Data mining and machine learning in the context of disaster and crisis management

Adam T. Zagorecki

Centre for Simulation and Analytics,
Cranfield University,
Defence Academy of the UK,
Shrivenham SN6, 8LA, UK
E-mail: zagorecki@gmail.com

David E.A. Johnson*

Department of Political Science,
Missouri State University,
901 S National Ave,
Springfield, MO 65804, USA
E-mail: davideajohnson@missouristate.edu
*Corresponding author

Jozef Ristvej

Department of Crisis Management,
University of Zilina,
Univerzitna 1, 010 26 Zilina, Slovakia
E-mail: Jozef.Ristvej@fsi.uniza.sk

Abstract: Disaster and crisis situations are characterised by high dynamics and complexity with human lives and substantial environmental and economic consequences at stake. The advances in information technology have had a profound impact on disaster management by making unprecedented volumes of data available to the decision makers. This has resulted in new challenges related to the effective management of large volumes of data. In this paper, we discuss the application of data mining and machine learning techniques to support the decision-making processes for the disaster and crisis management. We discuss the challenges and benefits of the automated data analysis to different phases of crisis management. Based on the literature review, we observe a trend to move from narrow in scope, problem-specific applications of data mining and machine learning to solutions that address a wider spectrum of problems, such as situational awareness and real-time threat assessment using diverse streams of data.

Keywords: disaster and crisis management; data analysis; data mining; machine learning.

Reference to this paper should be made as follows: Zagorecki, A.T., Johnson, D.E.A. and Ristvej, J. (2013) 'Data mining and machine learning in the context of disaster and crisis management', *Int. J. Emergency Management*, Vol. 9, No. 4, pp.351–365.

Biographical notes: Adam T. Zagorecki received his PhD in Information Science from University of Pittsburgh. He is currently a Senior Research Fellow at the Operational and Decision Analysis Group, Centre for Simulation and Analytics, Cranfield University, Defence Academy of the UK. He held Post-doctoral Fellow position at Graduate School of Public and International Affairs, University of Pittsburgh and Senior Research Programmer at Robotics Institute, Carnegie Mellon University. His research interests include probabilistic methods, decision support systems, operational analysis, and machine learning. His expertise lies in Bayesian networks: knowledge elicitation and learning from data.

David E.A. Johnson received his PhD from the University of Pittsburgh and received the David Oeler award for use of information technology. He is an IAEM Certified Emergency Manager and has over 35 years of emergency management experience at the national, state and local levels. He is now an Assistant Professor at Missouri State University focusing on emergency management and public policy. His research interests include vulnerability analysis, complex systems and the use of information technology in disaster response.

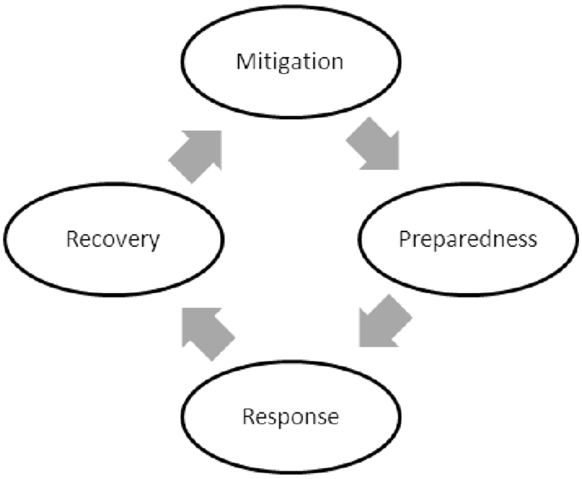
Jozef Ristvej, PhD, is an Assistant Professor at the Department of Crisis Management, University of Zilina, Slovakia. Since March 2011 is at the position of Vice-dean for Development and International Relations. For 2010 received the visiting scholar grant of Ruth Crawford Mitchell to the Center for Disaster Management at the Graduate School of Public and International Affairs, University of Pittsburgh. In his work he is focusing on supporting the information systems in the area of the decision making process in crisis management. He is a member of several scientific project teams and as well as of the ISCRAM Association in Brussels.

1 Introduction

Disaster and crisis situations are dynamic and complex events with human lives, environmental and economic consequences at stake. The combination of complexity and dynamics makes the understanding and predictions of the development of disaster and crisis situations extremely difficult. Additional difficulty relates to prediction of consequences of these events, as they are in the context of complex and interdependent social, infrastructure and natural environments. Disaster management is concerned not only with predicting the course and consequences of disasters, but mitigating those undesired consequences – undoubtedly a very challenging task made even more so under time pressure. Information technology (IT) has been proven to successfully support the decision making processes in managing many complex problem domains: medicine, logistics, air traffic control, military planning and operations, etc. In that sense, disaster management is no exception; however, we believe that it does present a particular challenge. One aspect in which the advances in IT have had a profound impact on the practise of all phases of the disaster management is making unprecedented volumes of data available to the decision makers. These large volumes of data have introduced new challenges related to the effective data management. In this paper, we discuss the roles and challenges of application of data mining (DM) and machine learning (ML) techniques to support the decision-making processes in the context of the crisis and disaster management. In this paper, we attempt to review the challenges and benefits of the automated data analysis to different phases of the crisis management.

