2012 European Intelligence and Security Informatics Conference

Estimating Citizen Alertness in Crises using Social Media Monitoring and Analysis

Fredrik Johansson, Joel Brynielsson
Swedish Defence Research Agency (FOI)
Stockholm, Sweden
Email: firstname.lastname@foi.se

Maribel Narganes Quijano
Tecnalia Research and Innovation
Zamudio, Spain
Email: maribel.narganes@tecnalia.com

***Abstract*—The use of social media for communication and interaction is becoming more and more frequent, which is also the case during crises. To monitor social media may therefore be a useful capability from a crisis management perspective, both for detecting new or emergent crises, as well as for getting a better situation awareness of how people react to a particular crisis. The work presented in this paper is part of the EU research project Alert4All, having the overall goal of improving the effectiveness of alert and communication toward the population in crises.**

***Index Terms*—Crisis management; social media monitoring; text analysis; web mining**

I. Introduction

Open source intelligence (OSINT) is quickly becoming an increasingly important source of information for many purposes, not least due to the enormous amounts of information communicated using various social media services such as Twitter and Facebook. Even though there are several issues with information collected from social media (such as a potential unreliability of the sources), there are also many potential benefits leading to an increased interest within many companies and organizations in order to be more actively involved in social media.

In the EU FP7 Alert4All research project [1], the focus is on improving the authorities' effectiveness of their alert and communication towards the population during crises. One component in a larger Alert4All system is therefore to monitor social media in order to find out how people perceive the crisis and how they react to communicated alert messages. This part of the system is referred to as the Screening of New Media (SNM) tool, which is the main focus of this paper.

Due to the enormous amounts of social media content being generated every day, it is not doable for human analysts to find, read, and analyze all relevant user generated content (UGC) from social media on their own. Instead, they need support from various automatic or semi-automatic tools and techniques to be able to filter out the content of relevance to them. Using the suggested SNM tool, we aim to be able to answer questions such as:

- Are people writing posts related to crises? (If so, what is the content of these postings?)
- Are people reacting to sent alert messages? (If not, this can be an indication of that the affected citizens have not been reached by the alerts.)
- Are people starting to "panic," or are they complaining about sent alerts or erroneous or unclear information? (If so, appropriate actions might be needed.)

To be able to answer such questions using the SNM tool, it must have the capabilities to retrieve and analyze posts in an automatic manner. For the data acquisition part of the tool, we survey existing techniques that can be used for such purposes, including both web harvesting techniques and open APIs to acquire the relevant data (blog posts, tweets, etc.) from relevant sources. For the data analysis part, we review existing natural language processing (NLP) and text mining techniques to automatically extract content, sentiment and emotions from the retrieved posts.

A basic assumption underlying the usefulness of the proposed tool is that parts of the population will use social media for communication in crises. The phases following the initial shock when a crisis occurs often involve people searching for and disseminating information about the crisis (e.g., information regarding their family, food, shelter, and transportation), and for such purposes the use of social media is becoming more and more common. As an example, social media was extensively used for such purposes in the aftermaths of the 2011 Tohoku earthquake, where the first tweet on the topic was written less than two minutes after the epicenter of the earthquake [2]. During the same crisis, tweeters asking for help with evacuating patients from the Kyoritsu hospital close to the Fukushima reactors lead to quick response after hours of fruitless attempts to get help via other channels. According to [3], monitoring of social media also made the UN World Health Organization (WHO) aware of that some people erroneously instructed others to drink wound cleaner with iodine to protect from radiation during the crisis. After finding this out, social media was used by WHO to inform people not to drink liquid iodine. Other examples of crises where social media have been used extensively are the California wildfires in 2007 and 2008, the Haiti earthquake in 2010, and the Tunisian uprising in 2011 [4].

Research articles covering the role of social media in crisis management do exist (see, e.g., [4], [5]), and various best practices for crisis communication have been suggested [6]. These best practices include listening to the public's concerns and understanding it through monitoring public opinions [4], [7], but to the best of our knowledge, no previous attempts

978-0-7695-4782-4/12 $26.00 © 2012 IEEE
DOI 10.1109/EISIC.2012.23

have been made to actually build tools that support awareness of how the crisis and communicated alert messages are perceived by the citizens. There are numerous of applications for sentiment analysis of tweets, but none that focus on emotion recognition/affect analysis in social media posts related to crises. Hence, there is a gap in research when it comes to affect analysis related to crisis alerting that we attempt to bridge in this work.

The rest of this paper is structured as follows. In Section II, we explain the overall scope of the Alert4All project and the concept of social media. The SNM part of the Alert4All project, which is the focus of this paper, is outlined schematically in Section III. In Section IV, we review existing tools and techniques that can be used as components for the data acquisition part of the tool. In the same manner, we are in Section V surveying available techniques and resources for sentiment and affect analysis. Some ethical aspects on social media monitoring and analysis are discussed in Section VI. Conclusions are provided in Section VII.

