates identifying a relevant set R = $\{c_1, c_2, c_3\}$ composed of the union of the events in the 3 identified clusters. Figure 7 identifies with red pins the events belonging to the 3 clusters: one cluster is located in the area around *Siena*, while the other two refer to *Forte dei Marmi* and *Pistoia – Prato* areas. Some events in the range of the clusters (e.g., the *Forte dei Marmi* one, upper left area of the figure) are in the NR set. These events are geographically close but temporally far from the others, thus they are excluded from the relevant clusters.

6.3 Relevance Labelling Strategy for Runtime Analysis

To prove the applicability of the relevance labelling strategy to online streams of events, we consider the data collected for the scenarios listed in Section 6.1 as dynamic streams of events, which are progressively submitted for the relevance labelling strategy. This allows us to realize an incremental knowledge, as new macro-events are progressively introduced. We execute our solution to such expanding set of events. We maintain the same setup of α, δ we used previously, to show how the outputs of the relevance labelling strategy evolve while macro-events are produced by the Secure! framework and stored. This way, we can understand how much data is needed to detect relevant subsets that adequately describe the crisis scenarios under investigation. As elaborated later in this section, depending on the scenario – and therefore on the stream of events we are considering – the relevance labelling strategy is converging quicker or slower to the final results that we discussed in Section 6.2.

We start discussing the *Political Manifestation* scenario. On the left side of Figure 8, we can observe that once the first 8 macro-events are available, two different relevant clusters are



a) b)

Fig. 8. Runtime Relevance Labelling for the “Political Manifestation” scenario. The figure a) is a zoomed version of the left-bottom corner of b).

identified. This is highlighted in Figure 8a, where (considering the first 8 macro-events) the two data series corresponding to the clusters are above the acceptability threshold identified by the continuous line, meaning that two relevant clusters are detected by our labelling strategy. This is different from the final cluster which is instead identified as the flow of events progress and that is described in Section 6.2 and that can be observed at the extreme right of Figure 8b. In particular, a cluster in the area of *Villa Borghese* is identified rouping three alerts delivered by authorities. These three alerts were delivered some hours before the crisis, constituting a relevant set for the time being. As the crisis progresses, and other events are identified, this cluster becomes non relevant: other macro-events related to different time periods and different areas of the city are instead relevant. It can be noted that the data series of the clusters often stuck at zero as the macro-events are delivered to the Secure! database (e.g., when we consider more than 10 events in Figure 8b). In such cases, the statistical check identifies a relevant cluster, and consequently the clustering step is not executed at all, resulting in the values of the clusters' data series we pointed out.

A similar trend can be observed in Figure 9 where the *Europa League Match* scenario is elaborated. If only few events are available (the first 20 events), the statistical check identifies only one relevant group of events with more than 50% (i.e., α) of events. This is represented by the line *mean range*, which is above the acceptability threshold. When the number of available events grows, the statistical check is no longer able to identify such relevant subset.The output of the relevance labelling strategy slowly converges to the result presented in Section 6.2, that are also depicted at the extreme right of Figure 9. Progressively, different clusters are identified, but the only one which is above the acceptability threshold is the one in the area of *Campo de' Fiori*.

Finally, in Figure 10 we report the results of the relevance labelling applied to the stream of macro-events of the *Weather Warnings* scenario. To simplify the visualization, considering that several sparse clusters are presented, we report a bar chart instead of a line graph. With the same reasoning for the discussion of Section 6.2, we can observe here that the statistical check is not able to find a relevant subset at any point of the stream. Moreover, we highlight that as the flow of macro-events progresses, the clustering step identifies up to five clusters. In particular, since most of the macro-events that were collected during the morning of the 4th

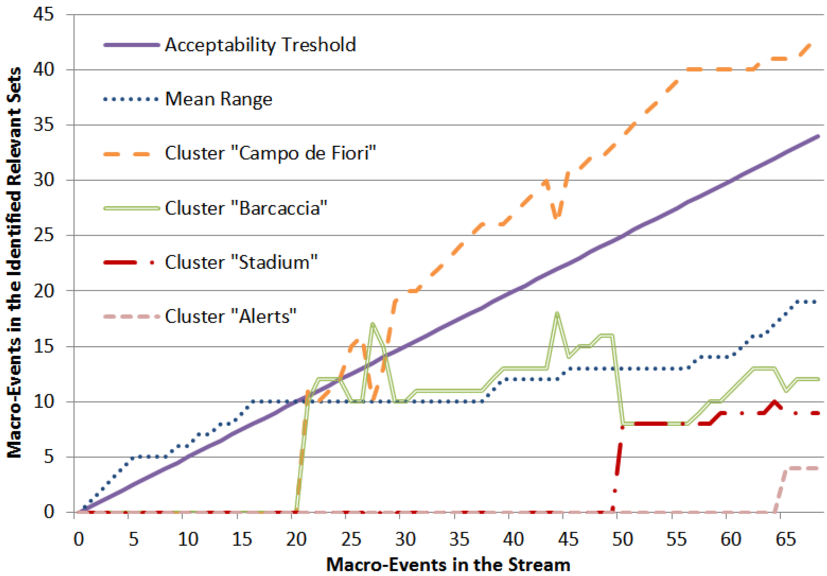


Fig. 9. Runtime Relevance Labelling for the "Europa League Match" scenario.