Disasters often undergo rapid substantial evolution; therefore, the disaster management is a non-uniform process characterised by phases, although these phases are not distinct in nature. Their delineation has evolved since first introduced by Carr (1932), with the transitions to other models as described by Neal (1997). Currently, the widely accepted concept of the disaster management cycle (Waugh and Tierney, 2007) divides the disaster management process into several distinct phases (Figure 1), although there is no clear consensus on the division into individual phases or even the number of phases.

Figure 1 Emergency management cycle



The advances in IT have had a profound impact on all phases of disaster and crisis management, though the degree of this impact varies between phases. These differences are mainly due to the ease of applying IT to particular problems. Historically, the first applications involved document sharing and local databases while with development of the IT infrastructure, more sophisticated applications became available. Examples of these applications include the use of the internet for agency coordination, computer modelling for complex tasks such as hazard analysis. Some specialised tools in the form of desktop software for addressing specific problems, such as chemical plume dispersion modelling (NOAA, 2004) or the earthquake, hurricane and flooding models used in HAZUS (FEMA, 2013) have become commonplace. The latest addition to the arsenal of tools is social media that further enhances communication between the agencies and the public sector and increases situational awareness (Zielinski and Bügel, 2012).

The advances in IT resulted in the ability to create new data and process this data in a manner that was unimaginable in the past (Pine, 2007). These technological advances led to the development of sophisticated emergency management information systems (EMIS) based on IT. These systems can not only provide the emergency manager with additional data, but also provide analysis in a timely manner. The collected data can also be used to simulate potential disasters exploring possible outcomes given various conditions to support mitigation and planning decisions. Ristvej and Zagorecki (2011) identified four key functions that a typical EMIS should provide. They are:

- Data collection – traditionally, data collection was done in the form of paper reports or questionnaires. The development of IT allowed for use of word processors, spreadsheets, and forms to enter data directly into the databases: initially at desktop computers and later online forms for which the data was stored on dedicated servers connected to network. Another increasingly important source of data is remote sensing technologies. In the past, remote sensing data was mostly related to expensive satellite systems and sophisticated geospatial tools (Nayak and Zlatanova, 2008). Sensing technologies are currently undergoing rapid advances, leading to affordable sensors that can communicate wirelessly, often without relying on external power sources, to an emergency Operations Center. Use of them significantly increases situational awareness, often in real time. In particular, use of widely available smart-phones, which typically have a rich suite of various sensors embodied (camera, global positioning system, accelerometers, magnetic field sensors, etc.), are currently being actively explored (Ehlert, 2012).
- Data transfer and distribution − developments in telecommunications, in particular computer networks, provide fast and reliable means of transferring data over large distances and large user groups. In the context of disaster management, the technological ability to share information between multiple organisations is of particular significance. Sharing important data is crucial not only to the situational awareness mentioned above but, in the development and distribution of incident action plans (IAPs) that coordinate response and recovery actions. In the last decade, wireless communication has become affordable, allowing new advances in disaster management. This has been achieved through the use of automated sharing of information with distributed users, such as remote sensing data sharing between the sensors and servers hosting software that provides advanced analyses which improve dynamic situational awareness. It also facilitates applications such as wireless cameras for real-time sharing of images for both response and damage assessment. The same improved communications also enables two-way instant communication and data sharing not only between the response organisations, but also with the general public equipped through items ranging from smart-phones to road-side signage directing the public, thereby facilitating the implementation of smart traffic management systems for challenges such as urban evacuations (Aved et al., 2006).
- Data storage – IT offers technology to efficiently store large volumes of data, and provide virtually instantaneous access to it. In the early years, hundreds of documents could be stored on computer hard drives; later relational databases were able to store large volumes of structured, interrelated data items. Although this technology is well established, it is continuously advancing to store new and more demanding data, such as video streams. Finally, in recent years, the *big data* term represents data storing techniques that allow efficiently accessing data items distributed in multiple locations exploiting the advantages of parallel data processing. This is critical in the many areas where regulation and consequence management are managed by different agencies.
- Data processing and analysis − the goal of an EMIS is to provide relevant information to the decision makers. The amount of data within an EMIS far exceeds human abilities to analyse it. Therefore, the EMIS should provide tools for manipulation of collected digital data. Although there are tools that allow for

> processing and analysis of large volumes of data efficiently, they have serious limitations, and considerable effort by human analysts and decisions makers is still required to produce sophisticated analysis of the available data including incident trends and interdependent relationships. Using the example of archived video data mentioned previously, when accompanied by change detecting analysis software, could facilitate the detection of environmental changes. These could indicate the development of potential event triggers such as buildings leaning or changes in high-hazard dams which could foretell of a catastrophic failure. These new capabilities not only allow detection for which analysts are difficult to process, but in some cases might be too small to be detected by human senses.