II. Background

According to a definition provided in [8], social media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user generated content (UGC). Hence, there is a clear focus on content that is publicly available and that has been created by end-users, including services and applications such as Facebook (social networks), Flickr (picture sharing), Twitter (microblogs), and YouTube (video sharing). In our view, the UGC aspect is more important than the actual technological foundation (will technology based on a hypothetical Web 4.0 count as belonging to social media?) For this reason, we prefer using the term Screening of New Media, rather than Screening of Social Media.

No matter what term is used, social media is changing the way people and organizations communicate. Sometimes the use of social media is referred to as strengthening the democracy (cf. Arabian Spring), while it under other circumstances has been blamed for supporting organized crime (cf. London riots). In any case, social media is likely to play a large role in future crises and disasters, not least due to the worldwide increase in use of smartphones and tablets. Although this may have been more evident than ever during the Fukushima crisis [2], [9], the use of social media has also been an important factor during many other recent crises [4], [10], such as the Katrina hurricane in 2005 and the Haiti earthquake in 2010. According to an online survey conducted by the American Red Cross, many people have used social media to get information about an emergency, and most people agree on that emergency response agencies should regularly monitor their websites and social media sites to promptly respond to any request for help posted there [11].

The Screening of New Media (SNM) module described in this paper is part of a larger Alert4All system [1], which aims at improving the effectiveness of alert and communication to the population in crises by means of investigating five particular areas: human behavior, role of new media, information management, communication technologies, and authorities'/responders' operations. Different modules in the Alert4All system address all these areas, and the SNM part implements a tool that provides an enhanced view of the crisis situation to the personnel in charge of managing the crisis, namely authorities and first responders, with regard to public opinions and citizen sentiments based on on-line UGC obtained from social media.

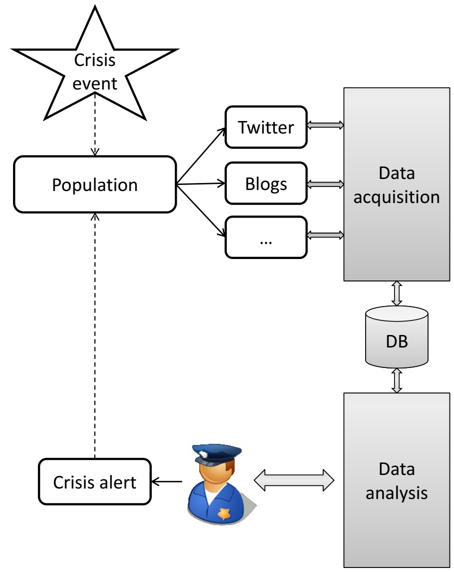

Figure 1. Overview of the intended use of the SNM tool.

III. Social media monitoring and analysis

A conceptual overview of the intended use of the SNM tool is shown in Figure 1. At a high abstraction level, the screening of new media process involves two clear and sequential subprocesses: firstly, social media monitoring for *data acquisition* and secondly, *data analysis* (including sentiment and affect analysis on the acquired data as well as visualization of relevant results). Using techniques such as web harvesting, the posts related to the crisis are retrieved in the data acquisition process and stored in a database. The retrieved posts are on regular intervals fetched from the database and become subject to various NLP techniques that are used to extract useful information such as emotions from the posts. The extracted information is then visualized to the user, in order to increase the awareness of the evolving crisis. Based on information from the SNM tool in combination with information from other sources, the user can decide on appropriate crisis alerts or messages to be communicated to citizens. By continuing monitoring social media, the effect of the communicated alert messages can be analyzed, and so on until the crisis is over.

Many companies and organizations have realized the potential of crowd-sourcing all information about personal opinions and interests that is publicly available online. Examples of purposes this is used for is to infer users' opinion about a brand, recognizing key trends in specific market sectors, engaging with online customers through focused campaigns, listening to existing dialogs about a company, and so forth. Many products and tools have recently appeared in the market providing this kind of service. Solutions such as Sysomos[1], Simplify360[2], Radian6[3], and Blogmeter[4] help companies and organizations to interpret and understand their social presence on the Web. However, when dealing with how the population is reacting to a crisis and to communicated alerts, new issues arise that no commercial or open source tools existing nowadays fulfill.

In the following, we are presenting the intended workings of the data acquisition and data analysis parts of the SNM tool.

A. Data acquisition (monitoring)

The data acquisition process is automatically launched during the crisis onset (i.e., as soon as the authorities get indications of a potential crisis) in order to find relevant UGC, providing input to the data analysis. This way, SNM users get quick results regarding the first reactions and emotions of citizens being affected by the crisis. The data acquisition process involves searching for and retrieving relevant data from many different information sources, and performing a first processing and filtering of the acquired content. The outcome of the data acquisition process is a stream of "posts" (i.e., chunks of meaningful text gathered from any information source), containing relevant conversations taking place on the Web about the crisis, ready to be processed by NLP algorithms in the data analysis part of the SNM process.

