AUTOMATED DATA ANALYSIS IN THE CONTEXT OF CRISIS AND DISASTER MANAGEMENT

Adam Zagorecki[1], Jozef Ristvej[2] and David Johnson[3]

ABSTRACT

Crisis and disaster situations are characterised by high dynamics and complexity, with human lives, substantial environmental and economic consequences at stake. The advances in information technology have had a profound impact on the practice of all phases of the 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 in the context of the crisis and disaster management. We attempt to review the challenges and benefits of the automated data analysis to different phases of the crisis management.

Key words:
Disaster and Crisis Management, Data Analysis, Data Mining, Machine Learning

1 INTRODUCTION

The advances in Information Technology (IT) have had a profound impact on all phases of disaster and crisis management. This is due to their ability to create new data and process it in a manner that was unimaginable in the past. These advances allowed the development of sophisticated Emergency Management Information Systems (EMIS) based on IT. In the literature [9] there are identified four key functions a typical EMIS should provide. They are:

[1] Adam Zagorecki, Senior Research Fellow, Department of Informatics and Systems Engineering, Cranfield University, Defence Academy of the United Kingdom, email: a.zagorecki@cranfield.ac.uk, phone: +44 (0) 1793 785293.

[2] Jozef Ristvej, Assistant Professor, Department of Crisis Management, University of Zilina, email: Jozef.Ristvej@fsi.uniza.sk, phone: +41 5136717.

[3] David Johnson, Assistant Professor, Department of Political Science, Missouri State University, phone: +1 417-836-6956

- Data collection − typically done in the form of paper or online forms, remote sensing, etc. In particular:
- Remote sensing technologies are undergoing rapid advances, leading to affordable sensors that can communicate wirelessly and significantly increase situational awareness, often in real time.
- Data transfer and distribution − developments in telecommunications, in particular computer networks, provide fast and reliable means of transferring data over large distances. In the last decade, wireless communication has become affordable allowing new advances in particular in remote sensing.
- Data storage − relational databases are able to efficiently store large volumes of data, and provide virtually instantaneous access to it. Although this technology is well established, it is continuously advancing to store new and more demanding data such as video streams.
- 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 analyze it. Therefore, the EMIS should provide tools to allow for manipulation of collected digital data. There are tools to provide processing and analysis, however more sophisticated analysis still must be done by human decision makers. Of all four functions, this function proves to be most challenging and the advances seem to not meet expectations.

In this paper we discuss the role of IT in the data processing and analysis function. Of particular focus are the advances in the automated analysis of the collected data. We are interested in advanced analysis – creating new knowledge from available data, rather than processing the data in a prescribed manner, even though it can involve very complex manipulations. The two fields that offer creating new knowledge from data are data mining [6] and machine learning [1]. They rely exclusively on data to discover new patterns, trends, or deliver predictions. Disasters are events that often result from unexpected and dramatic changes in a system. These create challenges that are far beyond the capabilities of response agencies to recover from them with the resources they have available. Disasters very rarely start with a massive disruption of the system -- rather they are rapid processes that propagate the initial effects (limited) thorough a complex system leading to major disruption of the system. These properties imply that disasters involve large scale effects, which are complex and dynamic. Each of these characteristics poses important challenges to modelling and analysis. To make the situation even worse, disasters are really rare events, each of them with some unique characteristics. Our knowledge of catastrophic system failures is typically much less than our understanding of the systems. While we can assume that it is possible to collect the data on normal operations of a system, the data on catastrophic events is always scarce, by the nature of these events. 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 and respond to the impact of the disaster on the system. Because of the scale, disaster response involves multiple organizations – governmental, public and private, often crossing multiple layers of authorities [2, 11]. This emphasizes the need for decision support systems -- there is no single human decision maker who can comprehend the complexity of the situation. Instead, problems such as situational awareness [12] building a common operating picture shared among multiple actors, who often have only partial view of the situation, are becoming the most urgent needs of the disaster management.

There are numerous ways that IT can enhance the practice of disaster management. The overview of technologies in the context of disaster management was discussed in [9]. 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 will begin with, what we believe is a necessary clarification, of the definition for the term *data*. In the Section 3 we introduce the data mining and machine learning. In Section 4, we review the applications of these techniques to disaster management, reviewing their role in different phases. We conclude the paper with the discussion of future directions and practical challenges.

2 DATA

Discussion of the application of DM or ML to the field of disaster management should start with the clarification of the very basic term -- the data. From the authors' experience, the understanding of the term data varies extensively, especially between the technical fields (such as IT, data mining 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.

Expectations of the data from the data mining community are as follows: the data should be well structured (tables, databases, etc.), translatable into variables (numeric or textual), and preferably large amounts (preferably, in thousands or more instances). In general, these requirements reflect the limitations of the methods used. Very often, 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 interpretation of the term data in the other fields is much broader. The data can range from the well structured form as discussed earlier 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 human experts, but are difficult to be exploited by the DM and ML techniques. Data that is available to disaster management comes typically in the following forms:

- Unstructured textual data: news articles and announcements
- Structured textual data: situational reports, forms, etc.
- Remote sensing data: numeric measurements, but this category increasingly include use of mobile devices that can produce images and video
- Spatial data: data within Graphical Information Systems (GIS), satellite imagery, etc.

Diversity of the data is a serious issue in applying analytical methods to disaster management, because most of the methods were originally developed to handle numerical, well structured data. However developments, especially in information retrieval (textual information) and methods for image and video data are considerable.

3 DATA MINING AND MACHINE LEARNING

Machine Learning is a branch of computer science (more precisely artificial intelligence) and 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.

