SOCIAL MEDIA INFORMATION ANALYSIS FOR CRISIS AND DISASTER MANAGEMENT

Hermann Huber, Georg Neubauer, Andrea Nowak, Joachim Klerx, Bettina Jager

AIT Austrian Institute of Technology[1]

Klaus Mak, Christian Meurers

Austrian Armed Forces, National Defence Academy[2]

Keywords

Social Media, Data Mining, Information Extraction, Crisis and Disaster Managment

Abstract

Mining Social Media content provides a lot of opportunities for various application domains. This paper investigates opportunities and challenges of mining social media data specifically for crisis and disaster management (CDM). Social Media content can be characterized as vast, noisy, distributed, unstructured and dynamic (Gundecha & Liu, 2012). In order to make use of the enormous amount of potentially useful content, efficient and scalable data extraction and analysis tools may be required. Hence, we introduce some promising basic techniques originating from the data mining research area that can be applied to process social media content in order to eventually generate useful and previously unknown insights for the CDM domain. Those insights may be used to create more accurate prediction models, support crisis managers in the decision making process, reveal influential peers in communication networks or improve the process and accuracy in public alerting scenarios. We refer to this process as Social Media Mining (SMM). Utilizing user-generated content in crisis situations induces some additional, rather non-functional issues like privacy, accuracy or reliability concerns. We propose some system design considerations in order to circumvent these issues by design. Furthermore, we describe some common CDM challenges that may be addressed with the use of SMM and emphasize the necessity of combining Social Media and Data Mining techniques by illustrating a practical showcase. Finally, we provide a discussion and assess potentials and risks.

Introduction

According to a 2011 IDC survey[3] the amount of existing data is estimated to approximately 1.8 zettabytes which is 1.8 trillion gigabytes. Furthermore, the mass of data is doubling every two years. In the same period of time the amount of IT staff is estimated to grow by factor 1.5. In order to make use of this huge mass of data, advanced analysis tools are needed to derive knowledge that is implicitly hidden in that data. During the last decade the popularity of social media as well as the number of participating users increased steadily, e.g. 73 % of

---

[1] Donau-City-Straße 1, 1220 Vienna, Austria{forename}.{surname}@ait.ac.at

[2] Stiftgasse 2a, 1070 Vienna, {forename}.{surname}@bmlvs.gv.at

[3] IDC Survey 2011, "Digital Universe Study: Extracting Value from Chaos", http://germany.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf, accessed 11.06.2013

wired US teens used social networks in 2010 reflecting an increase by 18% since 2006. Even though user statistics are still far from being representative regarding the entire society, the tremendous amount of available data encourages the idea of computer supported analysis for CDM purposes. Such data can be used to reveal indications of previously unknown societal parameters like opinions, trends or human behaviour in certain contexts that eventually generate new knowledge relevant for domains such as crisis and disaster management.

So the overall challenge is to collect raw data and transfer it to usable knowledge. Literature provides numerous definitions of the term knowledge. (Ackoff, 1989) describes the relation between data and knowledge in the so called Pyramid- or Data-Information-Knowledge-Wisdom-Model (DIKW, see Figure 1). The data layer is the most basic one. According to his definition, data is just non-interpreted symbols. Data observed in a certain context generates information (i.e. the digits "17" is raw data, 17°C is information). Adding context to information pushes it to the next level in the model which is knowledge. The boundary between information and knowledge is not well defined in literature. (Koohang, et al., 2008) describe knowledge as meaningful and useful information (i.e. in winter 17°C is quite much in Austria). The last layer describes wisdom which, according to (Ackoff, 1989), can be perceived as evaluated understanding. Paradoxically, wisdom, the most complex and diffuse DIKW concept is known to mankind since hundreds or thousands of years. At the same time, the rather easy to understand concept of data is young, at least in a linguistic sense. Today it seems to be impossible for ICT systems to handle the concept of wisdom. However, extracting information from data is a commonly applied, well understood task in various domains. With respect to Ackoff's model, we define the term data in the context of Social Media as highly structured content (e.g. database tables, markup languages, etc.) that can be used as input for algorithms without any further pre-processing. Though, the majority of Social Media content is human readable which is unstructured by nature. Plain text for instance carries a lot of information that is easily extractable by humans. Machines need to pre-process and analyse the symbols in order to extract valuable information like person names or emotions.