Considering the chronological order of tasks to be performed, the data acquisition module could be described by the following tasks:

1) *Configuration of search parameters*. For launching any data searching and capturing process it is essential to decide on the parameter settings regarding *what to search* (i.e., tags and keywords defining the theme of interest), *where to search* (i.e., which information sources to track), *when to search* (i.e., search process duration), and *how to search* (e.g., location-based search or any-location search). Ideally, each of these parameters should be configured for every new search process. However, for the practical case of Alert4All it would not be effective to do so since users will not always have time to configure the tool while involved in a crisis situation. Thus, the solution adopted in the SNM tool is to use default values for configuration parameters related to the specific context in which the overall Alert4All system is operating (e.g., a particular alerting plan describing the actions to be taken by authorities and first responders for a particular type of crisis). These default settings can then be changed to more well-adapted parameter settings, if the user has time to do so.
2) *Launch of opportunistic capturing processes according to the selected information sources*. Each type of information source has its specificities with regard to the type of information provided (i.e., raw, structured or unstructured data) and access method supported (e.g., whether there is a well-defined API or not). This means that different means of retrieval will be available in order to use any of the selectable information sources.
3) *Pre-processing of captured raw data*. The collected raw data will require the application of specific parsing algorithms before being submitted for data analysis. Specially for the case of data captured from non-structured sites, these parsing tasks will involve deducing the data structure (i.e., how content is organized within the web pages in terms of toolbars, ads, figure sections, and so on). Thus, depending on the information source, this pre-processing task can, e.g., involve to remove HTML tags and menu bars or to strip away irrelevant sections such as ads, header information or side bar columns. The overall aim is to find and extract relevant user generated content.

There are several critical tasks to be performed by the data acquisition module, namely: management of large data volumes, dealing with data source heterogeneity, real-time processing, and detecting and filtering out useless information.

B. Data analysis and visualization

When a post has been retrieved and stored in the database, it is time for the data analysis process in which NLP techniques are applied in order to extract entities and events, as well as for trying to classify what emotion states or moods that are expressed in the posts. The obtained results are then clustered and visualized in various ways, allowing the user to interact with the system in order to allow zooming in on interesting aspects of the data. Such visualizations include showing the number of posts containing specific emotions as a function of time, tag clouds with the most frequently occurring words acquired from the posts, and maps showing from which geographical area the posts have been sent (for applications sending the geographical position of the user).

IV. Data harvesting and open APIs for data acquisition

There are plenty of types of UGC available. However, only those that are really used for sharing personal experiences and opinions will be considered as relevant information sources for the SNM implementation. Information sources such as blogs and community forums are very relevant for the present case. Most quality blogs are interactive, allowing visitors to leave comments on the blogs, and it is in this way personal experiences and opinions are exchanged. Indeed, bloggers do not only produce content to post on their blogs but also build social relations with their readers and other bloggers. In that sense, blogging can be seen as a form of social networking.

Microblogs (such as Twitter) and social networks (such as Facebook) are of primary interest for the implementation.

[1]http://www.sysomos.com/
[2]http://simplify360.com/
[3]http://www.radian6.com/
[4]http://www.blogmeter.eu/

Twitter is nowadays the most popular online microblogging service that enables users to send and read text-based posts of up to 140 characters, known as "tweets." Since its creation in 2006, the service rapidly gained worldwide popularity, now generating over 300 million tweets and handling over 1.6 billion search queries per day. The most appealing feature of Twitter from an Alert4All perspective is that it is used for broadcasting messages to the public.

Considering "real" social network services, the number of existing websites is huge. Many of them are focused on specific topics such as academic research or sports, while those with a more general focus tend to be most popular. Facebook with more than 800 million active users has established its leadership position in most countries.

Finally, image and video hosting services such as Flickr and YouTube are also considered, which allow users to upload images and/or videos to a website. The website host stores the uploaded content on its server, and allow others to view it. It is common that users accessing photos and videos can write comments and establish a conversation with other users about the topic of the photo or video. Although the SNM tool will be implemented only for capturing content in text format, it might be interesting to consider this kind of websites because there may be relevant conversations taking place around a photo or video related to the active incident the SNM process is searching for. Therefore, for the scope of this paper, blogs, community forums, Twitter, Facebook, Flickr, and YouTube will be considered as the information sources from where content will be gathered. For some of these sources it exists well-defined APIs, while others require the use of web harvesting techniques for data acquisition purposes. Both cases are analyzed in more detail in Sections IV-A and IV-B.