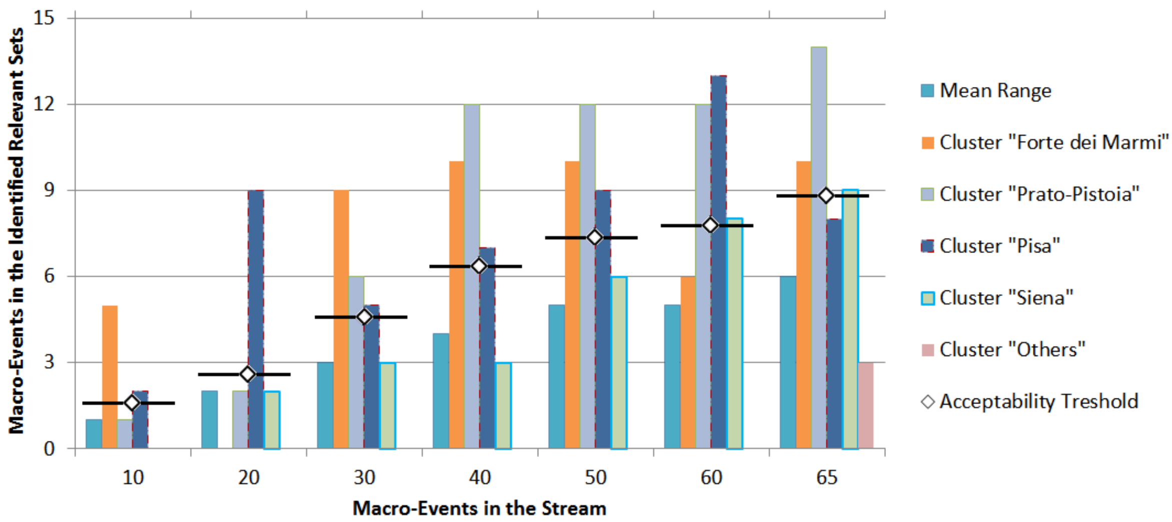


Fig. 10. Runtime Relevance Labelling for the "Weather Warnings" scenario.

of March regard a wide area around Pisa, at the beginning the cluster "Pisa" is the one cointaining the greater amount of macro-events. In the afternoon, the adverse weather events moved to the cities of Pistoia, Prato and Florence, and the cluster "Pistoia-Prato" becomes the biggest one. During the 5^{th} of March, other alerts regarded the area around Forte Dei Marmi, leading the clustering process to consider some of the events happened in the northen area of Pisa as belonging to the Forte de Marmi cluster. In addition, the cluster labelled as "Others" contains several sparse events in the area of Pisa – Viareggio. It is identified during the last execution of the relevance labelling strategy since it contains events that refer to a very specific time window of the morning of 4^{th} of March. Summarizing, in this case our strategy complies with the results discussed in section 6.2 only when the amount of information is almost the same. However, our approach was able to correctly describe how the situation evolved through the two days.

6.4 Sensitivity Analysis

We analyse the impact that the choice of the statistical index, the α parameter and the δ parameter have on the event labelling strategy. Since Secure! acquires most of the data from sensors in the city area of Rome, we select δ_{lat} and δ_{lon} such that $\delta_{lat}, \delta_{lon} \in L = \{0.01, 0.02, 0.03, 0.04, 0.05, 0.06\}$. The L set is selected considering that a difference of 0.01 points in latitude or longitude can be roughly approximated to 1 km, and that the area of Rome is approximately 1300 km^2. It follows that $\delta_{lat} = \delta_{lon} = 0.06$ identifies an area of radius 7 km around the value of the chosen statistical index, that is sufficiently wide for a city. We also considered T = {20m, 40m, 1h, 2h, 3h, 4h, 6h, 8h} as possible Δt values, while the A set is maintained as before. This resulted in 288 possible δ combinations ($|L| \times |L| \times |T|$) that we will examine with our sensitivity analysis considering mean value as reference statistical index.

Influence of the δ parameter. First, we analyse how changes in the δ influence the relevance labelling strategy. We trace the maximum α value in A that makes the strategy successful for different δ, identifying an R set that is not empty. We execute our relevance labelling strategy on each of the three scenarios considering all the 288 combinations of δ mentioned above. For a better understanding of the results, we split T in three subsets and L in two subsets (see Table IV). The tables only report the resulting 12 combinations averaging the results of the single analyses in the corresponding set.

Both for *Europa League* (Table V) and for *Political Manifestation* (Table VI) scenario we can observe that, as expected, high δ_{lat}, δ_{lon} or Δt values leads to higher average α to have a successful relevance labelling. Intuitively, considering "normal" a wide set of events, we can identify larger relevant sets. In particular, in both cases considering $\delta_{lat} \in L_2$, $\delta_{lon} \in L_2$, $\Delta t \in T_3$ always allows finding R sets that contains at least 50% of the events of the CE set.

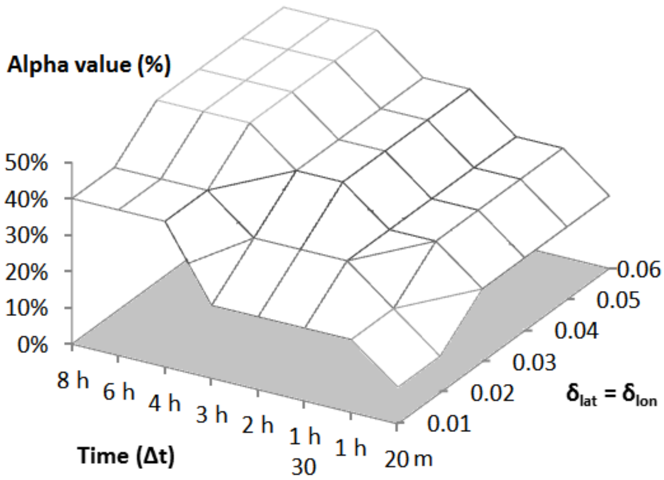


Fig. 11. Political Manifestation scenario: α values considering latitude, longitude ($\delta_{lat} = \delta_{lon}$) and time ($\Delta t$) parameters.