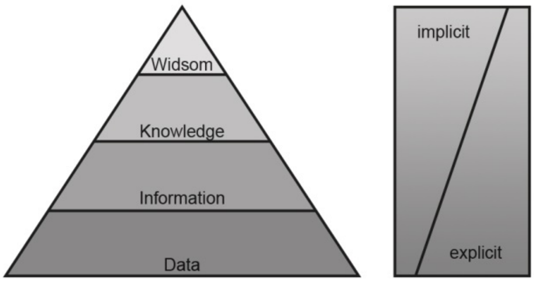


Figure 1: Relation between Data, Information, Knowledge and Wisdom (Ackoff, 1989). The representation of the respective concepts is explicit in lower layers and implicit in upper layers. The data mining research area focuses on revealing implicit knowledge from raw input data.

Data mining techniques can be used to perform transitions between several DIKW layers. Various research areas are already making use of such transitions. Astronomy for instance uses data mining techniques in order to analyse images taken by telescopes. Discovering supernovae on their early stages is a crucial success factor exploring Dark Energy. Today, astronomers detect about 100 supernovae a year. Once the Large Synoptic Survey Telescope (LSST) will be on duty in 2029 researchers expect to reveal 1000 supernovae per night which in terms of image data is about 30TB (Borne, 2009). Investigating this huge amount of data requires advanced analysis techniques and can hardly be done manually. The data mining research area aims to provide elaborate techniques to reveal knowledge that is implicitly hidden in explicit data.

According to (G.K.Gupta, 2009), Data Mining may be defined as

> *"… a collection of techniques for efficient automated discovery of previously unknown, valid, novel, useful and understandable patterns in large databases. The patterns must be actionable so that they may be used in an enterprise's decision making process."*

This definition comprises all major aspects of the data mining process. First of all it is important to notice that data mining is a collection of techniques, algorithms and software tools rather than a single technology or product. (Liu, 2007) describes data mining as a multidisciplinary field that involves machine learning, statistics, databases, artificial intelligence, information retrieval and visualization. So the implementation and the configuration of some chosen techniques may differ according to the given use case. Another aspect is to discover previously unknown, valid and novel patterns in large databases. What that means is that ideally, data mining reveals new insights that didn't exist previously rather than to prove or disprove a given hypotheses. The term valid refers to the fact that all newly discovered patterns have to be reproducible and must not be the result of random fluctuation in given data. This is where the amount of data comes into play. Increasing the amount of analysable data will lead to a better significance regarding the validity of a detected pattern.

A very basic data mining technique is the concept of *association rule mining*. Association rules are often applied in shopping chart analysis scenarios to detect items that are purchased together frequently in order to place them next to each in the shelf or release suitable marketing campaigns. A typical association rule may be of the type: "If the customer is male, between 20-30 years and if his average phone call length is less than 1:30min then he'd most probably terminate his contract within the next fortnight." An important fact is that the rule doesn't contain any justification nor does it prove any causal relation. More investigations might be necessary to determine the reason for the causal relations, in this case the sudden contract cancelations (i.e. competitor offerings, missing features, bad service, etc.). *Supervised Learning*, also referred to as *classification*, is probably the most prominent data mining technique. It aims to create a so called classifier that assigns new data items to previously defined classes. Incoming mail may be grouped in classes like "Business", "Family", "Confidential" and "Spam". Whenever a new mail arrives at the user's inbox the classifier studies its characteristics and assigns the mail to the most likely class. In order to do so, the classifier needs to learn the characteristics of each single class in advance (i.e. keywords, senders address, message length, time of arrival, attachments, etc.). This can be achieved with a representative set of training data mails that are already assigned to the correct class manually. *Unsupervised Learning* does not require the availability of a pre-classified training data set. This technique is also referred to as *clustering*, as it aims to group similar data items together in so called clusters. Data items within the same cluster share similar characteristics and provide a high degree of cohesion. In contrast, clusters are different to each other and are loosely coupled. With some text analysis algorithm in combination, the clustering technique is able to process given text documents (e.g. papers, web sites, tweets, etc.) and group them in clusters that share the same topic (e.g. radiology, sports, earthquakes, etc.). It is important to understand that clustering techniques are not designed for deriving a suitable cluster name. The resulting groups have to be interpreted by a domain expert who makes further decisions about the similarities found.