In addition, Dorner and Pfeifer (1993) found that stressed subjects, like emergency managers, focused on the general outline of the problem rather than in-depth analyses. EMIS can enhance decision making by providing this analysis when time-constrained situational conditions would not permit it. Of all four functions, this function proves to be most challenging and available tools seem to not meet the practitioners' expectations.

In this paper, we focus on the discussion of the role of IT in the data processing and analysis function and its utility to the emergency management community. This partnership can provide many useful tools to emergency managers already faced with daunting amounts of information and from an increasing number of sources. One of our particular interests is the advances in the automated analysis of the collected data. The focus of this paper is application of advanced techniques for data analysis. What we mean by the advanced analysis is creating new knowledge from available data, rather than processing the data in a prescribed manner, regardless of how complex these manipulations are. The two fields that offer techniques for creating new knowledge from data are data mining (Han and Kamber, 2006) and machine learning (Bishop, 2006). The techniques offered by these fields rely exclusively on data to discover new patterns, trends, or deliver predictions. They stand in contrast to systems that rely on subject matter input in order to define the knowledge base for reasoning about the domain.

Disasters are events that often result from unexpected and dramatic changes in a stable, but often precarious, system or interdependent systems that we know as our world. One of the approaches to the problem is proposed by Mileti (1999). He differentiated these into natural, human and constructed systems, but we acknowledge that there are many levels of these systems and subsystems. Disasters very rarely start with a massive disruption of the system (with a few exceptions such as earthquakes) – rather they are rapid processes that propagate the initial effects (limited) through a complex system leading to major disruption of this system (examples here can be Chernobyl or Hurricane Katrina). As Perrow (1999) points out, we often face complex systems in which a series of failures come together in a way that no one could anticipate. In the case of Hurricane Katrina, they included the failure of levees, failure of evacuation, shortcomings of the response, and finally, social and economic factors. These properties imply that disasters involve large scale effects, which are complex and dynamic. Each of these characteristics poses significant challenges to modelling and analysis to support the needs of emergency management decision makers. These challenges are part of the impetus for the call by the National Research Council for automated information fusion from diverse sources (National Research Council, 2007). To make the situation even worse, disasters are extremely rare events, each of them with some unique characteristics. Our knowledge of catastrophic system failures is typically much less than our understanding of the 'normal' states of the systems. If one considers that in the context of the automated systems that rely exclusively on data analysis, it leads to the conclusion that one can assume that it is possible to collect the data on normal operations of a system, but the data on catastrophic events will be always scarce, by the nature of these events. This was identified as one of basic challenges in disaster management, for example Cutter (2006) notes that basic data on the range and extent of the hazards has not kept pace with the needs. Therefore, the methods relying exclusively on the data, seemingly offer limited applicability in this context. We will discuss this problem, and show that it may not necessarily be the case.

Disaster management is a multifaceted process to avoid, reduce, respond to, and recover from the impact of the disaster on the system. Because of the scale of events, disaster response involves multiple organisations – governmental, public and private, often crossing multiple layers of authorities (Comfort et al., 2004; Waugh and Tierney, 2007). This emphasises the need for decision support systems, as it is virtually impossible for a human decision maker to comprehend the complexity of the situation. Instead, problems such as situational awareness (Zheng et al., 2011) and building a common operating picture, shared among multiple actors who often have only partial view of the situation, are becoming some of the most urgent needs of disaster management.

There are numerous ways that IT can enhance the practise of disaster management. The overview of technologies in the context of disaster management was discussed in Ristvej and Zagorecki (2011). In this paper, we focus on the data-driven methodologies within the frameworks of DM and ML and their potential and proven roles in supporting different phases of disaster management. We begin with what we believe is a necessary clarification, of the definition for the term *data*. Then, we introduce the data mining and machine learning. Consequently, we review the applications of these techniques to disaster management, focusing on their role in different phases. We conclude the paper with the discussion of future directions and practical challenges.

2 Interpretation of data

Data is a term that has two meanings in English, which is sometimes a source of confusion. The first meaning of data relates to a collection of information and to some degree its' encoding: it can be in form of facts, measurements, statistics, variables, etc, either quantitative or qualitative. It is the lowest level of abstraction, from which one can produce information and knowledge. The second meaning is related to processing of factual information: it is a body of facts that can be both assumed and known that form basis for reasoning. We can say that the former meaning is more scientific and technical, while the latter is more philosophical. The problems may arise in practise when in a conversation one party has in mind the first meaning and the other party the second meaning. For example: when discussing the availability of video data of disaster area, for the emergency personnel the data is the part of the content of the video such as extent of damage, description of events, or message recorded. For the data scientists, the data is a quantitative digital stream where the aspects such as data transfer, digital format and storage are of concern and interest. From the authors' experience, the understanding of the term data varies extensively, especially between the technical fields (such as IT, DM and ML communities) and the social sciences and practitioners in the field of disaster management. These different interpretations and the underlying assumption that the same term *data* carries the same meaning leads to misunderstandings or even conflicts, often affecting the performance of multidisciplinary projects. Discussion of the application of DM or ML to the field of disaster management should start with agreeing on understanding of the meaning of the very basic term – the *data*.