The results for *Weather Warnings* scenario, instead, are very different. Changing the δ value does not affect the α needed to find a relevant subset of the starting set (Table VII). The reason is the dimension of the geographical area. It follows that using $\alpha = 50\%$ the strategy fails considering the chosen δ: the strategy is only able to identify smaller subsets containing around 20% of the events in the scenario. To obtain a relevance labelling that finds a single relevant subset of more than 50% of events we need $\delta_{lat} = \delta_{lon} = 0.5$, that is far bigger (roughly 10 times) than the maximum value we considered in our sensitivity analysis.

Insights on Political Manifestation. An insight of the sensitivity analysis regarding the *Political Manifestation* scenario is reported in Figure 11 and Figure 12. In Figure 11 we report α values obtained varying δ_{lat}, δ_{lon} and Δt: to represent α values varying the three tolerance parameters in a 3D plot, we consider $\delta_{lat} = \delta_{lon}$ on the depth axis. This allowed to highlight the trend of the maximum α needed from our strategy to obtain a set R that is not empty. We can observe how raising the Δt parameter (i.e., moving from the right to the left part of the graph) leads to higher α values. Noticeably, we can observe in Figure 11 that for $\delta_{lat} = \delta_{lon} \geq 0.3$ the resulting α value does not change: this means that raising the tolerance of geographical parameters over 0.3 (i.e., 3.5 km in the investigated area) does not influence the minimum α needed from the strategy in this scenario. In Figure 12 we show a detailed view of the outcomes of the sensitivity analysis for $\Delta t = 2h$. Raising the value of only one between δ_{lat} and δ_{lon} does not always lead to a successful labelling. Instead, considering larger values for both of them allow the relevance labelling strategy to identify an R set

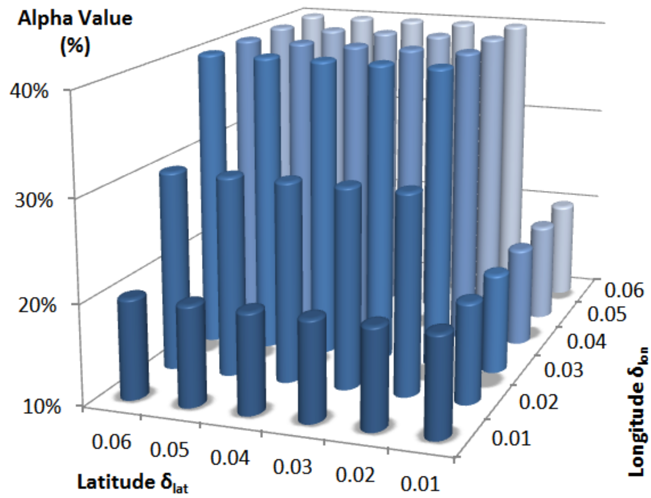


Fig. 12. Political Manifestation Scenario: in-depth view of resulting α values varying δ_{lat}, δ_{lon} with fixed Δt = 2h

using high values of α.

Selection of the Statistical index. The last remark is directed to the selection of the statistical reference index. In [34], we considered only the mean value as possible reference index. Unfortunately, the mean performs poorly when events are splitted in more clusters or when they are scattered on the map. Therefore as alternative our reference labelling strategy can consider also the median or the mode values. We repeated the relevance labelling on the targeted scenarios varying both the δ values and the statistical indexes, considering also the median and mode. The results are summarized below.

In the *Europa League* scenario no differences were observed choosing different statistical reference indexes. In the *Political Manifestation* scenario, out of 288 cases, in 124 cases the median allowed to find R without requiring the clustering step, while this drops to 120 cases with mean and mode. In the *Weather Warnings* scenario, where the relevant sets are always identified through clustering, the median allowed to consider on average 32.3 events out 44 as part of R set while with the mean only 31.1 events were considered part of the R set.

6.5 Algorithm Performance

To evaluate the capabilities of the algorithm to rapidly process information, we ran tests in which we simulate a large group of queries that an operator can execute on the Secure! database i) varying the values for the query clauses *fromDate* and *ToDate* (see Section 5.4), and ii) invoking the *getEventWithAndCondition* service in different dates and with different time window sizes. For each run of the tests, the output of the relevance labelling strategy was saved for statistical analysis. We conducted the experiments on a machine with an *Intel(R) Core(TM) i7-4510U* CPU @ *2*.00GHz, 8GB RAM. Each test is repeated 10 times keeping the system unaltered to optimize repeatability [26] conditions of each run. These results refer to the application of the relevance labelling strategy with A = {50%}.

In Table VIII we observe the average results of these tests, classified based on the outcomes of the relevance labelling strategy. In each experiment, we measured the time spent for the labelling activity, assuming that the data are already available on the machine and provided as input to the service. This allows evaluating the performances of our strategy without considering communication delays due to the network. In Table VIII we can notice that our strategy is able to identify a relevant subset for approximately 2/3 of the possible user queries to the Secure! database. Half of the successful executions of the relevance labelling strategy find an R set without requiring clustering, giving responses at average in less than two milliseconds. Otherwise the clustering step becomes mandatory, increasing its computational time that is still below 30ms, making it plenty adequate for the mentioned case study.