Frawley et al. define Data Mining as the nontrivial extraction of implicit, previously unknown, and potentially useful information from data [3]. 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 data sets. It is strongly related to large scale data management and processing (relational data bases, data warehousing, etc.). Unlike the ML, the DM emphasizes the practical aspects of data management and produced results. DM is expected to feed business intelligence and computer decision support systems.

There are obvious overlaps between DM and ML, but there are the key differences as well. The overlaps include: the same 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, while DM focuses on data.
- ML goal is more to reproduce existing knowledge, while DM is to discover new knowledge from the data.
- Data sources are of concern of DM, but not ML.
- DM requires large amounts of data, while ML does not.

DM and ML are widely recognized 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, email spam filters, online recommendation systems, and facial recognition for security applications, fraud detection, and many more.

4 DATA ANALYSIS IN DISASTER AND CRISIS MANAGEMENT

Disaster and crisis management poses a number of challenges. As discussed in Section 2, the availability of useful and comprehensive data is a serious challenge. But even if there would be available data, there are other problems. There are so many different ways a system can be subjected to a disaster, that it is impossible to have data on all of them. This is why the data-driven methods can more be useful to detect anomalies (out of normal) rather than interpret them. . There are inherent variations in the data values so the use of anomalies requires the determination of what level of value constitutes an anomaly.

4.1 PREVENTION AND MITIGATION

The most known examples of DM and ML for the prevention and mitigation phase of the disaster management have been applied to the prevention of man-made disasters in many areas: homeland security examples include detecting terrorist threats through analysis of social networks, fusion of sensor data for nuclear threat detection, face recognition at crowded settings, etc.

4.2 PREPAREDNESS

One of the key problems in the preparedness phase of the disaster management cycle is evacuation planning. Evacuation planning involves the need for spatial data and modelling evacuee behaviours [8]. The main research focuses on using DM methods to identify potential threat and safe areas.

ML finds its application in early warning systems [5], both for natural and man-made disasters. Examples include early warning systems for chemical and nuclear threats, floods, tsunami, etc. Conceptually, the early warning systems rely on the notion of 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.

4.3 RESCUE AND RELIEF

One of the intelligent technologies that has already gained recognition in the field of disaster management is the use of robots in search and rescue operations [10]. ML is closely related to robotics – the 'intelligence' of autonomous robots often originates from ML algorithms. The strength of ML algorithms is exploited in mapping new environments that were created after the disaster (for example rubble). It provides a clear example how algorithms that have ability to learn new knowledge from the data can be successfully applied in the real situations. Use of mobile devices that are capable of wireless communication and have substantial computing power during the rescue and relief operations is discussed by Zheng et al. [12]. They identified the problem of delivering repetitive information and information overload, and they used a DM technique to manage information presented to the users.

4.4 RECOVERY, REHABILITATION AND RECONSTRUCTION

The remaining phases of the emergency disaster cycle seem to receive substantially less attention from the ML and DM communities. One of the obvious explanations can be lack of the data on past events.

Despite the lack of previous attention does not mean there are not applications for the recovery phase. DM can be used for assisting in estimating economic damages while ML can be used for determining optimal debris management strategies.

5 CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we presented some selected examples of using DM and ML in disaster management. By no means do we claim that the examples are exhaustive and cover the problem comprehensively. More in-depth review can be found in [7]. In general, we observe the progression of using the DM and ML from problem-specific applications (e.g. rescue robots) to more direct applications such as building situational awareness and real-time threat assessment.

One of potentially very useful new directions 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 caught our attention. 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 realized in real settings.

REFERENCES

[1] BISHOP, C.M.,: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., 2006, Secaucus, NJ, USA.

[2] COMFORT, L.K., DUNN, M., JOHNSON, D., SKERTICH, R. and ZAGORECKI, A.,. Coordination in Complex Systems: Increasing Efficiency in Disaster Mitigation and Response. International Journal of Emergency Management (IJEM), 2004 vol. (1/2), pp. 62-80. 2004.

[3] FRAWLEY, W.J., PIATETSKY-SHAPIRO, G., and MATHEUS, C.J.: Knowledge Discovery in Databases: An Overview. AI Magazine Volume 13(3) (1992)

[5] GRASSO, V.F.: Early Warning Systems: State-of-the-Art Analysis and Future Directions – Report. United Nations Environment Programme (UNEP). Available: http://na.unep.net/geas/docs/Early_Warning_System_Report.pdf (Last accessed 28/06/2010)

[6] HAN, J, KAMBER, M.,: Data Mining – Concepts and Techniques, 2nd Ed. 2006. Morgan Kaufmann, San Francisco.

[7] HRISTIDIS, V., Chen, S-C., LI, T., LUIS, S., and DENG, Y.. 2010. Survey of data management and analysis in disaster situations. J. Syst. Softw. 83, 10 (October 2010), 1701-1714.

[8] MIAH, M., Survey of Data Mining Methods in Emergency Evacuation Planning. 2011 CONISAR Proceedings, v4 n1815. Wilmington NC.

[9] RISTVEJ, J., ZAGORECKI, A., Information Systems for Crisis Management – Current Applications and Future Directions, In: Communications: Scientific Letters of the University of Žilina, Vol. 13, No.2, 2011, pp. 59-63

[10] SHAH, B. and CHOSET, H. (2003), "Survey on Urban Search and Rescue Robotics", CMU, Pittsburgh, PA. 15213. Lee, C.H. and Kwak, Y.K.

[11] WAUGH, W. and TIERNEY, K. (Eds) (2007) "Emergency Management: Principles and Practice for Local Government, 2nd Edition, International City/County Managers Association. Washington, DC pp 7.

[12] 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. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD '11). ACM, New York, NY, USA, 283-291.

Článok recenzoval:
Ing. Katarína Kampová, PhD.