Social media potentially holds a lot of implicit information and is therefore a suitable subject to data mining analysis. Generally, Social Media comprises various digital media technologies to enable user interaction and the creation of multimedia content for sharing within a community or between individuals. By interacting socially, one-to-many communication is turned into the communication type "many-to-many" which is more familiar to human nature (Shirky, 2009). Researches benefit from extracting useful information from raw data retrieved from social media by discovering and exploring the information flow as a type of fieldwork experience. Structuring and classifying the unsystematic data is one of the key challenges in gaining relevant information for knowledge building.

The Haitian earthquake in 2010 was mentioned as the first use case of coordinating and engaging support for first responders via digital media, such as short message services and Social Media platforms. As summarized by (Fraustino, et al., 2012), Social Media played a key role in information dissemination in the Haitian earthquake 2010. Different types fulfilled various public needs. Twitter for example, was used to stay in contact with others (2.3 million tweets) and to coordinate disaster relief efforts by providing special skills voluntarily (e.g. technology skills). As an example, the appearance of the U.S. volunteers' initiative led by Tufts University developed the platform Ushahidi-Haiti, a crisis map based on incident reports of residents and volunteers. Crowdsourcing allows a qualified crowd to participate in different tasks such as provision or validation of information, but also editing in case of request (Gao, et al., 2011). Hurricane Sandy (2012) stimulated the appearance of Social Media in the disaster context. While Twitter served as a source of information and misinformation at the same, the hurricane achieved the second place on Facebook's hot topic and latest storm-pictures were provided to the public via Instagram. It is known that every Social Media type has its own purpose; while blogs are often used for emotional release and support, Twitter is used as a source for breaking news. Sometimes, content sharing platforms like YouTube have a negative connotation due to the opportunity to view and share shocking disaster videos (Fraustino, et al., 2012).

Processing social media content for CDM applications generates interesting opportunities but at the same time induces some additional issues that have to be taken into account. One of the major non-functional challenges is to use data without violating any privacy or legal constraints. Another important aspect is to detect invalid or unreliable content providers (i.e. persons, platforms, channels, etc.). This paper describes how to address these issues in terms of system design and assesses the applicability of several Social Media platforms regarding SMM purposes.

THESIS

An effective approach to circumvent the non-functional issues is to provide applications (i.e. social platforms, websites, mobile apps, etc.) dedicated to specific CDM use cases and accordingly inform users about its purpose as well as to clarify what is going to happen to their data. This approach is very close to the concept of crowd-sourcing. A lot of initiatives exist to foster the idea of utilizing user-generated content. The most prominent ones are Ushahidi[4] and Sahana[5]. These initiatives provide free and open-source software collections that facilitate geo-spatial information collection and visualization in crisis situations. The crowd-sourcing approach heavily depends on a high number of participants which are encouraged to self-adjust invalid information. (Neubauer et al, 2013) extended the user-generated content idea by explicitly asking specific users or groups of users to contribute some relevant but yet missing input to reduce fuzziness and gain additional reliability which is especially important for the CDM domain. However, determining reliable verifiers in the crowd as well as the development of trust management systems is still an open research question (Gao, et al., 2011). Advanced analysis of both anonymized user data and their provided content is a promising first step to establish automatic and highly scalable trust systems (Havlik, et al., 2013). Similar approaches have been applied to detect influential Social Media creators during disasters. (Mills, et al., 2009) analysed twitter feeds dedicated to the San Diego wildfires in 2007 and identified three feeds as most influential (one local broad casting news station and two local residents). One data point of influence, this case a tweet containing a link resource, induced 10.000 clicks conducted by followers. Therefore, identification of reliable and influential nodes in Social Media applications is also fundamental regarding public alerting during on-going crisis or disasters.

[4] http://www.ushahidi.com , last accessed on 2013/08/12
[5] http://sahanafoundation.org, last accessed on 2013/08/12

Basically there are two dimensions that have to be taken into account considering social media mining for crisis management (see Figure 2). The first dimension indicates whether the used material is structured or not. The second dimension describes whether individuals are identifiable or not. (Gundecha & Liu, 2012) argue that information provenance in social media is an important but yet unsolved research issue to differentiate rumors from truth.

