Case Study Assignment

Syed Muhammad Ahmed Zaidi

20972008

Articles chosen:

1) Crisis analytics: big data-driven crisis response Qadir et al. (2016)
2) Survey of data management and analysis in disaster situations (Hristidis et al., 2010).

Abstract:
Natural disasters have had long-lasting impacts on every affected society. However, with modern, state-of-the-art technology improvements, human response to tackling such phenomena has widely depended on various forms of data. The use of Big data management has allowed us to be ready for any future crises and respond during and after the hazards have passed. Crises response systems now incorporate data from various sources to provide in-depth valid insights that can be used to manage and mitigate possible risks. The purpose of this case study is to examine the overall role of data management with respect to managing environmental hazards. The aim is to congregate knowledge about the systems already in use for analysis of crisis-related data and also be able to reflect the future expectations we can keep from technology to be able to add more towards the protection of lives. To assist with the stated objective, multiple technologies are discussed in light of computer science disciplines, including information extraction from various sources, information filtering, and decision-making.

Introduction:

In order to effectively analyze any natural disaster, its process is broken down into different phases. The first stage is pre-crisis preparedness, which deals with preparing before a potential crisis and working towards reducing the disaster risk (Twigg, 2004). A few of the activities in this phase include regularly testing systems, planning, and warehousing emergency relief goods. The second phase is during-crises response which deals with all activities done by stakeholders in collaboration. These activities aim to reduce the impact caused by the disaster at the time it occurs. Lastly, the phase is called post-crises response, which covers all the analysis done to find deep insights allowing recovery and protection from such future crises. Numerous data sources are being considered while finding adequate information to form a functional analysis in all three phases. As an overview, these sources fall under categories of online activity, government data, sensing technologies, data exhaust, crowdsourced data, and many more. In order to evaluate this massive amount of data being generated on a daily basis (2.5 quintillion bytes per day (Makridakis, 2014).), systems need to be updated with the latest technology. Some examples of these are sentimental analysis using user-generated data, social behavior, and environmental analysis using satellites, UAVs, and wearable network devices. Other examples can include social network analysis, crisis maps, and Epidemiology. The challenges associated with this field of analysis are unique as mostly the terminologies being used within, defer from one scenario to another. The extraction and use of data to provide insights require highly time-sensitive systems to be able to aid in decision-making for effective results. The data itself is generated in heterogeneous formats, adding to the difficulty of standardizing data usage in the overall field (Hristidis et al., 2010)

Literature Review:

In the field of crisis management, it is highly significant to have access to timely, accurate, and effective data so that decisions can be taken to reduce devastating impacts. This is also reflected in the article by Henderson (1999), who says it is crucial to anticipate crises early and respond to them effectively. In this regard, there have been various publications of literature aiming to study past experiences in order to provide a better future. The discussions are around the data sources used over the years; for example, Haklay and Weber (2008) highlight the use of advancement in mapping technologies such as OpenStreetMap, allowing democratization to access sensing data. The Internet of Things (IoT) itself plays an essential role as it involves different sensors and actuators sharing a mutual network with various computing systems, paving the way toward providing actionable data (Manyika et al., 2013). Data is also important when recovering from any natural phenomenon, as Kirkpatrick, the director of UN Pulse innovation initiative, introduced a new terminology called "data philanthropy" (Kirkpatrick, 2013). This refers to the companies that gather mobile and social media data during a crisis to help disaster-affected people recover more quickly. The data itself is being produced at exponential rates; however, despite being heterogeneous, the key is finding insights and knowledge that enable correlation from multiple data sources (Foster & Grossman, 2003). Over the years, technology has improved so that even though data from sources related to disasters can be quite uncertain (Wachowicz & Hunter, 2005), it still manages to generate accurate figures for better-quality decision-making. Techniques such as Data-driven digital epidemiology with CDRs and social media have allowed in-depth social network analysis to improve response rates (Salathe et al., 2012). Urban analytics and population surveillance with the use of machine learning on internet data have also added to techniques like sentimental analysis, opinion mining, and search and rescue missions (Kamel Boulos et al., 2011).