We also generalize the observations above studying the computational complexity of our strategy. For a generic A set, let us consider $n = |A|$ and let $c = |CE|$ be the number of critical events. Considering i as the number of iterations needed from the clustering algorithm to complete the partitioning, in [43] authors defined the computational complexity of K-Medoids as $O(ic^2)$. Since the relevance labelling can execute the statistical check (that is $O(c)$) and the K-Medoids clustering in a row, the computational complexity of a relevance labelling with a given α is $O(ic^2)$. Moreover, this can be repeated for n times, therefore the computational complexity of the worst case execution of the relevance labelling strategy totalizes an $O(nic^2)$ quantity.

7. PROFILING THE STRATEGY FOR TYPICAL SCENARIOS

We conclude the presentation of our novel strategy for relevance labelling in crisis management system identifying general rules that support the human operator tailoring our relevance labelling strategy for his target scenario. A human operator that is observing a set of events (for both historical or runtime analysis) must be ready and able to change his focus. For example, if several events are grouped in a small area, the operator might be interested in restricting the windows (i.e., the tolerance) to observe the targeted area in detail. We examine i) in Section 7.1, how to select parameters of the *lf* function, in such a way that the operator can use the most adequate setup, and ii) in Section 7.2, how the availability of a severity feature related to crisis events can improve the significance of the whole relevance labelling.

7.1 Tailoring Parameters on the Considered Scenarios

In Table IX we report several setups of the parameters (statistical reference index, minimum α in A, δ) of the strategy that a user can adopt depending on the characteristics of the scenario. We consider relevant only events that happen in a time window of maximum 6 hours and we use the same A set as the one we used in Section 6. We choose two dimensions of analysis:

- *the spatial dimension of the investigated area*: we can observe a restricted area of a city (e.g., a circular area of radius 3 km) or a wider area, as example a whole country;
- *the critical events relation*: if the expected distribution of events on time and space is known, we can tune the strategy depending on this information.

If the crisis management operator is interested in a *single* subset of connected events (e.g., he is monitoring a concert in a stadium), he can tune the relevance labelling strategy with low δ and high α values. This means that the operator is looking for a consistent amount of events linked together in such a restricted area, e.g., near the musicians. Instead, when the crises could be *more than one* (e.g., multiple disorders spread across a city, or repeated disorders during time) we need to lower α to identify different relevant subsets. These will be presented to the crisis management operator as distinct groups of relevant events. The δ value, instead, need to be tuned depending on the geographical dimension of the scenario.

In case there are unknown relations between the events concerning the crises that may happen in the observed area (e.g., information on upcoming events is not available or too complex to analyse), the operator needs to select the greatest meaningful δ. The events could be scattered through space and time, making the mean reference index of limited use. In such conditions, we propose the use of the *median* or the *mode* as more suitable reference indexes. We also recommend using a low value of α, because secondary relations between subsets of events may exist and could be detected only with this particular setup. This may happen when a noticeable amount of tweets for a completely unexpected crisis are generated in a restricted time window about the same geographical area.

The same reasoning can be obtained considering a wide area, still keeping in mind that in this case the tolerance values for the geographical coordinates must be suitable for the extent of the investigated area.

For a crisis management operator, an approach to test and compare the results of labelling for different parameters' values is particularly useful, for example for his own training in configuring the relevance labelling for a specific scenarios and a specific crisis management system. Such approach, and the related operator's training, can be straightforwardly defined starting from the use cases of Section 6.1, their analysis in Section 6.2 and Section 6.3, and the sensitivity results reported in Section 6.4. The operator should first apply the relevant labelling strategy for different values of its configuration parameters, as we performed in Section 6.4, and for different stages of the crisis, as we performed during the runtime analysis in Section 6.3. This is intended as a test-retry approach, where the operator tries different configurations and verifies its output on his control panel. For example, the Secure! control panel would show different circles, of different dimensions, depending on the cardinality of the relevant sets, as discussed in Section 5.3. Since the macro-events for the scenarios considered in Section 6.1 are well-studied and the most relevant sets are known (they are identified in Section 6.2 and Section 6.3), it is easy for the operator to compare results, and consequently learn how to configure the relevance labelling strategy. Summarizing, despite an optimal a-priori configuration does not exist, this process would support training the operator on tailoring the different parameters for various scenarios, complementing the analysis elaborated in the previous part of this Section.

7.2 An Approach to Include Severity

Here we discuss how the availability of severity scores for each event in CE can improve the strategy described in this work.

The Role of Severity. In our study, the relevance labelling is executed considering only on the dispersion of the events in the space and time defined by the features in F. It follows that all sets composed by at least α events of CS with sufficiently similar feature values are considered relevant and contribute to the R set. A severity score associated to each event would allow to rate each detected subset of CS with a severity score instead of simply counting the cardinality of the set. In such a way, a small set of severe events can be rated as more relevant than a larger set composed by non-severe events.