IT is a field that was mostly driven by mathematics and engineering. It is a similar case for the data mining community which shares links to mathematics and engineering with additional strong statistics influence that clearly shape the interpretation and requirement for the data. The requirements of the data from the data mining community can be summarised as: the data should be well structured (tables, relational databases, etc.), translatable into variables (numeric or textual), and preferably large amounts (in thousands or more instances) that enable statistical methods to be justified. These requirements reflect the limitations of the methods used. Typically, it is desirable that the data is collected under the same conditions and samples are unrelated to each other (instances are independent and identically distributed). This requirement originates from assumptions required by the statistical methods that provide underlying mechanisms for many of the more advanced methods.

On the other hand, the expectations of the data in the other fields, especially from the practitioners' community, are much broader. The data can range from the well-structured form as desired by the data scientists, to concepts such as rosters, written reports, maps, newspaper articles, video clips, etc. Some of these forms can carry very rich and useful information to domain experts, but are difficult to be exploited by the DM and ML techniques. For example, a photograph of an affected area can deliver many clues for a human expert; however, there are yet to be any automated tools for interpreting arbitrary images.

As a part of our work, we surveyed 30 senior practitioners in the disaster management community and asked them how often they use different types of data in the process of decision making. They had an option to choose from the following general categories:

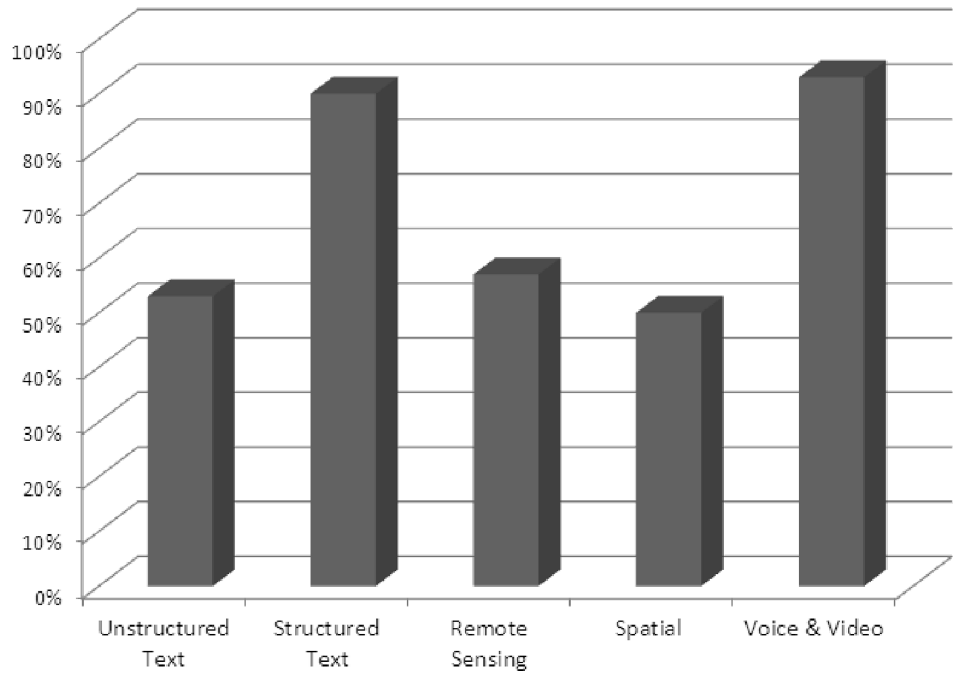
- Unstructured textual data: news articles, incident activity reports, and announcements, etc.
- Structured textual data: 9-1-1 CAD data, situational reports, damage assessment forms, etc.
- Remote sensing data: numeric measurements, but this category increasingly includes use of mobile image and video devices
- Spatial data: data within graphical information systems (GIS), satellite and aerial imagery, etc.
- Voice and video: radio communication, news broadcasts, etc.

The results of the survey are shown in Figure 2.

They were also asked which type of data they most commonly used. Structured text was chosen by 47% of the respondents followed by voice/video at 40% then 7% each for spatial and remote sensing.

As should be expected, this list follows the second meaning of the term data discussed earlier. In order to apply DM and ML methods, we should translate it into more 'technical' view. The challenge of applying DM and ML methods to disaster management starts at this point – the identification of useful data that the methods can exploit. Diversity of the data used by the disaster managers poses a considerable challenge in applying analytical methods because most of these methods were originally developed to handle numerical and well structured data. Although developments in information retrieval for the textual information (Manning et al., 2008) and methods for image and video data (Gonzales and Woods, 2007) are considerable, in reality, practical, automated and reliable text and image interpretation has not yet been achieved.