Data Management context through the application of the 5Vs

Both articles chosen for this case study reflected similar applications of 5vs as the underlying topic was interrelated. Regarding the volume of data, natural crisis management incorporates countless sources of information, making it very challenging to analyze and produce insightful results. Any limitations to data could cause adverse effects on rescue efforts, and at the same time, the abundance of data could slow down the process of response during crises. An example of this could be how there were 20 million tweets within a week with the hashtag "Sandy" that was referring to Hurricane Sandy (Qadir et al., 2016).

The variety of data is another element that can become a hurdle in getting quick analytical results. Firstly the models of any area are built on the information related to its climate, environment, population, and other geographical characteristics. Then these are expanded using numerous other repositories of data like mapping technologies, CDRs, and satellite figures. All of these are produced in heterogeneous formats making relationship-building another obstacle. Machine Learning and AI models, like NLP-based crisis informatics, also require data standardization to produce effective results. However, in respect to the natural crisis, the data is gathered from vastly spread lands making linguistic diversity another barrier that needs to be solved.

Considering the velocity of data, both articles reflect how the growth in exponential rates has caused the swiftness of data production to multiply over the years. Social media being of the major contributors to this pace of data production. The data points produced by this stream can be used in different ways; for example, Twitter is being used majorly for post-crisis period to locate people who may need further aid (Munro & Manning, 2012).

In certain scenarios, data can be produced that may misguide the authorities while making decisions. This shows how the 4th V, Varsity is another essential element that needs consideration. An example of this could be how people tend to exaggerate when facing stressful situations like a natural crisis (Qadir et al., 2016). To eliminate such deception, models like TweetCred are developed that use web- and ML-based plugins to autodetect non-credible tweets and forged images in real-time (Gupta et al., 2014).

Lastly, the value of data management in this regard can be judged based on the benefits extracted from applying the models produced. Web-based decision support system (Yong et al., 2001), simulation model for emergency evacuation (Pidd et al., 1996), and agent-based system for knowledge management (Wang et al., 2004) are all example models that have been created over the years that are highly effective and are being used to manage crisis situations. This shows how the value of data itself can be used to value humankind.

Similarities and Differences:

The articles chosen for this case study had multiple overlapping concepts and thoughts. The most interlinked theory was the disaster life cycle. The first article, called "Crisis Analytics: big data-driven crisis response" by Qadir et al. (2016), divides the whole life course of a disaster into three categories called Pre-Crises, During-Crises, and Post-Crises response. Whereas the second article, "Survey of data management and Analysis in disaster situations" (Hristidis et al., 2010), also mentions how the cycle should be divided into 4 phases for better targeting. They named these Preparedness, Mitigation, Response, and Recovery. They may have different names but present similar ideas of what is included in each and how they should be managed. Another similarity found in both articles is the list of sources mentioned as data inputs. Even though they both showed different usage and ways of analyzing data, they both made use of the same data sources, including sensing data, mapping technologies, satellite feeds, government data, and much more. Both of them also put extra emphasis on social media data as it could lead to far-sighted foresight.

Both articles may well be sharing the same gist for the topic, but in detail, they reflect different aspects of data usage. The first article by Qadir et al. (2016) focuses more on the overall sources of data usage and how they can be used in the modern-day era to find valuable insights related to the topic. It incorporates another wing to this subject matter that deals with how humanitarian data can also be vital when efforts are being made to aid the crisis-affected society. On the other hand, the primary focus of the article by Hristidis et al. (2010) was on the systems being used in relation to computer science. It elaborated on the scientific aspects of how the systems analyze data in collaboration with the Miami-Dade County emergency management office. Both articles were also distinct when considering future expectations. The first article made "real-time" data the highlight towards a securer society as Qadir et al. (2016) believe "real-time analytics can help prioritize the most urgent problems before any further damage is caused." In contrast to Article 2, which expects future systems to integrate better the wide variety of data formats produced by different sources. Hristidis et al. (2010) state how the focus should shift to "Intelligent querying on heterogeneous, multi-source streaming."