In other words, the *sf* parameter (defined in Section 3) can improve the significance of the relevance labelling results, although a CMS able to associate a severity score to the collected events is required. Most of the crisis management frameworks are not able to assign a severity score to the considered events. For instance, the Secure! CMS does include a "severity" field in the macro-events structure [23] but it does not yet implement mechanisms to compute its

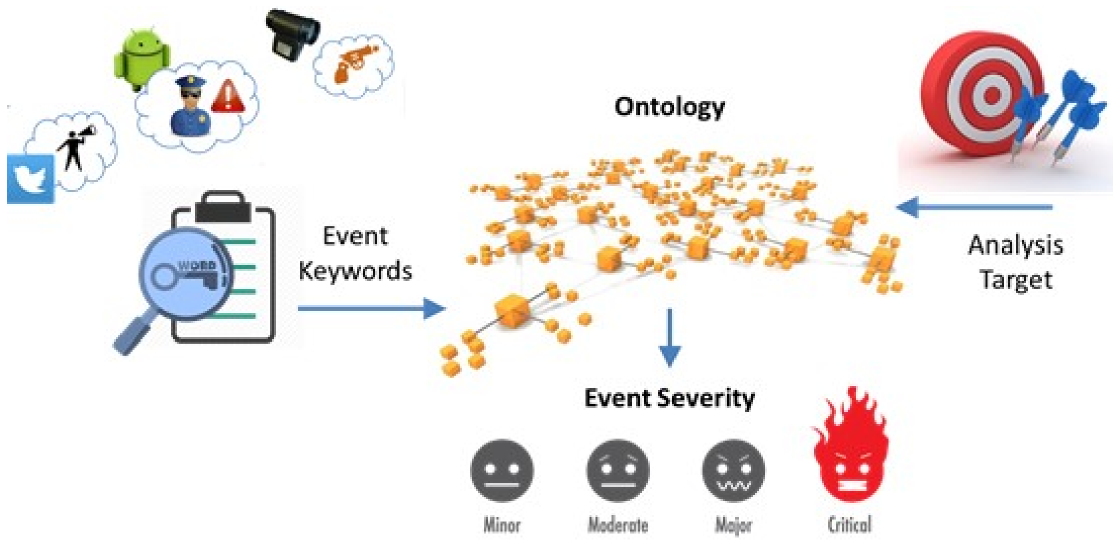


Fig. 13. Obtaining a severity score for an event in a generic crisis management framework.

value, making the usage of the *sf* parameter unfeasible.

A possible approach to compute severity. Nevertheless, including severity may bring relevant benefits to the analysis of events and the relevance labelling of clusters. A possible approach to include severity in a CMS is defined as follows. As described in Figure 13, if we consider a generic crisis management framework we need to assume that all the events regarding the crises match a defined vocabulary or ontology. For example, Secure! associates each macro-event to one or more terms from an ontology. A severity score can be assigned to the different terms defined in the ontology (e.g., terrorism, natural disaster): this way, each event could be matched to a severity score. This means that all the events in a cluster have an associated severity score, from which the severity of the clusters can be determined. For example, the severity of a cluster could be the highest severity score of its constituent events, or it could result from a weighted sum of the severity scores of the events. Further studies to include a severity score in Secure!, and successively in our relevant labelling strategy, are part of our future work. This requires to i) re-consider the data collected and the micro and macro-events generated, ii) define an approach to associate severity scores to the macro-events, and iii) modify the relevance labelling strategy to include such severity score.

8. CONCLUSIONS

This paper presented the design, development and assessment of a relevance labelling strategy for Crisis Management Systems that can retrieve and process significant amount of data from heterogeneous sources. It aims at supporting the human operator to decode and classify the crisis events delivered to the operative center of a crisis response team, assigning to each event a label that predicts its relevance for the operator. Events with higher relevance scores will be presented before the others, helping the operator to quickly clarify the context maximizing the effectiveness of the response.

Our instantiation of the relevance labelling strategy was implemented and integrated in the Secure! Crisis Management System in order to assign relevance scores to each of the events retrieved by the operator. Three real-world scenarios were investigated, showing the obtained relevance labelling results in association to a sensitivity analysis aimed to evaluate how changes in the parameters of our strategy can influence the final labelling result. Following this analysis, profiles defining values for typical situations were listed in order to help the crisis management operator tailoring the strategy on the specific scenario under investigation. All the input data used for this and results are available at [28], [35].

As future work, we will investigate how the proposed strategy scales for large scenarios (using datasets as example from *Ushaidi*) where a high number of events scattered on large areas is generated.

ACKNOWLEDGMENT

This work has been partially supported by the Regional Project POR-CREO 2007-2013 Secure!, the European FP7-IRSES DEVASSES and the TENACE PRIN Project (n. 20103P34XC) funded by the Italian Ministry of Education, University and Research.

REFERENCES

[1] Mehrotra, Sharad, Taieb Znati, and Craig W. Thompson. "Crisis management." IEEE Internet Computing 1 (2008): 14-17.

[2] Elwood, Sarah. "Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS." GeoJournal 72.3-4 (2008): 173-183.

[3] Vescoukis, Vassilios, Doulamis, Karagiorgou. "A service oriented architecture for decision support systems in environmental crisis management." Future generation computer systems (2012) 593-604.

[4] Fleischer, J., et al. "An integration platform for heterogeneous sensor systems in GITEWS–Tsunami Service Bus." Natural Hazards and Earth System Science 10.6 (2010): 1239-1252.

[5] Zook, Matthew, et al. "Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake." Available at SSRN 2216649 (2010).

[6] Farazmand, Ali. "Learning from the Katrina crisis: A global and international perspective with implications for future crisis management." Public Administration Review 67.s1 (2007): 149-159.

[7] Meier, Patrick. "New information technologies and their impact on the humanitarian sector." International review of the Red Cross 93.884 (2011): 1239-1263.

[8] De Longueville, Bertrand, et al. "Digital earth's nervous system for crisis events: real-time sensor web enablement of volunteered geographic information." Int. Journal of Digital Earth 3.3 (2010): 242-259.