Figure 2 Most commonly used data types



Disaster management is a field where there is a clear distinction between static and dynamic data. The static data is a type of data that is collected prior to the disaster event. Examples of static data include: spatial data capturing population characteristics, location of resources, emergency response plans, etc. The dynamic data is the real-time event data that is produced during the disaster and may include weather data, river states, response unit locations, resource distribution, etc. Managing these two types require different IT infrastructure, and from the perspective of DM they may be exploited differently: the static data offers more time for in-depth analysis, while managing dynamic data requires ML approaches that are suited for processing in real time (or near-real time) data that is highly unpredictable and uncertain.

Both the diversity of the data and its combination of static and dynamic data call for methods that allow combining heterogeneous data in order to produce useful information out of it. Data fusion is the process of combining multiple and often distinct data sources in order to achieve a more complete representation of the phenomena of interest. Data fusion is particularly useful in the case of incomplete and/or uncertain data as it allows using multiple data sources to compensate for the incompleteness and uncertainties (Michalowski et al., 2004). However, the concept of data fusion is difficult to implement in practise where the data sources are diverse and come from different sources that are beyond control of data scientists and developers. This is the case in disaster management – often the data is provided by different organisations, often with limited IT capabilities.

3 Data mining and machine learning

Machine learning is a branch of computer science (more precisely artificial intelligence) that is concerned with developing methods and algorithms that learn characteristics and patterns from available data in order to make predictions. The main focus of ML is to develop methods that can build models that describe data (and preferably underlying mechanisms) faithfully. Practical examples of ML application are e-mail spam filters: these software tools embodied into e-mail servers are able to automatically identify spam e-mails with great accuracy. The 'logic' behind these filters is learned automatically by analysing content of e-mails and the users' behaviours.

Frawley et al. (1992) define data mining (DM) as the nontrivial extraction of implicit, previously unknown, and potentially useful information from data (Frawley et al., 1992). In fact, DM is a wider concept than ML – it is concerned more about the analytical processes that lead to new knowledge discovery from large existing datasets. It is strongly related to large scale data management and processing (relational databases, data warehousing, etc.). Unlike the ML, the DM emphasises the practical aspects of data management and delivered results. DM is expected to feed business intelligence and computer decision support systems with information that should be useful for the decision makers.

There are obvious overlaps between DM and ML, but there are the key differences as well. The overlaps include: shared algorithms and methods (mostly statistical), both rely on data and try to draw conclusions from it. The differences include:

- ML is more concerned about the process of knowledge discovery, (algorithms that analyse data) preferably detached from given dataset characteristics, while DM focuses on extracting and exploiting useful information from available data (with focus on particular instance of data).
- ML goal is focused more on reproducing existing knowledge (in order to validate usefulness and correctness of the proposed algorithms), while DM is to discover and exploit new knowledge from the provided dataset.
- Data sources are of concern of DM: providing efficient data infrastructure in the form of databases and efficient retrieval algorithms is part of DM, while ML typically assumes datasets as input of the algorithms (though some algorithm specific data structures for efficient data access may be a part of ML).
- DM requires large amounts of data: large databases' infrastructure is the concern of DM, while ML may focus on methods and algorithms that are explicitly designed for the limited available data.

Both, DM and ML are widely recognised tools to support decision-making in many areas, including banking, insurance, aerospace and defence industries, etc. Some examples of applications include: search engines, bioinformatics and DNA sequencing, e-mail spam filters, online recommendation systems, and facial recognition for security applications, fraud detection, and many more.

4 Data analysis in disaster management

As discussed earlier, disaster and crisis management poses a number of challenges. The availability of useful and comprehensive data is one of these challenges. But even if the large volumes of data would be available, there would be other problems. If one considers a jurisdiction (which would be subject to a disaster) as a system in a 'normal' state, there are so many different ways this system can be subjected to a disaster, and it is impossible to have reasonably complete data on all of them (it may not be even possible to express all possible scenarios for disasters). However, it is theoretically possible to have rich data on the 'normal' state of the system. Part of the 'normal state' is the potential hazards and their repercussions. This is where the knowledge base that Cutter (1996) refers to would be useful. The history of incidents certainly is not comprehensive, but building that experience helps to turn possible events into probable and develop both normal system parameters and boundaries. Such process would help to inform the data-driven methods, which can be useful to detect anomalies (out of normal states). And indeed, there are already developed data-driven early warning systems that rely on the anomaly detection: examples include disease outbreak detection (Cooper et al., 2004), or fire detection (Bahrepour et al., 2010). The reason for the need for sophisticated methods to identify anomalies is because there are inherent variations in the data values and the determination of what level of value constitutes an anomaly requires taking into account multiple often correlated factors.