A. Web harvesting

When there is no specific API for accessing the information source, as is typically the case for blogs and forums, a generic and dynamic access mechanism is required for acquiring data. In this case it is necessary to harvest the data using a crawler, i.e., a computer program that browses the Internet in a methodical, automated manner. Given that it would not be effective to browse the whole web, the used approach will be to limit the crawler to visit only specific websites (i.e., blogs and forums with users' conversations about the crisis taking place). In order to first identify the websites that the crawler should visit, a search engine will be used. The URLs of the sites provided by the search engine will be the "seeds" that will be the entry point for the crawler. In next step, the crawler browses these seeds and checks whether the robots exclusion protocol allows access to the web server or not. If allowed, the crawler reads the content and follows the links found in these pages in order to retrieve their content and to find more new links. These links will be the new seeds, and the process is repeated cyclically until no more links are found or the link depth, previously defined, is reached. Once all this data is retrieved it is indexed and stored in the database.

To summarize, data harvesting of web sites with no specific API will be done following a two-steps technique for a crisis situation: 1) relevant websites are searched for sentiments and emotions related to the crisis, using search engines APIs, such as: Custom Search API, Blog Search API or Bing Search API, and 2) web crawling is undertaken for each website obtained as a result in the first step. Appropriate proposals for the implementation of the crawling task is to base it on Apache Nutch Crawler[5], Open Search Server[6], or Heritrix[7]. All three are open source crawlers written in Java.

B. Specific APIs

For the case of Twitter, Facebook, Flickr and YouTube, there are well-defined and documented APIs available that include functions for searching and retrieving data. For Twitter, there are two APIs available: the official Twitter API, known as Twitter Search API, and an unofficial Java library for the Twitter API, named Twitter4J. In both cases, it is possible to define the keywords, hashtags and users relevant to a query. It is important to note that private tweets are not accessible and those with few followers are also discarded. The Twitter Search API is a REST API that provides simple interfaces for most Twitter functionality. Twitter4J encapsulates its functionality and provides an easy way to access to it and to integrate it into an application. Therfore, this API seems to be used by most developers.

For the Facebook social network, there is also a comprehensive developers API, known as Graph API. This social graph represents objects (e.g., people, photos, events, and pages) and connections between them, which is not actually the focus of interest for the present case. However, it also provides a search functionality that allows to search over all public objects in the social graph, which would be the interesting part of the interface for obtaining relevant posts about a particular topic, such as a crisis taking place.

Flickr supports the Flickr API. For our purposes, there are some methods in the API that allows for requesting the list of photos matching some user-specified criteria and to retrieve the list of comments associated to a particular photo. YouTube provides the YouTube Data API that offers a similar functionality.

Another important aspect for the data acquisition is the relevance of location for some crisis scenarios. Hence, a geographical search criterion to be selectable in the configuration phase can also be of interest for the data acquisition process. However, when considering its practical implementation for the different information sources that have been analyzed above, several limitations arise that could prevent its fully realization. The main limitations are that users have to voluntarily select the "publication with location" option when posting in Twitter, Facebook, Flickr and YouTube, and this is not a common practice nowadays. Moreover, posts published in blogs and forums normally do not have location information

[5] http://nutch.apache.org/
[6] http://www.open-search-server.com/
[7] http://crawler.archive.org/

attached. Thus, geo-located posts will only be provided by the data acquisition process in cases where the users contribute with geographical information when posts are created and published.

V. Sentiment and affect analysis for the crisis domain

During the last decade, a lot of research efforts have been devoted to sentiment analysis (also known as opinion mining). As stated in [12], "the year 2001 or so seems to mark the beginning of widespread awareness of the research problems and opportunities that sentiment analysis and opinion mining raise." Most work within sentiment analysis has, according to [13], been focused on the classification of a text document or its sentences with respect to expressing a positive or negative (or sometimes neutral) opinion towards the topic of the text (assuming the document to have a single topic). This often holds true for various kinds of product reviews, such as book and movie reviews. Not only are such reviews mainly about a clearly specified topic, they are also popular for research since they in addition to the text description are also often associated with a human-annotated rating of the evaluated product, which can be used for supervised training of machine learning algorithms, as well as for evaluation purposes. In addition to judge whether a text document or sentence expresses a positive or negative opinion orientation (polarity), it is for many applications also of interest to determine the intensity of the opinion [13], i.e., how strong the opinion is. For some applications, such as the SNM tool, it is also needed to classify how strongly certain emotions such as joy or anger are expressed in a text. According to [14], there is a large intersection between the concepts of opinions and emotions, although they are not completely overlapping. Emotion classification or affect analysis can be considered to be a sub-problem of the broader sentiment analysis problem [15]. The overall problem of sentiment analysis is described in Section V-A, while we in Section V-B focus on the particular problem of affect analysis, for which less research exists.