So, in order to determine validity or reliability of certain users or data sources it is essential to be able to perform unique identification. We propose to use unique IDs or anonymized codes. Additionally, many social platforms (e.g. twitter) use nicknames for user identification which is absolutely sufficient for mining purposes. The important aspect is to be able to assign collected data to uniquely identifiable sources rather than knowing individual's personal information (i.e. name, date of birth, etc.). Implementing proper identification facilitates further assessment steps like verification conducted by either users or data mining algorithms. Depending on the use case, even more dimensions may play an important role in mining social media data for CDM (i.e. actuality, geo-spatial resolution, etc.).

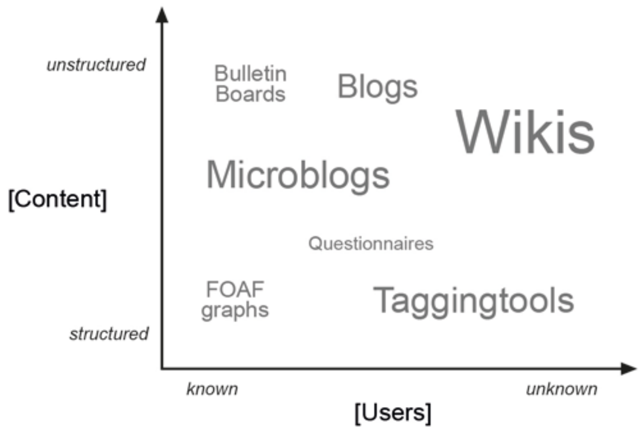


Figure 2: We determined two basic dimensions that one has to consider when planning a SMM application in the CDM domain. Certain applications may require identifying every participating user, others might not. Dealing with unstructured content may entail additional pre-processing steps.

The choice regarding the Social Media platform may not only depend on its provided content and the involved users it may also depend on the crisis situation. SMM may gain significant advantages in every single situation of a crisis:

Pre-crisis situations: SMM may be used to detect crisis in their early stages even before public mass media starts reporting about it. Adding a specific focus on dedicated events like earthquakes, flooding, etc., may additionally improve reliability and accuracy. *Example*: use twitters public API and process tweets by using the classification approach.

In-crisis situations: SMM may be used to analyze the current situation in terms of available resources or humanitarian needs. Social media content also comprises geo-related data which adds extra value for crisis coordination. *Example*: use Flickr API and search for photos, dedicated to a specific crisis event and apply clustering to their geo-position in order to reveal hot spots.

Post-crisis situations: Having collected a reasonable amount of data during on-going crisis, SMM may be used to analyze crisis activities in order to derive new models. These models help to better understand relations and can be applied in training scenarios or support pre-crisis analysis to provide more accurate predictions. *Example*: analyse data for specific vocabulary to create crisis related thesaurus.

Since different Social Media platforms provide different kinds of content we investigated several platforms and assessed their applicability regarding the CDM domain. On the basis of our findings, we propose a best practice data mining process for Social Media in crisis and disasters.

Application

A typical characteristic of Social Media is that users generate their own content. Some of the most used Social Media services, like Facebook, Twitter, Linkedin, Youtube and Flickr have between 0,5 and 1 Billion user. Even if only a very small proportion of the users are engaged in content creation, like on Youtube, they create a very large amount of content, which is a challenge for the data mining processes. From a data mining point of view, these services have two important basic data structures.

1. Almost all Social Media services are developed as network forming communities. Often the personal identity or the profile (not necessarily an approved identity) is treated as network node. Network links are formed by personal relationships, common interest or other network forming information. This implies that social network analysis (SNA) can be used as one of the main data mining methods to identify hubs and authorities. According to SNA hubs are specific nodes with a lot of outbound links. Authorities are nodes with a lot of inbound links. In this classification the directed links are understood as a vote for trust. In mass communication it is always an important task to identify communities of trust and communication multiplier in this community. This can be used in CDM to improve the efficiency and reliability of communication processes.
2. Almost all Social Media sites let users produce content, either as text, short text, picture, video, profile or full HTML content. This content can be used to understand the thematic structure of on-going communication processes and discover new trends and new discussion topics. In CDM this can be used for issue management, public communication and request for crowd support.