[9] Wong, Mr Weng-Keen, et al. "WSARE: what's strange about recent events?" Journal of Urban Health 80.1 (2003): i66-i75.

[10] SwiftRiver Project, https://github.com/ushahidi/SwiftRiver [last accessed 15th Nov. 2016]

[11] Ushahidi Project, http://www.ushahidi.com/ [last accessed 15th Nov. 2016]

[12] Annoni, Alessandro, et al. "Orchestra: developing a unified open architecture for risk management applications." Geo-information for disaster management. Springer Berlin Heidelberg, 2005. 1-17.

[13] Secure! Project, http://secure.eng.it/ [last accessed 15th Nov. 2016]

[14] Peng, Yi, et al. "An incident information management framework based on data integration, data mining, and multi-criteria decision making." Decision Support Systems 51.2 (2011): 316-327.

[15] Cameron, Mark A., et al. "Emergency situation awareness from twitter for crisis management." Proceedings of the 21st International Conference Companion on World Wide Web. ACM, 2012.

[16] Jennex, Murray E., and Murali Raman. "Knowledge Management in Support of Crisis Response." International journal of information systems for crisis response and management 1.3 (2009): 69.

[17] Kaufman, L. and Rousseeuw P.J. (1987). Clustering by means of medoids, Y. Dodge (ed.), Statistical Data Analysis Based on the L1-norm and Related Methods. Elsevier, North-Holland, pp. 405-416.

[18] MacQueen, J. "Some methods for classification and analysis of multivariate observations." Proceedings of the 5th Berkeley symposium on mathematical statistics and probability. Vol. 1. No. 14. 1967.

[19] Park, Hae-Sang, and Chi-Hyuck Jun. "A simple and fast algorithm for K-medoids clustering." Expert Systems with Applications 36.2 (2009).

[20] C. Dietrich, and P. Pawlak, "Crowd-sourcing - crisis response in the digital age", European Union Institute for Security Studies, Issue Alert 39. Nov 2013.

[21] Raghu K. Ganti, Fan Ye, and Hui Lei, "Mobile Crowdsensing: Current State and Future Challenges", IEEE Communications Magazine, 2011.

[22] M. Bishr and K. Janowicz. "Can we trust information?-the case of volunteered geographic information." Towards Digital Earth Search Discover and Share Geospatial Data Workshop at FIS. Vol. 640. 2010.

[23] Secure! Project, Technical Report "Tassonomia eventi" (Events Taxonomy), 2013, http://secure.eng.it/ontologySecure/category.owl [last accessed 15th Nov. 2016]

[24] "Feyenoord Fans damage the Barcaccia Fountain", http://www.ansa.it/english/news/2015/02/19/feyenoord-fans-damage-barcaccia-fountain-update_ac563070-1be5-4cb7-8473-c46351759703.html (19th Feb. 2015) [last accessed 15th Nov. 2016]

[25] “Shocking pictures show Feyenoord fans clash with riot police during long-running battle through Rome”, http://www.mirror.co.uk/news/world-news/shocking-pictures-show-feyenoord-fans-5193175 (19th Feb. 2015) [last accessed 15th Nov. 2016]

[26] Bondavalli, A. Ceccarelli, L. Falai, M. Vadursi, “A New Approach and a Related Tool for Dependability Measurements on Distributed Systems”, IEEE Transactions on Instrumentation and Measurement 59(4): 820-831 (2010)

[27] Ansoff, H. Igor. "Strategic issue management." Strategic Management Journal 1.2 (1980): 131-148.

[28] “Secure! database dump” (online repository), https://github.com/tommyippoz/Miscellaneous-Files/blob/master/UNIFI/161212_HASE_Journal/HASE_Journal_Archive_Secure_DB.rar [last accessed 15th Nov. 2016]

[29] Abel, Fabian, et al. "Semantics+ filtering+ search= twitcident. exploring information in social web streams." Proceedings of the 23rd ACM conference on Hypertext and social media. ACM, 2012.

[30] PublicSonar Project, http://publicsonar.com/ [last accessed 15th Nov. 2016]

[31] Lorincz, Konrad, et al. "Sensor networks for emergency response: challenges and opportunities." Pervasive Computing, IEEE 3.4 (2004): 16-23.

[32] A de la Asunción, Marc, et al. "SIADEX: An interactive knowledge-based planner for decision support in forest fire fighting." AI Communications 18.4 (2005): 257-268.

[33] Rogstadius, Jakob, et al. "CrisisTracker: Crowdsourced social media curation for disaster awareness." IBM Journal of Research and Development 57.5 (2013): 4-1.

[34] Zoppi, Tommaso, et al. "Presenting the Proper Data to the Crisis Management Operator: A Relevance Labelling Strategy," 2016 IEEE 17th International Symposium on High Assurance Systems Engineering (HASE), Orlando, FL, 2016, pp. 228-235.

[35] “Sensitivity analysis source data” (online repository), https://github.com/tommyippoz/Miscellaneous-Files/blob/master/UNIFI/161212_HASE_Journal/HASE_Journal_Archive_Sensitivity.rar [last accessed 15th Nov. 2016]

[36] “Roma, manifestazione anti-Salvini. Vie bloccate e scontri con le forze dell’ordine” http://www.ilfattoquotidiano.it/2015/02/27/roma-manifestazione-anti-salvini-vie-bloccate-scontri-forze-dellordine/1462548/ [last accessed 15th Nov. 2016]

[37] Bondavalli, Andrea, et al. "Experimental assessment of low-cost GPS-based localization in railway worksite-like scenarios.” Measurement 46.1, (2013): 456-466.