A. Sentiment analysis

There are various explanations or reasons for the sudden interest in sentiment analysis. One such reason is the growing amounts of opinion-rich text resources made available due to social media, giving researchers and companies access to the opinions of ordinary people. Before the development of Internet, not much opinionated text was easily accessible [14]. Another reason for the increased interest in sentiment analysis is the recent advances that have been made within the fields of text mining and NLP. There are several interpretations of what is covered by the term sentiment analysis, but here we will use the definition given in [14], i.e., sentiment analysis being the computational study of opinions, sentiment, and emotions expressed in text.

The most well studied sub-problem of sentiment analysis is that of sentiment classification [14]. As the name implies, the focus of sentiment classification is on classifying opinionated documents as expressing a positive, negative, or neutral opinion (sometimes referred to as document-level classification), or doing the same thing for a sentence (sentence-level sentiment classification). Another important problem is to find out whether a sentence contains any opinions at all, referred to as subjectivity classification or opinion identification throughout the literature.

An even harder problem is what in some work is referred to as feature-based sentiment analysis, including both the discovery of targets on which opinions have been expressed, and the determination of whether the opinions are positive, negative, or neutral [14]. The target of an opinion can be nearly anything: individuals or products, as well as organizations, services, and topics. For the scope of Alert4All, the sub-problem of opinion search and retrieval is of interest, i.e., to make a keyword search on a specific target, and to summarize the positive and negative opinions on the issue. Hence, sentiment analysis can have many other text mining problems such as topic-based classification as subcomponents. With this said, sentiment analysis can be considered to be a harder problem than topic-based classification, since topics often can be identifiable by keywords alone, while sentiments often are expressed in a more subtle manner [16]. Subtle opinions where no single words contain obvious positive or negative meaning are hard to detect with even the most state-of-the-art text analysis techniques present today.

One important point worth noticing is that most of the literature on sentiment analysis has been focused on texts written in English. Consequently, most of the developed resources available are therefore targeted on English. There have been attempts to perform machine translation from other languages to English as a preprocessing step to the sentiment analysis (see, e.g., [17]), but the results obtained from such methods can be expected to be worse than if working with resources developed directly for the language of interest. To develop such resources can however be costly, making it reasonable to make use of machine translation in some circumstances.

While much work has been done on deciding sentiment at the level of individual text documents or lower level, considerably less work has been done on an aggregate level [18]. Some exceptions to this are the works reported in [19], [20], and [17]. Results in [18] indicate that determining sentiment or mood at an aggregate level gives better results than when doing the same thing at a single document-level. This is good news for the work in Alert4All since we are more interested in how groups of people are reacting to a crisis message rather than individual citizens. Another work of interest for the Alert4All project is presented in [21]. There it is described how text and data mining techniques can be used for capturing the public's opinion towards governmental decisions, by automatically extracting topic and sentiment from online posts on governmental issues.

Much of the early research on sentiment analysis was based on manually constructed sentiment/discriminant-word lexicons [12]. However, a large shift towards data-driven approaches has since then been evident, not least due to the increased availability of opinionated text corpora (unlabeled as well as labeled). One motivation for this is that humans may not always have the best intuition for choosing discriminating words, as demonstrated by experiments presented in [16]. Access to unlabeled opinionated text documents may not seem very useful at first glance, but quite much work exists on the use of unsupervised learning for building the sentiment lexicon as a replacement for manually crafted lexicons. Typically, this is achieved by forming clusters of words that co-occur frequently together with seed words having a known polarity (and where the polarities of the terms in the cluster are assigned to be the same as the polarities of the known seed words). Some work (see, e.g., [22]) also analyze whether words are linked by conjunctions such as "and" (indicating that they should belong to the same cluster) or "but" (indicating that they should belong to different clusters). It is also usual to use the well-known lexical database WordNet [23] to propagate information about polarity to word synonyms and antonyms. Nevertheless, most existing techniques for document-level sentiment analysis rely on supervised learning [14]. There are many supervised learning algorithms to choose among, and most of them can easily be applied to sentiment analysis. Examples of such algorithms are naïve Bayes, maximum entropy, and support vector machines (SVMs), which are compared in [16]. All of these techniques performed quite well relative to a human-generated baseline, although the obtained accuracies ($\approx 80\%$ for a well-balanced movie-review corpus consisting of 700 positive and 700 negative opinions) are not as good as for standard topic-based categorization problems. Although quite good performance can be achieved when using a classifier on data originating from the same domain as it has been trained on, it is important to note that sentiment classifiers are often performing poorly when applied to data from another domain [14].