Social Media sites usually combine network data and content data, some of them with a stronger focus on network data (e.g. Facebook, Google +, LinkedIn) and other with a stronger focus on content data (e.g. Youtube, Twitter, Flickr and other). Table 1 shows the most popular Social Media sites, classified by network orientation or content orientation (bold is more important) and similarity of the data structure (separated by double lines).

Service	User[6]	Data aquisition	Network Data	Content Data
Facebook	1.5 bn	REST API, JSON and XML datasets about profiles, most of the data only visible in privat communities	**Network of friendship communities and relation ships**	Privat multimedia data, for friendship communities
Google+	343 m	REST API, JSON and XML datasets about profiles, most of the data only visible in privat communities	**Network of friendship communities and relation ships**	Privat multimedia data, for friendship communities
Youtube	1 bn	HTTP, only meta data, and video streaming supported	Similarity of videos	**Videos and comments, only streaming**
Twitter	500 m	REST API for full communication as twitter client and streaming API for high volume requests.	Network of followers	**Microblog text, links, hash tags**
Tumblr	216 m	HTTP requests and JSONP response	Network of followers	**Microblog text, links, hash tags**
Linkedin	238 m	REST API with html response to all data available to authenticated user	**Network of professionals**	Information about interests, expertise, and reputation
Flickr	87 m	REST API with JSON or XML response	Network data about contacts and groups	**Photos and comments**
Pinterest	70 m	No API, SDK only for upload of content	Sharing community	**Virtual pinboard to share content from web**
Wordpress	66 m	PHP Software for blogging, only partly centralised	Network by links to other homepages (not centralised)	**Webcontent, organised by posts**

Table 1: Amount of available data sets, data acquisition, data structure and service specific content. Double lines group platforms with a similar data structure. (Source: AIT)

[6] http://expandedramblings.com/index.php/resource-how-many-people-use-the-top-social-media/, last accessed 2013/08/12

The data analysis for CDM should use both types of data structure and all sorts of different services to generate a more complete situational awareness. However most of the data in social networks are restricted to limit access to friends. Therefore a Social Media strategy in CDM should include a strategy to build up a supporter community.

As depicted in the table above, some Social Media services offer similar information. In Facebook and Google+, e.g. private multimedia data and links to multimedia data are available. However, access is usually restricted to friendship communities. For all members, only attributes like ID, Name, User Name, Country, Gender are available. With authentication additional technical data for developers, like browser type, unique IDs and session information can be retrieved. For more useful data, user permission is necessary. With this, a complete access to profile is possible. In case of a crisis, in principle, pictures, videos, links, wall posts can support situational awareness. However, to get these data sets, it is necessary to convince the user to give their permission. Search is only allowed in public objects, which makes it very difficult to get discussions, with relevant information for a crisis.

LinkedIn is similar to Facebook and Google+, but put more emphasis on professional information about competences, interests and experience in a business context. The search function is more useful for CDM than at Facebook and Google+. It is possible to search for people, groups, companies and access company profiles. LinkedIn enables members to establish and grow their networks with invitations and messages by using the API. This makes it possible to develop a CDM application for a supporter community.

Microblogging services like Twitter and Tumblr, blogging services like Wordpress, dicussion forums and multimedia repositories provide valuable datasets for content analysis. Due to the limited length of the microblogging messages the main information of the post is often behind encrypted links to multimedia content stored in picture or video repositories, as well as content from other homepages. However, the full analytical power is reached with a combination from SNA results, results from content analysis, applied in a network-centric issue management for mass communication in CDM.

Social Media Monitoring Services usually concentrate on frequency statistics or semantic analysis. This has been proven by an internal AIT market research in 2011 which investigated about 90 monitoring services. There was no service or software available for what we call it, network- centric issue management (with issue identification, issue prioritisation and stakeholder identification according to a specific issue.) The AIT market research, as well as the presentation of top tools in traditional social media monitoring makes clear, that these tools are based on business needs.[7] Specifically in CDM the whole word vector for topic classification is superior to classical frequency analysis. However, this is only one aspect in appropriate text mining for CDM.