4.1 Mitigation

The mitigation phase of the disaster management cycle is concerned with reducing the risk of occurrence of the disaster and its possible consequences. Probably the most recognised examples of DM and ML for disaster mitigation relate to the prevention of different threats posed by man-made disasters. Examples include: detecting terrorist threats through analysis of computer networks (Last et al., 2006), social networks (Weinstein et al., 2009), fusion of sensor data for nuclear threat detection, face recognition at crowded settings, etc. DM and ML could be combined with static data to monitor changing conditions and their impact on static features of the community. This would enhance the ability to prioritise preventive actions in order to avert an incident.

4.2 Preparedness

One of the most challenging problems in the preparedness phase is evacuation planning (Crooks et al., 2008). Evacuation planning involves the need for combining spatial data and capturing evacuee behaviours (Miah, 2011). The main research focuses on using DM methods to identify potential threat and safe areas. DM could also be useful in monitoring information sources, such as public safety websites; to determine what information is being sought as both an indicator of public awareness and what resources may be required should an event occur (White et al., 2009). DM can also be used to estimate changing conditions. Yang et al. (2011) showed how DM applications to estimate tropical cyclone intensity changes. This could be valuable is the development of preparedness plans.

ML can also find application in early warning systems (Grasso, 2012), both for natural and man-made disasters. Examples include early warning systems for chemical and nuclear threats, floods, tsunami, and more. Conceptually, the early warning systems rely on anomaly detection – the lack of data on disasters is alleviated by the fact that anything out of normal can be labelled as potential threat, and be possibly subjected to human review. Another method, recently gaining a considerable of interest, is human review or potentially ground proofing the impact of early warning is the harvesting of social network data sources such as Facebook and Twitter.

One aspect related to early warning systems that is worth discussing is the role of simulated data in the process of their development and validation. Because of the limited actual event data (or sometimes lack of thereof) the researchers typically are forced to use simulated data to validate their work. This practise can suffer from several problems:

- Simulated data can be simplified – lacking the certain factors that exist in real data and were not included in the models that were used to generate simulated data. While preparing simulated data always trade-off between complexity of generative models and effort put in data generation must be done.
- Simulated data can be created based on actual event data that was used as a pattern to generate a larger set of simulated data. The danger in this approach is that the data generated in this manner may be biased toward the original event data.
- Simulated data may not be representative to the type of events that occur in practise.

Although using simulated data is a widely accepted practise, it can potentially lead to situations that the methods or systems provide good results on the simulated (synthetic) data, but their practical value remains in fact untested. Further evaluation requires a comparison of model performance with actual events and validation by experienced practitioners.

4.3 Response

The recent advances in mobile devices that are capable of wireless communication and have substantial computing power have caught attention of researchers and practitioners (Souza and Kushchu, 2005). Their role during response operations can include communication (both voice and digital), they can serve as automated sensors (typically equipped with GPS, motion sensors, etc.), and are capable of relatively high quality imaging and video recording. They can be used to enhance situational awareness by gleaning information from sites which can produce accurate results (Vieweg et al., 2010). However, if these data generation capabilities are used, the problems of delivering repetitive information and information overload arise. DM techniques to manage the content and amount of information presented to the users have been proposed (Zheng et al., 2011). Privacy concerns associated with use of mobile devices in the context of DM and a proposal of the solution to that problem based on aggregate data collection, rather than linked to individual users were discussed by Best (2012).

An intelligent technology gaining recognition in disaster response is the use of robots in search and rescue operations (Shah and Choset, 2003). ML is closely related to robotics – the 'intelligence' of autonomous robots typically originates from ML algorithms. The strength of ML algorithms is exploited in mapping new environments that were created in the result of the disaster (for example piles of rubble). Rescue robots provide a clear example how algorithms that have ability to learn new knowledge from the data can be successfully applied in the real situations.

4.4 Recovery

Recovery is the last phase of the disaster management cycle and can be referred as rehabilitation and reconstruction of the disaster affected systems. Recovery begins by clearing the community of any debris and re-establishing basic infrastructure.

This phase seems to receive substantially less attention from the ML and DM communities but it warrants additional attention. DM can be used to coordinate across levels of government, as mandated by the Hurricane Sandy Rebuilding Task Force, by gathering data from various agencies (Rubin, 2013).

But there are possibilities for DM and ML support. DM can be of assistance through the integration of social media that can be augmented by the use of data including, geo-tagged photos of damage transmitted from the wireless systems such as smart phones and then posted to websites such as Facebook and Flickr (Liu et al., 2008). This information would not only provide spatial data from which any associated textual data can be mined, but if augmented by analysis of video to determine the severity of impact, could add to the development of recovery plans.

A vital function that DM can provide is in the determination of recovery costs for specific projects or areas. In large-scale events funds may be dispersed through a distributed procurement section. These systems can generate large amounts of data. By mining this data the receipt of disaster recovery funds could be less labour intense (Hristidis et al., 2010). In a Congressional hearing (US Congress, 2007) the Government Accountability Office stated they were able to identify fraud by using data mining after Hurricane Katrina.