[38] Muhammad Imran, Carlos Castillo, Fernando Diaz, and Sarah Vieweg, “Processing Social Media Messages in Mass Emergency: A Survey”, ACM Computer Survey 47, 4, Article 67, 2015.

[39] Acar, Adam, and Yuya Muraki. "Twitter for crisis communication: lessons learned from Japan's tsunami disaster." International Journal of Web Based Communities 7.3 (2011): 392-402.

[40] Caragea, Cornelia, et al. "Classifying text messages for the Haiti earthquake." Information Systems for Crisis Response and Management, ISCRAM. 2011.

[41] “Maltempo, caos in Toscana”, http://iltirreno.gelocal.it/regione/toscana/2015/03/05/news/maltempo-caos-per-il-vento-in-toscana-scuole-chiuse-e-strade-ko-1.10983828 [last accessed 15th Nov. 2016]

[42] MacEachren, Alan M., et al. "Senseplace2: Geotwitter analytics support for situational awareness", Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on. IEEE, 2011.

[43] Velmurugan, T., and T. Santhanam. "Computational complexity between K-means and K-medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points." Journal of Computer Science 6.3 (2010): 363.

[44] Massimiliano L. Itria, Melinda Kocsis-Magyar, Andrea Ceccarelli, Paolo Lollini, Gabriele Giunta and Andrea Bondavalli. “Identification of Critical Situations via Event Processing and Event Trust Analysis” in press at Knowledge and Information Systems Journal.

TABLE I. CHARACTERISTICS OF EXISTING CRISIS MANAGEMENT FRAMEWORKS

		Related Crisis Management Frameworks					
		QOS-aware SOA [3]	**ESA-AWTM [15]**	**GITEWS [4]**	**SIADEX [32]**	**CrisisTracker [33]**	**Secure! [13]**
Data Collection and Integration	***Sensor Data***	Heterogeneous	Homogeneous	Heterogeneous	Heterogeneous	Homogeneous	Heterogeneous
	Sensor Type	Vehicle Position, Weather	Tweets	Tide Gauge, Seismology, GPS Ocean Observation, Weather	External GIS, Web Feeds	Tweets	Authority Alerts, Webcams, Tweets, Secure! App
	Sensor Data Check	Temporal and spatial down-sampling selectively discard data in order to reduce the amount of data transmitted over the network	Only tweets with specific hashtags, with relevance weights	A module acts as filter reducing the data coming from the sensor streams to relevant data only	Data must comply with ontologies stored in the integrated BACAREX Ontology Server	Data coherence checked using supporting metadata. Short (no more than 2 words after stop words removal) messages are discarded	Micro-events are filtered by a dedicated *Event Filtering* module
	Data Integration	Added-Value services layer	Not needed (homogeneous sensors)	Tsunami Service Bus (TSB)	SIADEX Web Center	Not needed (homogeneous sensors)	*Media Integration* layer
Information Processing	***Event Correlation***	Adding value to real time environmental data, predicting future states, providing operational guidance, …	Finding Geospatial links for each considered tweet (if not existing in the message, got from user's location)	Sensor integration from Tsunami Service Bus (TSB), including post processing and quality checks	Made through HTN planning, which involves all the available data	Indexing is made using hashtags. Then, a bag-of-words approach is used to compare different tweets and group them.	Micro-events are correlated to obtain macro-events. A macro event groups micro-events with similar characteristics
	Event Relevance Labelling	Missing	Missing	Missing	Missing	Not deepened. Stories deemed irrelevant can be hidden, preventing them from showing up in search results.	A relevance labeling strategy is the main contribution of this paper.
Human Interface	***Presentation Techniques***	Web Interface allows for a universal access and greatly reduces learning time and thus attracts more non-GIS professionals	Browser Presentation component: uses web services to retrieve and display information to a watch officer	Not provided: need of external applications that fetch data from Tsunami Service Bus (TSB)	Not described: however, the output is composed by one (or more) approximate temporal plans ready for execution	Not described: however, it is mentioned that both expert and volunteer participants found the system to be intuitive and easy to use	Secure! Control Panel, presented in this paper. It shows the macro-events to the operator on a map with their detailed characteristics

TABLE II. SCENARIOS AND PARAMETERS DETAILS

<table>
<tr><th rowspan="2">Scenario</th><th colspan="2">Query Clauses</th><th colspan="3">δ</th><th rowspan="2">A</th><th rowspan="2">Statistical Index</th></tr>
<tr><th>fromDate</th><th>toDate</th><th>δ_{lat}</th><th>δ_{lon}</th><th>Δ_t</th></tr>
<tr><td>Europa League Match</td><td>18th Feb</td><td>19th Feb</td><td>0.06</td><td>0.06</td><td>4h</td><td>{50%, 40%, 30%, 20%}</td><td>MEAN</td></tr>
<tr><td colspan="8">Query: getEventWithAndCondition?Area=Rome&fromDate=18022015&toDate=19022015</td></tr>
<tr><td>Political Manifestation</td><td>27th Feb</td><td>28th Feb</td><td>0.06</td><td>0.06</td><td>4h</td><td>{50%, 40%, 30%, 20%}</td><td>MEAN</td></tr>
<tr><td colspan="8">Query: getEventWithAndCondition?Area=Rome&fromDate=27022015&toDate=28022015</td></tr>
<tr><td>Weather Warnings</td><td>4th Mar</td><td>5th Mar</td><td>0.06</td><td>0.06</td><td>4h</td><td>{50%, 40%, 30%, 20%}</td><td>MEAN</td></tr>
<tr><td colspan="8">Query: getEventWithAndCondition?Area=Tuscany&fromDate=04032015&toDate=05032015</td></tr>
</table>