The left column of Table 2 lists features, which are state of art in topic mining within existing social media monitoring solutions.The right column lists features which do not exists yet but would be beneficial in crisis communication.

State of the Art feature in Social Media	Beyond state of the Art features for CDM
Search for keywords	Search concepts
Filter (time, relevance, geography, resources)	Emerging topic detection with expectation
Relevance assessment	Emerging topic detection without expectation
Frequency analysis at item-level	Weak signal detection
Frequency analysis on subject level	Exception reporting using word vectors
Frequency analysis of trends	Issue management process control
Author profiles	Buzz tracking (tracking of discussion)
Author interactions in social media	Influencer networks
Tag cloud	Automated cluster detection

Table 2: Beyond state of the art features for topic mining in CDM (Source: AIT)

[7] "50 Top Tools for Social Media Monitoring, Analytics, and Management", http://socialmediatoday.com/pamdyer/1458746/50-top-tools-social-media-monitoring-analytics-and-management-2013, last accessed 2013/08/12

With these new features, Social Media monitoring can address in particular the information overflow of crisis managers. For authorities it is very important to be informed about all the topics, discussed in social networks and to have a prioritization of these topics, so that they do not miss important information and that they are aware of escalation triggers. Therefore, the main focus on development is to concentrate on the most important information for the crisis manager and to provide this information in a clear and reliable manner.

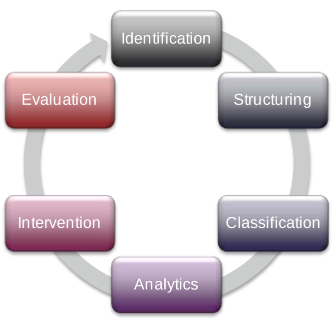
Emotion mining is a quite new topic and somehow it is a kind of buzz word. It is very difficult to identify real emotions in written text. It is possible to identify emotions in sound files and from body signals. However, even for humans it is not easy to identify emotions from email and SMS. Thus, emoticons and abbreviations are used in social media to express emotions. This was one of the first methods for emotion mining in messages. Later, specific words were used in text mining, to identify emotion in messages. In the AIT market research, we found that existing Social Media monitoring companies often use dual sentiment classification. This means, that classifiers only differentiate between right or wrong, positive emotion or negative emotion.

Table 3 gives an overview, of what emotion mining is about in actual media monitoring software solutions and what emotion mining can be in CDM specific monitoring software.

State of the Art features	Beyond state of the Art features for CDM
Tonality with key words, manual tagging	multi emotion classification
Association graph	identify emotions useful for CDM communication
Sentiment with SVM	emotional clusters
Search with Search Clouds	sentiment in each emotional cluster
Tonality with SVM	georeferencing of emotional cluster
Sentiment with key words	tracking of sentiments over time

Table 3: Beyond state of the art features of Social Media monitoring solutions facilitating emotion mining for CDM (Source: AIT)

A best practice data mining process for CDM will start with a process to identify relevant channels of communication, for a specific use case. It can be expected, that there are sources for network data and content data in the results. Second, a social network analysis will produce a structure of the relevant community. It would be very helpful, to have the network geo-referenced. However, most of the social networks have unreliable information about the location in the entity profile. We expect, that this will improve in the future. Third, based on the network of relevant users, content data acquisition can start and should continue over the period of crisis. Topic mining and emotion mining is used to extract the relevant content. Finally, a network-centric issue management can be set up for effective mass communication in CDM. The network-centric issue management uses the same processes, than classical issue management, as presented in the graphic, but rely on results from SNA to make communication more effective.



- **Identification** of relevant communication channels
- **Structuring** of communication content
- Automatic **classification** of communication content
- **Analysis** of content to identify
  - new and emerging topics
  - influencer groups
  - and opinion leader
- **Intervention** e.g.. with information campaign, moderation or participation in active social media discussions
- **Evaluation** in social media is different (two way communication)

Figure 3: Network- centric issue management in CDM (Source: AIT)

The approach to combine both data mining and Social Media content generates a lot of very beneficial use cases for the CDM domain. Anyway, processing Social Media content requires a high amount of social responsibility as well as legal and respectful administration of user's privacy. SMM applications may exclusively utilize data that is explicitly available for the general public. This excludes data being available by accident (e.g. misconfiguration, opt-out trap, etc.). However, the CDM domain may gain a lot of interesting insights in various scenarios in order to understand people's needs and their behaviour and foster proper communication between all stakeholders.