DM can also be used for assisting in estimating the effects on economic recovery of the affected areas (Zheng et al., 2011), while ML can be used for determining optimal debris management strategies. Both DM and ML can be used in planning post-event recovery creating a more disaster resilient community (Hristidis et al., 2010).

5 Conclusions and future directions

In this paper, we discussed the potential and challenges in applying DM and ML to disaster management. We presented some selected examples of applications of DM and ML in different phases of the disaster management cycle. We had no intention of providing an exhaustive list of applications or to cover the problem comprehensively. Our intention was to highlight and discuss the challenges and indicate future trends. In general, we observe the progression of using the DM and ML from problem-specific applications (e.g., plume dispersion modelling, rescue robots, etc.) to somewhat broader applications, such as building situational awareness and real-time threat assessment.

One potentially useful new direction would be applying social network data mining to real-time analysis of data delivered from mobile applications used during the response phase of disaster management. During this review, the problem of the use of simulated data has been discussed. The lack of actual data is often addressed by using simulated data. Such synthetic data can lead to performance evaluations of the proposed algorithms and methods that are not realised in real settings and this practise should be scrutinised by the community.

Acknowledgement

This work has been partly supported by the Slovak Research and Development Agency under the contract No. APVV-0043-10.

References

Aved, A., Ho, A., Hua, K., Hoang, L. and Hamza-Lup, G. (2006) 'A mobile computing approach to automatic traffic evacuation management', *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, pp.85–90.

Bahrepour, M., Meratnia, N., Poel, M., Taghikhaki, Z. and Havinga, P.J.M. (2010) 'Distributed event detection in wireless sensor networks for disaster management', *Proceedings of 2nd International Conference on Intelligent Networking and Collaborative Systems (INCOS)*, pp.507–512.

Best, E. (2012) 'Collective models of disaster: making a case for using collective mobile phone location data in disaster science', *37th Annual Natural Hazards Research and Application Workshop*, Broomfield, CO, http://www.colorado.edu/hazards/awards/paper-competition/best.pdf, pp.1–19.

Bishop, C.M. (2006) *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, New York Inc., Secaucus, NJ.

Carr, L. (1932) 'Disaster and the sequence-pattern concept of social change', *American Journal of Sociology*, Vol. 38, pp.207–218.

Comfort, L.K., Dunn, M., Johnson, D.E., Skertich, R.L. and Zagorecki, A. (2004) 'Coordination in complex systems: increasing efficiency in disaster mitigation and response', *International Journal of Emergency Management*, pp.62–80.

Cooper, G.F., Dash, D.H., Levander, J.D., Wong, W-K., Hogan, W.R. and Wagner, M.M. (2004) 'Bayesian biosurveillance of disease outbreaks', *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI '04)*, AUAI Press, Arlington, Virginia, USA, pp.94–103.

Crooks, A., Castle, C. and Batty, M. (2008) 'Key challenges in agent-based modelling for geo-spatial simulation', *Computers, Environment and Urban Systems*, Vol. 32, No. 6, pp.417–430.

Cutter, S.L. (1996) 'Vulnerability to environmental hazards', *Progress in Human Geography*, Vol. 20, No. 4, pp.529–539.

Cutter, S.L. (2006) *Hazards, Vulnerability, and Environmental Justice*, Earthscan, Sterlng, VA.

Dorner, D. and Pfeifer, F. (1993) 'Strategic thinking and stress', *Ergonomics*, Vol. 36, No. 11, pp.1345–1360.

Ehlert, J.F. (2012) 'Smartphones, disasters, and knowledge management: an examination of field data collection, analysis and dissemination', *NATO RTO Proceeding of Human Factors and Medicine Panel (HFM) Specialists' Meeting*, Tallinn, Estonia, 16–18 April, 2012.

Federal Emergency Management Agency (FEMA) (2012) *Hazus Applications*, Retrieved from http://www.fema.gov/hazus-applications

Frawley, W.J., Piatetsky-Shapiro, G. and Matheus, C.J. (1992) 'Knowledge discovery in databases: an overview', *AI Magazine*, Vol. 13, No. 3, pp.57–70.

Gonzales, R.C. and Woods, R.E. (2007) *Digital Image Processing*, 3 ed., Prentice-Hall, ISBN: 978-0-131-68728-8.

Grasso, V.F. (2012) *Global Resource Information Database – Sioux Falls*, Retrieved from United Nations Environment Programme: http://na.unep.net/geas/docs/Early_Warning_System_Report.pdf

Han, J. and Kamber, M. (2006) *Data Mining – Concepts and Techniques*, 2nd ed., Morgan Kaufman, San Francisco.

Hristidis, V., Chen, S-C., Li, T., Luis, S. and Deng, Y. (2010) 'Survey of data management and analysis in disaster situations', *Journal of System Software*, Vol. 83, No. 10, pp.1701–1714.