Irrespectively of the sentiment lexicon being manually crafted, or if it is created using some kind of statistical analysis, it typically consists of a large set of terms associated with a positive or negative state, and some indication of the strength of the word. Examples of sentiment words that can indicate positive opinions are "great" and "excellent," while "horrible" and "bad" are examples of words indicating negative opinions. Especially adjectives have been shown to indicate subjectivity and opinion [14], [21], [22], but also other word classes are useful, e.g., adverbs [24]. In information retrieval and text mining applications, it is common to look at how frequently terms are occurring (using the well-known term frequency-inverse document frequency, TF-IDF [25]). TF-IDF is used in many sentiment analysis approaches as well, although term frequency has by some researchers been argued to not be as important for sentiment analysis. Actually, experiments in [16] indicate that better performance can be achieved if replacing term frequency with term presence (i.e., replacing the number of instances of a term with a 1 if the term has been present, and keeping the 0 otherwise). Another, even more important, difference from traditional information retrieval is the importance of negation. While inclusion of the word "not" has a very limited impact in a traditional bag-of-word representation, it can flip the sentiment of a sentence completely.

When it comes to availability of opinionated documents in general, this is clearly not an issue since there are huge amounts of opinionated weblogs, product reviews, discussion forums, etc. available on Internet. However, these are not labeled in the general case, and neither are they well-suited for the crisis management domain. An important question hence becomes how to construct a sentiment classifier well suited to the this particular domain. Our approach is to make use of "standard" resources based on general discriminative-world lexicons based on supervised and unsupervised learning, and to extend these using considerably fewer manually labeled samples from earlier crises (e.g., tweets from the Fukushima crisis). An example of a resource that can be used for the more general sentiment analysis is SentiWordNet [26], which to each WordNet [23] synset assigns three numerical scores (describing the positivity, negativity, and objectivity of the synset). These numerical scores are based on the classifications from an ensemble of classifiers that have been trained in a semi-supervised fashion.

B. *Affect analysis*

Although only a few types of basic emotions exist[8], there are very many different ways in which these can be expressed in written text. Consequently, the problem of affect analysis is a challenging issue, as has been observed by [26]. Interestingly enough, this seems to hold true also for humans. In an experiment reported in [28], manual annotation of affective content in children stories resulted in quite low inter-annotator agreement. In the same manner, low accuracy for a human assessor when classifying affect in blog posts was reported in [29].

Closely related to our interest in classifying the emotional reactions in groups of people on communicated alert messages is the work on mood classification reported in [29] and [18]. According to the authors, this was the first published work on computational analysis of affect in blogs. One attempt to analyze public opinion and mood in tweets is presented in [30]. As for sentiment analysis in general, most of the work on emotion analysis has been carried out for English [26].

A good definition of emotion classification of a linguistic unit (i.e., affect analysis in our vocabulary) is given in [28]. They define it as a multi-class classification problem, where T denotes a text, and where $s \in T$ is an embedded linguistic unit such as a sentence. By letting k represent the number of emotion classes $E = \{em_1, em_2, \ldots, em_k\}$, the goal of the classification becomes to determine a mapping function $f : s \rightarrow em_i$ resulting in an ordered pair (s, em_i). This definition can of course also be used for sentiment classification, where the possible classes are positive, neutral, and negative. In the general case where multiple emotion classes can characterize

[8]There seems to be no consensus on which the basic or primary emotions are, but most available research on affect analysis seems to focus on anger, disgust, fear, happiness, sadness, and surprise, as suggested in [27].

a linguistic unit s, we instead search an ordered pair (s, E'), where $E' \subset E$.

According to [15], a central component for much work on affect analysis is manually created affect lexicons, containing words or phrases associated with specific emotions that can be used for estimating the intensity of various affect in texts. An example of work that relies on a manually constructed affect lexicon is presented in [31], used for affect analysis of a text document by tokenizing the document and looking up word for word in the lexicon. If a token is present in the lexicon, its affect categories and their associated centrality and intensity scores are retrieved. These are then combined using fuzzy logic for estimating the overall presence of various emotions in the text. Manually constructed affect lexicons are also used in the work presented in [30], where they are used to analyze the Iranian public opinion and mood after the 2009 Iranian presidential election, based on Twitter feeds.

In [28], it is argued that it is necessary to go beyond using simple rules and hand-crafted emotion lexicons for affect analysis. As an alternative, they suggest the use of supervised machine learning-techniques (in their case, using the SNoW architecture for learning a linear classifier in the used feature space). Another way to automatically construct a discriminant-word list based on labeled data is presented in [29]. There, SVMs are used to classify the mood of blog posts. A number of features were used as input the SVM, e.g., frequency counts of unigrams, the sentiment of the words, and the pointwise mutual information (PMI) between the words in the blog entry and moods in training data. The presented results are modest for identifying specific moods in the blog posts. Somewhat surprisingly, aggregating the mood sets into two single clusters (positive moods and negative moods) and thereby transforming it into a sentiment analysis problem rather than an affect analysis problem did not yield much better results. However, it was observed that an increase of training set size yielded improved performance, indicating that better performance can be achieved with even larger training data sets. In a follow-up experiment described in [18], the SVM was changed to a simple Pace regression classifier (a form of linear regression analysis), and the task was changed to capturing aggregate (global) mood levels in blog posts rather than classifying each individual post. The obtained results indicate that predicting the intensity of moods over a time span can be done with quite high degree of accuracy. Of relevance to Alert4All is also the discovery that the proposed method failed to discover the high percentage of bloggers reporting moods such as "sadness" after the large-scale terror attack taking place in London on the 7th of July 2005. Frequently used words were, e.g., "terror" and "al-Qaeda," and since these were not part of the training data, the method failed to discover the changed mood. This shows the importance of having a representative training data set for learning.