Use Case Scenario: Flooding

Flooding scenarios can usually be anticipated for a certain time period ahead by weather forecasts and repeated monitoring of river gauges, so that the forces (such as disaster relief units, fire fighters, or emergency response organisations) can be put on alert on time. Unfortunately, disaster management is still always a race against time, including a set of developments, which – due to the nature of crisis, cannot be foreseen for all possible cases. Residents affected by disastrous events are likely to share their experiences and exchange their thoughts via Social Media platforms. Gaining data from this exchange of information certainly contributes to receive a better picture about the most up to date developments from the various, existing acute flooding scenario. On the other hand channels are established, enabling direct communication between the residents of the affected regions and the forces. This helps to better predict and react to the developments of flooding scenarios in place. Not only the typical attributes and characteristics of the relevant flooding scenarios can be explored, also the evolving patterns of the crisis and the disaster development can be recognized by the forces in an enhanced manner. We developed a model that distinguishes between three phases in flooding scenarios, each resulting in different tasks and challenges for the integrated crisis and disaster management and describes possible reactions of the operating emergency response organisations.

During the **pre-phase**, weak signals and first impacts of possible flooding events can be detected. Flood monitoring measures and local explorations are the basis for decisions of the operation staff which triggers warnings and flood alerts. Besides that, first instructions are suggested to coordinate the local residents. Social Media information helps to detect and analyse additional signals and developments earlier, and can be used to predict and identify possible flood incidents ahead. The **integrated operation control phase** starts with flood intervention by different emergency response and military organisations such as the Austrian Red Cross, fire fighters, or the Austrian Army and other emergency response or non-governmental organisations, being operative in different fields of intervention (i.e. Medical Care, Psychological Care, Supply & Logistics or Technical Infrastructure). A high level of coordination and therefore a sophisticated information management is necessary to establish a „shared awareness" to be able to optimize and synchronize every action over all fields of intervention. After the flooding, after-action– and incident-reviews are held in a **post-phase**, while recovery measures are triggered by the results and gained experience out of the operation. SMM applications may be used to assist the Lessons Learned process which helps to optimize the detection and handling of flooding events in the future and closes the feed-forward-loop to the next pre-phase.

Discussion

The use of data from social media in the domain of crisis management offers on one hand multiple chances, but is also associated with many challenges on the other. Opportunities encompass the availability of real time data on crisis evolution to help crisis managers to sharpen their decisions, new or additional approaches to co-ordinate emergency forces in the field or psychological assistance of affected people (Rainer, et al., 2013). Challenges encompass legal aspects such as data protection, liability and privacy issues and in some cases copy right aspects. Moreover, ethical aspects like misuse of data by a despotic regime have to be taken into account, too. Finally, cultural aspects caused by the degree of availability of and the willingness to use the internet by specific populations need to be considered (Rainer, et al., 2013). Data provided from social media can be used in different ways by crisis managers. One approach is to use them in order to obtain information from social media in order to enrich the picture on the development of a crisis obtained from traditional sources of information. The other approach is to use social media to establish communication paths between the multiple actors in crisis, e.g. crisis managers, emergency forces, affected population, volunteers providing help or representatives of mass media. The communication processes can be uni – or bidirectional, they can be open or can be controlled by the crisis manager. In any case there is need to extract relevant information out from the huge amount of data by applying methodologies such as those described in chapter Data Mining. The process required to generate relevant knowledge from data obtained by social media is described by (Neubauer, et al., 2013) in more detail. A common requirement on all methodologies such as association rule mining is to identify relevant data in order to extract crisis – relevant information, because the unstructured, not corrected data do not contain all demanded and correct data. Another relevant aspect is the credibility of the obtained data. Before to generate knowledge supporting crisis managers, the reliability of the extracted information has to be validated by approaches such as peer reviews of information or credibility checks. As already mentioned above major questions are liability and privacy aspects, that may vary from state to state being involved in on specific crisis. In case of wrong decision causing damages or even human losses, the question arises who is liable. Nevertheless, if all those aspects are considered, social media offer brilliant perspectives for crisis management.