Last, M., Markov, A. and Kandel, A. (2006) 'Multi-lingual detection of terrorist content on the web', *Intelligence and Security Informatics, Lecture Notes in Computer Science*, Vol. 3917, pp.16–30.

Liu, S., Palen, L., Sutton, J., Hughes, A. and Vieweg, S. (2008) 'In search of the bigger picture: the emergent role of on-line photo sharing in times of disaster', *Proceedings of the 5th International ISCRAM Conference*, ISCRAM, Washington DC.

Manning, C.D., Raghavan, P. and Shütze, H. (2008) *An Introduction to Information Retrieval*, Cambridge University Press, Cambridge, UK.

Miah, M. (2011) 'Survey of data mining methods in emergency evacuation planning', *CONISAR Proceedings*, Wilmington, NC, p.v4 n1815.

Michalowski, M., Ambite, J.L., Knoblock, C.A., Minton, S., Thakkar, S. and Tuchinda, R. (2004) 'Retrieving and semantically integrating heterogeneous data from the web', *IEEE Intelligent Systems*, pp.72–79.

Mileti, D. (1999) *Disasters by Design: A Reassessment of Natural Hazards in the United States*, National Academy of Sciences, Washington DC.

National Research Council (2007) *Improving Disaster Management: The Role of IT in Mitigation, Preparedness, Response and Recovery*, National Academies Press, Washington DC.

Nayak, S. and Zlatanova, S. (Eds.) (2008) *Remote Sensing and GIS Technologies for Monitoring and Prediction of Disasters*, Springer.

Neal, D. (1997) 'Reconsidering the phases of disaster', *International Journal of Mass Emergencies and Disasters*, Vol. 15, No. 2, pp.239–264.

NOAA (2004) *ALOHA User's Manual*, NOAA/HMRAD and EPA/CEPPO, 7600 Sand Point Way N.E., Seattle, WA 98115.

Perrow, C. (1999) *Normal Accidents: Living with High-Risk Technologies*, Princeton University Press, Princeton, NJ.

Pine, J. (2007) *Technology in Emergency Management*, Wiley and Sons, Hoboken, NJ.

Ristvej, J. and Zagorecki, A. (2011) 'Information systems for crisis management – current applications and future directions', *Communications: Scientific Letters of the University of Zilina*, Vol. 13, No. 2, pp.59–63.

Rubin, C. (2013) *Sandy Task Force Issues Recommendations for Long-Term Recovery*, Emergency Management, Retrieved from http://www.emergencymgmt.com/disaster/Sandy-Task-Force-Recommendations-Long-Term-Recovery.html

Shah, B. and Choset, H. (2003) *Survey on Urban Search and Rescue Robotics*, CMU, Pittsburgh, PA. 15213.

Souza, F. and Kushchu, I. (2005) 'Mobile disaster management system applications – current overview and future potential', *Proc. of the First European Mobile Government Conference MGOV2005*, Brighton, UK, 10–12 July, pp.455–466.

US Congress (2007) United States Congress, Senate, Committee on Homeland Security and Governmental Affairs, Hurricanes Katrina and Rita: Outstanding Need, Slow Progress, *110th Cong., 1st Sess*., GPO, Washington, Print.

Vieweg, S., Hughes, A., Starbird, K. and Palen, L. (2010) 'Microblogging during two natural hazards events: what twitter may contribute to situational awareness', *Proceeding of the Twenty-Eighth International Conference on Human Factors in Computing Systems*, Association for Computing Machinery, Atlanta, GA, pp.1079–1088.

Waugh, W. and Tierney, K. (Eds.) (2007) *Emergency Management: Principles and Practice for Local Government*, 2nd ed., International City/County Managers Association, Washington DC.

Weinstein, C., Campbell, W., Delaney, B. and O'Leary, G. (2009) 'Modeling and detection techniques for counter-terror social network analysis and intent recognition', *Aerospace Conference, 2009 IEEE*, pp.1–16.

White, C., Plotnick, L., Kushma, J., Hiltz, S. and Turoff, M. (2009) 'An online social network for emergency management', *Proceedings of the 6th International ISCRAM Conference*, Gothenburg, Sweden.

Yang, R., Tang, J. and Sun, D. (2011) 'Association rule data mining applications for Atlantic tropical cyclone intensity changes', *Weather and Forecasting*, Vol. 26, pp.337–353.

Zheng, L., Shen, C., Tang, L., Li, T., Luis, S. and Chen, S-C. (2011) 'Applying data mining techniques to address disaster information management challenges on mobile devices', *Proceedings of the 17th Association of Computing Machinery SIGKKD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, pp.283–291.

Zielinski, A. and Bügel, U. (2012) 'Multilingual analysis of twitter news in support of mass emergency events', *Proceedings of the 9th International ISCRAM Conference*, Vancouver, Canada, pp.1–5.