A conditional random field (CRF) classifier is used in [26] for emotion classification of blogs as well as news headlines. The training and test data has been annotated and verified manually. Affect words lists for English and Bengali are together with a number of features such as POS tags, sentence length, and emoticons used as input to the classifier. An obtained accuracy of approximately 60–70% is reported for the emotion classification task.

A blog resource that has been used in some experiments on affect analysis [18], [29] is the weblog service LiveJournal. Their web interface allows the blog authors to indicate their mood when writing the blog update by selecting a mood from a predefined list of 132 moods (or by entering a mood of their choice in free-text). As observed in [29], this resource has the disadvantage of having a potentially low level of annotator agreement, since it is likely that their definitions of various moods will differ. Another potential problem is that many of the moods are not real emotions or moods (e.g., "working"). With this said, the resource is a non-artificial large-scale resource complete with labels of mood indications, making it a very interesting data source. Furthermore, to remove blog posts with unwanted moods is an easy task, making it easy to reduce the dataset to consist only of emotion categories of interest. Weblogs are also used in [32] and [33], where emoticons are used as labels of emotion categories for blog posts in the Yahoo! Kimo Blog service. WordNet-Affect [32] is a linguistic resource for lexical representation of affective knowledge, derived from WordNet. WordNet-Affect can be used as a basis for an affective lexicon, since it assigns affective labels (a-labels) to a large number of WordNet synsets and words representing affective concepts (e.g., emotions, cognitive states, behaviors, and feelings). As for many other resources, WordNet-Affect is limited to English.

VI. Discussion

Since the tool described in this paper deals with techniques for automatic retrieval of UGC and the analysis of the retrieved text, it becomes important to consider ethical aspects of such work. It is not very hard to come up with a scenario in which these kinds of techniques can be misused by the wrong type of regime (e.g., for tracking people that have a negative sentiment against the current regime in a country). However, this is obviously not the purpose of the work presented here. The purpose is instead to contribute to the overall goal of Alert4All, i.e., to improve the effectiveness of alert and communication towards the population during a crisis. Neither are these types of techniques completely novel. Rather, it is the application of the techniques to crisis management that is new. Hence, the kind of ethical issues related to the SNM tool are no different from those related to other types of tools dealing with social media monitoring and analysis.

That said, it is obvious that this kind of application has to take data protection and privacy issues into consideration, since individuals' have a right to privacy. Processing personal data with respect to this right is regulated in the EU directive 95/46/EC. In order not to interfere with this right, we will take a number of actions for reducing the impact on privacy issues. These actions include the use of strict data retention periods, constraining the user's direct access to the data, and to not store any user IDs in the database. We are also currently looking into available techniques for privacy-preserving data mining, in order to see if such techniques can also be applied on unstructured data to further strengthen the citizens' privacy.

VII. Conclusions and future work

In this paper we have presented a tool for social media monitoring and analysis, which we refer to as the Screening of New Media (SNM) tool. The purpose of this tool is to be used in crises for providing authorities with an understanding of how citizens' are reacting to communicated alert messages. This will be achieved using web harvesting techniques as well as open APIs to retrieve relevant posts, and by using sentiment and affect analysis techniques for classifying the emotion states expressed in the acquired posts. In order to build a state-of-the-art system for this task, we have made a literature survey on existing research on sentiment and affect analysis, as well as identified relevant information sources to include in the data acquisition phase. The results from this study have influenced our design of the system which is now finished. In the next phase of the project the system will be implemented and integrated in the overall Alert4All system.

Acknowledgements

This work has been supported by the Alert4All research project which is funded by the European Union Seventh Framework Programme under contract no 261732, and by the R&D program of the Swedish Armed Forces.

References

[1] C. Párraga Niebla, T. Weber, P. Skoutaridis, P. Hirst, J. Ramírez, D. Rego, G. Gil, W. Engelbach, J. Brynielsson, H. Wigro, S. Grazzini, and C. Dosch, "Alert4All: An integrated concept for effective population alerting in crisis situations," in *Proceedings of the Eighth International Conference on Information Systems for Crisis Response and Management (ISCRAM 2011)*, Lisbon, Portugal, May 2011.

[2] S. Doan, B.-K. H. Vo, and N. Collier, "An analysis of Twitter messages in the 2011 Tohoku eartquake," in *Proceedings of the 4th ICST International Conference on eHealth*, 2011.