References

Ackoff, R. L., 1989. From data to wisdom. *Journal Of Applied Systems Analysis*, Volume 16, pp. 3-9.

Borne, K. D., 2009. Scientific Data Mining in Astronomy. *Next Generation of Data Mining*, pp. 91-114.

Fraustino, J. D., Liu, B. & Jin, Y., 2012. Social Media Use during Disasters: A Review of the Knowledge Base and Gaps. *Final Report to Human Factors/Behavioral Sciences Division*, U.S. Department of Homeland Security

G.K.Gupta, 2009. Data Mining with Case Studies. New Delhi: PHI Learning Private Limited

Gao, H., Barbier, G. & Goolsby, R., 2011. Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. *Intelligent Systems*, 26(3), pp. 10-14.

Gundecha, P. & Liu, H., 2012. Mining Social Media: A Brief Introduction. *TutORials in Operations Research, INFORMS*, Volume 9, pp. 1-17.

Havlik, D., Egly M., Huber H., Kutschera P., Falgenhauer M., Cizek M., 2013. Robust and trusted crowd-sourcing and crowd-tasking in the Future Internet, in: Hřebíček, J., Schimak, G., M. Kubasek M., Rizzoli A. (eds.) *Environmental Software Systems. Fostering Sharing Information. IFIP Advances in Information and Communication Technology*, Springer, Heidelberg (2013)

Koohang, A., Harman, K. & Britz, J., 2008. Knowledge Management: Theoretical Foundations. Santa Rosa, California: Informing Science Press

Liu, B., 2007. Web Data Mining. Berlin Heidelberg New York: Springer

Mills, A., Rui, C., JunKyu, L. & H., R., 2009. Web 2.0 Emergency Applications: How useful can Twitter be for Emergency Response. *Journal of Information Privacy & Security*, 5(3)

Neubauer G., Nowak A., Jager B., Havlik D., Foitik G., Kloyber C., Flachberger C., 2013. Crowdtasking for crisis and disaster management - opportunities and challanges, *Interdisciplinary Information and Management Talks*

Rainer, K. Grubmüller V., Pejic I., Götsch K., Leitner P., 2013. Social Media Applications in Crisis Interaction. *systems. connecting matter, life, culture and technology* 1(1), pp. 110-127.

Shirky, C., 2009. Here Comes Everybody: The Power of Organizing Without Organizations. s.l.:Penguin Books.

Author Biography

Hermann Huber studied Information & Knowledge Management at the University of Technology in Vienna. He was employed as assistant for business intelligence and did research on data mining in the cultural heritage domain. Today, he is employee of AIT and is involved in several research projects related to CDM.

Georg Neubauer is thematic coordinator for “ICT solutions for crisis and disaster management” at AIT. He was lecturer at the University of Technology in Graz from 2002-2010. Since 2004 he gives lectures at the University of Technology in Vienna. Mr. Neubauer worked as a consultant for several national and international organisations like the EU, WTO or BMVIT.

Andrea Nowak joined AIT in 1992 and is Head of Business Unit “Informationmanagement & eHealth”. She studied Mathematics at the University of Vienna and was employee of Siemens for several years. Mrs. Nowak successfully managed R&D projects in cooperation with the Austrian Army and is member of diverse international research boards.

Joachim Klerx is predominantly active in the research field of innovation systems with a strong focus on innovative groups in society since 14 years. He is an expert in the research fields of data mining, artificial intelligence and multi-agent systems and was part of numerous research projects in the European Union.

Bettina Jager studies political science at the University of Vienna and is part of the Crisis and Disaster Management Team at AIT since 2010. Mrs. Jager focuses on politics and administration, good governance, technological impact assessment related to CDM.

Klaus Mak is Head of Central Documentation at the National Defence Academy of the Austrian Armed Forces. He is lector on several educational institutes for information and knowledge management. Mr. Mak is responsible for conceptual planning and R&D projects with strong focus on knowledge management at the Austrian Armed Forces.

Christian Meurers studied Informatics at the University of Technology in Vienna and is responsible for the Situation Awareness Center at the National Defence Academy. His main research areas are cyber & information war and technology & society.