TABLE III. RELEVANCE LABELLING FOR THE SCENARIOS WITH $\alpha = 50\%$. RELEVANT (R), NON-RELEVANT (NR) AND WRONG (W) EVENTS ARE REPORTED

Scenario	**R**	**NR**	**W**	**Label**
Europa League Match	43	25	0	Clustering Successful
Political Manifestation	32	18	5	Statistical Check Successful
Weather Warnings	-	-	21	Unsuccessful Labelling

TABLE IV. SETS LEGEND

Sets Legend
$L_1 = \{0.01, 0.02, 0.03\}$
$L_2 = \{0.04, 0.05, 0.06\}$
$T_1 = \{20m, 40m, 1h\}$
$T_2 = \{2h, 3h, 4h\}$
$T_3 = \{6h, 8h\}$

TABLE V. AVERAGE α VALUE AND STANDARD DEVIATION FOR EUROPA LEAGUE SCENARIO

Europa League		δ_{lat} , δ_{lon}			
		$\delta_{lat} \in L_1$ $\delta_{lon} \in L_1$	**$\delta_{lat} \in L_1$ $\delta_{lon} \in L_2$**	**$\delta_{lat} \in L_2$ $\delta_{lon} \in L_1$**	**$\delta_{lat} \in L_2$ $\delta_{lon} \in L_2$**
Δt	**$\Delta t \in T_1$**	28.06% (0.27)	29.17% (0.41)	29.17% (0.41)	32.50% (0.69)
	$\Delta t \in T_2$	36.15% (0.81)	38.33% (0.91)	38.33% (0.91)	45.00% (0.89)
	$\Delta t \in T_3$	45.58% (0.25)	46.67% (0.23)	46.67% (0.23)	50.00% (0.00)

TABLE VI. AVERAGE α VALUE AND STANDARD DEVIATION FOR POLITICAL MANIFESTATION SCENARIO

Political Manifestation		δ_{lat} , δ_{lon}			
		$\delta_{lat} \in L_1$ $\delta_{lon} \in L_1$	**$\delta_{lat} \in L_1$ $\delta_{lon} \in L_2$**	**$\delta_{lat} \in L_2$ $\delta_{lon} \in L_1$**	**$\delta_{lat} \in L_2$ $\delta_{lon} \in L_2$**
Δt	**$\Delta t \in T_1$**	24.14% (0.38)	27.27% (0.57)	26.00% (0.44)	30.00% (0.50)
	$\Delta t \in T_2$	35.00% (0.86)	41.67% (0.75)	36.67% (0.67)	45.00% (0.19)
	$\Delta t \in T_3$	43.33% (0.22)	50.00% (0.00)	43.33% (0.22)	50.00% (0.00)

TABLE VII. AVERAGE α VALUE AND STANDARD DEVIATION FOR WEATHER WARNINGS SCENARIO

Weather Warnings		δ_{lat} , δ_{lon}			
		$\delta_{lat} \in L_1$ $\delta_{lon} \in L_1$	**$\delta_{lat} \in L_1$ $\delta_{lon} \in L_2$**	**$\delta_{lat} \in L_2$ $\delta_{lon} \in L_1$**	**$\delta_{lat} \in L_2$ $\delta_{lon} \in L_2$**
Δt	**$\Delta t \in T_1$**	20.00% (0.00)	20.00% (0.00)	20.00% (0.00)	20.00% (0.00)
	$\Delta t \in T_2$	20.00% (0.00)	20.00% (0.00)	20.00% (0.00)	20.00% (0.00)
	$\Delta t \in T_3$	20.00% (0.00)	20.00% (0.00)	20.00% (0.00)	20.00% (0.00)

TABLE VIII. RELEVANCE LABELLING PERFORMANCE

Outcome Label	**Outcome Result %**	**Time (*ms*)**		
		Mean	**Clustering**	**Total**
Mean Successful	*50.0*	*1.87*	*Not Needed*	*1.87*
Clustering Successful	*16.7*	*4.20*	*14.63*	*18.83*
Unsuccessful Labelling	33.3	3.50	22.85	26.35
All	**100.0**	**2.80**	**20.11**	**12.86**

TABLE IX. PROFILES FOR RELEVANCE LABELLING DEPENDING ON SCENARIOS

		Spatial Dimension of the Investigated Area									
		Restricted (City Area)					Wide (Country)				
		Index	**α**	**δ_{lat}**	**δ_{lat}**	**Δt**	**Index**	**α**	**δ_{lat}**	**δ_{lat}**	**Δt**
Critical Events Relations	**Single - Close in Time**	MEAN	50%	0.03	0.03	2h	MEAN	50%	0.5	0.5	2h
	Single - Close in Space	MEAN	50%	0.02	0.02	6h	MEAN	50%	0.2	0.2	6h
	Multiple - Close in Time	MODE	30%	0.03	0.03	2h	MODE	30%	0.5	0.5	2h
	Multiple - Close in Space	MODE	30%	0.02	0.02	6h	MODE	30%	0.2	0.2	6h
	Not Existing	MEDIAN	20%	0.05	0.05	6h	MEDIAN	20%	0.8	0.8	6h
	Unknown	MEDIAN	40%	0.05	0.05	6h	MEDIAN	40%	0.8	0.8	6h