[3] S. Sternberg, "The world to the rescue – Japan crisis showcases social media's muscle." [Online]. Available: http://www.usatoday.com/printedition/news/20110412/1ajapansocialmedia12_cv.art.htm

[4] S. R. Veil, T. Buehner, and M. J. Palenchar, "A work-in-process literature review: Incorporating social media in risk and crisis communication," *Journal of Contingencies and Crisis Management*, vol. 19, no. 2, pp. 110–122, 2011.

[5] S. Nilsson, J. Brynielsson, M. Granåsen, C. Hellgren, S. Lindquist, M. Lundin, M. Narganes Quijano, and J. Trnka, "Making use of new media for pan-european crisis communication," in *Proceedings of the Ninth International Conference on Information Systems for Crisis Response and Management (ISCRAM 2012)*, Vancouver, Canada, Apr. 2012.

[6] M. W. Seeger, "Best practices in crisis communication: An expert panel process," *Journal of Applied Communication Research*, vol. 34, no. 3, pp. 232–244, 2006.

[7] H. Artman, J. Brynielsson, B. J. E. Johansson, and J. Trnka, "Dialogical emergency management and strategic awareness in emergency communication," in *Proceedings of the Eighth International Conference on Information Systems for Crisis Response and Management (ISCRAM 2011)*, Lisbon, Portugal, May 2011.

[8] A. M. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," *Business Horizons*, vol. 53, no. 1, pp. 59–68, 2010.

[9] Beat Communication, "The effect of social media on the disaster relief effort following the March 11 earthquake in Japan." [Online]. Available: http://www.prweb.com/releases/2011/7/prweb8667174.htm

[10] L. Palen and S. B. Liu, "Citizen communications in crisis: Anticipating a future of ict-supported public participation," in *Proceedings of ACM CHI 2007 Conference on Human Factors in Computing Systems*, 2007.

[11] (2010, August) American Red Cross. [Online]. Available: http://www.redcross.org/www-files/Documents/pdf/other/SocialMediaSlideDeck.pdf

[12] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, pp. 1–135, 2008.

[13] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Language Resources and Evaluation*, vol. 1, 2005.

[14] B. Liu, *Handbook of Natural Language Processing*, 2010, ch. Sentiment Analysis and Subjectivity.

[15] A. Abbasi and H. Chen, "Affect intensity analysis of dark web forums," in *Proceedings of the 5th IEEE International Conference on Intelligence and Security Informatics*, 2007.

[16] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2002.

[17] M. Bautin, L. Vijayarenu, and S. Skiena, "International sentiment analysis for news and blogs," in *Proceedings of the 2nd International Conference on Weblogs and Social Media*, 2008.

[18] G. Mishne and M. de Rijke, "Capturing global mood levels using blog posts," in *Proceedings of the 2006 Spring Symposia on Computational Approaches to Analyzing Weblogs*, 2006.

[19] R. Tong, "An operational system for detecting and tracking opinions in online discussion," in *Proceedings of the SIGIR 2001 Workshop on Operational Text Classification*, 2001.

[20] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proceedings of the 14th International Conference on World Wide Web*, 2005.

[21] G. Stylios, D. Christodoulakis, J. Besharat, M. Vonitsanou, I. Kotrotsos, A. Koumpouri, and S. Stamou, "Public opinion mining for governmental decisions," *Electronic Journal of e-Government*, vol. 8, no. 2, pp. 203–214, 2010.

[22] V. Hatzivassiloglou and K. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, 1997.

[23] G. Miller, "Wordnet: A lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[24] P. Turney, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 2002.

[25] G. Saltong, A. Wong, and C. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[26] D. Das and S. Bandyopadhyay, "Analyzing emotion in blog and news at word and sentence level," in *Proceedings of the 4th Indian International Conference on Artificial Intelligence*, 2009.

[27] P. Ekman, "Biological and cultural contributions to body and facial movement," *Anthropology of the Body*, pp. 34–84, 1977.

[28] C. Ovesdotter Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in *Proceedings of the 2005 Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005.

[29] G. Mishne, "Experiments with mood classification in blog posts," in *Proceedings of the 1st Workshop on Stylistic Analysis Of Text For Information Access*, 2005.

[30] S. Beth Elson, D. Yeung, P. Roshan, S. R. Bohandy, and A. Nader, "Using social media to gauge iranian public opinion and mood after the 2009 election," RAND Corporation, Tech. Rep., 2012.

[31] P. Subasic and A. Huettner, "Affect analysis of text using fuzzy semantic typing," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 483–496, 2001.

[32] C. Strapparava and A. Valitutti, "Wordnet-affect: an affective extension of wordnet," in *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 2004.

[33] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proceedings of the 4th International Workshop on Semantic Evaluations*, 2007.