Discussion and Reflection:

The articles chosen for this case study were a practical representation of what we have studied in the classroom. Each article discussed the attributes of data in detail, allowing a better understanding of how Volume, Variety, Velocity, Varasity, and Value can be used to find insights for real-world issues. This case study also widened my knowledge about the different types of sources used in different industries, specifically in Crisis management. In the future, this field could improve the system's intake for heterogeneous formats, which can pave the way towards a more Machine Learning and Artificially intelligent reporting system that deals with real-time data. This will help make the decision-making more effective as the results from the analysis will be highly accurate and timely. This can be highly useful for the future as the Earth's climate changes, extreme weather events such as droughts, heat waves, floods, and hurricanes are becoming more common and intense, posing a emerging threat to human health and safety, infrastructure and ecosystems (Ogunbode et al., 2020). For the upcoming data management plan, learnings from these articles can be incorporated into data relationships, data formats, and data sources to work towards building a data model for a specified area affected by natural disasters.

References

Foster, I., & Grossman, R. L. (2003). Data integration in a bandwidth-rich world. *Communications of the ACM*, *46*(11), 50-57.

Gupta, A., Kumaraguru, P., Castillo, C., & Meier, P. (2014). Tweetcred: Real-time credibility assessment of content on twitter. Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11-13, 2014. Proceedings 6,

Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, *7*(4), 12-18.

Henderson, J. C. (1999). Managing the Asian Financial Crisis: Tourist Attractions in Singapore. *Journal of Travel Research*, *38*(2), 177-181. https://doi.org/10.1177/004728759903800212

Hristidis, V., Chen, S.-C., Li, T., Luis, S., & Deng, Y. (2010). Survey of data management and analysis in disaster situations. *Journal of Systems and Software*, *83*(10), 1701-1714. https://doi.org/https://doi.org/10.1016/j.jss.2010.04.065

Kamel Boulos, M. N., Resch, B., Crowley, D. N., Breslin, J. G., Sohn, G., Burtner, R., Pike, W. A., Jezierski, E., & Chuang, K.-Y. S. (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International journal of health geographics*, *10*(1), 1-29.

Kirkpatrick, R. (2013). Big data for development. *Big Data*, *1*(1), 3-4.

Makridakis, S. (2014). Siegel, Eric. 2013. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. *Interfaces*, *44*(4), 435-437.

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E. A. (2013). Open data: Unlocking innovation and performance with liquid information. *McKinsey Global Institute*, *21*, 116.

Munro, R., & Manning, C. D. (2012). Short message communications: users, topics, and in-language processing. Proceedings of the 2nd ACM Symposium on Computing for Development,

Ogunbode, C. A., Doran, R., & Böhm, G. (2020). Exposure to the IPCC special report on 1.5 C global warming is linked to perceived threat and increased concern about climate change. *Climatic Change*, *158*, 361-375.

Pidd, M., De Silva, F., & Eglese, R. (1996). A simulation model for emergency evacuation. *European Journal of operational research*, *90*(3), 413-419.

Qadir, J., Anwaar, A., ur, R. R., Zwitter, A., Sathiaseelan, A., & Crowcroft, J. (2016). Crisis analytics: big data-driven crisis response. *Journal of International Humanitarian Action*, *1*(1). https://doi.org/https://doi.org/10.1186/s41018-016-0013-9

Salathe, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., Campbell, E. M., Cattuto, C., Khandelwal, S., & Mabry, P. L. (2012). Digital epidemiology.

Twigg, J. (2004). Disaster Risk Reduction, Mitigation and Preparedness in Development and Emergency Programming, Good Practice Review 9. *Humanitarian Practice Network (HPN), London: ODI*.

Wachowicz, M., & Hunter, G. J. (2005). Dealing with uncertainty in the real-time knowledge discovery process. *Geo-information for disaster management*, 789-797.

Wang, L., Qin, Q.-M., Wang, D.-D., Ghulam, A., Zuo, C., & Zhao, J. (2004). Decision support system of flood disaster for property insurance: theory and practice. IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium,

Yong, C., Chen, Q., Frolova, N., Larionov, V., Nikolaev, A., Pejcoch, J., Suchshev, S., & Ugarov, A. (2001). Web based decision support tool in order to response to strong earthquakes. *Proceedings of TIEMS2001, Oslo, Norway*.